Predicting City Poverty Using Satellite Imagery

Simone Piaggesi[*†] Laetitia Gauvin[*] Michele Tizzoni[*] Natalia Adler[‡] Stefaan Verhulst[§]
Andrew Young[§] Rihannan Price[¶] Leo Ferres[‖] Ciro Cattuto[*] André Panisson[*]

Abstract

Reliable data about socio-economic conditions of individuals, such as health indexes, consumption expenditures and wealth assets, remain scarce for most countries. Traditional methods to collect such data include on site surveys that can be expensive and labour intensive. On the other hand, remote sensing data, such as high-resolution satellite imagery, are becoming largely available. To circumvent the lack of socio-economic data at high granularity, computer vision has already been applied successfully to raw satellite imagery sampled from resource poor countries.
In this work we apply a similar approach to the metropolitan areas of five different cities in North and South America, starting from pre-trained convolutional models used for poverty mapping in developing regions. Applying a transfer learning process we estimate household income from visual satellite features. The urban environment we consider is characterized by different features with respect to the resource-poor training environment, such as the high heterogeneity in population density. By leveraging both official and crowd-sourced data at city scale, we show the feasibility of estimating the socio-economic conditions of different neighborhoods from satellite data.

1. Introduction

For years, estimating economic growth and development to assess human well-being has been a central issue for international research and policy [3, 11]. Accurate data about human development primarily come from surveys and censuses. Collecting socioeconomic information can be hard-working, expensive and might suffer from reporting errors. In the last years, the huge amount of remote sensing data, combined with recent developments in machine learning techniques, led to methods for estimating socio-economic indicators from geospatial data, such as *nightlights* [1, 17] and *satellite imagery* [26]. In particular, such methods have been used to overcome the limitations or lack of development data and to estimate poverty at large scale in developing countries [22, 9, 16].

In this work we apply a machine learning approach similar to the one developed by Jean et al. [16] for predicting economic outcomes in 5 African countries from satellite imagery, to the smaller scale of a urban environment. In particular, we explore the feasibility of predicting household income at various municipality levels in a city. To this aim, we examined the Metropolitan Area of Santiago in Chile and other five big cities in the USA: Los Angeles, Philadelphia, Boston, Chicago, and Houston. Our work tackles three main research questions:

1. *Is it possible to extend machine learning methods, previously applied to resource poor settings, to estimate poverty levels in a city of a developed country?*
2. *Given different aggregation levels in a city – usually corresponding to different administrative subdivisions – can a model trained on a lower spatial resolution yield information about a more granular aggregation level?*
3. *What is the out-of-sample predictive power of such a model, when tested on a new city?*

Our main findings related to the above questions are:

- Starting from pre-trained deep computer vision models we show the feasibility of predicting household income in a city through a regression task, with best results on settings with only urban areas. We show that in this framework there is no need to fine tune existing models or to leverage on proxy variables (such as night-time lights data) to achieve good performances.
- Just considering municipalities in which the regressor is better predictive about the target, we can also improve the estimation in more fine-grained levels of aggregation.

[*]ISI Foundation, Turin, Italy
[†]University of Bologna, Bologna, Italy
[‡]UNICEF, New York, USA
[§]The GovLab, New York, USA
[¶]Digital Globe, Westminster, CO, USA
[‖]Universidad Del Desarrollo & Telefónica R&D, Santiago, Chile
E-mail addresses: simone.piaggesi2@unibo.it, laetitia.gauvin@isi.it, michele.tizzoni@isi.it, nadler@unicef.org, stefaan@thegovlab.org, andrew@thegovlab.org, rhiannan.price@digitalglobe.com, lferres@udd.cl, ciro.cattuto@isi.it, andre.panisson@isi.it

- Testing our regression model on a new city, never seen before by the algorithm, results in a good performance if compared to a null model.

2. Related Literature

In the last decade the development of new machine learning techniques, such as Deep Learning [19] and Convolutional Neural Networks [18], has allowed to extraordinarily improve many Computer Vision applications [25, 7, 12]. Recently, deep learning has been used in combination with remote sensing data for various tasks, such as Scene Classification [21, 20, 4], Urban Planning [2] and Crop Yield Prediction [27]. A recent application concerns the estimation of socioeconomic indicators, such as assets, consumption and wealth indexes, from satellite imagery. In [16] the authors proposed a transfer learning process, where nightlights intensities are used as an intermediate proxy [1] to map poverty in five African countries. Other works have used similar transfer learning approaches to estimate other variables [14], with different proxies [23] or deep models [6]. Also, other studies [24, 8, 15] have trained a deep neural network to predict poverty from satellite images without proxies, or used other types of remote sensing data for the same task [10]. All the latter models were trained and tested in resource poor settings (Africa, India, Bangladesh, and Sri Lanka) mainly considering their rural areas. To our knowledge no previous work has shown the application of similar techniques to the urban areas of a developed country.

3. Data

3.1. Economic Variables

In this work, we focus on the *household income* as the main economic indicator to be predicted. Data about this indicator comes from different sources, depending on the city under study, and the same indicator has to be available at different levels of granularity for validation. The following cities were selected, due to data availability:

Santiago, Chile The household income is obtained from the EOD (*Encuesta Origen Destino de Viajes*), a mobility survey realized from July 2012 to November 2013 by assignment of the Chilean Ministry of Transport and Telecommunications[1]. The survey refers to a random sample of 18,264 households coming from the Santiago Metropolitan Area, for a total of 60,054 people involved. Household income information is averaged at the municipality-level of *comunas*, but also at a more fine-grained city subdivision, called *zonas*.

USA cities Households income data are available from Census Reporter[2], a website through which the United States Census Bureau provides socio-economic and demographic data on the population of the United States. The geographic levels we considered include the *ZIP codes* and the more fine grained *census tracts*.

In the following, we will refer to the different levels of aggregation (*comunas*, *zonas*, *ZIP codes* or *census tracts*) with the term *clusters*.

3.2. Satellite Imagery

Santiago, Chile Satellite images are downloaded as a mosaic of 34 big tiles from the DigitalGlobe web platform[3], to entirely cover the area of the city. Tiles are a mixture of pansharpened and natural color RGB images, taken during daytime between September 2017 and February 2018, with 50 cm resolution and maximum cloud coverage of 3%. We generated a grid of not overlapping images from the mosaic of tiles covering the city area. The grid has a $1km$ step and each image is cropped from the original mosaic and resized to 400 x 400 pixels.

USA cities Satellite images are downloaded using the Google Static Maps API, each of them with 400 x 400 pixels at zoom level 16, resulting in images with a resolution of 2.5 m/pixel ($1km^2$ area per image) not overlapping each other.

Table 1 provides a summary of the geographic elements and the different spatial resolutions considered in our study. A description of the resolution scales that we are covering is reported in Fig. 1, with the distribution of the surfaces areas of the geospatial elements for each aggregation level.

3.3. Urban Areas Boundaries

Shapes and boundaries of urban areas of each city are obtained from official[45] sources and then merged with OpenStreetMap[6] (OSM) crowd-sourced data to have a more comprehensive description of the urban landscape. From OSM map elements, we selected urban spaces joining those with the tag key `landuse` with possible values: *residential*, *commercial*, *retail*, *recreation_ground*, *construction*, *college*, *university*, *public*, *allotments*, *churchyard*, *depot*.

4. Methods

4.1. Convolutional Features Extraction

To estimate socio-economic indicators from satellite imagery, we leverage the transfer learning approach intro-

[1] http://www.sectra.gob.cl/encuestas_movilidad/encuestas_movilidad.htm

[2] https://censusreporter.org/topics/income/

[3] https://services.digitalglobe.com/

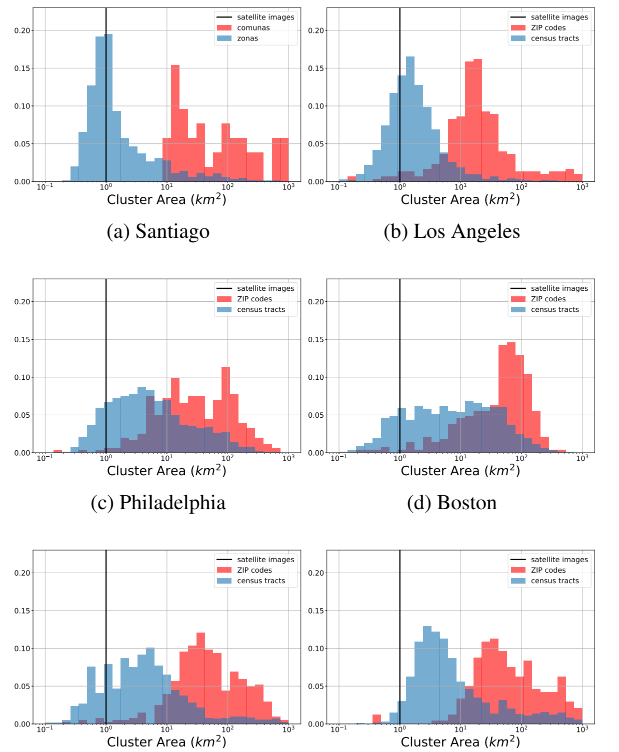
[4] http://www.rulamahue.cl/mapoteca/presentaciones/chile_regiones.html

[5] https://www.census.gov/geo/maps-data/data/cbf/cbf_ua.html

[6] http://download.geofabrik.de/

Geographic Level	Santiago	Geographic Level	Los Angeles	Philadelphia	Boston	Chicago	Houston
Images	12444	*Images*	16774	19528	14766	31042	37353
Zonas	588	*Census Tracts*	1772	1334	907	1829	1061
Comunas	52	*ZIP Codes*	290	359	282	403	237

Table 1: Number of satellite images and geospatial clusters for each city.



(a) Santiago (b) Los Angeles

(c) Philadelphia (d) Boston

(e) Chicago (f) Houston

Figure 1: Distribution of different cluster-level areas for the considered cities.

duced and elaborated by [26, 16]. Such approach consists in using a Convolutional Neural Network (CNN), trained on a nighttime light intensity prediction task, as a feature extractor that maps from each input image to a vector representation, which incorporates nightlights features. Such features are proved to be a good proxy for economic development [17, 1], and can be used at the bottom of a linear regression model to predict wealth-related indicators.

To perform the forward pass from raw satellite images to visual features we have considered three CNN models:

- `ResNet50`. Residual Network model [13] with 50-layers and initialized with ImageNet weights.
- `VGGF`. The 8-layers model [5] also considered in [26] with ImageNet weights.
- `VGGF+nightlights`. The fully convolutional variant of the previouos one, fine-tuned with nightlights intensity labels on some African countries (Uganda, Tanzania, Nigeria, Malawi).

In essence, we compared a model fine-tuned on the nighlights prediction task with other two models (`VGGF` and `ResNet50`) just pre-trained on the ImageNet dataset, with the aim of quantifying the goodness of fit when using general features not related with night-time lights.

For each tile, we extract vector representations from the top layer of the model, before the softmax classifier. The size of our images is the same respect to the input of the `VGGF` fine-tuned model, while the pre-trained neural networks take as input 224x224 pixels images. In this case, we averaged the feature vectors extracted from the four 224x224 overlapping quadrants of each image to get vector representations. In addition, each image is rotated by multiples of 90 degrees and flipped horizontally/vertically before the mapping, to obtain an augmented dataset.

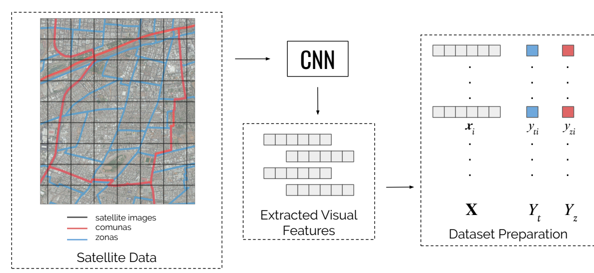


Figure 2: Diagram with the collected datasets, feature extraction and data preparation processes for the experiments. The feature vector $\mathbf{x}_i$ is obtained from a CNN model using a satellite image as input. y_{ti} and y_{zi} are the socio-economic variables extracted from two different levels of census area boundaries, associated to the original satellite image.

The diagram on Fig. 2 shows the collected datasets, feature extraction and data preparation processes. For each city, satellite images are divided in a grid and each tile is associated to socio-economic variables at different levels of census area boundaries. A vector representation for each tile is extracted with different CNN models, as explained above. For each experiment in Section 5, a different part of this dataset is used as training and validation sets.

4.2. Regression for Spatial Prediction

As observed in previous applied computer vision works, visual features can be used to estimate socio-economic indicators. Here we test the prediction of the household income variable on multiple aggregation levels in different cities.

To explore how features inferred from the urban environment influence the final estimations, we can train and validate our models on different subsets of images, belonging or not to urban areas. In this section every image is assigned to an urban area if its center falls inside the shape of the area, but we have observed that assigning to urban regions also images with marginal overlaps with the latter (less restrictive statement respect the one above) does not change considerably final the results. Urban boundaries are obtained following the methodology explained in Sec. 3.

We perform each experiment with a Ridge Regression model, using image vector representations as predictor variables and income values as target. Hyperparameters are estimated with a 10-fold cross-validation, using as metric the determination coefficient R^2 and evaluating it on different validation sets (specified in each experiment).

We distinguish two different cases of prediction:

Image-level Estimation We use features of image-level embeddings as set of predictors and assign to this the target value up-sampled to an individual resolution level (*comunas*, *zonas*, *ZIP code* or *census tract*). The absence of superimposition among images prevents the information leaking between training and test sets in the learning process. Because of the fine-grained property of the set of images, in this case it is possible to tune the model on points that belong or not to urban areas. Then, we can investigate how much the presence of urban areas may influence the final performance of the model, but also see if a global training improves the estimation only in urban spaces.

Cluster-level Estimation For each cluster we use as predictor variables the average of image-level embeddings computed on the images which are part of the cluster itself. This is the same approach used in previous published works. As done by [16], we neglect spatial clusters with less than 10 images for the regression task.

5. Experiments

5.1. Income Prediction within Cities

The first experiment concerns the income estimation at the level of a single city, to understand two significant aspects: which convolutional model yields more informative features from images, and what is the importance of the level of urbanization in terms of regression performance. Table 2 shows the results for the municipality of Santiago, for briefness we omit outcomes for all the other cities because they are equally demonstrative. We can observe that globally the pre-trained ResNet50 model performs better at every considered resolutions, besides the VGGF model fine-tuned on nightlights has low performances in any task (meaning that this model is not reliable if not applied to the original data space where it is fine-tuned). For this evidence, we perform all the other experiments presented with satellite features extracted with the ResNet50 model.

Table 3 illustrates the analysis of urban areas, using only the ResNet50 features (having observed that it is the best model among all those considered). Here, features are image-level and we compute the regression score in two validation sets: a first one with all the images and a second one including only those related to urban spaces. We can observe that in most cases the regressor can better explain the variance of the target in urban neighborhoods. Fig. 3 displays the spatial heatmaps comparing the true income distribution with the predicted one in the city of Santiago.

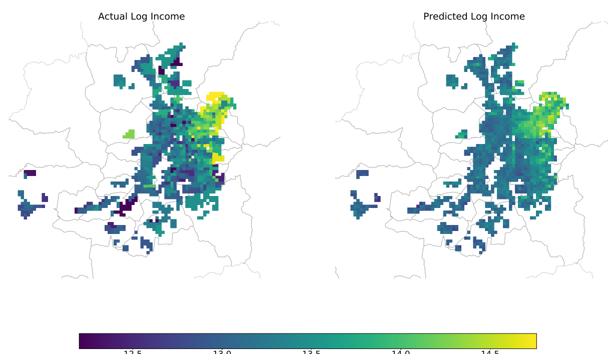


Figure 3: Spatial representations for the image-level prediction task performed in Santiago. Visual features are extracted with the ResNet50 model, the target variable is the *zonas*-level household income. **Left:** Distribution of real values. **Right:** Distribution of predicted values.

5.2. Inference of Higher Resolution Estimates

In Table 3 we have seen that we can estimate the distribution of the household income, taking individual image features as input, at a more fine-grained resolution with respect to the target itself (each image is $1km^2$). In this section we investigate more in depth this aspect, asking if with a training step on the more coarse-grained level the regression model can also assess the distribution of the target at a more fine-grained level.

To this aim, we consider two sets of aggregation levels T and Z of the same geographical area (suppose T to be more

Santiago					
Features Level	Target Level	Train.-Val.	ResNet50	VGGF	VGGF+nightlights
Zonas	*Zonas*	All Areas	0.477	**0.484**	0.356
		Urban Areas	**0.591**	0.523	0.433
Comunas	*Comunas*	All Areas	**0.643**	0.598	0.553
		Urban Areas	**0.737**	0.711	0.623
Images	*Zonas*	All - All	**0.454**	0.408	0.258
		All - Urban	**0.520**	0.506	0.314
		Urban - Urban	**0.584**	0.542	0.358
Images	*Comunas*	All - All	**0.667**	0.613	0.342
		All - Urban	**0.691**	0.625	0.429
		Urban - Urban	**0.772**	0.713	0.503

Table 2: Image-level and cluster-level (*zonas* and *comunas*) cross validation performance, measured with R^2 scores, of the household income regression task in Santiago. Here we compare features coming from different convolutional models (ResNet50, VGGF, VGGF+nightlighs). The ResNet50 features outperform the others in almost all tasks. Training and/or validation sets are coupled with different neighborhoods of the city (urban areas or not).

		Validation set	
City	Target Level	All	Urban
Santiago	*Zonas*	0.454	**0.520**
	Comunas	0.667	**0.691**
Los Angeles	*Census Tracts*	**0.657**	0.569
	ZIP Codes	**0.569**	0.458
Philadelphia	*Census Tracts*	0.360	**0.460**
	ZIP Codes	0.358	**0.443**
Boston	*Census Tracts*	**0.384**	0.374
	ZIP Codes	0.367	**0.399**
Chicago	*Census Tracts*	0.301	**0.361**
	ZIP Codes	0.309	**0.382**
Houston	*Census Tracts*	0.250	**0.327**
	ZIP Codes	0.266	**0.340**

Table 3: Image-level cross validation performance, measured with R^2 scores, with different environments in the validation set. The training set includes image-level features extracted with the ResNet50 model, from both urban and rural areas. We observe that the validation score is higher on urban areas, regardless of the granularity level of the target.

granular than Z). In our case, T is a set of *census tracts* (or *zonas* in Santiago), and Z is a set of *ZIP codes* (*comunas* in Santiago). Our framework consists in two sets of labelled features $L_T = \{(\mathbf{x}_t, y_t),\ t \in T\}$ and $L_Z = \{(\mathbf{x}_z, y_z),\ z \in Z\}$, representing two training sets at different spatial scales. After training a regression model to learn a mapping on the higher scale Z

$$\mathbf{x}_z \mapsto y_z = f_Z(\mathbf{x}_z)$$

we want to show that the function f_Z is also predictive of

City	score	baseline
Santiago	0.411	**0.483**
Los Angeles	0.531	**0.618**
Philadelphia	0.631	**0.684**
Boston	0.523	**0.626**
Chicago	0.447	**0.706**
Houston	0.506	**0.614**

Table 4: Cross-validation performance for the cluster-level income prediction task, when the the Ridge Regression is trained to predict *ZIP codes*-level (or *comunas*) target values, but validated on *census tracts*-level (or *zonas*) set of target values. Scores are compared with a baseline computed assigning to each *census tract* (or *zona*) the target of the corresponding *ZIP code* (or *comuna*).

the target on the lower scale T. To do so, we define a baseline on this task, which assigns to each $t \in T$ the target y_{z_t}, where $z_t \in Z$ is the cluster which spatially contains t.

Our purpose consists in training the regression model with cluster-level features on the lower resolution level (i.e. the set Z), but maximizing the validation score on the higher one (i.e. the set T). From Table 4 we can argue that this procedure does not improve the baseline.

We can go beyond, validating only on areas in which the model predictions are better. To do so, firstly we sorted the lower level clusters according to a local regression loss (the squared error between the actual target and the predicted one), and secondly we introduced an additional hyperparameter q (optimized during the cross-validation) which tunes the fraction of clusters taken into account for the validation.

In this way, we can also estimate the number of ele-

City	q	score@q	baseline@q
Santiago	0.35 (39%)	**0.484**	0.481
Los Angeles	0.1 (0%)	0.548	**0.616**
Philadelphia	0.25 (28%)	**0.628**	0.617
Boston	0.1 (0%)	0.645	**0.761**
Chicago	0.2 (0%)	0.454	**0.598**
Houston	0.4 (39%)	**0.578**	0.542

Table 5: Cross-validation performance for the cluster-level income prediction task, when the the Ridge Regression is trained as in Table 4, but validated on *census tracts* (or *zonas*) belonging to the fraction q of best predicted *ZIP codes* (or *comunas*). The choice of the hyperparameter q is optimized during the cross-validation, between values from 0.1 and 0.5 with step 0.05. In parentheses are shown corresponding percentages of higher-level clusters (*census tracts* or *zonas*) for which the prediction is improved respect to the baseline by the regression model. Baseline values are different respect to Table 4 because are evaluated on a different subset of clusters.

		Our Model		Null Model	
City	Target Level	`All`	`Urban`	`All`	`Urban`
Los Angeles	*Census Tracts*	**0.435**	0.424	-0.038	-0.032
	ZIP Codes	**0.362**	0.349	-0.070	-0.084
Philadelphia	*Census Tracts*	0.256	**0.295**	-0.201	-0.186
	ZIP Codes	0.402	**0.442**	0.025	0.036
Boston	*Census Tracts*	0.028	**0.029**	-0.977	-0.985
	ZIP Codes	0.172	**0.185**	-0.586	-0.541
Chicago	*Census Tracts*	0.193	**0.262**	-0.013	-0.015
	ZIP Codes	0.172	**0.400**	0.008	0.019
Houston	*Census Tracts*	0.255	**0.298**	-0.037	-0.031
	ZIP Codes	0.226	**0.327**	-0.178	-0.141

Table 6: Performance of the cluster-level household income prediction task, in R^2 scores, when for each test city the model is trained only on the others. We performed also 500 experiments with a null model, reporting the average of the scores for each city.

ments of the higher resolution level T for which we can improve the estimation with respect to the baseline. Table 5 shows that such approach gets positive results for a significant fraction of *census tracts* (or *zonas* in Santiago).

5.3. Income Predicition Among Cities

The last experiment is related to the application of a *leave-one-out* approach for the household income prediction, i.e. using information gained on a training set of multiple cities to estimate the economic variable in a new city that has never been seen by the algorithm. In this section, we apply this method only to the set of US cities, since they share the same aggregation levels. Results are shown for the cluster-level task in Table 6, and for each city is reported the outcome from the model trained on the others. To test the statistical significance of this method, we also report scores when the prediction is performed by randomly assigning cluster-level features among training cities, keeping constant the number of clusters for each city. From the scores discrepancy between our outcomes and the null model we can figure out that our model's out-of-sample predictive power does not derive from an accidental case.

6. Conclusions

In this work, we investigated the poverty prediction task in the urban environment of two developed countries. We showed that methods used for poverty mapping in resource poor settings can be also applied in this context. Specifically, we have shown that a model pre-trained on the ImageNet dataset can explain, about the target, a significant fraction of the variance with no fine-tuning procedure or proxies. Moreover, we showed that a regression model trained with respect to a given aggregation of the target can infer spatial properties of more granular resolution levels. Finally, we demonstrated the predictive power of these methods if applied to infer the target in new test cities.

Acknowledgments

This project was funded in part by the Data2X Initiative through the Big Data for Gender Challenge Awards. SP acknowledges support from the Lagrange Project and CRT Foundation (`http://www.isi.it/en/lagrange-project/project`). AP acknowledges support from Intesa Sanpaolo Innovation Center. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Douglas M Addison and Benjamin P Stewart. Nighttime lights revisited: the use of nighttime lights data as a proxy for economic variables. 2015.

[2] Adrian Albert, Jasleen Kaur, and Marta C. Gonzalez. Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 1357–1366, New York, NY, USA, 2017. ACM.

[3] Sudhir Anand and Amartya Sen. Human development index: Methodology and measurement. 1994.

[4] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv:1508.00092 [cs]*, Aug. 2015. arXiv: 1508.00092.

[5] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[6] Derek Chen. Temporal Poverty Prediction using Satellite Imagery. page 9.

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.

[8] Tony Duan, Elliott Chartock, Haque Ishfaq, Paul Novosad, Sam Asher, Marshall Burke, David Lobell, and Stefano Ermon. Predicting Poverty with Satellite Imagery in Bangladesh and India. page 8.

[9] Christopher D Elvidge, Paul C Sutton, Tilottama Ghosh, Benjamin T Tuttle, Kimberly E Baugh, Budhendra Bhaduri, and Edward Bright. A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8):1652–1660, 2009.

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

[11] James E Foster, Luis F Lopez-Calva, and Miguel Szekely. Measuring the distribution of human development: methodology and an application to mexico. *Journal of Human Development*, 6(1):5–25, 2005.

[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.

[13] 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.

[14] Andrew Head, Mlanie Manguin, Nhat Tran, and Joshua E. Blumenstock. Can Human Development be Measured with Satellite Imagery? In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development - ICTD '17*, pages 1–11, Lahore, Pakistan, 2017. ACM Press.

[15] Jeremy Irvin, Dillon Laird, and Pranav Rajpurkar. Using Satellite Imagery to Predict Health. page 7.

[16] Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, Aug. 2016.

[17] Souknilanh Keola, Magnus Andersson, and Ola Hall. Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, 66:322–334, 2015.

[18] 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.

[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[20] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657, Feb. 2017.

[21] Keiller Nogueira, Otvio A. B. Penatti, and Jefersson A. dos Santos. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61:539–556, Jan. 2017.

[22] Abdisalan M Noor, Victor A Alegana, Peter W Gething, Andrew J Tatem, and Robert W Snow. Using remotely sensed night-time light as a proxy for poverty in africa. *Population Health Metrics*, 6(1):5, 2008.

[23] Shailesh M. Pandey, Tushar Agarwal, and Narayanan C. Krishnan. Multi-Task Deep Learning for Predicting Poverty From Satellite Images. In *Thirty-Second AAAI Conference on Artificial Intelligence*, Apr. 2018.

[24] Anthony Perez, Christopher Yeh, George Azzari, Marshall Burke, David Lobell, and Stefano Ermon. Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning. Nov. 2017.

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

[26] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. *arXiv:1510.00098 [cs]*, Sept. 2015. arXiv: 1510.00098.

[27] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. Deep gaussian process for crop yield prediction based on remote sensing data. In *AAAI*, pages 4559–4566, 2017.