Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identifying patterns in urban housing density in developing countries using convolutional networks and satellite imagery

Rahman Sanya [*], Ernest Mwebaze

AI and Data Science Lab, Makerere University, Kampala, Uganda

ARTICLE INFO

Keywords:
Computer science
Housing classification
Urban areas
Developing countries
Convolutional neural networks
Satellite imagery

ABSTRACT

The use of Deep Neural Networks for remote sensing scene image analysis is growing fast. Despite this, data sets on developing countries are conspicuously absent in the public domain for benchmarking machine learning algorithms, rendering existing data sets unrepresentative. Secondly, current literature uses low-level semantic scene image class definitions, which may not have many relevant applications in certain domains. To examine these problems, we applied Convolutional Neural Networks (CNN) to high-level scene image classification for identifying patterns in urban housing density in a developing country setting. An end-to-end model training workflow is proposed for this purpose. A method for quantifying spatial extent of urban housing classes which gives insight into settlement patterns is also proposed. The method consists of computing the ratio between area covered by a given housing class and total area occupied by all classes. In the current work this method is implemented based on grid count, whereby the number of predicted grids for one housing class is divided by the total grid count for all classes. Results from the proposed method were validated against building density data computed on OpenStreetMap data. Our results for scene image classification are comparable to current state-of-the-art, despite focusing only on most difficult classes in those works. We also contribute a new satellite scene image data set that captures some general characteristics of urban housing in developing countries. The data set has similar but also some distinct attributes to existing data sets.

1. Introduction

Knowledge of spatial housing characteristics in an area has implications and applications in many domains including public and environmental health, utility and urban planning, and emergency or humanitarian disaster response. For example, the spread of airborne infectious diseases such as Tuberculosis (and Covid-19 is a good example too) is closely associated with human overcrowding in many kinds of places including homes. High population growth rates and rapid urbanization are causing housing shortage in many developing country cities. For example in Uganda, the housing shortage in major urban areas is estimated at over two million units and, is expected to reach three million by 2030 if not addressed. To put that into context, 47 per cent of households in Uganda comprising an average of 5 persons shared a bedroom in 2014, according to the Uganda National Bureau of Statistics (Uganda Bureau of Statistics, 2016). The housing shortage coupled with economic deprivation has led to sprawling of informal settlements with squalid living conditions in urban centers, posing serious public health risk (Ezeh et al., 2016; Elsey et al., 2016; Lilford et al., 2016; Riley et al., 2007).

The United Nations Sustainable Development Goals (SDG) goal number 11 urges cities to address existing housing inadequacies if they are to ever become sustainable (United Nations, 2015). We believe that a first step to addressing the problem of poor quality urban housing in developing countries is to understand how extensively it is spread across geographic space. To this end we propose the use of Deep Neural Network (DNN) methods and remote sensing imagery to identify patterns in urban housing quality. In this work, the phrase "housing quality" refers to the extent of building congestion (or crowding) per unit geographic area.

DNN are increasingly being applied for classifying remotely sensed land use scene imagery and have achieved high accuracy levels. Most of the existing literature in this area however, is focused on classifying scenes based on class definitions that may be considered semantically low-level. Secondly, majority of the remote sensing data sets used for this purpose relate to developed country settings where land use is generally carefully planned and regulated. In most developing countries however,

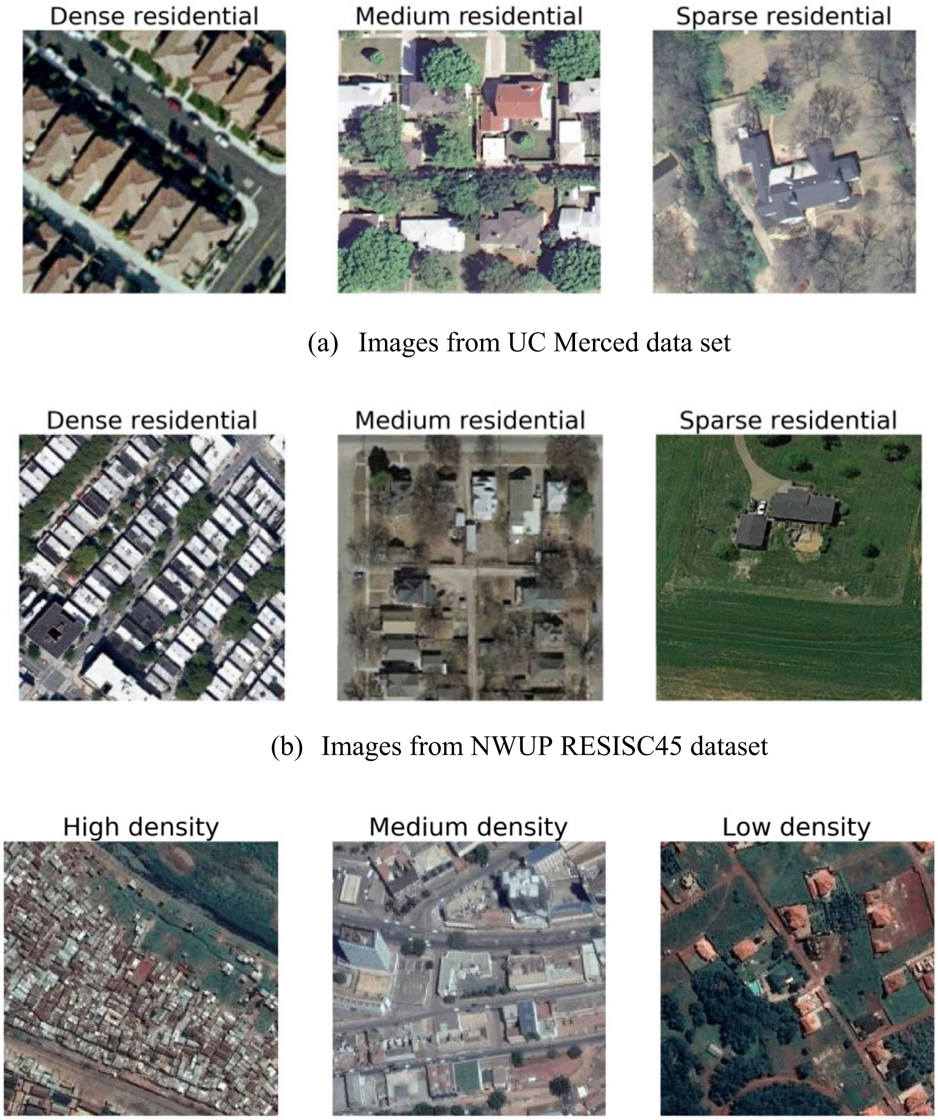
[*] Corresponding author.
E-mail address: hbasanya@gmail.com (R. Sanya).

https://doi.org/10.1016/j.heliyon.2020.e05617
Received 24 April 2020; Received in revised form 15 August 2020; Accepted 24 November 2020
2405-8440/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

a large proportion of land use for human purposes (for example, residential and commercial use) is informal and poorly regulated. A consequence of this is that scenes constituting semantically similar land use types vary significantly across developed and developing country settings. Figure 1 shows example images taken from two existing remote sensing land use data sets. In the top row are images from UC Merced data set (Yang and Newsam, 2010) and in the middle row are images from NWPU-RESISC45 data set (Cheng et al., 2017). In the bottom row are example images from our data set. As can be seen there are stark differences in what is considered, for example, dense residential or high density housing in the three data sets.

Two important questions arise from the scenario described in the previous paragraph. First, we think that classifying land use scenes comprising multiple geographical features using low-level semantic class definitions does not render the output to many relevant applications in certain domains. For example, classifying remotely sensed scene images as building, airplane, road, etc such as in (Cheng et al., 2017; Kang et al., 2018; Castelluccio et al., 2015; Basu et al., 2015) may not be of much relevant application in identifying neighborhoods at high risk of infectious diseases associated with indoor overcrowding. Secondly, existing DNN techniques have not been benchmarked on representative remote sensing scene image data sets. In particular, data sets from developing country settings have not been considered for benchmarking DNN algorithms due to absence of such data sets in the public domain. It is demonstrated in (DeVries et al., 2019) that the performance of machine learning algorithms will generally degrade when trained using geographically unrepresentative data sets.

In the current work we employed DNN methods for land use scene image classification in a developing country context to address some of the gaps identified above. Specifically, we used Convolutional Neural Networks (CNN) and satellite imagery for identifying housing density patterns in urban areas based on high-level semantic class definitions. Housing density is used here to mean the extent to which building units are congested or crowded. A method is also proposed for quantifying spatial extent of urban housing classes that provides useful insight into settlement patterns. The method involves computing the ratio between area covered by a particular urban housing class and total area occupied by all classes. In the current work we computed this ratio based on grid



(a) Images from UC Merced data set

(b) Images from NWUP RESISC45 dataset

(c) Images from our data set

Figure 1. Example scene images taken from three remote sensing data sets. Some differences can be seen between scenes constituting semantically similar land-use types across geographic regions. See, for example, images for the class Dense residential/High density.

count whereby we divided the number of grids predicted by the CNN model as belonging to one class by the total number of grids for all classes. Results from the proposed method were validated against building density data computed on OpenStreetMap data. Our results for scene image classification are comparable to current state-of-the-art, despite focusing only on most difficult classes in those works. We also showed that estimating spatial extent of urban housing classes using our method is at par with an alternative approach based on OpenStreetMap buildings data. A new satellite scene image data set that captures some general characteristics of urban housing in developing countries has been proposed. The data set has similar but also some distinct attributes to existing data sets. The contributions of our work therefore includes,

1. Our experimental results demonstrate that CNN are useful for identifying housing density patterns in developing country urban settings.
2. An end-to-end deep learning workflow based on fine-tuning pre-trained CNN models is proposed for identifying housing density patterns in urban areas of low-income countries.
3. A method for quantifying spatial extent of identified housing density classes has also been proposed. The method is comparable to an alternative approach based on OpenStreetMap buildings data.
4. We also contribute a first of its kind, relatively large satellite scene image data set based on a sub-Sahara Africa land-use for building. This high resolution data set is suitable for analysis of phenomena such as high resolution population distribution mapping. It will also be of benefit for benchmarking new machine learning algorithms.

The rest of this paper is organized as follows. Sections 2, 3, and 4 present related work, materials, and methods, respectively, while results are found in section 5. Discussion and conclusion and future work are provided in section 5, 6 and 7, respectively.

2. Related work

Although modern DNN have a fairly recent history, the volume of research utilizing these methods is growing rapidly. In this section we highlight some of the recent works that employ DNN, especially CNN, for semantic analysis of remotely sensed scene images.

Kang et al. (Kang et al., 2018) proposed a framework for classifying images of individual buildings based on functionality by utilizing CNN and street view images combined with remote sensing imagery. The study by Albert et al. (Albert et al., 2017) employed CNN to analyze urban physical environments across European cities based on ten classes derived from Urban Atlas land-use classification data set. Cheng et al. (Cheng et al., 2017) on the other hand reviewed recent progress in the field of aerial scene image classification, proposed a new data set, and then benchmarked state-of-the-art machine learning algorithms on it. Zhang et al. (Zhang et al., 2016) suggested a gradient boosting random Convolutional Network (GBRCN) framework that can be used to combine multiple different DNN for scene image classification. They report higher accuracy for their ensemble framework than methods that use single models. Jean et al. (Jean et al., 2016) posted high accuracy when they deployed CNN on remote sensing data to predict poverty in developing countries. Yuan (2018) developed CNN architecture with a final stage that integrates activations from multiple preceding stages for pixel-wise building footprint extraction from remote sensing imagery and geographic information systems (GIS) data. The work of Romero et al. (Romero et al., 2015) proposes a method for unsupervised deep feature extraction based on learning sparse features for aerial image classification. They demonstrate the superiority of deep architectures over shallow ones based on their method using the UC Merced data set (Yang and Newsam, 2010). Ajami et al. (Ajami et al., 2019) used CNN and very high resolution (VHR) images for identifying degree of deprivation in slums.

It is worth noting that all the above studies represent one research direction of utilizing aerial imagery directly for scene classification. However, another line of research uses the approach of object detection and extraction from aerial imagery for the same task of semantic scene classification. One of the earliest works along this line is that of Minh (Mnih, 2013), who developed a framework based on CNN for automated detection and labeling of roads and buildings from aerial images. Similar to this work is that of Vakalapoulou et al. (Vakalapoulou et al., 2015). Wurm et al. (Wurm et al., 2019) used CNN and remote sensing imagery for segmenting slums in satellite images. The work in (Yao et al., 2020) attempts to solve the problem of weakly supervised object detection (WSOD) from remote sensing imagery using only image-level annotations (no object location information is required) during model training. They propose a dynamic curriculum learning strategy that progressively learns an object detector by feeding training images of increasing difficulty that matches current detection capability. Further improvements to WSOD can be found in (Feng et al., 2020) and (Cheng et al., 2020). Other methods have also been explored for improving the discriminative capability of CNN. For example, in (Cheng et al., 2018; Cheng 2018) a method is proposed based on learning discriminative CNN for improving image scene classification. Effectiveness of the proposed method over state-of-the-art is demonstrated using multiple remote sensing benchmark data sets. Part-based CNN have also been suggested for discriminating between objects belonging to similar categories under the framework of fine-grained visual categorization (FGVC), see for example work in (Han et al., 2019; Zhang et al., 2014; Krause et al., 2014).

One of the currently persisting challenges in the field of semantic scene classification from remote sensing imagery using DNN is lack of large labeled data sets for evaluating algorithms. Some researchers suggest that existing data sets are too small in terms of total image count and number of classes and, they are saturated on algorithm accuracy, calling for development of new and larger data sets. This situation is said to limit development of new DNN algorithms. Current data sets available in the public domain include two large satellite image sets put together by Basu et al. (Basu et al., 2015), one of which we used in the present work to train a pre-trained model as a preceding stage to fine-tuning on our own data set. For the interested reader, a recent review of existing data sets can be found in (Cheng et al., 2017). The lack of labeled satellite scene image data sets for training ML algorithms is compounded by the huge human labor costs and inefficiencies associated with manual annotation of such images. To address this problem, a number of solutions are being proposed. For example, the work of Yao et al. (Yao et al., 2016) proposes a unified framework for automated semantic annotation of high resolution optical satellite images. The method combines discriminative high-level feature learning with weak, supervised feature transfer.

We would like to observe that despite the surging interest in scene classification using aerial imagery owing to its potential applications, no study has curated and/or utilized globally inclusive data sets. Of particular concern is the fact that scene image data sets from developing country settings are conspicuously absent in the public domain for evaluation and benchmarking of DNN algorithms. The goal of this research is therefore, to partly address this gap by suggesting a scene image data set extracted exclusively over a developing country. Our data set has some distinctive characteristics over some of the existing data sets. For example, it consists of three classes made up of geographical object type "building" that includes some unconventional built structures in addition to having extreme crowding in one class. As a use case, we applied our data set to the task of classifying and mapping housing density.

3. Materials

3.1. Study area

The geographical area of study is the country of Uganda, which lies between 10 29 South and 40 12 North latitude, 290 34 East and 350 0 East longitude. Uganda had a population of 35 million people in 2014 and covers geographical area of 241,551km^2 (Uganda Bureau of Statistics, 2016). The current analysis is based on twenty one districts (out of

113) that host the most populated urban centers. A map of the study area is shown in Figure 2.

3.2. Data sets

Our satellite scene image data set consists mostly of land-use type, "buildings". In other words, we are interested in building or housing built structures as geographical objects of interest. A land-use type refers to the activity or activities taking place on the land at a particular point in time (Shapiro, 1959). The data set has three classes namely high-, medium-, and low-density housing. These density classes were constructed based on spatial distribution of buildings in a uniform size area by grouping together images with similar building density, after eliminating all others. For this task we used a clustering tool developed using a pre-trained deep CNN that hierarchically clusters images based on content https://elcorto.github.io/imagecluster. The choice of housing density classes and their appearance in satellite scene images was guided by both intuition and works in the benchmark data sets shown in Figure 1. There was no need to label the images individually since we used a feature of Keras (Chollet, 2015) that extract class labels directly from directory names. Each composite image consists of red, green, and blue spectral bands of size 224 by 224 pixels resolution or 250m by 250m coverage for an image on the earth's surface. First of its kind from a sub-Sahara Africa setting, the data set is large on both per class (between 7,000 and 12,000 images) and total image count (31,000 images). It has high variability on object form, occlusion, orientation, cloud cover, background, and illumination. It also has high within-class diversity and between-class similarity i.e., classes in the data set overlap with each other, especially the high and medium density classes (see sample images in Figure 3). These characteristics make our data set more challenging compared to other similar data sets. Example images taken from the three housing density classes of our data set are shown in Figure 3.

As can be seen, there are observable differences in scene objects among images across the study area. For example, some housing structures in the northern region are small huts constructed of mud and wattle wall with grass thatch roof while all structures from the central region (as well eastern and western regions) have iron sheet roof. Table 1 provides details of image count by region of study area and by housing density class.

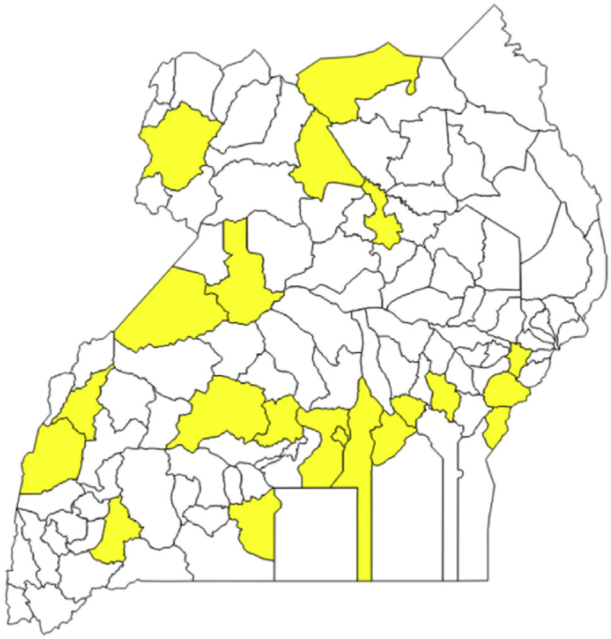


Figure 2. Map of Uganda showing twenty one districts (in yellow) over which satellite images were extracted for training a CNN model for identifying patterns in urban housing density. Within each district images were sampled over selected urban administrative units.

4. Methods

4.1. Sampling and satellite image acquisition

Our strategy to acquire high quality data samples for scene image classification was partly inspired by the work of Albert et al. (Albert et al., 2017). The strategy is based on use of free data sources for which we selected Google Maps Static API. This service allows free API requests per month up to a certain limit which was sufficient to generate enough images for our needs. Based on sample locations defined by latitude-longitude pair, image patches of size 224 by 224 pixels at zoom level 17 (1.2m per pixel spatial resolution or approx. 250m^2 coverage for a satellite scene image) were extracted. Our process for defining locations to extract images involved partitioning a shape-file polygon of each district into equal-sized grids of 250m^2. We then retrieved centroid information (longitude-latitude pair) of each grid and used it to extract satellite images. Data acquisition was carried out over a period of one month in August/September, 2018.

4.2. Model training

The task was to classify housing scene images into three classes, namely high density housing, medium density housing, and low density housing. We adopted a model development strategy common in the literature to ensure we got high classification results with minimal effort. Our strategy involved two comparable training modalities, 1) fine-tuning VGG16 and ResNet-50 directly on our data set and, 2) pre-training VGG16 and ResNet-50 on DeepSat data (Sat-6) (Basu et al., 2015) before fine-tuning on our data set as in (Albert et al., 2017). VGG16 (Simonyan and Zisserman, 2014) and ResNet-50 (He et al., 2016) are models that have been pre-trained on the ImageNet data set (Deng et al., 2009). We chose VGG16 and ResNet-50 for their superior and comparable performance on similar data sets as shown in recent literature, see (Cheng et al., 2017) and (Albert et al., 2017).

The model with highest validation accuracy during training under each modality was saved for evaluation. The strategy adopted in this work (of fine-tuning a pre-trained model) has many benefits. For example, it eliminates network architecture design time and requires less computational resources. It also provides better accuracy by order of magnitude over training a model from scratch, or extracting features from a data set using pre-trained models to be used for training a new classifier.

4.3. Model evaluation

We used multiple standard metrics in machine learning research to evaluate classification performance of our CNN model namely overall accuracy, AUC (area under Receiver Operating Characteristic Curve ROC), confusion matrix, precision-recall, and F1-score. Overall accuracy, often expressed as a percentage, is defined as the count of correctly classified samples (regardless of class they belong to) divided by total count of samples. The ROC is a graphical plot to illustrate the diagnostic capability of a classifier against varying discrimination threshold. The AUC is the percentage of area under ROC curve, which ranges between 0 and 1. The confusion matrix on the other hand summarizes classifier predictions with respect to individual classes. Precision measures a classifiers ability to label all samples correctly, recall is its ability to retrieve all positive samples while F1-score is a weighted average of precision and recall with best value at 1 and worst at 0. The evaluation protocol we used involved setting aside (hold-out) a validation set on which we evaluated the trained model.

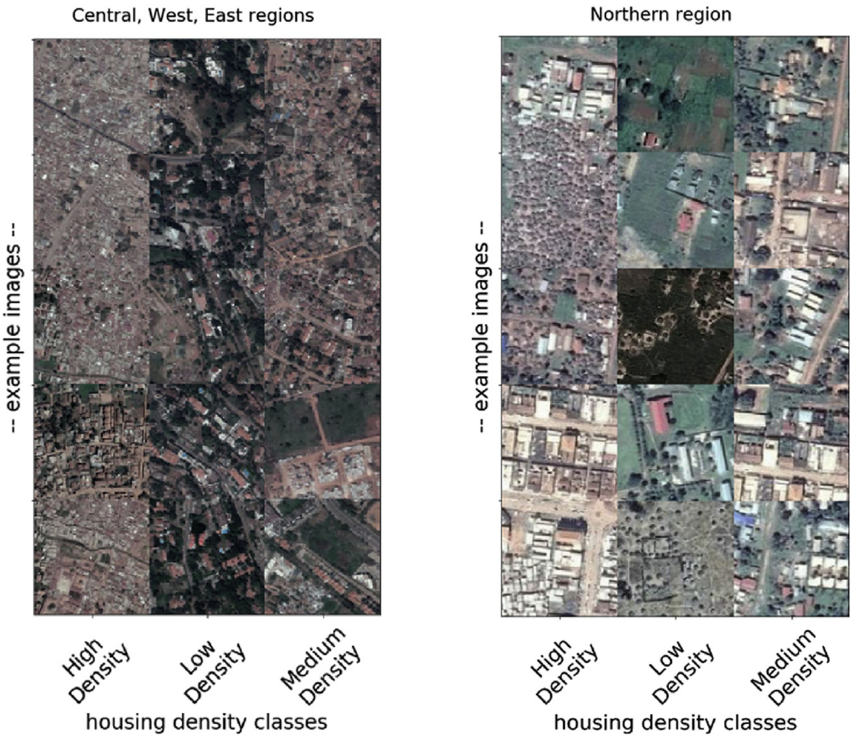


Figure 3. Example satellite scene images taken from the three housing density classes in our data set. Images from central, eastern, and western regions of the study area (left) have similar appearance of housing density classes while those from the northern region (right) have some unique building artifacts.

4.4. Experimental setup

Our experiments were implemented in Keras (Chollet, 2015), an open source deep learning framework, using TensorFlow (Abadi et al., 2015) as back-end. Common regularization techniques namely data augmentation and drop out were utilized during training to improve model generalization capability. Data augmentation techniques used include random rotation (15° maximum either direction), shearing (up to 0.1 radians), zooming (0.2), and horizontal/vertical flipping. The input images were of size 224 by 224 pixels composed of red, green, and blue (RGB) spectral bands. Adadelta (with variable learning rate) was used to optimize network loss function (categorical cross entropy) in all experiments. We used a ratio of 80:20 per cent to split our data set into training and test sets. A total of five train-test cycles were completed. The models were trained for at most 100 epochs. Our hardware set up is a remote virtual machine consisting of 1 GPU + 8 CPUs and 30 GB RAM.

4.5. Mapping urban housing patterns

We deployed the VGG16 model fine-tuned on our data set to predict housing density on fresh remote sensing data collected across the study area. Fresh data extraction was done over urban centers where training data was previously collected and, over other urban centers where training data was not collected. Housing density predictions were made on this new data set.

Class predictions for this new data were retrieved and used to create housing density maps, which are presented as raster maps for easy visual interpretation.

We adopted a qualitative approach to evaluate the predictions by visually analyzing and interpreting raster maps of predicted housing distribution patterns against ground truth data obtained from Google Static Maps API. We also estimated the spatial extent of each housing density class and present the results as percentage. This is done by dividing the area of land covered by a given housing density class by the total land area of all classes. Our method for estimating spatial extent (i.e., proportion of land area) e for a housing density class i was given by Eq. (1),

$$e_i = \frac{a_i}{\sum_1^k a} \tag{1}$$

where e_i is spatial extent for housing density class i, a_i is land area for class i, and $(\sum_1^k a)$ sums up land area for all classes from 1 up to k (in our case, $k = 3$).

In the current work we implemented a simple method for calculating e based on grid count. For example, to calculate spatial extent for high density housing e_h, we used Eq. (2),

$$e_h = \frac{\sum_0^k n_h}{\sum_1^k (n_h + n_m + n_l)} \tag{2}$$

Table 1. Satellite scene image count by regions of our study area and by housing density class.

Class/region	North	West	Central	East	Class Total
High density housing	814	567	8,111	984	10,476
Medium density housing	1,061	937	4,580	1,357	7,935
Low density housing	1,958	925	8,077	1,742	12,702
Regional Total	3,833	2,429	20,768	4,083	31,113

Table 2. Classification accuracy for CNN models based on VGG16.

Training modality	Accuracy (%)
VGG16 model fine-tuned directly on our data set	79.9
VGG16 model trained on DeepSat, fine-tuned on our data set	75.0

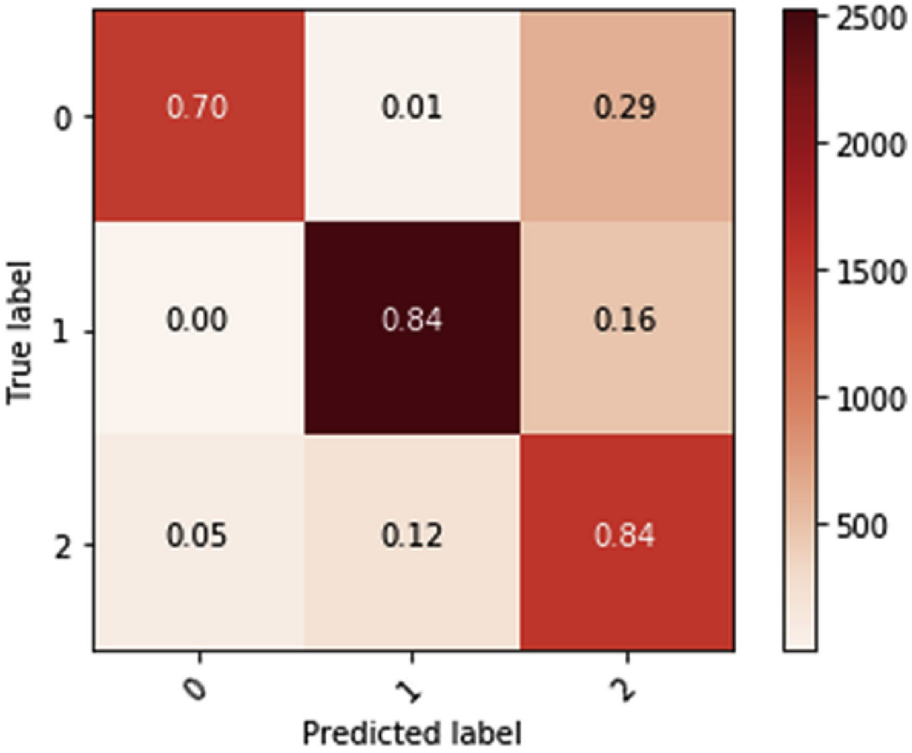


Figure 4. Confusion matrix plot for VGG16 model fine-tuned on our data set. Key: 0 for high-, 1 for low-, and 2 for moderate density housing.

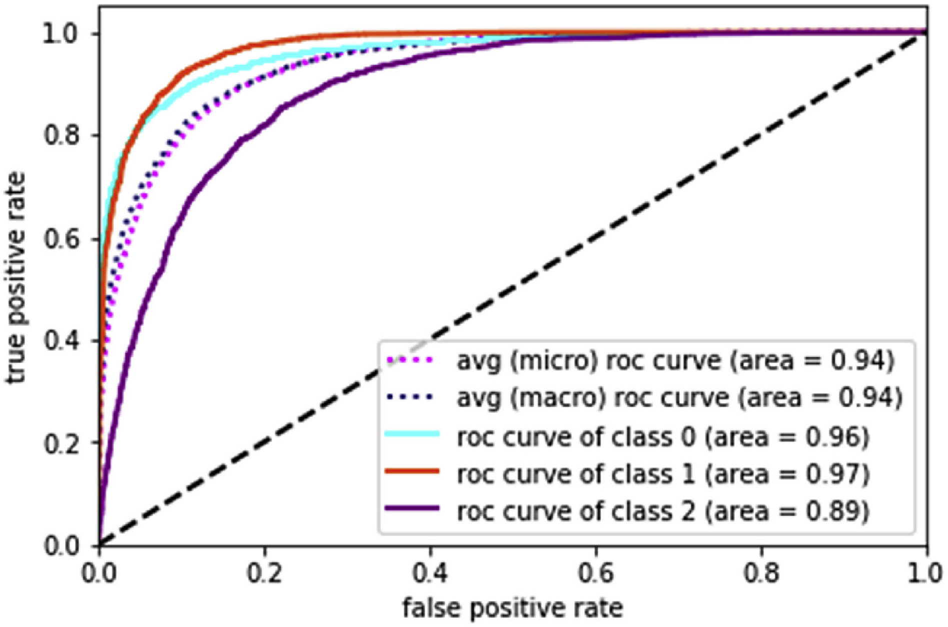


Figure 5. AUC curve for VGG16 model fine-tuned on our data set. Key: 0 for high-, 1 for low-, and 2 for moderate density housing.

Table 3. Precision, recall, and f1-score values for CNN model fine-tuned on our data set.

Class	Precision	Recall	F1-Score
High density housing	0.94	0.70	0.80
Low density housing	0.91	0.84	0.88
Medium density housing	0.59	0.84	0.69
Average/Total	0.84	0.80	0.81

where n_h is count for predicted high density housing grids, n_m is count for predicted medium density housing grids, and n_l is count for predicted low density housing grids. The value of e_h will range from 0 for no crowding, to 1 for completely crowded housing.

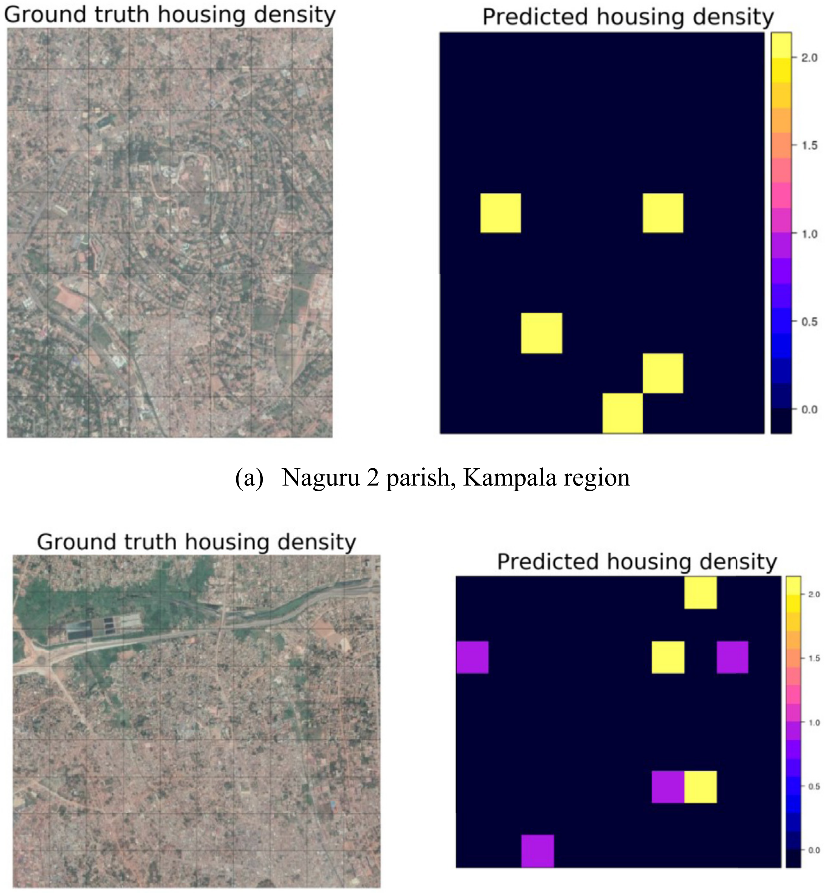
To demonstrate the effectiveness of our approach for quantifying spatial extent of urban housing classes, it was necessary to validate the results. For this purpose we used a different method for estimating building density B_d given by Eq. (3),

$$B_d = \frac{\sum_0^k b}{A} \tag{3}$$

where b is count of buildings in an area while A is size of the area in km^2.

To implement the method in Eq. (3), we used OpenStreetMap (OSM) Project's buildings data (OpenStreetMap, 2017) for Uganda. This data consists of Geographic Information Systems (GIS)-based vector data whereby buildings are represented using polygons. The OSM data is publicly available for free under the Open Database 1.0 License. We downloaded this data from the Geofabrik website at http://download.geofabrik.de as ESRI shapefiles. OSM data is updated on a daily basis hence, the analysis in this work used latest data as of August 5th, 2020. Pre-processing of the buildings shapefile involved transforming the Coordinate Reference System (CRS) from EPSG:4326 - WGS 84 - Geographic, whose units is degrees, to EPSG:32636 - WGS 84/UTM zone 36N - Projected, whose units is meters. This transformation was necessary for two reasons. To convert the CRS into relevant one for the geographical region of the world being analyzed (i.e., Uganda) and allow for calculation of area in square kilometers. Another pre-processing we performed involved clipping the buildings shapefile to our geographical units of interest based on relevant information from administrative boundaries shapefiles for Uganda. Both tasks were accomplished using QGIS (version 3.8.1-Zanzibar), an open source GIS software.

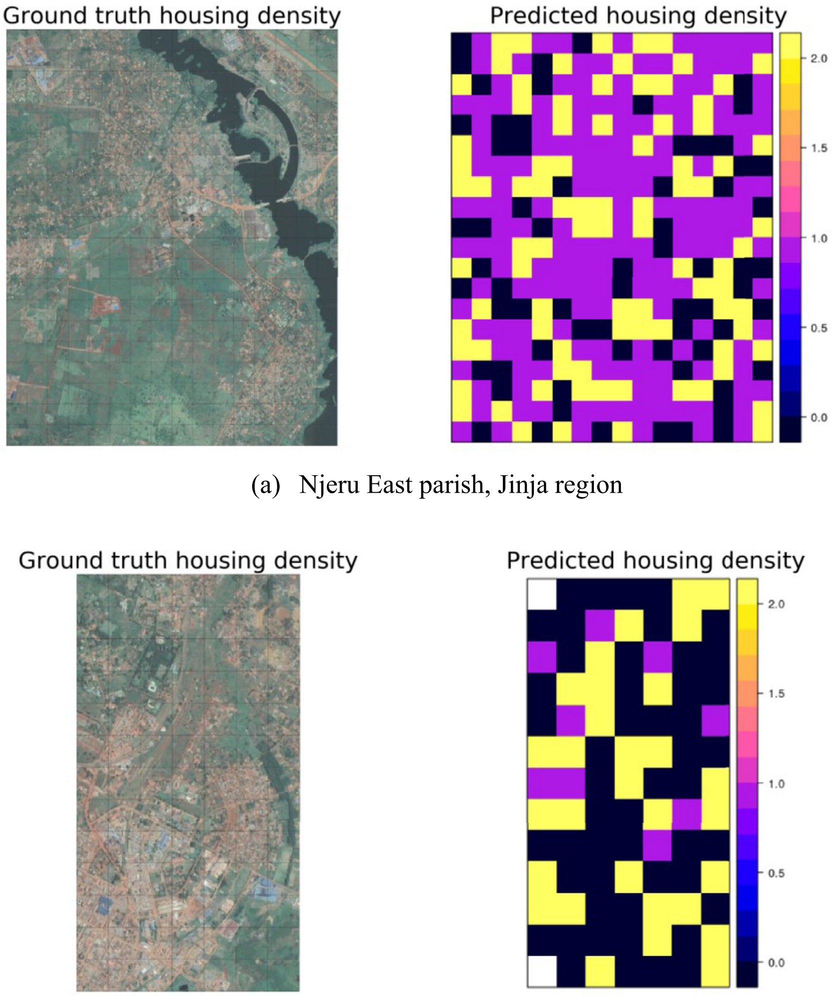
Calculating building density for a geographical region of interest based on Eq. (3) using OSM buildings data involved two steps. First, we counted all building polygons available in buildings shapefile of the region in question. Secondly, we divided the building polygon count by the area (in km^2) of that region.


Ground truth housing density
Predicted housing density
2.0
1.5
1.0
0.5
0.0
Ground truth housing density
Predicted housing density
2.0
1.5
1.0
0.5
0.0

(a) Naguru 2 parish, Kampala region

(b) Kasubi parish, Kampala region

Figure 6. Two regions of Kampala showing ground truth (Google imagery) and corresponding predicted housing density estimated using method in Eq. (1). Color key: deep blue (or 0) for high density housing area, purple (or 1) for low, and yellow (or 2) for medium.


Ground truth housing density
Predicted housing density
2.0
1.5
1.0
0.5
0.0
Ground truth housing density
Predicted housing density
2.0
1.5
1.0
0.5
0.0

(a) Njeru East parish, Jinja region

(b) Walukuba West parish, Jinja region

Figure 7. Two regions of Jinja showing ground truth (Google imagery) and corresponding predicted housing density estimated using method in Eq. (1). Color key: deep blue (or 0) for high density housing area, purple (or 1) for low, and yellow (or 2) for medium.

Table 4. Spatial extent (%) of housing density classes for selected parishes in Kampala and Jinja estimated using method in Eq. (1).

Class	Naguru 2	Kasubi	Njeru East	Walukuba West
High density housing	94	92	21	56
Low density housing	00	05	51	10
Medium density housing	06	03	28	34
Total	100	100	100	100

Table 5. Spatial extent of housing density classes in Kampala and Jinja estimated using method in Eq. (1).

Class/Proportion (%)	Kampala	Jinja
High density housing	70	80
Low density housing	16	14
Medium density housing	14	06
Average/Total	100	100

Table 6. Building density classes (column 2) used in method 2 (Equation 3).

Class	Density range	Population range
High density	>4,000	>40,000
Medium density	2,000–4,000	10,000–40,000
Low density	0-1,999	0–9,999

Table 7. Results of classifying some Kampala parishes using method 1 (Equation 2) and method 2 (Equation 3).

Parish	Method 1	Method 2
Kasubi	**High density**	**High density**
Naguru 2	**High density**	**Low density**
Bwaise 3	High density	High density
Kisenyi 1	Medium density	Low density
Kololo 2	Low density	Low density
Mulago 2	High density	High density

5. Results

5.1. Classification

Classification accuracy for the two training modalities is provided in Table 2. Since the ResNet-50 model performed worse on our data set than the VGG16 model, we report the better results only. As can be seen in the table, the VGG16 model fine-tuned directly on our data set gave better overall accuracy (79.9 per cent) on the test set than the VGG16 model pre-trained on DeepSat data prior to fine-tuning on our data set (75 per cent). Therefore, prediction and analysis of housing patterns presented in the next section is based on this model only. The confusion matrix and AUC plots for the VGG16 model fine-tuned on our data set are displayed in Figures 4 and 5, respectively. Table 3 provides details of precision, recall, and f1-Score values.

5.2. Housing density mapping and analysis

We present results of mapping and analyzing housing density patterns using both quantitative and qualitative approaches. To enable interpretation of results using the latter approach, we visualized housing density predictions as raster maps. Examples of ground truth (Google Static Maps API imagery) and predicted housing density patterns at $250m^2$ grid are shown in Figures 6 and 7.

Figure 6 (a) and (b) show two sites in Kampala while Figure 7 (a) and (b) are other two sites in Jinja. The ground truth images are overlaid with a grid layer (each grid is also $250m^2$ to aid in identifying corresponding grids on the predicted housing density raster. Our model predicts the two selected parishes in Kampala as predominantly high density housing places with more than 90 per cent. One region in Jinja is predicted to be low-density housing area (Njeru East parish, at 51 per cent) while Walukuba West is predicted to be high-density (56 per cent) with significant proportion of medium density housing (34 per cent). These results may be qualitatively understood by visually inspecting and interpreting the ground truth and predicted housing density raster maps. For example, the dominance of deep blue colored grids shows that the two parishes in Kampala are high density areas while dominance of purple colored grids suggests Njeru East parish is generally low-density housing area.

Results of quantitatively analyzing housing density patterns in the four parishes are shown in Table 4. On the other hand, Table 5 shows estimated proportion of each housing density class in Kampala and Jinja. As can be seen in the latter table, Kampala and Jinja are composed mostly of high density housing with 70 and 80 per cent, respectively.

5.3. Comparing housing classification methods

Here, we compare housing/building density classification generated using method 1 (Eqs. (1) and (2)) and method 2 (Equation 3). To help with the comparison, we generated building density classification for 22 (out of a total 89, thus representing 24%) of Kampala City parishes using method 2. We grouped results for the buildings classification into three classes: high, medium, and low building densities to correspond with the number and naming convention for classes used in method 1, as shown in Table 6.

Results of classifying some Kampala parishes are given in Table 7. Results for Kasubi and Naguru 2 (the two parishes referenced in Table 4 and Figure 6) are highlighted in bold text. It can be seen that Kasubi is classified as a high density housing/building area by both methods. However, Naguru 2 is classified as high density by method 1 but low density by method 2.

To give some context to the building density analysis, we provide population data derived from Uganda's 2014 census (Uganda Bureau of Statistics, 2016), for each of the 22 Kampala parishes. This is visualized as a scatter plot in Figure 8. This analysis shows that Kasubi is a high density building area, as well as a high population area. Naguru 2 on the

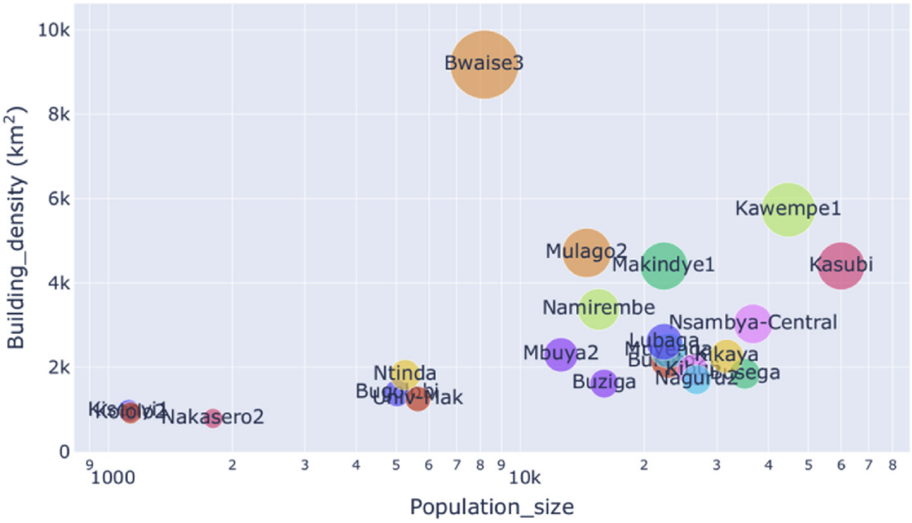


Figure 8. Building density vs. population size for 22 parishes of Kampala City.

other hand, is a low density building and medium population parish. These results (buildings/housing density predictions) are intuitive and consistent with ground truth data (Google images) for the two parishes in Figure 6.

6. Discussion

We had expected the VGG16 model pre-trained on DeepSat data prior to fine-tuning on our data set to outperform the one fine-tuned directly on our data set in conformity with results in (Albert et al., 2017), for the reason that the former data set has similar characteristics as ours. This was not the case however, suggesting there was no learning gain from pre-training the VGG16 model on the DeepSat data. At the moment we cannot understand why this is so, but speculate that it has something to do with differences in image size whereby our data set has image size 224 by 224 pixels while the DeepSat data has 28 x 28 pixel images. Image size might have implications for the amount and diversity of features available for learning to distinguish high level semantic objects as is the case in our work.

The classification accuracy we obtained (approx. 80 per cent) is not as high as state-of-the-art results (90 per cent) for this kind of data set. This however, can be understood given the high-level semantic class definitions used in our work which comprises only the most difficult three classes in the state-of-the-art (Cheng et al., 2017) and in (Albert et al., 2017). Despite this difficulty, our results are comparable to those in the two works. Another limitation of our work comes from subjective definition and construction of housing density classes, because a single scene image could potentially be placed in one or the other class, especially in the case of high- and medium density classes. This is manifested by the fact that most of the errors made by our model involve these two classes.

We also failed to distinguish the kind of neighborhood in a scene image based on building type/density for example, industrial, commercial, upscale residential, slum, etc, which would have been more informative in understanding current spatial patterns in building/housing distribution. For example, Zhang et al. (Zhang et al., 2017) showed that commercial areas tend to have higher density than other types of land use for building, even though this may not necessarily be the case in developing countries. We indirectly validated spatial extent of housing classes estimated using our method in Eqs. (1) and (2). This was done using an alternative approach based on OSM data. While the results of validation were promising, we could not establish how complete and accurate the OSM buildings data set for our study area was. Such a validation would have obviously benefited better from an established, suitable quantitative methodology and data set which currently are not available to us. Lastly, we have not evaluated our model's ability to generalize to data sets from other developing regions outside of the study area. In view of these and other limitations we may not have identified, we can only advise cautious interpretation of our results, especially the housing density maps and any analyses based on them.

With respect to estimating housing density from remote sensing data, work most similar to ours is that of Zhang et al. (Zhang et al., 2017) . The authors propose a method for estimating building density BD in a $240m^2$ (200 by 200 pixels) grid centered at (i, j) as shown in Eq. (4),

$$BD_{(i,j)} = \frac{S_{building}(i,j)}{S_{land}(i,j)} \tag{4}$$

where S_{building} is total area occupied by buildings in a given grid and S_{land} is total area size of the grid.

While the goal of their method is simply to estimate the total area occupied by buildings in a square grid, the goal of our method (Equation 1) is to estimate the area occupied by a specific housing density class in a much larger geographic region of interest. In our case this area is computed by summing up pixels that have been identified by the CNN model to belong to a particular housing density class based on features learned from raw images. In their case the total area occupied by buildings is derived from the sum of areas of polygons falling within boundaries of a grid.

7. Conclusion and future work

We set out to investigate CNN for high-level semantic scene image classification based on housing density prediction in developing countries as case study. The results we have obtained are encouraging given the challenging nature of the task owing to subjective class definition and data set construction.

Our contribution to existing body of knowledge is several folds. We make a contribution to the field of applied Machine Learning by deploying a CNN model to the task of identifying patterns in urban housing. For this task we proposed an end-to-end workflow for using a CNN model for housing density mapping based on fine-tuning pre-trained models on training data. Furthermore, the results obtained by a method we proposed for estimating spatial extent of housing density classes gives insight into current state of human settlement patterns in the study area, which knowledge is useful for urban planning and development, among other uses. Lastly, our work contributes a new remote sensing data set on spatial housing patterns that will be of benefit for benchmarking machine learning algorithms. As future work, we plan to investigate implications of current housing characteristics in the study area together with other related factors on urban phenomena such as infectious disease distribution in geographic space.

Declarations

Author contribution statement

R. Sanya: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

E. Mwebaze: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

R. Sanya was supported by the African Development Bank Higher Education Science and Technology (AFDB-HEST), Uganda.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

R. S would like to thank the African Development Bank Higher Education Science and Technology (AFDB-HEST) project for financial support towards his PhD study.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., 2015. TensorFlow: large-scale machine arXiv preprint.
Ajami, A., Kuffer, M., Persello, C., Pfeffer, K., 2019. Identifying a slums' degree of deprivation from VHR images using convolutional neural networks. Rem. Sens. 11 (11), 1282.

Albert, A., Kaur, J., Gonzalez, M.C., 2017. Using convolutional networks and satellite imagery to identify patterns in urban environments at large scale. In: ACM SigKDD 2017 Conference.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. DeepSat: a learning framework for satellite imagery. In: SIGSPATIAL '15: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM Digital Library, pp. 1–10.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land Use Classification in Remote Sensing Images by Convolutional Networks. CoRR.

Cheng, G.L., 2018. Exploring hierarchical convolutional features for hyperspectral image classification. IEEE Trans. Geosci. Rem. Sens. 6712–6722.

Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: benchmark and state-of-the-art. In: Proceedings of the IEEE, 105. IEEE, pp. 1865–1883.

Cheng, G., Yang, C., Yao, X., Guo, L., Han, J., 2018. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative cnns. IEEE Trans. Geosci. Rem. Sens. 2811–2821.

Cheng, G., Yang, J., Gao, D., Guo, L., Han, J., 2020. High-quality proposals for weakly supervised object detection. IEEE Trans. Image Process. 29 (2020), 5794–5804.

Chollet, F., 2015. Keras. GitHub.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image dataset. IEEE Int. Conf. Comp. Vis. Patt. Recogn. 248–255.

DeVries, T., Misra, I., Wang, C., van der Maaten, L., 2019. Does Object Recognition Work for Everyone? CoRR abs/1906.02659.

Elsey, H., Manandah, S., Sah, D., Khanal, S., MacGuire, F., King, R., et al., 2016, September. Public health risks in urban slums: findings of the qualitative 'healthy kitchens healthy cities' study in kathmandu, Nepal. (M. Kirk, Ed.). PloS One 11 (9).

Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y.-F., Ndugwa, R., Sartori, J., 2016. October). The history, geography, and sociology of slums and the health problems of people who live in slums. Lancet 389 (10068), 547–558.

Feng, X., Han, J., Yao, X., Cheng, G., 2020. April). Progressive contextual instance refinement for weakly supervised object detection in remote sensing images. IEEE Trans. Geosci. Rem. Sens. 58 (11), 8002–8012.

Han, J., Yao, X., Cheng, G., Feng, X., Xu, D., 2019. August). P-cnn: Part-based convolutional neural networks for fine-grained visual categorization. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.

Jean, N., Burke, M., Xie, M., Davis, M.W., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Sciencemag 790–794.

Kang, J., Korner, M., Wang, Y., Taubenbock, H., Zhu, X.X., 2018. November). Building instance. ISPRS J. Photogrammetry Remote Sens. 145 (Part A), 44–59.

Krause, J., Gebru, T., Deng, J., Li, L., Fei-Fei, L., 2014. Learning features and parts for fimne-grained recognition. In: 22nd International Conference Pattern Recognition, pp. 26–33.

Lilford, R.J., Oyebede, O., Satterthwaite, D., Melendez-Torres, G.J., Chen, Y.-F., Mberu, B., 2016, October. Improving the health and welfare of people who live in slums. Lancet 389 (10068), 559–570.

Mnih, V., 2013. Machine Learning for Aerial Image Labeling.

OpenStreetMap, 2017. Planet Dump.

Riley, L.W., Ko, A.I., Unger, A., Reis, M.G., 2007. March). Slum health: diseases of neglected populations. BMC Int. Health Hum. Right 7 (2).

Romero, A., Gatta, C., Camps-Valls, G., 2015. Unsupervised Deep Feature Extrac-. IEEE.

Shapiro, I.D., 1959. Urban land use classification. Land Econ. 35 (2), 149–155.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR.

Uganda Bureau of Statistics, 2016. The National Population and Housing Census 2014 - Main Report. Uganda Bureau of Statistics, Kampala.

United Nations, 2015. Sustainable Development Goals.

Vakalapoulou, M., Karantzalos, K., Komodakis, N., Paragios, N., 2015. Building Detection in Very High Resolution Multispectral Data with Deep Learning Features. IEEE IGARSS, pp. 1873–1876.

Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenbock, H., 2019, April. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS J. Photogrammetry Remote Sens. 150 (2019), 59–69.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. *GIS '10*. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM Digital Library, pp. 29–270.

Yao, X., Feng, X., Han, J., Cheng, G., Guo, L., 2020, May. Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning. In: IEEE Transactions on Geoscience and Remote, pp. 1–11.

Yao, X., Han, J., Cheng, G., Qian, X., Guo, L., 2016. Semantic annotation of high-resolution satellite images via weakly supervised learning. In: IEEE Transactions on Geoscience and Remote Sensing, 54. IEEE, pp. 3660–3671.

Yuan, J., 2018. November). Learning building extraction in aerial scenes with convolutional networks. IEEE Trans. Patt. Anal. Mach. Int. 40 (11), 2793–2798.

Zhang, F., Du, B., Zhang, L., 2016. Scene classification via gradient boosting random convolutional network framework. IEEE Trans. Geosci. 54 (3), 1793–1802.

Zhang, F., Du, B., Zhang, L., 2017. A Multi-Task Convolutional Neural Network Fo Rmega-City Analysis Using Very High Resolution Satellite Imagery and Geospatial Data. CoRR abs/1702.07985.

Zhang, N., Donahue, J., Girshick, R.B., Darrell, T., 2014. Part-based r-cnns forfine-grained category detection. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. 8689. Springer.