

Kiln-Net: A Gated Neural Network for Detection of Brick Kilns in South Asia

Usman Nazir¹, Usman Khalid Mian¹, Muhammad Usman Sohail¹, Murtaza Taj¹, and Momin Uppal¹, Senior Member, IEEE

Abstract—The availability of high-resolution satellite imagery has enabled several new applications such as identification of brick kilns for the elimination of modern-day slavery. This requires automated analysis of approximately 1 551 997 km² area within the “Brick-Kiln-Belt” of South Asia. Although modern machine learning techniques have achieved high accuracy for a wide variety of applications, problems involving large-scale analysis using high-resolution satellite imagery requires both accuracy as well as computational efficiency. We propose a coarse-to-fine strategy consisting of an inexpensive classifier and a detector, which work in tandem to achieve high accuracy at low computational cost. More specifically, we propose a two-stage gated neural network architecture called *Kiln-Net*. At the first stage, imagery is classified using the ResNet-152 model which filters out over 99% of irrelevant data. At the second stage, a YOLOv3-based object detector is applied to find the precise location of each brick kiln in the candidate regions. The dataset, named *Asia14*, consisting of 14 000 Digital Globe RGB images and 14 categories is also developed to train the proposed kiln-net architecture. Our proposed network architecture is evaluated on approximately 3,300 km² region (337 723 image patches) from 14 different cities in five different countries of South Asia. It outperforms state-of-the-art methods employed for the recognition of brick kilns and achieved an accuracy of 99.96% and average F1 score of 0.91. To the best of our knowledge, it is also 20 x faster than existing methods.

Index Terms—Brick Kiln, ResNet-152, sustainable development goals, you only look once (YOLO).

I. INTRODUCTION

ACCORDING to the global slavery index of 2019, 40.3 million people across the globe are trapped in forced labor [1]. An estimated 24.3 million of these are within the so-called “Brick-Kiln-Belt” of South Asia comprising an area of approximately 1 551 997 km² between Afghanistan, Pakistan, India, Bangladesh, and Nepal (see Figs. 1 and 2). Keeping in view the UN’s sustainable development goal (SDG) 8.7, which specifically aims to address forced labor, mapping brick kilns in

Manuscript received February 28, 2020; revised April 27, 2020; accepted May 29, 2020. Date of publication June 12, 2020; date of current version June 25, 2020. This work was supported in part by the National Agriculture Robotics Lab at LUMS under National Center of Robotics and Automation, Pakistan. (Corresponding author: Usman Nazir.)

The authors are with the Department of Computer Science, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan (e-mail: 17030059@lums.edu.pk; 20100061@lums.edu.pk; 20100053@lums.edu.pk; murtaza.taj@lums.edu.pk; momin.uppal@lums.edu.pk).

This article has supplementary downloadable material available at <https://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2020.3001980

South Asia is an essential first step toward eliminating modern-day slavery in the region. However, the progress toward this goal is hampered by persistent lack of data. Whatever little data are available it is not only spatially sparse, but also infrequently collected primarily due to the high cost associated with manual surveys.

As a means to map brick kilns in a region of interest (ROI), a promising alternate to manual surveys is automated detection using satellite imagery. This has become possible due to recent advancements in machine learning, particularly deep convolutional neural networks [2], as well as public availability of high-resolution satellite imagery. These advancements have paved the way for automated analysis in a large number of applications such as object detection [3], classification [4], and image segmentation [5]). These techniques, in turn, have enabled large-scale automated spatial surveys for a wide variety of problems including, but not limited to, poverty estimation [6], crop yield estimation [7], and damage assessment due to natural and man-made disasters [8].

The problem of mapping brick kilns in limited regions of the “Brick-Kiln-Belt” of South Asia has been addressed in the past, albeit to a limited extent, using both low-resolution [10] and high-resolution satellite imagery [13]. Some of these solutions are based on manual surveys [13] in which a crowd-sourced methodology is utilized for manually annotating brick kilns in remote satellite imagery. On the other hand, while deep convolutional neural networks (CNN) have enormous potential for detection of objects in satellite imagery, their use requires a significant amount of training data, which is nonexistent for most of the cities in the ROI. The existing solutions for automatically detecting brick kilns (such as those based on faster R-CNN [9] or normalized difference vegetation index (NDVI) [10]) are limited in their spatial coverage (see Table I) due to lack of annotated data. In addition, these methods lack scalability which is essential for the large-scale problem at hand.

To develop an efficient and a scalable solution for automated surveying of brick kilns, we propose Kiln-Net: a gated neural network-based strategy that can efficiently detect and map brick kilns in a given ROI using satellite imagery. Most existing object detection techniques [3], [4], [20] perform simultaneous classification and localization and are thus computationally infeasible for application to high-resolution satellite imagery. It is to avoid this that Kiln-Net relies on a gated structure which decouples classification and localization. This results in a coarse-to-fine search in which fine-grained localization via object detection is

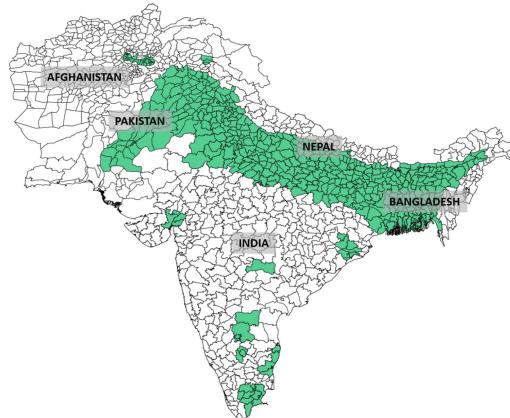
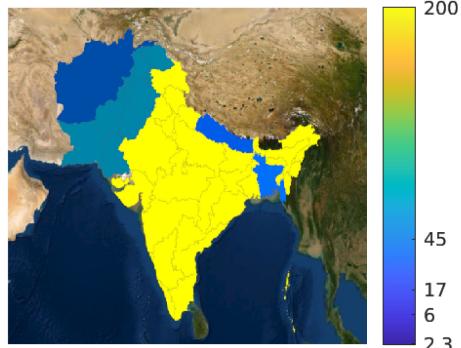


Fig. 1. Geographical distribution of brick kilns in “Brick-Kiln-Belt” of South Asia [14]–[19].



Country	Number of enterprises	Production (billion bricks/year)
India [14], [15]	35,000	200
Pakistan [14], [15]	11,500	45
Bangladesh [15], [16]	4,500	17
Nepal [15], [17], [18]	450	6
Afghanistan [19]	-	2.3
Total	51,450	270.3

Fig. 2. Total production of bricks (in billions) in South Asia. (Base map courtesy ArcGIS).

performed as a second stage only on less than 10% of the total region. This results in a $3.8 \times$ improvement in speed in addition to improvement in accuracy. We tested our algorithms on all five countries within the region of $1\,551\,997\text{ km}^2$ without retraining and showed that it is scalable as well as generalizable to varying structural, environmental, weather, and terrain conditions. The issue of scarcity of data is addressed by building a novel satellite image dataset that comprises images from the South Asian region for 14 different categories including 14 000 annotations for brick kilns. Consequently, with the help of Kiln-Net, we were able to generate an annotated dataset identifying and mapping brick kilns across a large swathe of the South Asian Brick-Kiln-Belt—a dataset that we believe will be extremely beneficial for future researchers. This dataset is available via.¹

The remainder of this article is organized as follows. In Section II, we provide a survey of related work while Section III provides an overview of the challenges associated with

automated detection of brick kilns using satellite imagery. This is followed by Section IV, which describes the proposed method based on neural networks, whereas results and evaluation are provided in Section V. Section VI conclude the article and also provides potential avenues for future research.

II. RELATED WORK

Due to recent advancements and availability of computational power, recent years have seen an ever-increasing utilization of deep neural network architectures [21], which surpass the performance of shallow architectures [22]. Deep neural networks learn manifold stages of representation and abstractions of high-dimensional data enabling one to train machine learning models without using hand-crafted features and with minimal preprocessing. In addition, the use of convolutional layers in these architectures, especially for image data, has given rise to what are called CNNs [23]. CNNs help in reducing the computational complexity as well as enabling the network to take into account the spatial context of imaging data. The great potential of artificial intelligence and remote sensing toward social science research and development purposes has been noted by many [2], [13], [24], [25]. Indeed, neural network-based machine learning algorithms mimicking brain-like behavior are among the most successful in solving such problems [2].

One area that has taken advantage of the power of such neural network-based machine learning methods is poverty studies. For instance, machine learning has been used to identify poverty stricken populations and their geographic location through the analysis of high-resolution satellite images [26]. The work is based on the premise that night-time luminosity is strongly correlated with economic activity and development. Using satellite images, such efforts compare daytime and night-time images of a specified location (e.g., a country or a continent) to identify which factors are associated with greater night lighting e.g., paved road networks. Classification of proposals for predicting socio-economic indicators along with their methods are summarized in Table I.

Landuse *classification* of satellite imagery into known classes such as road, houses, vegetation, etc. particularly using high-resolution imagery is another area that has greatly benefited from the power of CNNs [27]–[32] and residual neural networks (RNNs) [33]. Among these architectures, RNNs [34] based approach [33] have been shown to produce favorable results for the problem of landuse classification. Similarly, Mask-R-CNN [4] is another such algorithm that is used 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. Likewise, 3-D residual network [35] has been proposed for spatio-temporal analysis of remote sensing data. On the other hand, discriminative CNNs (D-CNNs) [36] have been developed to boost the performance of scene classification in remote sensing images by using different loss functions. It allows to map same-scene images closer to each other and different-scene images further apart.

The state-of-the-art *object detection* networks have been playing an important role in the field of remote sensing for

¹Online. [Available]: <https://cvlab.lums.edu.pk/kiln-net/>.

TABLE I
CLASSIFICATIONS OF PROPOSALS FOR PREDICTING SOCIO-ECONOMIC INDICATORS

Proposals	Indicators	Prediction	Method	Spatial Coverage
Foody, Giles M., et al. (2019) [9]	Brick Kilns	Slavery	Faster R-CNN	East of Jaipur, India
Misra, Prakhar, et al. (2019) [10]	Brick Kilns	Brick Kilns Locations	Transfer Learning + Pixel Classification	Delhi state surrounding areas, India
Nazir, Usman, et al. (2019) [11]	Brick Kilns	Slavery	CNN	Punjab, Pakistan and Punjab, India
Piaggesi, Simone, et al. (2019) [12]	Household Income	Poverty	Transfer Learning	5 cities in North and South America
Boyd, Doreen S., et al. (2018) [13]	Brick Kilns	Slavery	Crowdsourcing	Rajasthan, India
Xie, Michael, et al. (2016) [6]	Nighttime Light Intensity	Poverty	Transfer Learning	Uganda

quite a while and can ensure state security, construction, port management, cargo transportation, maritime rescue, and ship detection [37]–[41]. Similarly, [36] has proposed cloud detection from optical satellite imagery.

Recently, remote sensing images have also been used to analyze the extent of modern slavery [9], [10], [13], [42]. The “Slavery from Space” project [13] proposed a crowd-sourced procedure to manually detect brick kilns from satellite imagery. They randomly sampled the potential kiln areas into 320 cells of 100 km^2 each. However, they were only able to manually annotate 30 geographic cells (i.e., only 2% of the entire Brick-Kiln-Belt). As a result, the manual crowd-sourced method lacks scalability as is evident from the resulting annotated maps that are extremely sparse. On the other hand, low-resolution multispectral satellite data have also been used to classify brick kilns in the region surrounding the Delhi state in India [10]. The analysis in this work was based on low NDVI, which unfortunately is prone to generate a large number of false detections. In contrast, high-resolution satellite imagery has also been used to detect brick kilns to the east of Jaipur, which is the capital city of India’s, Rajasthan state [9]. This work used faster R-CNNs to automate the process of brick kiln identification in the given tile of images. However, owing to the large computational complexity, this approach is difficult to apply at a large scale. Moreover, it yields a very high false-positive rate for which they proposed to train a two-staged R-CNN classifier to achieve acceptable performance which further increased the processing time.

In this article, we propose a novel gated neural network-based method (Kiln-Net) that is composed of a classifier followed by an object detector. The first stage corresponds to an inexpensive classifier that attempts to identify brick kilns albeit with a high false positive rate. This is followed by an object detector that operates only on the images positively identified by the classifier. From among this subset, the object detector localizes detected brick kilns resulting in a coarse-to-fine strategy. It is due to this that our approach, in contrast to existing works, is scalable while at the same time effectively identifying brick kilns in the South Asia region with an accuracy of 99.96% and average F1 score of 0.91.

III. CHALLENGES AND SOLUTIONS

Brick Kilns are typically identifiable through a visual inspection of satellite imagery. However, while considering a large geographic area, several inherent complexities in satellite imagery make automated detection of brick kilns a challenging task. This includes, but are not limited to, first, variations in imaging sensors, second, differences in kilns’ structure across

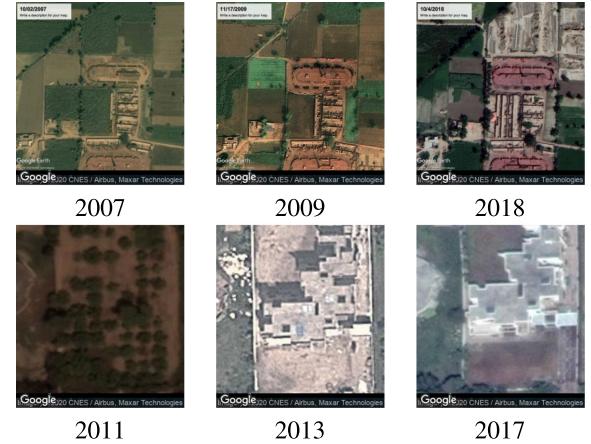


Fig. 3. Example satellite imagery from the same spatial location at different times showing (Row 1) sensor variations, (Row 2) sensor and environmental variations and dynamic surroundings. (Satellite images courtesy Google Earth).

the countries, third, dynamic kilns’ surroundings, and finally, variations in luminosity, seasonal changes, and pollution levels, etc. We discuss these challenges in detail in the following sections.

A. Sensor Variations

Satellite imagery is obtained through various active or passive remote sensors² installed on satellites. Well-known satellites include GeoEye’s GeoEye-1, DigitalGlobe’s WorldView-2, WorldView-3, QuickBird, Pleiades, RapidEye, and EROS. Sensor variations across the imaging devices on these satellites result in the same locations being captured differently over time as shown in Fig. 3. When spatial analysis is spread across multiple countries, these sensor variations manifest themselves as huge differences in the quality, resolution, and color profile of the imagery. To handle this challenge, we tried to incorporate the sensor variations by diversifying the training data across spatial locations. For the most part, our training data comprises annotated images from the Indo-Pak region. For example, the top image shown in Fig. 4(a) is from Punjab, Pakistan while the bottom two belong to New Delhi, India.

B. Structural Variations

Commonly used kilns in South Asia are continuous kilns (sometimes called tunnel kilns). Such kilns include the traditional fixed chimney Bull’s trench kiln, Hoffmann kiln, and natural draught zig-zag kiln [43]. These kilns may significantly

²Online. [Available]: <https://earthdata.nasa.gov/learn/remote-sensors>

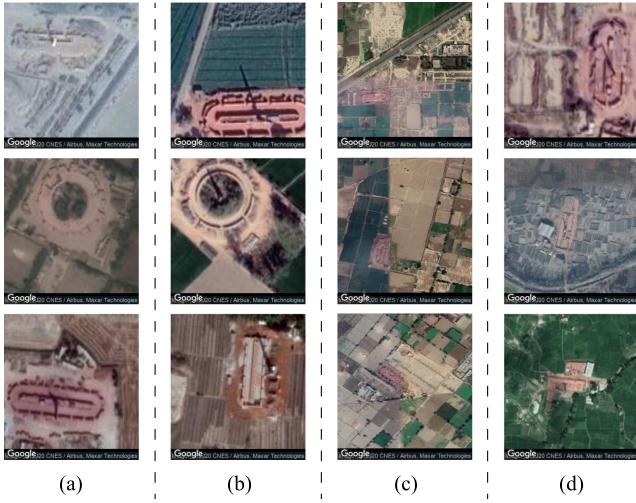


Fig. 4. Example satellite imagery of brick kilns from different spatial locations showing variation in quality, structure and color profile. (a) Sensor variations, (b) Structural variations. (c) Environmental variations. (d) Dynamic surroundings. (Satellite images courtesy Google Earth).

vary in structure and shape and are found to be of oval, circular, and rectangular shapes, respectively. These variations are illustrated in Fig. 4(b) in which the top image is of a traditional fixed chimney Bull's trench kiln from Pakistan. On the other hand, the middle image in Fig. 4(b) (row 2) is from Indian regions; it is circular in shape and is mostly likely a Hoffmann kiln. Finally, the bottom image of Fig. 4(b) shows a natural draught zig-zag kiln from India. These variations are usually enforced by geographical locations, environmental conditions, manufacturing technologies, building material, and local building regulatory authorities. These variations impose an obvious challenge for any machine learning approach which must learn a generic set of visual features for each type of kilns. Although, we used all shapes of kilns in our training dataset, our training set is dominated by oval-shaped kilns as they are the most commonly used. This may result in an unavoidable intraclass data imbalance.

C. 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 contribute to confusion in the classifier. An obvious effect of these changes is illustrated in Fig. 4(c). More specifically, the three images in the figure correspond to different luminosity levels, most likely indicating different times of the day when the images were captured. Given that the object detector, we utilize in Kiln-Net you only look once (YOLOv3 [3]) is known to be robust to changes in luminosity and other parameters, we expect our method to perform well even in the face of these variations.

D. Dynamic Surroundings

Brick kiln sites can have different features in their immediate surroundings. Examples include fields, barren lands, and

houses. These variations are illustrated in Fig. 4(d), which show images from Pakistan (top), Bangladesh (middle), and Nepal (bottom). One can observe these are surrounded by barren land, houses, and orchards, respectively. Previous literature [9], [10] attempted to solve the kiln/non-kiln problem as a binary classification. However, these variations in kilns' surrounding pose a significant challenge for binary classifiers resulting in deteriorated performance. To deal with these issues, we used a 14-class classification strategy so that nonkiln classes can be filtered out more effectively. We derive the list of these 14 classes from a manual observation in the Indo–Pak region. For the given images in this area, we note that the brick kiln sites are surrounded by one of these 14 classes: brick kilns, houses, roads, tennis courts, farms, grass, sparse forest, dense forest, parking lots, parks, ponds, tanks, mosques, and barren lands. Based on our experimental results, we were able to verify that this multiclass classification is better able to cater to the challenge of dynamic surroundings as compared to binary classification, thus justifying the choice of the former over the latter.

IV. PROPOSED METHOD

Existing pixel classification in low-resolution satellite imagery is based on spectral or texture properties. Spectral properties like NDVI, normalized difference built-up index, normalized difference moisture index, enhanced vegetation index, and burn area index have been extensively used in a wide variety of remote sensing problems [10], [44]–[46]. Although classification using these properties is usually computationally efficient, it is highly inaccurate due to inherent uncertainty of low-resolution remote sensing data [47]. On the other hand, while more precise measurements can be obtained from high-resolution satellite imagery, classification using these is significantly challenging due to the high memory footprint and processing requirement.

The high processing requirements can be mitigated by designing a coarse-to-fine search strategy whereby a significant portion of the data can be first filtered out using computationally efficient techniques, which may very well result in overestimation [48]. The resulting overestimated data can subsequently be further filtered using a more accurate and fine-grained localization strategy. For example, deep learning-based classification [49] on high-resolution satellite imagery is found to be 4x faster than object detection [20], [50], [51]. It is to be noted that a classifier, in contrast to object detection, will not provide precise localization information such as a bounding box. However, it can be used to quickly eliminate portions of data that is highly likely not to contain a brick kiln. For these reasons, we propose a two-stage neural network named as Kiln-Net (see Figs. 5 and 6) which performs classification followed by object detection to achieve the coarse-to-fine search of kilns. Using our proposed two-stage method, we first perform 14 class classification and generate potential candidates for brick kilns. Object detector is then used only on these filtered regions to generate precise bounding boxes around each kiln. This coarse-to-fine search strategy thus not only addresses the overestimation problem, but also results in lower computational overhead as compared to single-stage object detection methods.

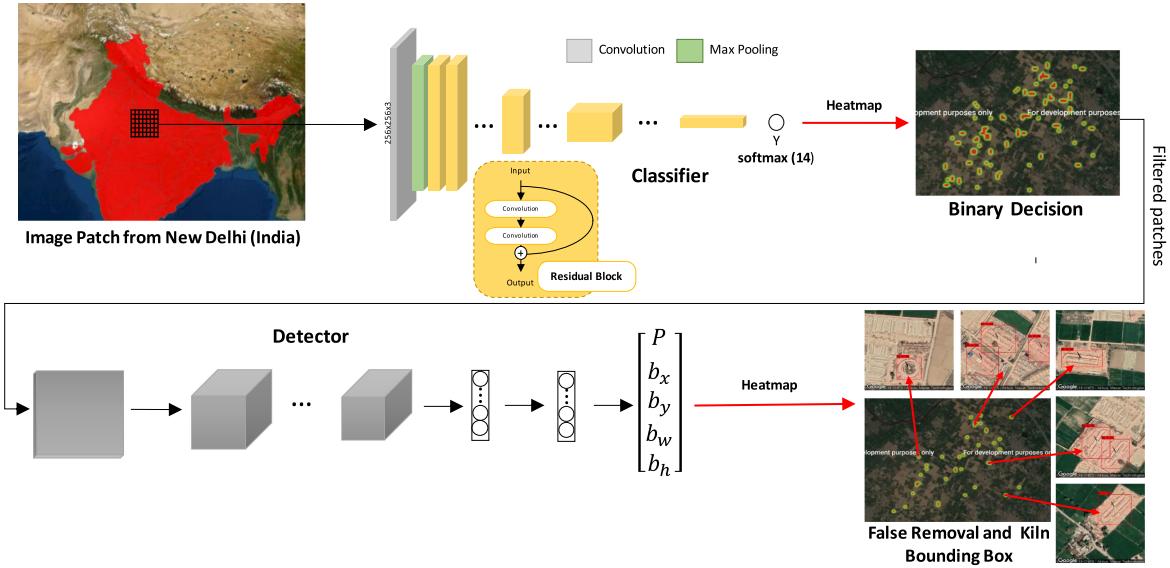


Fig. 5. Illustrative example of working of the proposed Kiln-Net architecture. (Input base map courtesy ArcGIS and satellite images courtesy Google Earth).

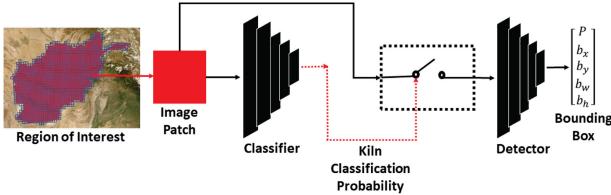


Fig. 6. Proposed gated neural network. (Input base map courtesy ArcGIS).

In the following sections, we first discuss the semi-automatic strategy adopted for annotating imagery for the 14 classes described earlier. This is followed by a discussion on patch-based classification and detection. Then, we discuss the criteria for selecting the combination of appropriate classifier and detector.

A. Dataset Generation

Unlike existing methods [10], [13] that analyze only one specific region, our goal in this work is to identify brick kilns across the entire South Asia. This requires a training dataset that contains samples under varying conditions as discussed in Section III. Furthermore, unlike street imagery, landuse is subject to significant variations in satellite imagery. To cater for this, we develop a 14-class dataset named *Asia14* using a two-stage data generation strategy. We first employed existing sources such as OpenStreetMap³ and obtained a small dataset consisting of Digital Globe RGB band images from 2016 and 2017 of resolution 256×256 at zoom level 20 (corresponding to 0.149 pixel per meter on the equator). We collected 300 images for each of the 14 classes including brick kilns, houses, roads, tennis courts, farms, grass, sparse forest, dense forest, parking lots, parks, ponds, tanks, mosques, and barren lands. The issue of sensor variations is handled by diversifying the training data across several spatial locations within the Indo-Pak region of

South Asia. In the second stage, we adopted a semi-automatic strategy for the generation of additional annotated data. We trained a ResNet-152 [34], which is a CNN-based multiclass classifier. Using this land use model, we automatically classified satellite imagery from 2017 of Punjab, Pakistan region of South Asia. An additional 700 annotations per class is then generated by selecting highly probable locations for each class. The resulting data of 14 000 annotated samples is then manually verified for any misclassifications. In addition, in order to generate training data for our object detector, we also manually annotated bounding boxes for each one of the 1300 brick kiln images in the dataset.

B. Kiln-Net: Gated Neural Network

Kiln-Net is a two-stage gated neural network. In the first stage, it filters the bulk of the data using an inexpensive classifier while the detector is only applied on small amounts of positive detections to generate localization information while filtering false positives. For the overarching goal of identifying and mapping brick kilns in South Asia, we utilized high-resolution satellite imagery patches from Google Earth. We first fed each patch of 256×256 pixels from the selected region to the classifier stage to get a corresponding binary decision for kiln detection. If kiln classification probability of patch is greater than a threshold value, we pass this patch to the detector using a gate as shown in Fig. 6. In order to make sure that the proposed classifier does not miss detecting a kiln with very high probability, we set a very small threshold value (0.1 in our experiments). While this does yield a very high false-positive rate in the first stage, the second object detector stage effectively filters out these falsely identified patches. An example illustration of this is shown in Fig. 5 for an image patch from New Delhi, India. The positively labeled outputs of the classifier could either be a true positive (TP) or false positive (FP), whereas the rest of the 13-class outputs are considered as nonkiln or true negative (TN). A heatmap is

³Online. [Available]: <https://www.openstreetmap.org/>

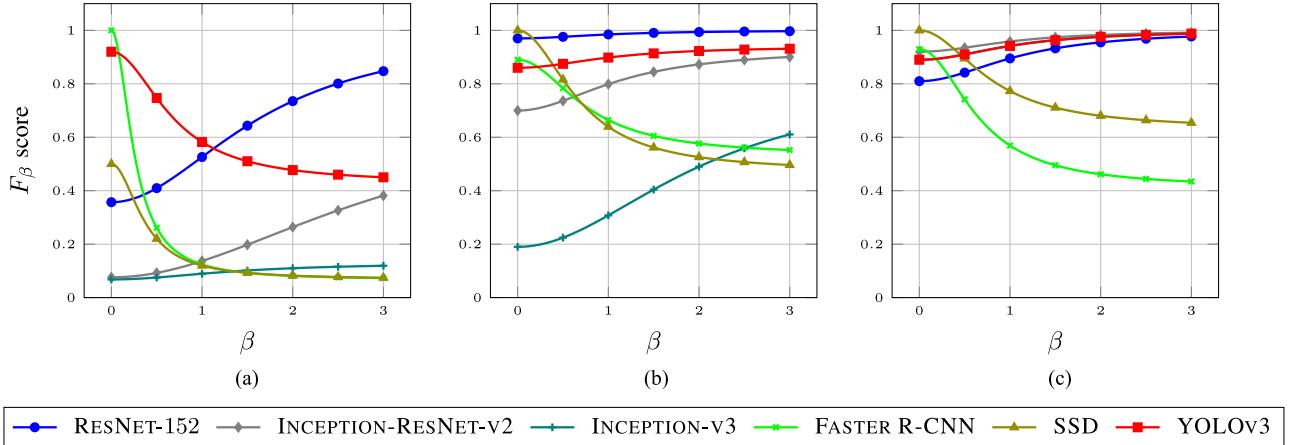


Fig. 7. F_β Measure for state-of-the-art classifiers and detectors. (a) Afghanistan. (b) Pakistan. (c) India.

shown to depict the probability of presence of kiln at specific geographical points. Areas of higher intensity (red) are kilns or (TP/FP) while the areas of lower intensity (green) belong to the nonkiln class TN. To verify the results of the classifier on TP and to improve the results on FP, the filtered patches (red spots on the heatmap) are input to the object detector, which predicts bounding boxes around what it thinks are brick kilns. The bounding box prediction has five components: (b_x, b_y, b_w, b_h, P) . The (b_x, b_y) coordinates represent the center of the box, relative to the grid cell location, and (b_w, b_h) coordinates represent the width and height of the box, respectively. On the other hand, the variable P represents the confidence score which is defined as the intersection over union (IOU) between the predicted box and the ground truth [3]. The detector removes the FP and predicts the bounding boxes for each kiln candidate as shown in Fig. 5.

C. Selection of Classifier and Detector

We employed a systematic approach based on F_β measure to select the most appropriate classifier and detector combination. The F_β score is the weighted harmonic mean of precision and recall, reaching its optimal value at 1 and its worst value at 0. The nonnegative parameter β determines the weight of recall in the combined score: $\beta < 1$ lends more weight to precision, whereas $\beta > 1$ favors recall ($\beta \rightarrow 0$ considers only precision, $\beta \rightarrow \infty$ only recall). The score is defined as

$$F_\beta = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \quad (1)$$

In our analysis, we used $0 \leq \beta \leq 3$ thus favouring recall. This helps in addressing class imbalance as we have a smaller number of Kiln samples in our dataset and recall is the measurement of coverage of the minority class [52].

We trained several classifiers and detectors on the proposed *Asia14* dataset. More specifically, we trained some of the state-of-the-art classifiers including ResNet-152 [34], Inception-ResNet-v2 [53], and Inception-v3 [54]. Similarly, among detectors we trained YOLO[51], single shot detector (SSD [20]), and

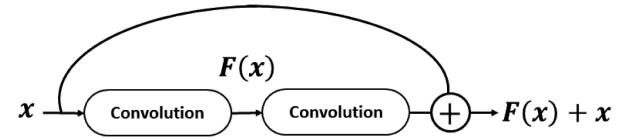


Fig. 8. Residual connection used in all variants as short skip connection.

faster R-CNN [50] and selected the best classifier and detector using the F_β score. This aspect is further elaborated in the following.

ResNet versus inception-based classifier: Residual [34] and inception [53] blocks are two popular improvements that were proposed to improve the performance of CNNs. Residual connections were introduced to address the issue of vanishing gradients in deeper networks. ResNet block [34], [53], [54] is a direct connection of a layer with a layer deeper in the network, skipping one or more layers (see Fig. 8) and can be defined as

$$\mathbf{r}(\mathbf{x}) = F(\mathbf{x}) + \mathbf{x} \quad (2)$$

where \mathbf{x} is input to the residual connection, $F(\mathbf{x})$ is residual mapping to be learned. Inception block [53] consists of convolutions of different sizes that allow the network to process features at different spatial scales, which are lumped and fed to the next layer for further processing.

It can be seen in Fig. 7 that the models based on inception block—Inception-ResNet-v2 and Inception-v3—have a lower F_β score as compared to ResNet-152 in almost all major regions (Pakistan, India, and Afghanistan). This indicates that the inception block does not help us in learning to see beyond domains and regions due to sensor and environmental effects on satellite imagery. Employing residual connections, as those employed in ResNet, assists us in reducing these effects. It is because of this that we choose ResNET-152 as the classification stage in Kiln-Net.

YOLO versus faster R-CNN versus SSD: Following the classification stage is that of the detector. The detector YOLO [51] is capable of performing both classification as well as regression of bounding boxes. It takes an input image and learns the

TABLE II
TABLE SHOWING DETAILS OF THE EVALUATION DATASET

Countries	Cities	Total Images	Area (km^2)	Top Left (Lat, Lon)	Bottom right (Lat, Lon)
Afghanistan	Deh Sabz	4224	41.25	34.64516, 69.25096	34.5944, 69.33046
	Surkh Rod	3584	34.99	34.44128, 70.30233	34.41003, 70.40979
	Kasur	3110	30.17	31.25986, 74.08455	31.22117, 74.15761
Pakistan	Multan	6848	66.87	30.1316, 71.3364	30.06226, 71.42692
	Faisalabad	181448	1769.86	31.60504, 72.95836	31.22606, 73.44482
	Mirpurkhas	16128	157.49	26.7966, 68.26489	26.74511, 68.34145
	Perumbedu	1280	12.50	13.36811, 80.2379	13.33868, 80.27313
India	New Delhi	4482	43.74	28.75963, 77.42762	28.7141, 77.51618
	Uttar Pradesh	11649	113.74	29.04141, 77.13713	28.97769, 77.29965
	Tamil Nadu	1536	14.99	13.24967, 80.22785	13.21921, 80.26849
	Barisal	63552	620.38	22.66222, 90.157	22.48017, 90.45498
Bangladesh	Sylhet	14720	143.74	24.88985, 91.91529	24.81484, 92.08395
	Dhaka	19548	190.88	23.87492, 90.11505	23.76341, 90.2654
Nepal	Kathmandu	5824	56.87	27.7198, 85.35858	27.6666, 85.4557

class possibilities with bounding box coordinates using IOU, nonmax suppression and anchor boxes. On the other hand, faster R-CNN [50] uses region proposal and performs regression of bounding boxes on each ROI. In contrast to faster R-CNN and YOLO, SSD [20] runs a CNN on input image only once and computes a feature map. Then, it runs a small 3×3 sized convolutional kernel on this feature map to predict the bounding boxes and distribution probability. As shown in Fig. 7, YOLOv3 [3] performs better than the faster R-CNN and SSD across all regions of South Asia. In Afghanistan, YOLOv3 shows an average F_β score of 0.6 for $\beta > 1$ and helps us in detecting kiln from bleak and barren lands. Consequently, the detector of our choice is YOLOv3.

V. RESULTS AND EVALUATION

In this section, we present the results of detailed experimentation of the proposed Kiln-Net architecture. While existing approaches are limited in spatial coverage (see Table I), the generalization of the proposed approach is demonstrated by performing analysis on 14 different cities from five different countries of South Asia. These include the cities of Deh Sabz and Surkh Rod from Afghanistan; Kasur, Multan, Faisalabad, and Mirpurkhas from Pakistan; Perumbedu, New Delhi, Uttar Pradesh, and Tamil Nadu from India; Barisal, Sylhet, and Dhaka from Bangladesh; and Kathmandu from Nepal. The details of the images used from each of these regions are shown in Table II. Satellite imagery is downloaded for year 2019 (Pakistan and India) and 2020 (Afghanistan, Nepal, and Bangladesh). The high-resolution satellite imagery (zoom level 20) from these locations is first divided into nonoverlapping patches of size 256×256 , each one of which is then passed through various neural network architectures. The scalability, efficiency, and quality of the proposed method is measured by providing performance comparisons with respect to six different state-of-the-art classifiers and detectors namely ResNet-152 [34], Inception-ResNet-v2 [53], Inception-v3 [54], faster R-CNN [50], SSD [20], and YOLOv3 [3].

A. Experimental Setup

We used the *Asia14* dataset for training of classifiers and 1300 brick kiln images with bounding box annotations for the training of object detectors (see Section IV-A). We divided the training data into 70% training, 20% validation, and 10% testing. The trainings were performed using the same hyperparameter values in all our classification experiments. All the networks

were optimized using stochastic gradient descent with an initial learning rate of 0.01 and a rate decay of 10^{-6} . Instead of using fixed number of epochs, we used early stopping criteria which terminates the training process in case there is no improvement for 10 consecutive epochs. All the experiments were conducted on a system with an Intel Core i7 – 7500 U CPU 2.70 \times 2 GHz processor, 8 GBs of RAM, and a NVIDIA GeForce GTX 950MX GPU.

B. Performance Metrics

The performance metrics used for performance comparisons are the usual accuracy, precision, recall and F1 scores. For the sake of completeness, these are defined below.

- 1) *Accuracy* is a metric that quantifies the percentage of correctly classified instances. It is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

- 2) *Precision* is a metric that quantifies the fraction of correct positive predictions made. It is given as

$$P = \frac{TP}{TP + FP}. \quad (4)$$

- 3) *Recall* is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. It is defined as

$$R = \frac{TP}{TP + FN}. \quad (5)$$

- 4) *F1* score provides a way to combine both precision and recall into a single measure that captures both properties. It is given as

$$F1 = \frac{2}{P^{-1} + R^{-1}} \quad (6)$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively.

C. Quantitative Evaluation

We evaluated our proposed Kiln-Net on unseen dataset consisting of almost 337933 zoom-20 images from different South Asian regions (see Table II) that are shown in Tables III, IV, and V.

Quantitative evaluation of ResNet-152 classifier in Kiln-Net: The performance evaluation of the first stage of Kiln-Net, i.e., ResNet-152 classifier is shown in Table III. In order to ensure that our classifier does not miss any kiln, we utilize a very low threshold of probability (beyond which a kiln is declared). However, this results in several misclassifications in the form of FPs (which is subsequently corrected by the detector stage). The classifier's lowest accuracy of 94.70 is obtained for Deh Sabz, Afghanistan whereas its accuracy reached approximately 100% for Faisalabad, Pakistan. The relatively low accuracy for Afghanistan is perhaps due to the fact that the region mostly corresponds to barren land which is under represented in the training data. On the other hand, the Faisalabad region is mostly fertile with much of the training data also corresponding to similar surroundings. The overall accuracy for each region remained

TABLE III
TABLE SHOWING QUANTITATIVE EVALUATION OF THE RESNET-152 CLASSIFIER IN PROPOSED KILN-NET ARCHITECTURE

Countries	Cities	Classification Score						Precision	Recall	F1 score
		TP	TN	FP	FN	Accuracy (%)				
Afghanistan	Deh Sabz	122	3882	220	0	94.70	0.36	1	0.53	
	Surkh Rod	51	3471	62	0	98.27	0.45	1	0.31	
	Kasur	39	3070	1	0	99.96	0.98	1	0.99	
Pakistan	Multan	70	6678	100	0	98.54	0.41	1	0.58	
	Faisalabad	206	181236	6	0	100	0.97	1	0.98	
	Mirpurkhas	12	16114	2	0	99.98	0.86	1	0.46	
India	Perumbedu	10	1280	4	0	99.69	0.71	1	0.83	
	New Delhi	42	4430	0	0	100	1	1	1	
	Uttar Pradesh	89	11555	5	0	99.95	0.95	1	0.97	
India	Tamil Nadu	14	1495	27	0	98.24	0.34	1	0.51	
	Barisal	38	63506	8	0	99.98	0.83	1	0.91	
	Sylhet	18	14676	26	0	99.82	0.41	1	0.58	
Bangladesh	Dhaka	16	19525	7	0	99.96	0.70	1	0.82	
Nepal	Kathmandu	36	5761	27	0	99.53	0.57	1	0.73	
Overall Score		763	336665	495	0	99.85	0.72	1	0.84	

TABLE IV

TABLE SHOWING QUANTITATIVE EVALUATION OF THE YOLOV3 DETECTOR (IN PROPOSED KILN-NET ARCHITECTURE) ON POTENTIAL KILN LOCATIONS (TP + FPs) GIVEN BY CLASSIFIER TO DECREASE FP RATE

Countries	Cities	Classification Score						Annotations
		TP	TN	FP	FN	Accuracy (%)		
Afghanistan	Deh Sabz	85	191	29	37	80.70	122	
	Surkh Rod	48	36	26	3	74.34	51	
	Kasur	19	21	0	0	100	19	
Pakistan	Multan	50	107	13	0	92.35	50	
	Faisalabad	206	6	0	0	100	6	
	Mirpurkhas	12	1	1	0	92.85	12	
India	Perumbedu	8	4	0	2	85.71	10	
	New Delhi	40	0	2	0	95.23	40	
	Uttar Pradesh	83	5	0	6	93.62	89	
India	Tamil Nadu	13	27	0	1	97.56	14	
	Barisal	34	7	1	4	89.13	38	
	Sylhet	16	21	5	2	84.09	18	
Bangladesh	Dhaka	15	6	1	1	91.30	16	
Nepal	Kathmandu	26	22	5	10	76.19	36	
Overall score		655	454	83	56	85.92	321	

above 96.38%, whereas the ResNet-152 classifier achieved an overall accuracy of 99.85% for the entire region. This high accuracy is due to the high TN rate which indicates that 14-class classification helped in putting nonkiln categories into their respective bins. Thus, despite the extremely relaxed threshold, the classification stage resulted in correctly filtering 99.68% of the data (3 36 665 samples), and only 0.32% of the patches (TP and FP) (1 058 samples) were passed to the next stage.

Quantitative evaluation of YOLOv3 detector in Kiln-Net: Analyzing the numbers presented in Table III, one finds that of the samples that reach the detector stage of Kiln-Net, approximately 47% are FPs. The YOLOv3 reduced the FP rate to only 7.64%, of which 68% are from the Afghanistan region. This can be explained by the fact that Afghanistan typically unfolds as a bleak and barren land in most satellite images (see Fig. 9). It is difficult to classify brick kiln from bleak land because the kilns and the surrounding land itself have similar texture properties. An additional reason is that samples from Afghanistan were also missing in the training set. However, our selected detector YOLOv3 performs better than all other state-of-the-art classifiers and detectors in Afghanistan. This can also be seen from the F_β measurement (see Fig. 7). The best performance of overall YOLOv3 achieved an overall accuracy of 85.92% (see Table IV). The proposed two-stage Kiln-Net when tested on approximately 3 300 km² region (3 37 933 image patches) from 14 different cities resulted in 83 FP and 66 FN, whereas it was able to correctly identify 655 TP (see

Table IV) and 3 37 129 TN (3 37 933 – 655 – 83 – 66). Thus, the proposed Kiln-Net obtained an overall accuracy of 99.96% and average F1 score of 0.91.

D. Performance Comparison With State-of-the-Art

We compared our proposed two stage methods with three other classifiers namely ResNet-152 [34], Inception-ResNet-v2 [53] and Inception-v3 [54]. Similarly, we also compared our methods with three other detectors namely YOLOv3 [51], SSD[20], and faster R-CNN [50]. These comparisons are shown in Table V for three cities namely Kasur (Pakistan), New Delhi (India) and Deh Sabz (Afghanistan). This table also shows the number of identified duplicate detection which are introduced when a particular kiln is partially visible in more than one image patch. Among classifiers ResNet-152 obtained best performance both in-terms of precision, recall and F1 score whereas YOLOv3 proved to be a best performing object detector. It can also be seen that the proposed method outperformed all the other methods on all three cities in terms of precision, recall as well as F1 score. The proposed Kiln-Net obtained 100% score on Kasur and New Delhi, however its precision, recall and F1 score dropped to 0.75, 0.70, and 0.72, respectively for Deh Sabz (Afghanistan). This is due to low contrast between kiln and surrounding barren land.

E. Compute Cost Comparison With State-of-the-Art

Our selected classifier: ResNet-152 takes around 50 ms per image and it is four times faster than Inception-ResNet-v2 and 68 times faster than Inception-v3 as shown in Table V. Table V also shows that faster R-CNN takes approximately 1 s per image as compared to only 50 ms taken by ResNet-152, and since due to our two-stage strategy, YOLOv3 only gets 0.32% of the data thus it can be concluded that the proposed architecture is almost 20x faster than existing literature [9].

1) *Compute Cost Comparison of Different Detectors:* Our detailed experimentation and evaluation intimate that directly using an object detector on a large dataset is computationally prohibitive. In our experiments, we noted that the average compute time per image of resolution 256 × 256 for faster R-CNN, SSD, YoloV3 is 1, 0.6 and 0.2 s, respectively. This indicates that the overall compute cost of these detectors over the entire

TABLE V
TABLE SHOWING QUANTITATIVE EVALUATION OF THE PROPOSED TWO STAGE NETWORKS WITH STATE-OF-THE-ART ARCHITECTURES

Testing Datasets	Network Architectures	Classification Score									
		TP	TN	FP	FN	Duplicates	Precision	Recall	F1 score	Time (seconds)	
Pakistan (Kasur)	ResNet-152	39	3070	1	0	20	0.88	1.00	0.99	154.5	
	Inception-ResNet-v2	28	3068	12	2	9	0.70	0.93	0.80	661.26	
	Inception-v3	17	3017	72	4	0	0.19	0.81	0.31	10506	
	Faster R-CNN	8	3094	1	7	0	0.89	0.53	0.67	3090	
	SSD	9	3091	0	10	0	1	0.47	0.64	1854	
	YOLOv3	18	3088	3	1	0	0.86	0.94	0.90	618	
	Two Stage Faster R-CNN	13	3090	1	6	0	0.93	0.68	0.79	195.5	
	Two Stage SSD	12	3090	1	7	0	0.92	0.63	0.75	179.5	
India (New Delhi)	Kiln-Net (Two Stage YOLOv3)	19	3091	0	0	0	1	1	1	162.5	
	ResNet-152	42	4430	10	0	2	0.81	1	0.90	224	
	Inception-ResNet-v2	113	4359	10	0	73	0.92	1	0.96	958.72	
	Inception-v3	57	4418	7	0	17	0.89	1	0.94	15232	
	Faster R-CNN	13	4449	1	19	0	0.93	0.41	0.56	4480	
	SSD	25	4442	0	15	0	1	0.63	0.77	2688	
	YOLOv3	40	4437	5	0	0	0.89	1	0.94	896	
	Two Stage Faster R-CNN	37	4441	1	3	0	0.97	0.93	0.95	276.1	
Afghanistan (Deh Sabz)	Two Stage SSD	38	4442	0	2	0	1	0.95	0.97	255.3	
	Kiln-Net (Two Stage YOLOv3)	40	4442	0	0	0	1	1	1	232.8	
	ResNet-152	122	3882	220	0	0	0.36	1	0.53	211.2	
	Inception-ResNet-v2	84	3078	1024	38	0	0.076	0.69	0.14	903.94	
	Inception-v3	16	3883	219	106	0	0.07	0.13	0.09	14361.6	
	Faster R-CNN	8	4102	0	114	0	1	0.07	0.12	4224	
	SSD	1	4163	1	59	0	0.5	0.02	0.038	2534.4	
	YOLOv3	52	4098	4	70	0	0.92	0.426	0.58	844.8	
Nepal	Two Stage Faster R-CNN	100	4097	5	22	0	0.95	0.82	0.88	553.2	
	Two Stage SSD	90	4094	8	32	0	0.92	0.74	0.82	416.4	
	Kiln-Net (Two Stage YOLOv3)	85	4073	29	37	0	0.75	0.70	0.72	279.6	

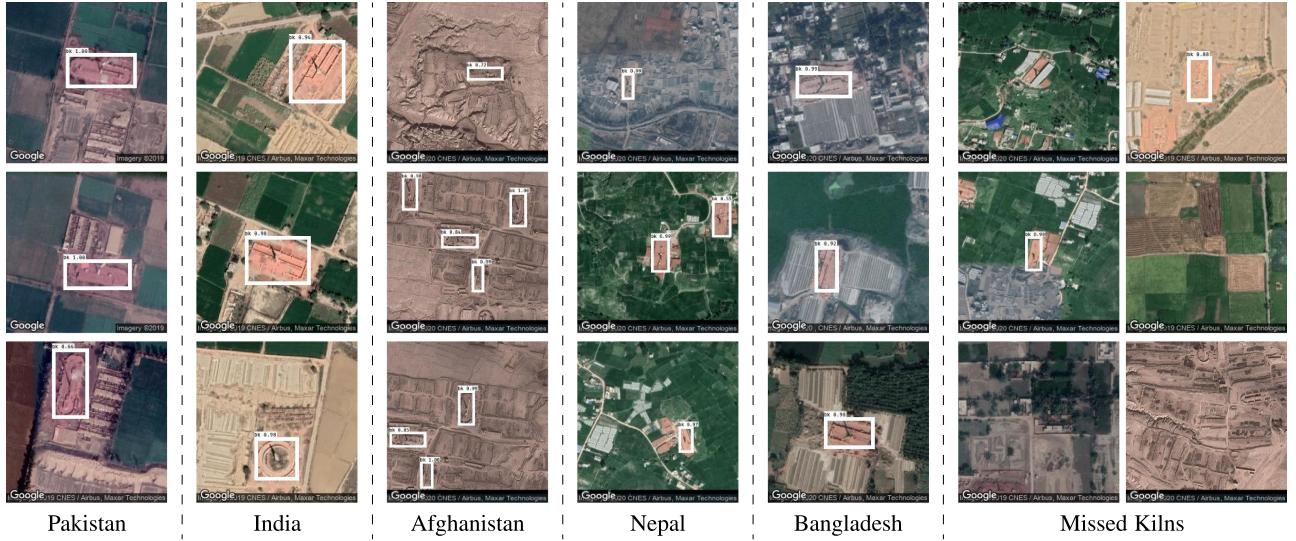


Fig. 9. Illustrative examples of our proposed Kiln-Net on South Asian countries. (Satellite images courtesy Google Earth).

Brick-Kiln-Belt of South Asia will be 1840, 1104, and 368 days, respectively as shown in Fig. 10. In other words, if we follow the SSD based approach it will take 1104 days as compared to only 95 days using the proposed two-stage strategy. Although any combination of detector and classifier can be used in the proposed two-stage strategy, we performed a systematic study on tradeoff between computational cost and accuracy as elaborated in Figs. 10 and 11.

F. Qualitative Evaluation

Fig. 9 shows the qualitative evaluation of our proposed Kiln-Net architecture. It can be seen that despite sensor, structural, dynamic surroundings, and environmental variations, the

proposed architecture has successfully identified all three types of kilns. Detection of traditional fixed chimney Bull's trench kiln is shown in Fig. 9 (column 1 and 3) for the Pakistan and Afghanistan regions. Afghanistan region exhibits extremely low contrast between kiln and nonkiln regions; however, the proposed Kiln-Net was able to correctly detect multiple kilns in each image. Similarly, Hoffmann Kiln and natural draught zig-zag kiln are detected for India, Bangladesh, and Nepal regions [Fig. 9 (column 2, 4 and 5)] along with a few Bull's trench kilns.

The FN or missed kilns are shown in columns 6 and 7 of Fig. 9. The FN in column 6 is due to either partial occlusion or abrupt sensor variation between neighboring pixels. Some of the FNs shown in the last column of Fig. 9 are due to low

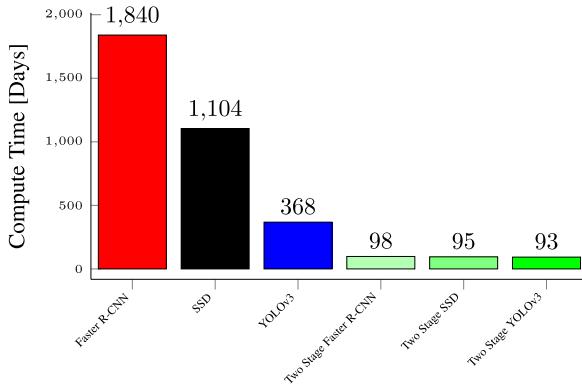


Fig. 10. Estimated compute cost of state-of-the-art detectors and our proposed two stage strategy on Brick-Kiln-Belt of South Asia. (Two-stage means ResNet-152 classifier followed by detector).

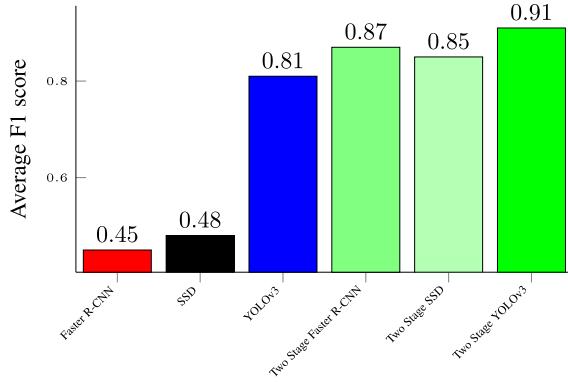


Fig. 11. Average F1 score of state-of-the-art detectors and our proposed two stage strategy on testing regions of South Asia. (Two-stage means ResNet-152 classifier followed by detector).

contrast between terrain and kiln pixels and one of the kilns is an abandoned kiln hence its appearance has significantly changed over time. The low-contrast images are from the Afghanistan region which is not represented in the training set.

VI. CONCLUSION AND FUTURE WORK

We proposed a gated neural network composed of ResNet-152 and YOLOv3 for identification and detection of brick kilns in South Asia. We tested our algorithm on a 3300 km^2 region (337933 image patches) from five countries including India, Pakistan, Afghanistan, Nepal, and Bangladesh. As compared to object detection-based strategy, our two-stage gated neural network is not only 20x faster, it also achieved 99.96% accuracy and average F1 score of 0.91 making it feasible for large-scale analysis. We also proposed a semi-automated method for dataset generation from satellite images and prepared a novel 14-class dataset consisting of 1000 images per class. Furthermore, the testing followed by semi-automated verification also contributed annotations for 337933 locations. Our proposed solution would not only enable regional monitoring and evaluation mechanisms for the SDGs, it can also be extended for other applications. Exact numbering of brick kilns and their locations is needed in order

to understand the brick sector's pollution and climate impacts and address black carbon and other short-lived climate pollutant emissions. We also aim to further reduce the computational cost by developing a hybrid solution using both low-resolution multispectral imagery and high-resolution satellite imagery.

REFERENCES

- [1] T. Landman and B. W. Silverman, "Globalization and modern slavery," *Politics Governance*, vol. 7, no. 4, pp. 275–290, 2019.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, Art. no. 436.
- [3] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.
- [5] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.
- [6] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 3929–3935.
- [7] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian process for crop yield prediction based on remote sensing data," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4559–4565.
- [8] S. Cotrufo, C. Sandu, F. Giulio Tonolo, and P. Boccardo, "Building damage assessment scale tailored to remote sensing vertical imagery," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 991–1005, 2018.
- [9] 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 Sens.*, vol. 11, no. 3, 2019, Art. no. 266.
- [10] P. Misra, W. Takeuchi, and R. Imausu, "Brick kiln detection in north India with sentinel imagery using deep learning of small datasets," *Proc. 40th Asian Conf. Remote Sens.*, 2019, pp. 2594–2601.
- [11] U. Nazir, N. Khurshid, M. Ahmed Bhimra, and M. Taj, "Tiny-Inception-ResNet-v2: Using deep learning for eliminating bonded labors of brick kilns in South Asia," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 39–43.
- [12] S. Piaggesi *et al.*, "Predicting city poverty using satellite imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 90–96.
- [13] 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 J. Photogrammetry Remote Sens.*, vol. 142, pp. 380–388, 2018.
- [14] S. Maithel, "Evaluating energy conservation potential of brick production in India," *SAARC Energy Centre, Islamabad, Pakistan*, 2013.
- [15] D. Mitra and D. Valette, "Environment, human labour & animal welfare," *Int. Lab. Org.*, Geneva, 2017. [Online]. Available: https://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_542925.pdf
- [16] L. Croitoru and M. Sarraf, "Benefits and costs of the informal sector: The case of brick kilns in Bangladesh," *J. Environmental Protection*, vol. 3, no. 06, 2012, Art. no. 476.
- [17] U. Manadhar and S. Dangol, "Study on evaluating energy conservation potential of brick production in SAARC countries: A report on Nepal," *SAARC Energy Centre, Islamabad, Pakistan*, 2013.
- [18] S. Nepal *et al.*, "A comparative study of stack emissions from straight-line and zigzag brick kilns in Nepal," *Atmosphere*, vol. 10, no. 3, 2019, Art. no. 107.
- [19] P. Suyesh and M. Soumen, "Vertical shaft brick kiln (VSBK) technology for small and medium brick entrepreneur and operational manual," 2019. [Online]. Available: http://www.ecobrick.in/resource_data/KBAS100050.pdf
- [20] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [22] D. Tian, "A review on image feature extraction and representation techniques," *Int. J. Multimedia Ubiquitous Eng.*, vol. 8, pp. 385–395, 01 2013.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

- [24] D. Blumberg and D. Jacobson, "New frontiers: Remote sensing in social science research," *Amer. Sociol.*, vol. 28, no. 3, pp. 62–68, 1997.
- [25] 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 *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 1217–1220.
- [26] 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.
- [27] I. Demir *et al.*, "DeepGlobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 172–17209.
- [28] S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, "Feature pyramid network for multi-class land segmentation," in *Proc. IEEE Conf. Comput. Visi. Pattern Recognit. Workshops*, Jun. 2018, pp. 272–2723.
- [29] A. Raklin, A. Davydow, and S. Nikolenko, "Land cover classification from satellite imagery with U-net and Lovasz-Softmax loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 257–2574.
- [30] C. Tian, C. Li, and J. Shi, "Dense fusion classmate network for land cover classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 262–2624.
- [31] T.-S. Kuo, K.-S. Tseng, J.-W. Yan, Y.-C. Liu, and Y.-C. Frank Wang, "Deep aggregation net for land cover classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 247–2474.
- [32] Q. Liu, R. Hang, H. Song, and Z. Li, "Learning multiscale deep features for high-resolution satellite image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 117–126, Jan. 2017.
- [33] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [35] M. A. Bhimra, U. Nazir, and M. Taj, "Using 3-D residual network for spatio-temporal analysis of remote sensing data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 1403–1407.
- [36] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.
- [37] K.-M. Kang and D.-J. Kim, "Ship velocity estimation from ship wakes detected using convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4379–4388, Nov. 2019.
- [38] C. M. Neagu *et al.*, "The role of remote sensing in combating terrorism and ensuring state security," *LESIJ-Lex ET Scientia Int. J.*, vol. 24, no. 1, pp. 97–104, 2017.
- [39] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.
- [40] F. Meyer and S. Hinz, "Automatic ship detection in space-borne SAR imagery," *Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 1, pp. 1682–1750, 2009.
- [41] U. Algancı, M. Soydas, and E. Sertel, "Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 458.
- [42] 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. Modern Slavery*, vol. 4, pp. 169–199, 2018.
- [43] R. Pokhrel and H. Lee, "Integrated environment impact assessment of brick kiln using environmental performance scores," *Asian J. Atmospheric Environ.*, vol. 8, pp. 15–24, 2014.
- [44] V. J. Pasquarella, C. E. Holden, and C. E. Woodcock, "Improved mapping of forest type using spectral-temporal landsat features," *Remote Sens. Environ.*, vol. 210, pp. 193–207, 2018.
- [45] B. Matsushita, W. Yang, J. Chen, Y. Onda, and G. Qiu, "Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: A case study in high-density cypress forest," *Sensors*, vol. 7, no. 11, pp. 2636–2651, 2007.
- [46] G. M. Gandhi, S. Parthiban, N. Thummala, and A. Christy, "NDVI: Vegetation change detection using remote sensing and GIS—a case study of vellore district," *Procedia Comput. Sci.*, vol. 57, pp. 1199–1210, 2015.
- [47] Q. Zhang and P. Zhang, "An uncertainty descriptor for quantitative measurement of the uncertainty of remote sensing images," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1560.
- [48] F. Fleuret and D. Geman, "Coarse-to-fine face detection," *Int. J. Comput. Vis.*, vol. 41, nos. 1/2, pp. 85–107, 2001.
- [49] Z. Shi *et al.*, "A deep CNN based transfer learning method for false positive reduction," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 1017–1033, 2019.
- [50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [51] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [52] H. He and Y. Ma, *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed. Hoboken, NJ, USA: Wiley, 2013.
- [53] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. Assoc. Advancement Artif. Intell.*, 2017, vol. 4, Art. no. 12.
- [54] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.



Usman Khalid Mian received the B.S. degree in computer science from the Lahore University of Management Sciences, Lahore, Pakistan, in 2020. He was part of a senior year research project which this research paper refers to. He has also constructed a model, "sentiment observer," for a course project which analyzes the sentiments of social media posts in real time. He has worked on a course project for the digital preservation of historical monuments as well.

His research interests include intersection of computer vision, remote sensing, and deep learning. He is passionate for building scalable machine learning systems and algorithms to learn socio-economic indicators from remote sensing data at large scale.



Muhammad Usman Sohail received the B.S. degree in computer science from the Lahore University of Management Sciences, Lahore, Pakistan, in 2020.

Soon after graduation, he joined Afiniti as a Data and AI Analyst. He has been a part of a senior year research project which this research paper refers to. During his research project, he earned experience in working on Geographic Information System. His research interests include machine learning, deep learning, data science, computer vision, and computer graphics. Apart from that, he is also interested in design, digital art, and AR/VR. His aim is to combine both technical and creative aspects of computer science in his professional career.



Usman Nazir received the B.S. degree in electrical engineering from COMSATS University Islamabad, Islamabad, Pakistan, in 2012 and the M.S. degree in electrical engineering from Lahore University of Management Sciences (LUMS), Lahore, in 2016. He is currently working toward the Ph.D. degree with the Technology for People Initiative Laboratory, Department of Computer Science, LUMS.

His research interests include intersection of computer vision, remote sensing, and deep learning. He is passionate for building scalable machine learning systems and algorithms to learn socio-economic indicators from remote sensing data at large scale.



Murtaza Taj received the Ph.D. and M.Sc. degrees in electronic engineering and computer science from the Queen Mary University of London, London, U.K., in 2009 and 2005, respectively.

He is currently an Assistant Professor with the Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences (LUMS), Lahore, Pakistan. He is also an Adjunct Faculty with the Ontario Tech University, Oshawa, ON, Canada. At LUMS, he is a Director of Computer Vision Laboratory (a research group within LUMS Computer Science Department) and a Director of Technology for People Initiative (a Research and Development Group at LUMS that develop solution to leverage technology to catalyze development in the public sector and improve data accessibility to facilitate good governance) His research interest include in the area of computer vision, graphics, and image processing. In particular, he is interested in detection and tracking of object in 2-D and 3-D scenes and in automatic generation of 3-D models from raw point cloud data.



Momin Uppal (Senior Member, IEEE) received the B.S. degree (Hons.) in electronic engineering from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Khyber Pakhtunkhwa, Pakistan, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 2006 and 2010, respectively. Since 2010, he has been associated with the Department of Electrical Engineering, Lahore University of Management Sciences, Lahore, Pakistan, where he is currently an Associate Professor. His research interests include signal processing for communications, nonorthogonal multiple access, coding for cooperative communications, prototyping using software-defined radios, and sensing using radio signals.