The Egyptian Journal of Remote Sensing and Space Sciences 25 (2022) 569–577

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Instance segmentation scheme for roofs in rural areas based on Mask R-CNN

Mark Amo-Boateng [a,b,c], Nana Ekow Nkwa Sey [a,c,*], Amprofi Ampah Amproche [a,c], Martin Kyereh Domfeh [a,b,c]

[a] Earth Observation Research & Innovation Centre (EORIC), University of Energy & Natural Resources, Sunyani, Ghana
[b] Civil & Environmental Engineering Department, University of Energy & Natural Resources, Sunyani, Ghana
[c] Regional Centre for Energy & Environmental Sustainability (RCEES), University of Energy & Natural Resources, Sunyani, Ghana

ARTICLE INFO

Article history:
Received 26 October 2021
Revised 31 January 2022
Accepted 29 March 2022
Available online 7 April 2022

Keywords:
Deep learning
Mask R-CNN
Instance segmentation
Rural rooftops
Object detection

ABSTRACT

Rooftop detection has numerous applications such as change detection in human settlements, land encroachments, planning routes to rural areas and estimation of solar generation potential of cities. Detecting the number, type and shape of building roofs form part of preliminary procedures to perform a variety of tasks for making decisions. Assessment of rooftops in rural areas is an important task in the estimation of potential solar generation and sizing of solar PV systems which has proven to be very challenging due to the different quality, lighting conditions and resolution of aerial and satellite images. In this research, we implement a mask RCNN algorithm using TensorFlow Object Detection API to detect the rooftop of buildings in a typical rural settlement. The average precision and recall values (@ IoU = 0.5:0.05:0.95) of the trained model were 85% and 88.2% respectively. The results of the experiment show that the approach can effectively and accurately detect and segment rural rooftops from high-resolution aerial images.

1. Introduction

Buildings are a prominent man-made feature in high-resolution satellite/aerial images. Rooftop detection has numerous applications such as change detection in human settlements, land encroachments, planning routes to rural areas and estimation of solar generation potential of cities. Almost 1.3 billion people are still without electricity in Sub-Saharan Africa and Asia. The main obstacle to universal access to electricity is the electricity supply to rural areas which are not connected to the electrical grid (Arranz-Piera et al., 2018). In Sub-Saharan Africa, more than 620 million people are living without electricity. This region is the only place in the world where the majority of people are living without electricity (Alfaro et al., 2017; Azimoh et al., 2016; Eder et al., 2015). Kantro, Boreso and Henekrom, rural communities in the Bono Region of Ghana are examples of such areas still without electricity. Therefore, assessing the unique roof types in a typical Ghanaian rural community is an essential task in the estimation of solar generation potential as a viable alternative.

Detecting the number, type and shape of rooftops of buildings form part of preliminary procedures to conduct for a variety of other tasks for decision making. Assessment of rooftops in rural areas is an important task in the estimation of solar generation potential and sizing of solar Photovoltaic (PV) systems. This task has proven to be very challenging due to the different quality, lighting conditions and resolution of aerial and satellite images. Another reason why rooftops are not easily spotted is that they have complex shapes, sizes and colours which can be easily confused with features such as dusty and unpaved roads.

The use of remote sensing technology for automatic detection of buildings has been widely acknowledged as an effective method to provide timely and useful data on various buildings in large areas (Vetrivel et al., 2018). Using high-resolution images from aerial drones shows high detailed features of rural buildings which facilitates accurate detection of the location of different rooftops (Fernandez Galarreta et al., 2015). A computational solution for rooftop detection is a process utilized for processing images with complex algorithms. The basic approach is to identify true roof types in rural areas by using various image segmentation techniques. Machine learning methods have become widely used in recent studies. Building detection and segmentation methods such as Support Vector Machine (SVM), Maximum Likelihood Classifier (MLC) and Random Forest (RF) rely on training samples to predict

Peer review under responsibility of National Authority for Remote Sensing and Space Sciences.

* Corresponding author at: Earth Observation Research & Innovation Centre (EORIC), University of Energy & Natural Resources, Sunyani, Ghana.
E-mail address: nkwa.sey@uenr.edu.gh (N. Ekow Nkwa Sey).

https://doi.org/10.1016/j.ejrs.2022.03.017
1110-9823/

certain features. However, these techniques can only perform well with more complex data sets and dimensionality (Zhong et al., 2018). A Deep Convolutional Neural Network (DCNN) is a promising method in remote sensing for extracting features that outperform previous methods including scene detection, object detection and semantic segmentation (Boonpook et al., 2018). A CNN-based neural network was proposed by (Zeggada et al., 2017) to classify multi-labelled images of UAVs.

Currently, the most widely used target detection algorithms are RCNN (Girshick et al., 2014), Fast RCNN (Girshick, 2015), Faster RCNN (Ren et al., 2017) and SSD (Liu et al., 2016). However, due to the large amount of training data that these frameworks require, they cannot provide end-to-end detection capabilities and their ability to position the feature in the detection frame is limited by the number of convolution layers as gradient explosion can occur during feature extraction (Zhang et al., 2020). ResNet was proposed by He et al. (2016), which addresses these drawbacks by helping the model to converge by combining the Mask RCNN's detection model with the learning process of the neural network to greatly improve the accuracy of object detection and segmentation. Mask RCNN is a deep learning framework that combines the capabilities of one network to identify and segment individuals in the same image. It can also perform target detection and segmentation (He et al., 2020).

Compared to existing segmentation algorithms, Mask RCNN does not only produce fine instance segmentation results but also provides great accuracy in detecting small targets. This framework has been applied in agriculture (Yu et al., 2019), medicine (Khan et al., 2021; Masood et al., 2021), construction (Attard et al., 2019), etc. The Mask RCNN algorithm has been applied in a wide range of areas, but according to our literature survey, no one has used it to detect rooftops in a typical Ghanaian rural community.

The objective of this paper is, thus, to implement a mask RCNN algorithm to detect the rooftop of buildings in a typical rural settlement. This research is important and can serve as a preliminary task in the field of human settlement change detection and estimation of solar generation potential. Mask RCNN is applied in this paper to detect and segment rooftops in aerial images.

2. Study area and data

2.1. Study area

To test the proposed framework's performance, we collected data samples from high-resolution aerial images covering Kantro, Henekrom and Boreso rural settlements within the Bono Region of Ghana (See Fig. 1). These villages are predominantly farming communities with scattered settlements located within a relatively complete forest landscape. Cassava, maize, cashew and cocoa are among the most grown crops in the area. Aside from Boreso, which was recently connected to the national grid, Kantro No.2 and Henekrom are still without electricity. Most of the settlers resort to mini solar photovoltaic (PV) systems or commute to Boreso and other adjoining communities to power their devices. The detection and measurement of rural rooftops are valuable for understanding rural development and estimating the solar generation potential as a viable alternative.

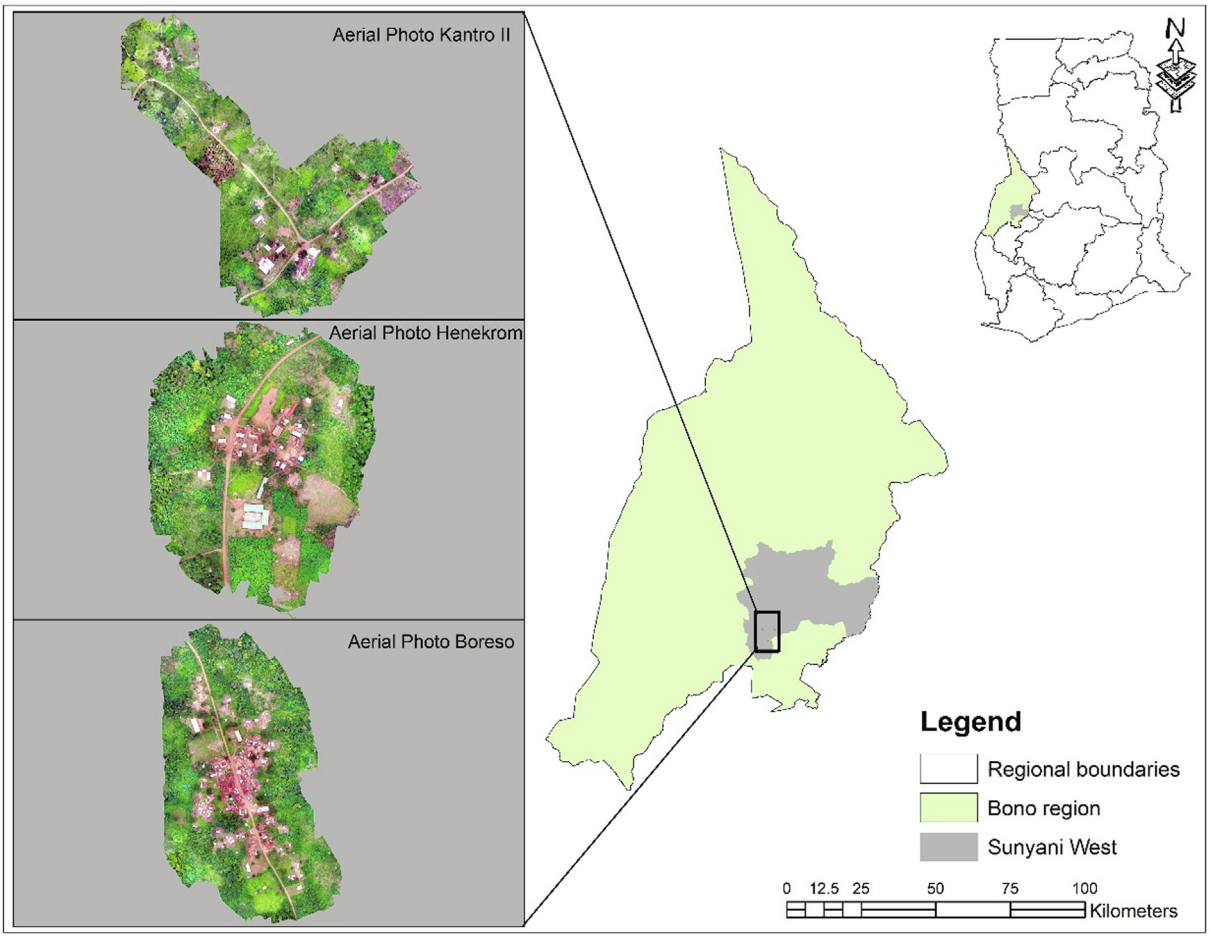

Fig. 1. Aerial map of study area.

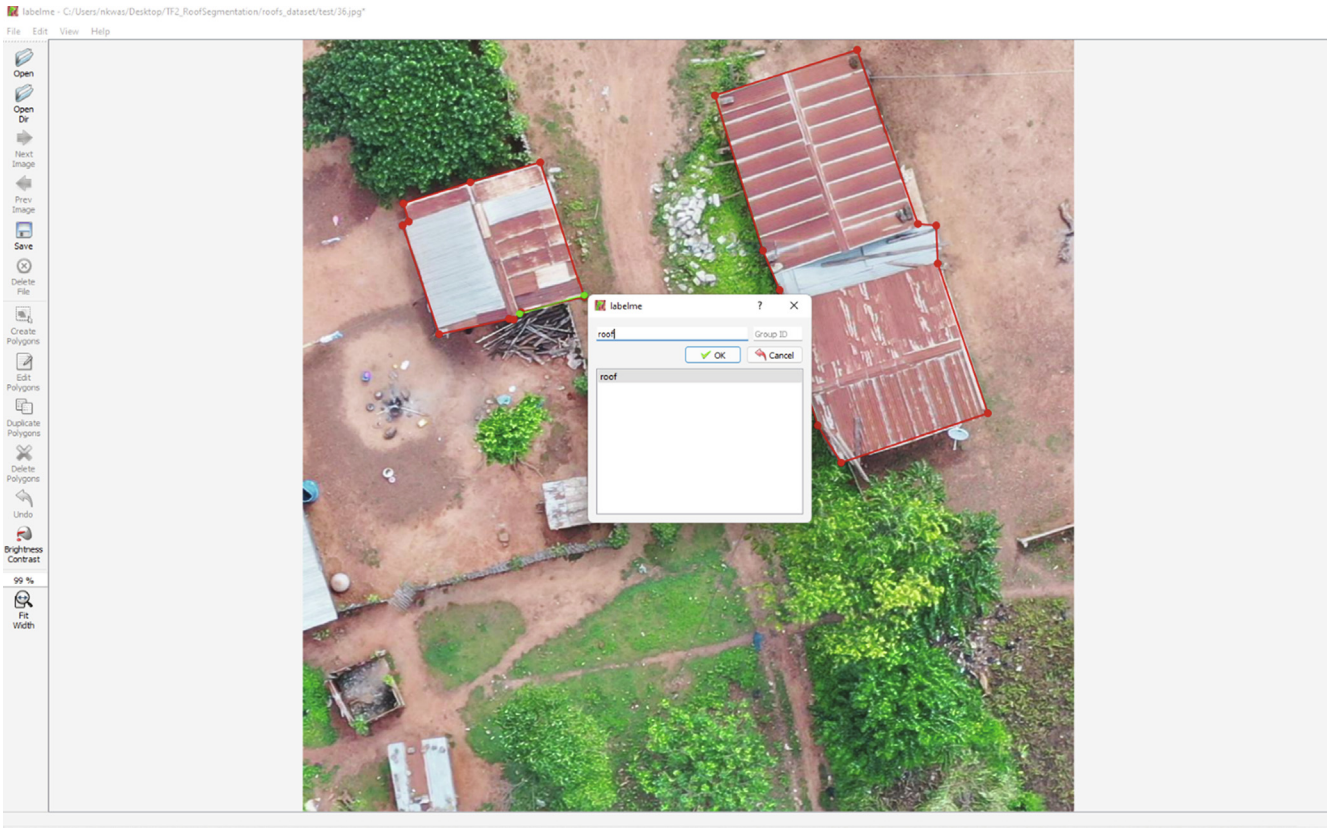

Fig. 2. LabelMe- Annotated rooftop image.

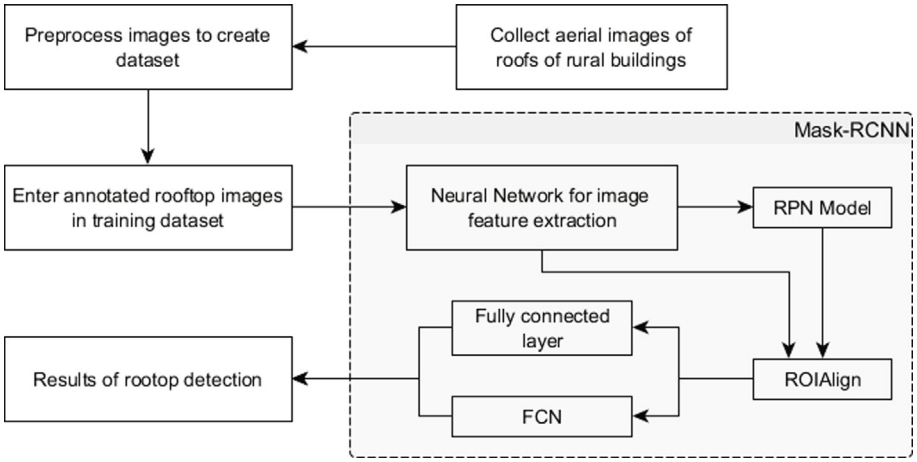

Fig. 3. Rural-rooftop-detection and segmentation system framework.

2.2. Data collection and annotation

The main objective of this research paper is to identify and segment the rooftops of rural buildings. In all, about 450 aerial images of different roof types of rural buildings within Kantro No.2, Henekrom and Boreso using a DJI Mavic 2 Pro equipped with a 20MP camera were taken. High-resolution images were captured with the drone at about 85 m above the ground. Each image has a size of 5472 × 3078 pixels in the RGB colour space. The rooftop images were cropped and normalized to a uniform size of 1024 × 1024 pixels.

An open-source image annotation tool, LabelMe (Wada, 2016), was used to delineate the footprints of the roofs in each picture. An image of the labelling interface is shown in Fig. 2. Out of 450 images, 405 were randomly split into training and validation datasets which were then annotated. The rest of the images were set aside as the test dataset.

3. Methods

3.1. Rural-rooftop-detection system architecture

Fig. 3 shows the detection and segmentation system framework for rural rooftop using a Mask RCNN model developed for this paper.

Table 1
Config file experimental parameter table.

Parameter	Value
model {num_classes}	1
model {height, width}	612, 612
train_config {batch_size}	1
train_config {num_steps}	10,000
train_config {learning_rate_base}	0.008
train_config {fine_tune_checkpoint_type}	detection
eval_config {metrics_set}	coco_detection_metrics
eval_config {metrics_set}	coco_mask_metrics

Table 2
Definition of true positives (TP), false positives (FP), false negatives (FN):

Definition	*gt* compared to *pd*	Confusion score, *conf_sc*
TP (correct detection)	IoU > 0.5 (roof present)	*conf_sc* $\geq$ *thres* (roof detected)
FP (invalid detection)	IoU $\leq$ 0.5 (no roof present)	
FN (Missed *gt*)	Missed roof, but roof present	*conf_sc* < *thres*

Aerial images of rooftops of rural buildings were marked by the LabelMe annotation tool. The annotated images were split into training and validation datasets which were then converted into TFRecord format, a binary format suitable for the TensorFlow Object Detection API. The Mask R-CNN Inception ResNet V2 1024x1024 model was chosen from among the pre-trained models available in the TensorFlow 2 Detection Model Zoo (Yu et al., 2020). The models have been trained on the COCO dataset and contain pre-trained weights that allow for transfer learning. Models that are heavily trained on diverse large datasets can be trained on more specialized data which can help improve their detection capabilities on distinct class sets. Transfer learning enables models to develop novel task-solving abilities by building on previous

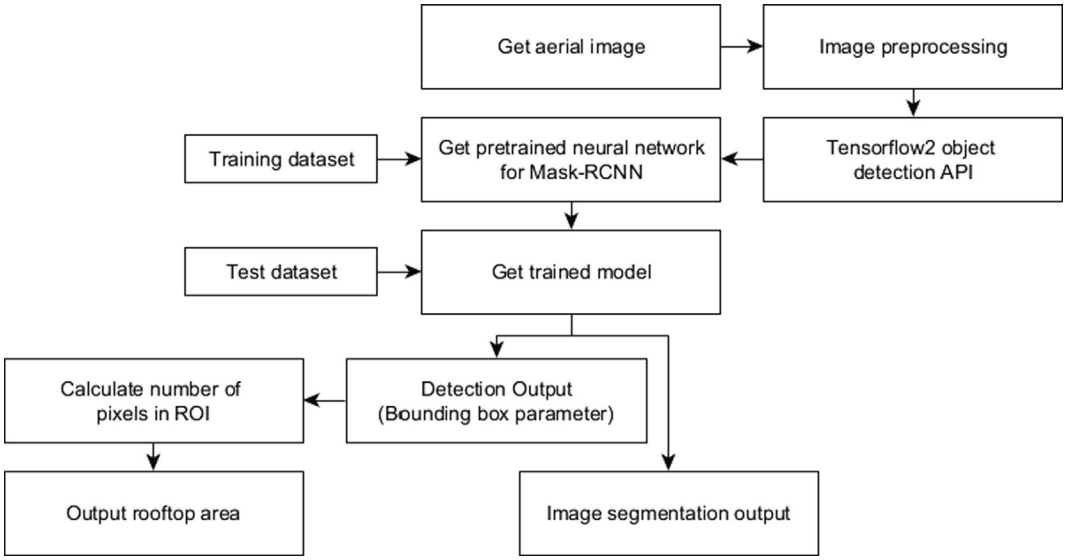

Fig. 4. Workflow of rooftop segmentation framework.

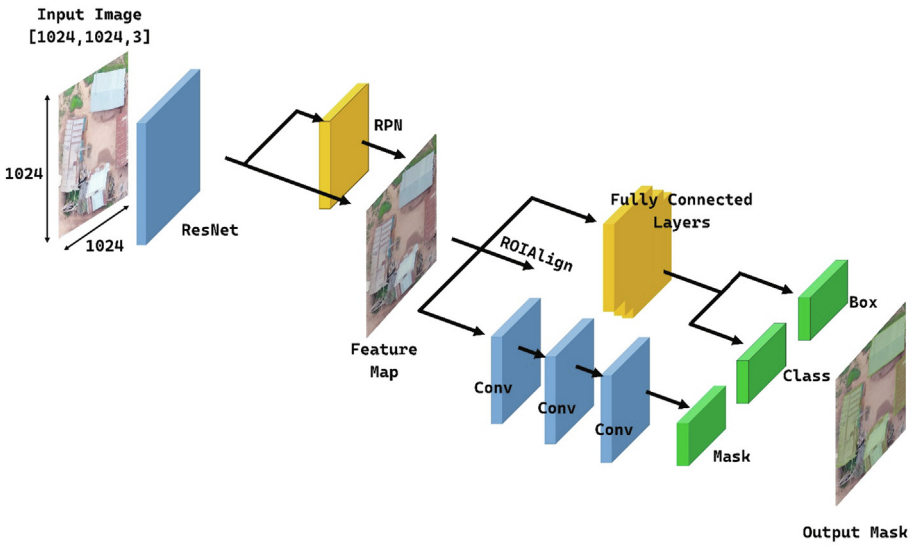

Fig. 5. Workflow block diagram of the Mask RCNN algorithm.

knowledge. The model goes together with a configuration file for defining the training procedure. The model is fine-tuned and further trained on images containing rooftops of rural buildings. The parameter settings tuned for training the model are shown in Table 1. The rural-rooftop detection outputs the masked roof and its segmented size. The overall flowchart is shown in Fig. 4.

3.2. Mask R-CNN

Mask R-CNN is an improved version of the Faster R-CNN framework that can perform target identification, classification, and instance segmentation within a neural network. It has better accuracy and time efficiency than its predecessor. This diagram shows the three main stages involved in preparing an aerial dataset for use with the Mask RCNN algorithm. Before we can implement the Mask R-CNN algorithm on an aerial dataset, we need to preprocess it using a grid to split the image into tiles of patches. Firstly, a feature extractions step is carried out using the trained deep learning model. The images are then input into the ResNet 101 to obtain the feature map. ResNet 101 (He et al., 2016) is a convolutional neural network (CNN) for feature extraction that can reduce hyperparameters while increasing its complexity and improving accuracy. After the CNN calculation, anchors of different sizes are chosen uniformly across the feature map. The sizes of the regions of interest (ROIs) for each anchor are computed and linked to the original image. This feature map shows a large number of candidate frame locations (e.g., regions of interest, or ROI), and it uses the softmax classification to identify the background and foreground of a given frame. Irrelevant Bounding Boxes (BB) are filtered from these ROIs and classified into the object and background label by applying a Regional Proposal Network (RPN). Binary classification is performed by RPN to identify the background and target object of interest to some extent. A Bounding-Box regression is then utilized to identify the object's real contour. A ROIAlign layer inputs the feature map and remaining feature map to generate a fixed-size feature map. The ROIAlign operation is a method to improve the efficiency of the ROI pooling process during feature extraction in Faster RCNN. It eliminates the pixel offset caused by the quantization process. Lastly, it goes through two branches, the first for object classification and the second for frame regression. It uses a fully connected layer and a Full Convolutional Network (FCN) to generate a mask (Fig. 5).

3.3. TensorFlow Object Detection API

The TensorFlow Object Detection API (Huang et al., 2017) is an open-source framework that simplifies the evaluation and deployment of object detection models. This framework is based on TensorFlow which is a programming language that can be used for training and analysing deep neural networks among a wide range of applications. It works seamlessly across various platforms like Windows, Linux, macOS and computational devices. Multidimen-

Average Precision (AP):	
AP	% AP at IoU=.50:.05:.95 (primary challenge metric)
$AP^{IoU=.50}$	% AP at IoU=.50 (PASCAL VOC metric)
$AP^{IoU=.75}$	% AP at IoU=.75 (strict metric)
Average Recall (AR)	
$AP^{max=1}$	% AR given 1 detection per image
$AP^{max=10}$	% AR given 10 detections per image
$AP^{max=100}$	% AR given 100 detections per image
AP/ AR across Scales	
AP^{small}	% AP for small objects: area < 32^2
AP^{medium}	% AP for medium objects: 32^2 < area < 96^2
AP^{large}	% AP for large objects: area > 96^2

Figure 6: COCO detection metrics

Fig. 6. COCO detection metrics.

sional typed arrays referred to as tensors perform TensorFlow computations; input–output operations are executed on tensor's edges while mathematical operations are performed at the tensor's nodes (Abadi et al., 2016).

3.4. Evaluation of model

The performance of the experiment is evaluated through two phases: detection performance and image segmentation performance. The two phases are respectively evaluated by the following metrics: the Average Precision (AP) value and the Average Recall (AR) value. The COCO object Detection metrics are used by TensorFlow object detection API. The mean and average precision metrics are the most popular ones used to evaluate object detection algorithms. They were also used to evaluate submission for competitions like the PASCAL VOC and COCO challenges (Nguyen et al., 2020). Additionally, the performance of the model to determine the location of rooftops was evaluated by comparing it to the ground truth using the Intersection of Union (IoU) metrics. The intersection between the ground truth (*gt*) and the predicted mask (*pm*) is known as the intersection over union. IoU is a threshold that is used to mark a bounding box as successful in detecting the location of a roof. It can also be used to predict the results of the segmentation mask. A threshold of 0.5 is often used to distinguish a valid detection from one that is not (He et al., 2020; Huang et al., 2017; Liu et al., 2016; Redmon and Farhadi, 2016). The true positives (TP), false positives (FP) and false negatives (FN) were determined according to Table 2. A confidence score was then calculated for varying thresholds (*thres*) to determine if the model detected a roof.

Precision and Recall metrics were calculated using Equations (1) respectively.

$$P = TP/(TP + FP)$$
$$R = TP/(TP + FN) \quad (1)$$

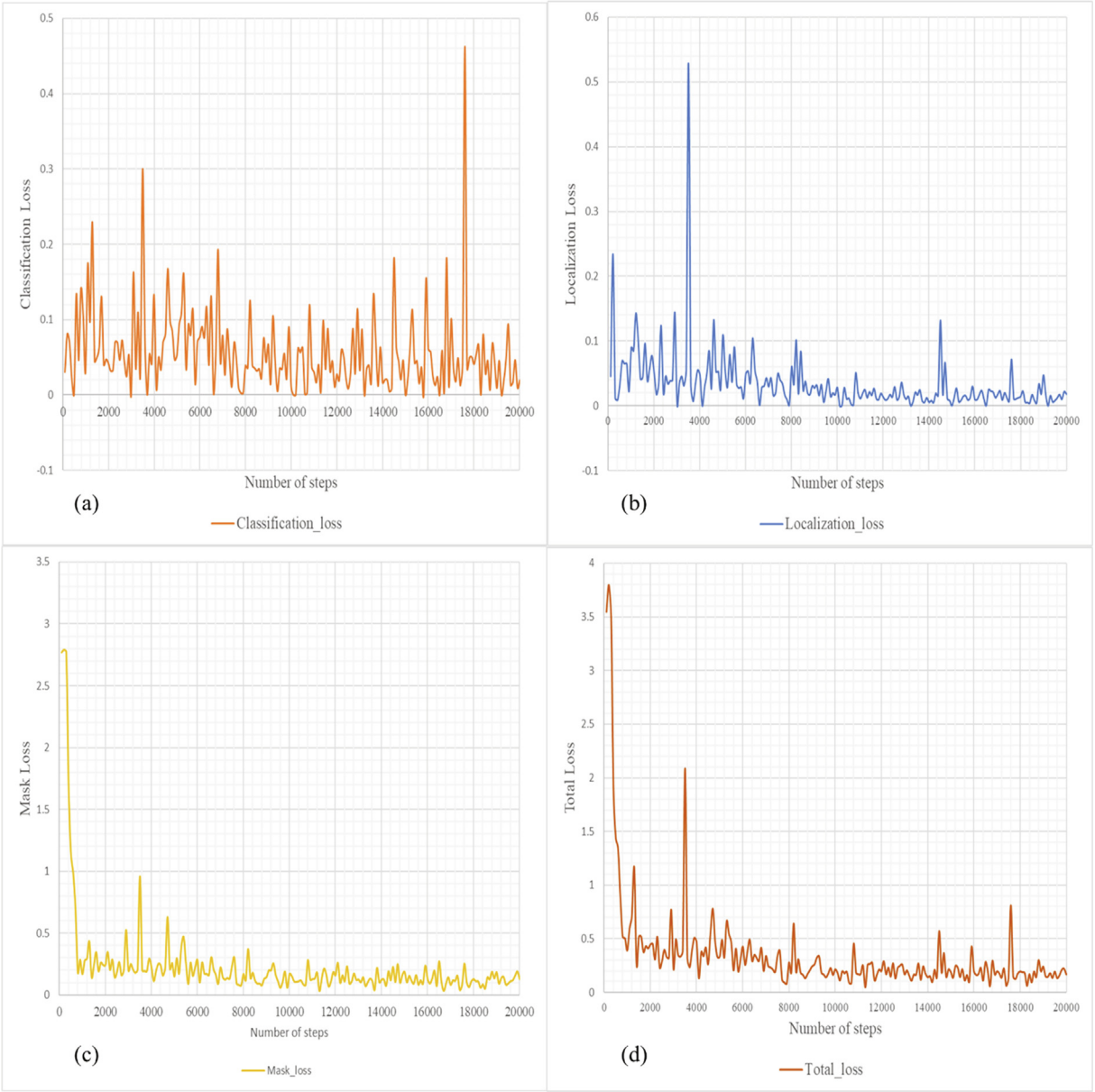

Fig. 7. The Losses per steps (a) Classification loss (b) Localization loss (c) Mask Loss (d) Total loss.

The COCO detection metrics define several average precision and recall values using different IoU thresholds and are also calculated across different object scales; small, medium and large. Fig. 4 shows the metrics used for measuring the performance of the detector (Lin et al., 2014) (Fig. 6).

3.5. Loss function

The multi-loss function is applied to the Mask-RCNN during the learning stage to evaluate the model's fitting to new datasets and is calculated as the total weighted sum of three (3) losses i.e. classi-

Fig. 8. Examples of rooftop detections on the test dataset.

Table 3
Evaluation metrics of rural-roof detection model.

Metric	Area	@ IoU	MaxDets	Value
Average Precision (AP)	All	0.50:0.95	100	0.848
		0.5		0.960
		0.75		0.950
	Small	0.50:0.95		−1.000
	Medium	0.50:0.95		0.302
	Large	0.50:0.95		0.870
Average Recall (AR)	All	0.50:0.95	1	0.296
			10	0.878
			100	0.882
	Small			−1.000
	Medium			0.471
	Large			0.903

fication error (L_{cls}), regression error (L_{bbx}) and segmentation error (L_{msk}) of the model on each proposal ROI during various phases of the training as shown in Equation (2).

$$Loss = L_{cls} + L_{bbx} + L_{msk} \quad (2)$$

The classification error (L_{cls}), regression error (L_{bbx}) and segmentation error (L_{msk}) are also referred to as the loss of classification, the loss of localization or bounding box and the loss of mask or segmenting the predicted mask respectively.

4. Results

4.1. Segmentation of rooftops

To monitor the performance of the model, the loss function values at each step of the learning process of the model on both the training and validation datasets are shown in Fig. 7. The learning of the model is carried out through the training dataset while the validation dataset is used to calibrate the model's efficiency and minimize overfitting (Kim et al., 2018).

Fig. 7 shows that the convergence of the various loss output values is observed to be approaching 0 as the steps increase, indicating the accuracy of the framework. Fig. 8 shows the segmentation results achieved by the trained Mask-RCNN model on data that was not part of the training and validation dataset. Some of the results of the segmentation masks do not fit or marginally overlap the object (rural rooftops) however, the model performed well in detecting roofs in the aerial images. Various factors such as increasing the number of training sets, changing some training parameters, and increasing the data volume can also help improve the model's performance. The environmental conditions (such as hillshades, shaded buildings, soil colour etc.) of an image data set should also be taken into account during the training as it can potentially affect the model's decisions when it comes to detecting and segmenting objects while training and testing.

4.2. Evaluation metrics

In our research, good performance was obtained for the model framework with regard to the accuracy of the detections and segmentation made. The average precision and average recall (@ IoU = 0.50:0.95) for 100 max detections resulted in values of 0.848 and 0.882 respectively. The evaluation metrics of the trained Mask RCNN model on the full dataset are shown in Table 3.

5. Conclusion

In this study, a Mask R-CNN deep learning model, implemented with TensorFlow Object API was applied to detect and segment rooftops of typical rural communities from aerial images. Transfer learning was performed while the starting weights were computed based on the model pre-training performed on the Microsoft common objects in context (COCO) dataset. The rural-roof detection model was able to reach high average precision and recall scores for the various segmentation masks and boxes. These components were tested against various IoU thresholds. The Mask-RCNN model can be applied to detect rural rooftops from aerial images with good accuracy. Even though the segmentation masks do not fit on the roofs entirely in all cases, the object is accurately detected and can help identify which pixels in the image make up a rural rooftop. The method employed in this paper can serve as a preparatory step to measuring the sizes of rooftops in rural areas using drones in order to estimate the solar generation potential of such areas.

6. Declarations of interest

None.

Acknowledgements

The authors are grateful for the immense assistance received from the Earth Observation Research and Innovation Centre (EORIC), University of Energy and Natural Resources.

Conflicts of interest

The authors declare no conflict of interest.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Alfaro, J.F., Miller, S., Johnson, J.X., Riolo, R.R., 2017. Improving rural electricity system planning: an agent-based model for stakeholder engagement and decision making. Energy Policy 101, 317–331. https://doi.org/10.1016/J.ENPOL.2016.10.020.

Arranz-Piera, P., Kemausuor, F., Darkwah, L., Edjekumhene, I., Cortés, J., Velo, E., 2018. Mini-grid electricity service based on local agricultural residues: feasibility study in rural Ghana. Energy 153, 443–454. https://doi.org/10.1016/j.energy.2018.04.058.

Attard, L., Debono, C.J., Valentino, G., Di Castro, M., Masi, A., Scibile, L., 2019. Automatic crack detection using mask R-CNN. In: International Symposium on Image and Signal Processing and Analysis, ISPA. IEEE, pp. 152–157. https://doi.org/10.1109/ISPA.2019.8868619.

Azimoh, C.L., Klintenberg, P., Wallin, F., Karlsson, B., Mbohwa, C., 2016. Electricity for development: mini-grid solution for rural electrification in South Africa. Energy Convers. Manag. 110, 268–277. https://doi.org/10.1016/J.ENCONMAN.2015.12.015.

Boonpook, W., Tan, Y., Ye, Y., Torteeka, P., Torsri, K., Dong, S., 2018. A deep learning approach on building detection from unmanned aerial vehicle-based images in riverbank monitoring. Sensors 18, 3921. https://doi.org/10.3390/s18113921.

Eder, J.M., Mutsaerts, C.F., Sriwannawit, P., 2015. Mini-grids and renewable energy in rural Africa: How diffusion theory explains adoption of electricity in Uganda. Energy Res. Soc. Sci. 5, 45–54. https://doi.org/10.1016/J.ERSS.2014.12.014.

Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. Nat. Hazards Earth Syst. Sci. 15, 1087–1101. https://doi.org/10.5194/nhess-15-1087-2015.

Girshick, R., 2015. Fast R-CNN. 2015 IEEE Int. Conf. Comput. Vis. 2015 Inter, 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2020. Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. 42, 386–397. https://doi.org/10.1109/TPAMI.3410.1109/TPAMI.2018.2844175.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2016.90.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K., 2017. Speed/accuracy trade-offs for modern convolutional object detectors. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 3296–3297. https://doi.org/10.1109/CVPR.2017.351.

Khan, M.A., Akram, T., Zhang, Y.D., Sharif, M., 2021. Attributes based skin lesion detection and recognition: a mask RCNN and transfer learning-based deep learning framework. Pattern Recognit. Lett. 143, 58–66. https://doi.org/10.1016/j.patrec.2020.12.015.

Kim, D., Arsalan, M., Park, K., 2018. Convolutional neural network-based shadow detection in images using visible light camera sensor. Sensors 18, 960. https://doi.org/10.3390/s18040960.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8693 LNCS, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single Shot MultiBox Detector. In: Lecture Notes in Computer Science (Including

Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
Masood, M., Nazir, T., Nawaz, M., Mehmood, A., Rashid, J., Kwon, H.-Y., Mahmood, T., Hussain, A., 2021. A novel deep learning method for recognition and classification of brain tumors from MRI images. Diagnostics 11, 744. https://doi.org/10.3390/diagnostics11050744.
Nguyen, N.D., Do, T., Ngo, T.D., Le, D.D., 2020. An evaluation of deep learning methods for small object detection. J. Electr. Comput. Eng. 2020, 1–18. https://doi.org/10.1155/2020/3189691.
Redmon, J., Farhadi, A., 2016. YOLO9000: Better, Faster, Stronger. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 6517–6525. https://doi.org/10.1109/CVPR.2017.690.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS J. Photogramm. Remote Sens. 140, 45–59. https://doi.org/10.1016/j.isprsjprs.2017.03.001.
Wada, K., 2016. Labelme: Image Polygonal Annotation with Python.
Yu, H., Chen, C., Du, X., Yeqing, L., Rashawn, A., Hou, L., Jin, P., Yang, F., Lui, F., Kim, J., Li, J., 2020. TensorFlow Model Garden. URL https://github.com/tensorflow/models (accessed 9.1.21).
Yu, Y., Zhang, K., Yang, L., Zhang, D., 2019. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Comput. Electron. Agric. 163, 104846. https://doi.org/10.1016/j.compag.2019.06.001.
Zeggada, A., Melgani, F., Bazi, Y., 2017. A deep learning approach to UAV image multilabeling. IEEE Geosci. Remote Sens. Lett. 14, 694–698. https://doi.org/10.1109/LGRS.2017.2671922.
Zhang, Q., Chang, X., Bian, S.B., 2020. Vehicle-damage-detection segmentation algorithm based on improved Mask RCNN. IEEE Access 8, 6997–7004. https://doi.org/10.1109/ACCESS.2020.2964055.
Zhong, Y., Ma, A., Ong, Y.S., Zhu, Z., Zhang, L., 2018. Computational intelligence in optical remote sensing image processing. Appl. Soft Comput. 64, 75–93. https://doi.org/10.1016/j.asoc.2017.11.045.