

Tiny-Inception-ResNet-v2: Using Deep Learning for Eliminating Bonded Labors of Brick Kilns in South Asia

Usman Nazir, Numan Khurshid, Muhammad Ahmed Bhimra, Murtaza Taj
Department of Computer Science, Syed Babar Ali School of Science and Engineering
Lahore University of Management Sciences (LUMS), Lahore, Pakistan
{17030059, 15060051, 17030015, murtaza.taj}@lums.edu.pk

Abstract

This paper proposes to employ a Inception-ResNet inspired deep learning architecture called Tiny-Inception-ResNet-v2 to eliminate bonded labor by identifying brick kilns within “Brick-Kiln-Belt” of South Asia. The framework is developed by training a network on the satellite imagery consisting of 11 different classes of South Asian region. The dataset developed during the process includes the geo-referenced images of brick kilns, houses, roads, tennis courts, farms, sparse trees, dense trees, orchards, parking lots, parks and barren lands. The dataset¹ is made publicly available for further research. Our proposed network architecture with very fewer learning parameters outperforms all state-of-the-art architectures employed for recognition of brick kilns. Our proposed solution would enable regional monitoring and evaluation mechanisms for the Sustainable Development Goals.

1. Introduction

According to the Global Slavery Index of 2018 24.9 million people are trapped in forced labor globally [1]. An estimated 12.7 million are within “Brick-Kiln-Belt” of South Asia (1, 551, 997 km^2 between Afghanistan, Pakistan, India, Bangladesh and Nepal), because indebted labor is a major contributing factor to the prevalence of slavery in the Asia-Pacific constituting 55% of all forced labor-mapping brick kilns and estimating their labor force is an essential step in eliminating slavery from the region. The UN’s Sustainable Development Goal (SDGs) 8 specifically refers to forced labor and governs the world over aims for ending modern slavery by 2025 [2]. They are, however, faced with a lack of access to reliable, up-to-date and actionable data on slavery activity. Such data needs to be spatially explicit and scalable to allow governments to monitor activities and implement strategies to emancipate individu-



Figure 1. Example satellite imagery of brick kilns from different spatial locations showing variation in quality, structure and color profile

als trapped within institutions of slavery. The absence and insufficiency of data compromises evidence-based action and policy formulation, as has been highlighted by various sustainable development data-gap analysis reports. Thus, to meet this challenge, new and innovative approaches are needed.

One of the key challenges faced by all development efforts is selecting target populations, or the beneficiaries of programs. In order to ensure that interventions reach those who need them, and have an impact on economic and human development, one must first identify those who are vulnerable and most in need of help. This problem only increases in complexity when we seek to identify individuals trapped in slavery or forced labor. In the past, household surveys have generally been the norm in identifying those in need of development assistance. However, surveys are costly and cumbersome, and rarely have large enough samples to be representative at the village or town level. Data is

¹<https://cvlab.lums.edu.pk/?p=1779>

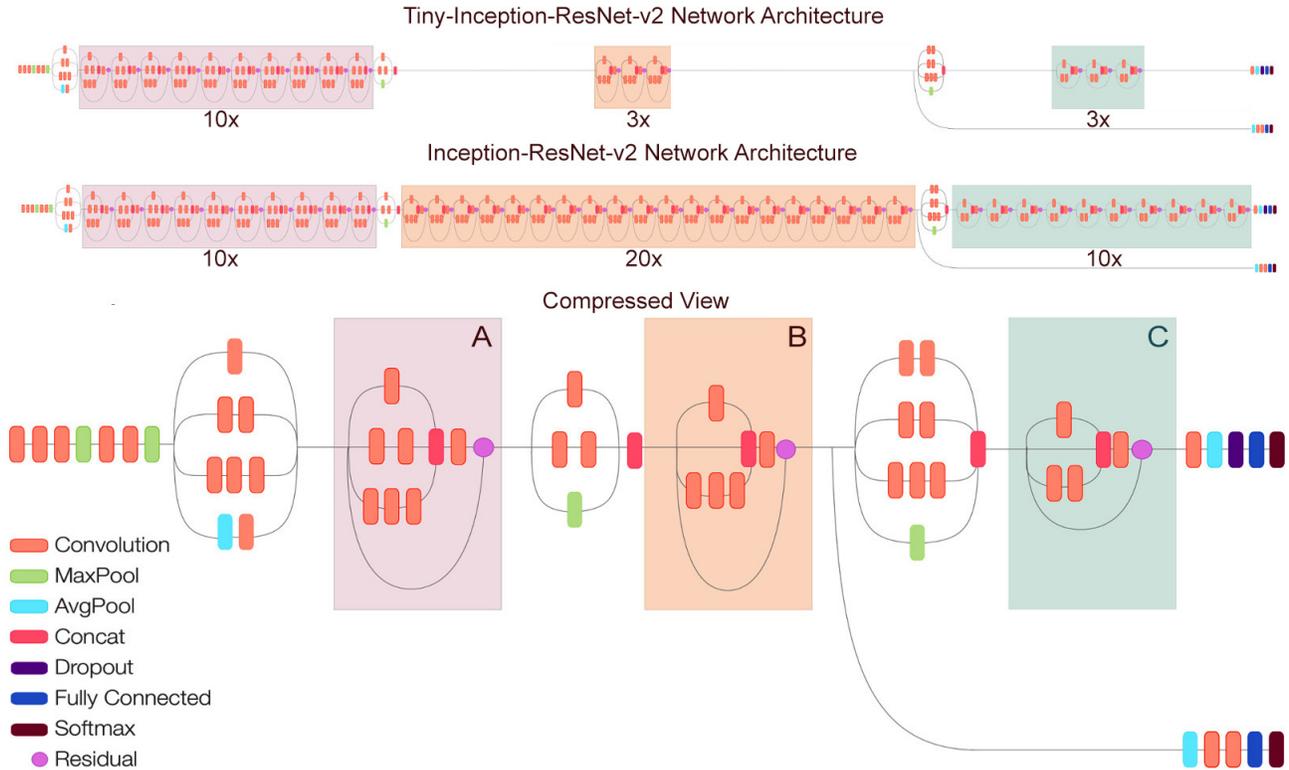


Figure 2. Proposed Tiny-Inception-ResNet-v2 Architecture. In Tiny-Inception-ResNet-v2, A inception block (in compressed view) is repeated 10 times, B is 3 times and C is 3 times as shown in top subfigure.

thus usually aggregated at the national, regional, provincial, or district level. Statistics estimated at the national or district level, however, may mask the fact that there are pockets of space which are severely in need of assistance. In addition to this, slave owners may hinder data collection in the areas that they control, or enslaved individuals may face numerous pressures in disclosing themselves as enslaved to enumerators or authorities.

Remote sensing is one such method that has been applied to various social and developing issues to help better inform policy. Remote sensing refers to the use of satellite or other sensor technologies to obtain information about the physical characteristics of an area from afar without conducting expensive and strenuous surveys. Moreover, new and improved techniques in computer vision and deep learning use large datasets with somewhat similar proxy variables as are collected in surveys to gauge vulnerability on a more minute scale and provides solution with minimal human intervention.

To develop an efficient automated solution for brick kiln identification, we propose Tiny-Inception-ResNet-v2 network applied to satellite imagery. We also developed a new satellite image dataset comprising the images of the South Asian region for 11 different categories. The dataset will

be made publicly available and will help inform the policy of governments within the region, and allow them to gauge the success of past interventions, as well as serve as a baseline by which to assess future progress. To the best of our knowledge, this will be the first of its kind effort that goes beyond the geographic boundaries of a specific country and aims to provide a comprehensive survey of bonded labor at kilns along with their accurate geolocation.

1.1. Related Work

Availability of high resolution satellite imagery along with recent advancements in machine learning particularly deep convolutional neural networks (CNN) have paved a way for large scale analysis of wide variety of parameters across the globe. High resolution satellite imagery has been used by [3] to estimate per capita expenditure in Africa. Recently, remote sensing images have also been used to analyze the extent of modern slavery. For instance, [1] have mapped fish farms suspected of using child slave labor in the Sundarbans Reserve Forest in Bangladesh. Similarly, [1],[4], through their “Slavery from Space” project proposed a crowd-sourced procedure to manually detect brick kilns from satellite imagery. However, they were only able to manually annotate 320 geographic cells (i.e. only

2% of the entire Brick-Kiln-Belt). This study also does not account for workforce size of each kiln.

The potential inherent within remote sensing and machine learning for social science research and development purposes has been noted by many [1, 5, 6, 7]. Residual Neural Network [8] have been shown to produce favorable results on satellite imagery for the problem of classifying imagery into known classes such as road, houses, vegetation etc. Similarly, [9] proposed algorithm to classify each pixel of satellite imagery into known classes thus obtaining an exact boundary of the object of interest in the given satellite image.

One area that has especially benefited from such techniques is poverty studies. Machine learning has been used to identify poverty stricken populations and their geographic location through the analysis of high resolution satellite images. One such technique utilizes the premise that nighttime luminosity is strongly correlated (if noisily) with economic activity and development [10]. Using satellite images, such efforts compare daytime and nighttime images of a specified location (a country, a continent) to identify which factors are associated with greater night lighting e.g. paved road networks. [10] find in their study that their method of estimating consumption expenditure and asset wealth through machine learning is able to explain 75% of the variation in local level economic outcomes. Though such methodologies cannot detect poverty at the level of the individual or household, they can inform policy makers about which sub-populations or communities are economically marginalized or vulnerable, and help them target their interventions at particular geographical locations where they are most needed. Recent contribution of [2] analyzed various machine learning techniques to automate the process of brick kiln identification in the given tile of images. However, their approach outputs very high false positive rate. To cope with the scenario they proposed to train a two-staged R-CNN classifier to achieve acceptable performance.

Our approach contributes by proposing Tiny-Inception-ResNet-v2, an improved architecture inspired from Inception-ResNet-v2 [11] to identify kilns in the satellite images. We showed that with fewer learnable parameters of Tiny-Inception-ResNet-v2 we could easily classify brick kilns with about 95% accuracy.

2. Challenges

Identifying brick kilns from satellite imagery involves the following challenges:

2.1. Structural Variations

Brick kilns may significantly vary in structure, shape, and size as shown in Fig. 1. They come in circular, oval, rectangular, and elongated structures. These variations are

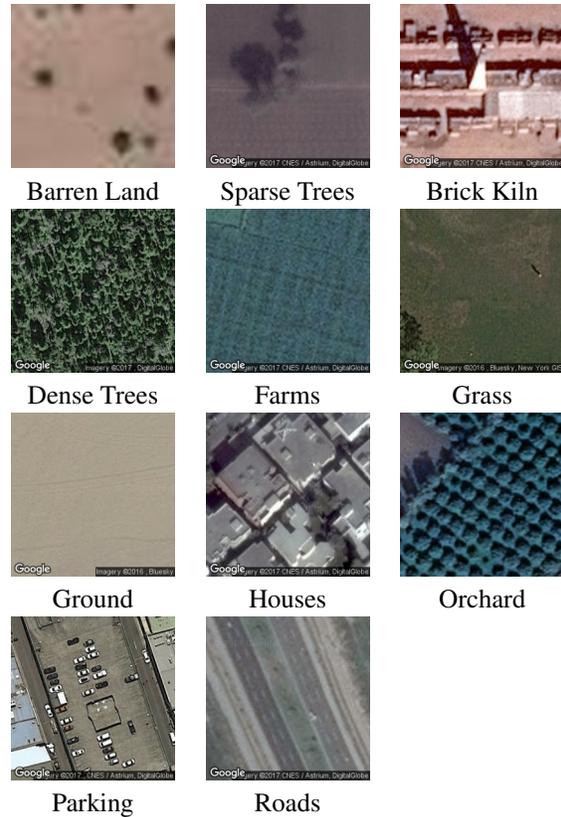


Figure 3. Example satellite image from every class in dataset (Satellite image courtesy Google Maps)

usually enforced by geographical locations, environmental conditions, manufacturing technologies, building material, and local building regulatory authorities. This is considered as one of the obvious challenges, for any machine learning approach, to learn a generic set of visual features for each type of kilns.

2.2. Environmental Variations

Satellite images are prone to atmospheric variations including but not limited to cloud cover, pollution, variation in luminosity, and seasonal changes in the environment for which the imagery has been acquired. These variations in images contributes in confusing the classifier. An obvious effect of these changes could be observed between the top-left and top-right images of Fig. 1.

2.3. Sensor Variations

Satellite images are acquired through various imaging devices and sensors installed on satellites. Some of the famous ones include WorldView, Pleiades, GeoEye-1, and QuickBird. Sensor variations are usually caused by difference in parameters of imaging devices capturing same imagery differently over time. When spatial analysis is spread across multiple cities, these changes in sensors can be ob-

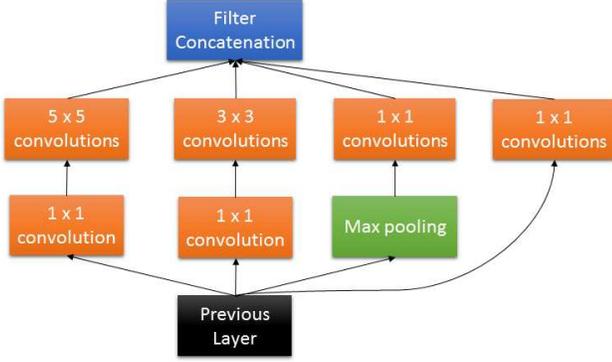


Figure 4. Inception Block

served as huge variations in the quality, resolution and color profile of the imagery due to the different satellite data (see Fig. 1).

3. Proposed Method

3.1. Dataset Generation and Pre-processing

We developed a dataset of satellite imagery from South Asian region by acquiring images at zoom level 20 with their geo-locations from OpenStreetMaps (OSM). These images were then manually labelled through crowd-sourcing into 11 different classes each having around 600 images. All these 6827 images were used to train the proposed classifier for categories of kilns, houses, road, tennis courts, farms, sparse trees, dense trees, orchards, parking, parks and barren lands (see Fig. 3).

To identify brick kilns in the "Brick-Kiln-Belt" we used the same network, testing it on an entirely different images from digital globe imagery (787,100 images) for the entire Afghanistan-to-Nepal region 1, 551, 997 km^2 for the year 2018. These images acquired at zoom-level 17 ($1.193 \frac{meters}{pixel}$ on Equator) were first converted to zoom-level 20 (about 50 million images) using following pre-processing steps:

1. Finding the midpoint *latitude-longitude* (lat-lon) of each image tile of zoom-level 17 using following equations:

$$lat^{z_{17}} = \left[\left((Y_{tile}^{z_{17}} - A) \times B \right) + C \right], \quad (1)$$

where

- $A = 42295$ is a reference tile #. It can be chosen randomly; however, we selected 42295 throughout our experiments.
- $B = 0.0054931640625$ is difference between adjacent tiles of zoom-level 17.

- $C = 52.33612060546875$ is a relative lat and lon with respect to A .

$$lon^{z_{17}} = \left[\left((X_{tile}^{z_{17}} - A) \times B \right) + C \right], \quad (2)$$

where A , B and C are given above.

2. Converting zoom-level 17 lat-lon to zoom-level 20 lat-lon using following equations:

$$cornerLat^{z_{20}} = lat^{z_{17}} + 4 * D, \quad (3)$$

$$cornerLon^{z_{20}} = lon^{z_{17}} - 4 * D, \quad (4)$$

where

- $D = 6.866455078125 * 10^{-4}$ is the difference between lat/lon of adjacent zoom-level 20 tiles.

In our work each image is associated with its geo-location which makes it very easy for images classified as brick kilns class to be identified on the maps. The geo-location of the identified brick kiln could then be linked to the census data provided by Punjab-Brick-Kiln-Census-dataset² which is publicly available containing all the relevant information regarding number of workers, number of children, age of each child and schooling status associated with each kiln.

We proposed a classifier that is inspired from the well-known Inception-ResNet-v2 network. The performance gain of the network as compare to other deep learning architectures encompassing the inception block incorporated in the network.

3.2. Inception Block

The inception module consists of convolutions of different sizes that allow the network to process features at different spatial scales. They are then lumped and fed to the next layer for further processing as shown in Fig. 4. For dimensionality reduction, 1x1 convolutions are used before the more expensive 3x3 and 5x5 convolutions. In many remote-sensing problems, we need the deeper network to process features at different spatial scales. To cope with our challenges, such flexibility can be incorporated in convolutional neural networks by introducing inception blocks.

3.3. Tiny-Inception-ResNet-v2

An improved form of Inception-ResNet-v2 called as "Tiny-Inception-ResNet-v2" with much less inception blocks have been proposed to classify satellite images. We suppose that even fewer number of inception blocks used in Inception-ResNet network are able to classify the satellite

²http://202.166.167.115/brick_kiln_dashboard/

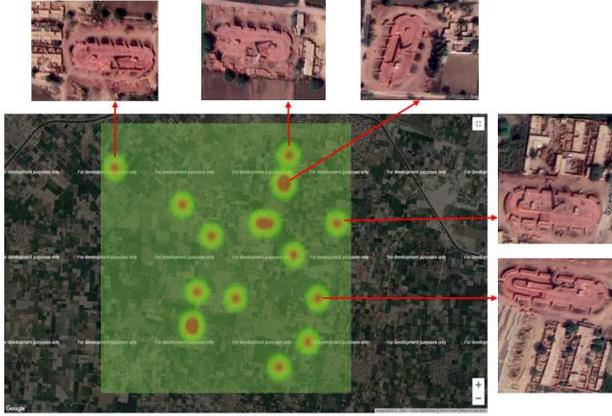


Figure 5. Qualitative Analysis on Kot Radha Kishan, in Punjab Province, Pakistan. It can be seen within the green region-of-interest; proposed network correctly classified all brick kilns. (Satellite image courtesy Google Maps)

images in classes. For this purpose optimal number of inception blocks are to be identified. Fig. 2 shows the detailed architecture of our proposed network with a block level architecture at the top of the figure while the expended version at the bottom. Three main blocks namely A, B, and C containing different number of stacked inception blocks are shown. To identify the optimal number of inception modules to be used in each block, we tried various combinations given in Table 1 and found that network with 10 inception block in A, 3 blocks in B, and 3 blocks in C outperforms the other combinations. It has also been observed that this network has much fewer learnable parameters as compare to other combinations. More inception blocks have been used in A than B and C, to learn the features of the image responsible for visual appearance. The later ones focus on discriminating the learned features.

4. Results and Evaluation

We evaluated our trained network of Tiny-Inception-ResNet-v2 on unseen dataset consisting of 700 images of zoom 20 for quantitative analysis and 787,100 images of zoom 17 acquired from Digital Globe Imagery for qualitative analysis. The classifier with output kiln is considered as 1 whereas the rest of the class outputs are considered as non-kiln (0). Following benchmark performance metrics have been calculated to validate the performance of our approach.

4.1. Performance Metrics

Precision, Recall, and F1-Score have been calculated for each of the networks as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (5)$$

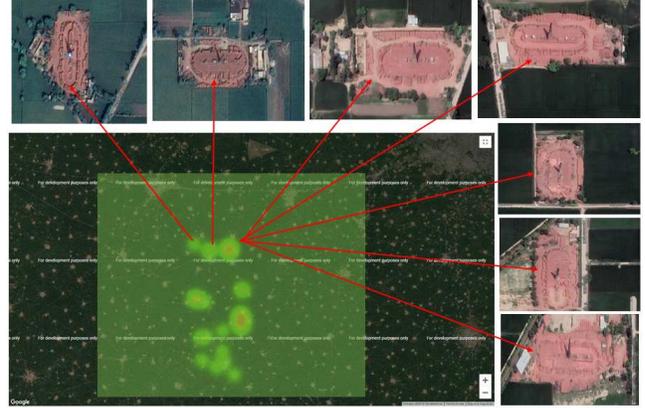


Figure 6. Qualitative Analysis on Province Punjab, India. It can be seen within the green region-of-interest; proposed network correctly classified all brick kilns. (Satellite image courtesy Google Maps)

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (6)$$

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (7)$$

4.2. Quantitative Results

The proposed method explained in the previous section to identify brick kilns has been tried for various standard deep learning residual architectures. Performance metrics with their number of learnable parameters have been shown in Table. 2. It is evidently clear from the table that even though deep ResNets and Inception networks may consists of many layers and much more learnable parameters, still they fail to perform for our problem. Our proposed Tiny-Inception-ResNet-v2 may lag behind in terms of precision value; however, it leads the list in recall, F1-score, and number of learnable parameters. One of the major problems with the two staged R-CNN approach [2] was, huge number of false positives identified by the binary classifier. The issue was resolved by adapting the two-staged classification approach. Our proposed multi-class classifier has very less false positive results with a recall of about 90.52% and F1-Score of 94.35% with 19m parameters only.

4.3. Qualitative Results

For the qualitative analysis of our approach we showed the results of Kot Radha Kishan, Punjab, Pakistan in Figures 5 and Punjab, India in 6 respectively. The images clearly show that brick kilns with various structural differences have been identified by our proposed classifier. Moreover, our classifier efficiently recognizes the brick kiln in the image despite the variation in size, spatial location and orientation of the kiln in the image. The heatmap generated pin points each and every brick kiln identified in block at zoom 17 adapting the proposed method.

Table 1. Comparison between different versions of Tiny-Inception-ResNet-v2

A-block	B-block	C-block	Val Loss	Precision	Recall	F1 Score	Parameters
11	20	10	0.00037	0.9950	0.8550	0.9200	54,353,643
20	5	5	0.00016	0.9940	0.2990	0.4597	27,243,579
10	3	3	0.00022	0.9854	0.9052	0.9435	19,682,011
10	1	1	0.00041	0.993	0.3130	0.4760	13,354,523

Table 2. Table showing quantitative evaluation of the proposed network with state-of-the-art architectures

Network Architectures	Precision	Recall	F1 Score	# of Parameters
Two Staged R-CNN [2]	0.9494	0.9494	0.9494	-
ResNet-152	0.9906	0.8166	0.8952	41,407,238
ResNet-50	0.9909	0.8416	0.9102	21,312,267
ResNet-34	0.9892	0.8841	0.9337	21,312,267
Inception-v3	0.9846	0.7413	0.8458	21,815,078
Inception-ResNet-v2	0.9955	0.8552	0.9200	54,345,958
Tiny-Inception-ResNet-v2 (Proposed)	0.9854	0.9052	0.9435	19,682,011

5. Conclusion

We proposed a novel network architecture: “Tiny-Inception-ResNet-v2” to identify the brick kilns in satellite images. The network was trained on zoom-level 20 images (782 images) of brick kilns of Lahore, Pakistan while tested on the zoom-level 17 images of South Asia (about 700). Despite the structural and color variation in the images of kilns, our proposed network, with very fewer learning parameters, outperforms all the state-of-the-art network architectures achieving F1-Score of around 94%. We also provided detailed geo-referenced dataset and annotations for 11 classes, which will serve as a valuable resource for further such analysis.

References

- [1] D. S. Boyd, B. Jackson, J. Wardlaw, G. M. Foody, S. Marsh, and K. Bales, “Slavery from space: Demonstrating the role for satellite remote sensing to inform evidence-based action related to un sdg number 8,” *ISPRS journal of photogrammetry and remote sensing*, vol. 142, pp. 380–388, 2018.
- [2] G. M. Foody, F. Ling, D. S. Boyd, X. Li, and J. Wardlaw, “Earth observation and machine learning to meet sustainable development goal 8.7: Mapping sites associated with slavery from space,” *Remote Sensing*, vol. 11, no. 3, p. 266, 2019.
- [3] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, “Transfer learning from deep features for remote sensing and poverty mapping,” in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [4] B. Jackson, K. Bales, S. Owen, J. Wardlaw, and D. S. Boyd, “Analysing slavery through satellite technology: How remote sensing could revolutionise data collection to help end modern slavery,” *J. Mod. Slavery*, vol. 4, pp. 169–199, 2018.
- [5] D. Blumberg and D. Jacobson, “New frontiers: remote sensing in social science research,” *The American Sociologist*, vol. 28, no. 3, pp. 62–68, 1997.
- [6] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, “Adversarial synthesis learning enables segmentation without target modality ground truth,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1217–1220.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, p. 436, 2015.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [10] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, “Combining satellite imagery and machine learning to predict poverty,” *Science*, vol. 353, no. 6301, pp. 790–794, 2016.
- [11] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” in *Association for the Advancement of Artificial Intelligence*, vol. 4, 2017, p. 12.