ZANCO Journal of Pure and Applied Sciences
The official scientific journal of Salahaddin University-Erbil
https://zancojournals.su.edu.krd/index.php/JPAS

ISSN (print):2218-0230, **ISSN (online):** 2412-3986, **DOI:** **http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Building Footprint Extraction from UAV Imagery Using Deep Learning.

Hoshang J. Khdir, Haval A. Sadeq[2]

[1]Department of Geomatics(Surveying), College of Engineering, Salahaddin University-Erbil, Kurdistan Region, Iraq

ABSTRACT:

Building footprint extraction from UAV (Unmanned Aerial Vehicle) imagery data has been an active research topic in the photogrammetric and remote sensing communities in the past two decades. The traditional methods for building extraction from high-resolution imagery data are time-consuming and may not provide desired results. Recently, effective high-level approaches have been developed for buildings footprint extraction. However, their efficiency must be balanced by reducing the processing time required to obtain acceptable results using high-resolution imagery. This paper introduces an automatic method to extract building footprints from UAV imagery using deep learning algorithms. A Mask R-CNN (Region Based Convolutional Neural Network) model has been applied to building footprint extraction. For the building extraction, three experiments have been achieved. For the algorithm testing, two study areas have been selected. An orthophoto has been produced for each study area using photogrammetric software based on UAV imagery. Three experiments have been achieved for building extraction from the study area. The first experiment was based on using the pre-trained model only. In the second trial, the raw model was trained based on the study area only was used. While in the third trial, the model was trained based on fine-tuning (based on satellite imagery) and the pre-trained model with UAV training data. The analysis showed that the highest accuracy rate of the building footprint extraction increased to around 95% through using fine-tuning and more data sets in the model training, specifically with similar data sets to the study area.

KEY WORDS: Building Footprint, Deep Learning, Mask R-CNN, UAV, Fine-Tune.
DOI: http://dx.doi.org/10.21271/ZJPAS.35.SpA.3
ZJPAS (2023) , 35(SpA);26-34 . .

1. Introduction

Building footprint extraction is essential for creating base maps, analyzing urban planning, development workflows, change detection, 3D modelling, infrastructure planning, etc. Traditional manual building digitizing from images is considered to be a time-consuming and expensive task. Meanwhile, the available automatic methods based on processing satellite and aerial images are considered to be inaccurate and specified to be low accuracy (Gavankar and Ghosh, 2018, Wang et al., 2006, Shackelford et al., 2004, Lafarge et al., 2008). Object detection based on deep learning can be classified into semantic and instance segmentation. Semantic segmentation occurs when each pixel in an image is classified as belonging to a specific class. Instance segmentation is a more precise object detection technique that first instantly draws each object's boundary and then extracts features.

Different articles have been focused on the field of building footprint extraction based on deep learning. They are highly capable of learning these complex semantics and producing more accurate results than automatic methods such as image processing techniques. Thus, it is considered to be

* Corresponding Author:

Hoshang Jaafar Khdir
E-mail: Hoshang.jaafar@gamail.com

Article History:
Received: 17/12/2022
Accepted: 27/03/2023
Published: 1/11 /2023

more efficient than the manual process of building extraction.

Based on the literature, different approaches have been developed for object recognition from images, such as Mask R-CNN, U-Net, YOLO ,etc.

Mask R-CNN has been used extensively for extracting building footprint from satellite imagery (Chitturi, 2020, Stiller et al., 2019, Tiede et al., 2021, Raghavan et al., 2022, Tejeswari et al., 2022). Furthermore, researchers evolved the Mask R-CNN for extracting building footprints from satellite or aerial images (Li et al., 2021, Zhang et al., 2020, Li et al., 2019, Zhao et al., 2018, Wen et al., 2019). Moreover, it is also applied to UAV image datasets that have been used for building footprint extraction using Mask R-CNN or modified new approaches-based Mask R-CNN by researchers (Wang et al., 2022, Li et al., 2021, Chen et al., 2020).

On the other hand, the U-Net, a semantic segmentation network based on Fully Convolutional Networks, also has been applied in building extraction. It has been used for extracting building footprints from satellite and aerial images (Ronneberger et al., 2015, Li et al., 2019, Alsabhan and Alotaiby, 2022). Some approaches, based on the U-Net, are being designed to extract buildings from satellite images (Alsabhan et al., 2022, Rastogi et al., 2022). U-Net segmentation or new approaches based on U-Net algorithms have been used for extracting building footprints from UAV images (Liu et al., 2021, Chafiq et al., 2021, Daranagama and Witayangkurn, 2021, Liu et al., 2019b).

Additionally, some researchers have continued developing various algorithms such as Mask R-CNN and U-Net approaches for their specific studies to extract building footprints from high-resolution satellite/aerial images (Zhao, 2019, Idris et al., 2021, Abdollahi et al., 2020, Wei et al., 2019, Liu et al., 2019a, Ran et al., 2021, Aung et al., 2022, Xiong et al., 2022). Moreover, a cutting-edge deep learning algorithm was developed using UAV image datasets for building footprint detection (Chen et al., 2016, Boonpook et al., 2018, Zhou et al., 2022) based on using SegNet and EDSANet and applying UAV in the training the model.

Based on the literature, using high-resolution UAV imagery is considered a challenge in extracting building boundaries, and the obtained accuracy was very low compared to the satellite and aerial imagery (Boonpook et al., 2018, Ammour et al., 2017, Sheppard and Rahnemoonfar, 2017).

To date no research have been conducted on Fine-Tuning method based on using Mask R-CNN for the building extraction from UAV imagery, thus in this research, Mask R-CNN has been applied for the building extraction from UAV imagery.

The paper is organized as follows, the introduction in this section. The study area and orthomosaic generation are discussed in section2. Section 0 is related to the implemented deep learning model and training dataset. The accuracy assessment and result are shown in section 4. Finally, the conclusion is discussed in section 0.

2. Study Area and Orthomosaic generation

UAVs have been considered an essential tool in obtaining accurate geospatial data, particularly for photogrammetric products such as DSMs and orthoimages with very high resolution up to centimetres resolution. For the algorithm testing, two different study areas have been selected. The areas are located in Erbil governorate in Iraq. Both sites contain residential buildings with a nearly similar pattern. The first study area contains around 280, while the second study area includes around 120

buildings.

For image acquisition, The Quad-Rotor UAV (DJI Phantom 4 Pro plus) has been used. The sensor of the used camera is based on CMOS with a 20 MP resolution and a lens with a FOV (Field of View) of 84°. The camera's focal length is 9 mm/24 mm (35 mm format equivalent) with an aperture of 2.97 and an image size of 5472 × 3648 pixels. The implemented UAV has onboard a positioning system (GPS/GLONASS) which aids the photogrammetric processing (Wolf et al., 2014, Sadeq, 2019). Initially, Self-camera calibration has been achieved using the Pix4D software based on using self-calibration for determining the interior orientation parameters of the images using Pix4d software. Later the images are acquired from both study areas which are counted to be 176 and 380 images, for the first and second study areas respectively. The end and side overlapped were the same and selected to be 80%. The UAV height is selected to be 100 m for the first study area and 74 m for the second study area.

For photogrammetric image processing, Pix4D software has been used for orthophoto generation. In the processing stage, 5 and 4 GCPs were used in the first and second study areas respectively. The Ground control points (GCP) have been distributed according to photogrammetric principles and collected using Global Navigation satellite systems (GNSS) as shown in the Figure *3*, which later have been used in exterior orientation. The obtained orthophoto was georeferenced to the WGS1984 UTM Zone 38N with a resolution (Ground Sample Distance (GSD)) of 2.67cm and 1.95 cm for the first and second study areas respectively. RMSE values for both areas have been calculated as 0.006 and 0.003 for the first and second areas respectively, which is considered to be ideal in both cases.

3. Methodology and Experiments

In this section, the implemented models, the process of the model training, and their application to the study area will be explained. The steps are illustrated in the flowchart as shown in the Figure 1.

3.1 Mask R-CNN model

Mask R-CNN is a state-of-the-art model developed on top of Faster R-CNN with an additional branch for predicting segmentation masks on each Region of Interest (RoI). The Faster R-CNN is a region-based convolutional neural network that returns bounding boxes for each object and its class label with a confidence score. Based on the literature the best algorithm for instance segmentation is Mask R-CNN. Therefore, in this paper, it has been selected for the building extraction, which is applied to the obtained ortho mosaics that consisted of 3 bands (RGB). Deep learning algorithms consist of neural networks. Prior to applying the deep learning model, the neural should be trained by assigning weights to each neuron through using training data.

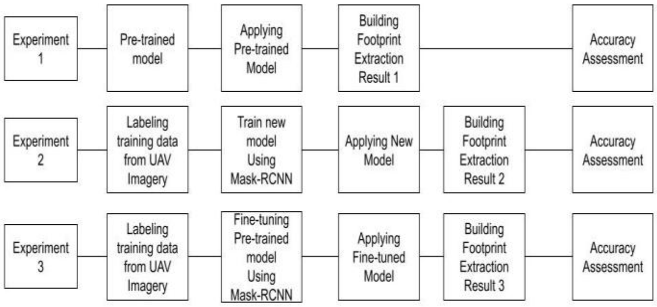
Experiment 1
Pre-trained model
Applying Pre-trained Model
Building Footprint Extraction Result 1
Accuracy Assessment
Experiment 2
Labeling training data from UAV Imagery
Train new model Using Mask-RCNN
Applying New Model
Building Footprint Extraction Result 2
Accuracy Assessment
Experiment 3
Labeling training data from UAV Imagery
Fine-tuning Pre-trained model Using Mask-RCNN
Applying Fine-tuned Model
Building Footprint Extraction Result 3
Accuracy Assessment

Figure 1: Flowchart for methodology and experiments.

As shown in the left image in Figure 2, Mask R-CNN implementation consists of two stages:

The first stage consists of two networks, a backbone, and a region proposal network. These networks run once per image to give a set of region proposals. Region proposals are based on selecting regions in the feature map that contain the object. In the second stage, the network predicts bounding boxes and object classes for each of the proposed regions obtained in the first stage. Each proposed region can be of a different size, whereas fully connected layers

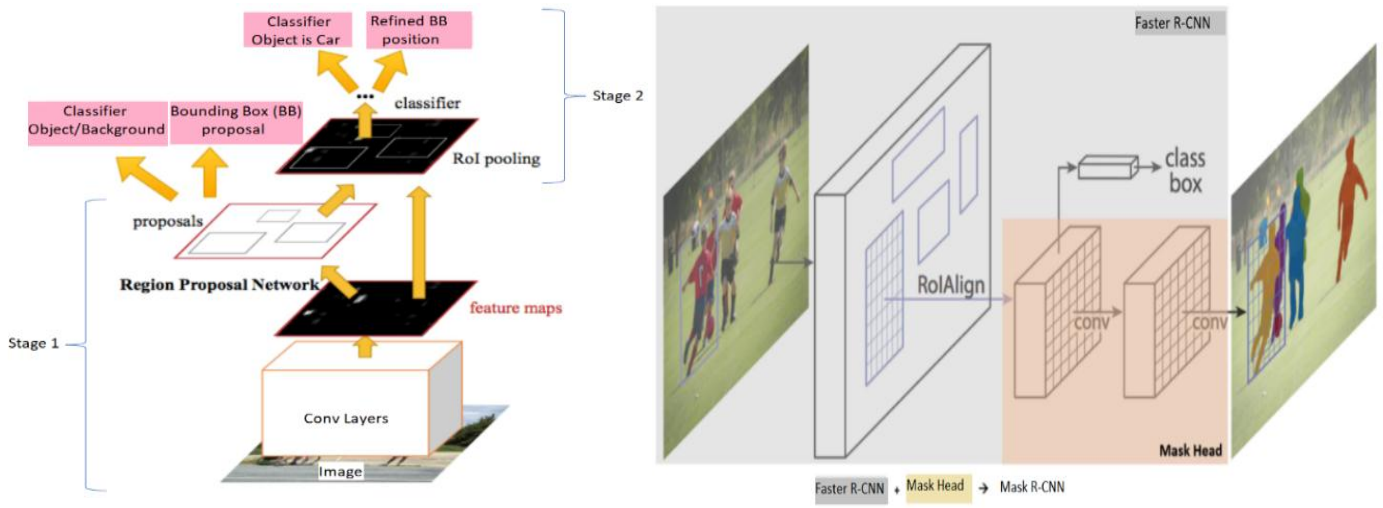

Figure 2: Mask RCNN Architecture and Mask RCNN masks for each Region of interest *(He et al., 2017, ESRI, 2022b)*

in the networks always require fixed-size vectors to make predictions. The size of these proposed regions is fixed using either the RoI pool or the RoIAlign method (an operation for extracting a small feature map from each RoI in detection and segmentation segmentation-based tasks).

The RoIAlign removes the severe quantization of the RoI Pool and properly aligns the extracted features with the input). As a result, faster R-CNN predicts object class and bounding boxes. See the right image in Figure 2.

The replacement of the RoI pool by RoI Align, in the second stage, will lead to preserving spatial information which gets misaligned region in the case of the RoI pool. RoIAlign uses binary interpolation to create a feature map of fixed size, e.g., 7 x 7. The output from the RoI Align layer is then fed into the Mask R-CNN head, which consists of two convolution layers. It then generates a mask for each RoI, thus segmenting an image based on pixel-to-pixel type (He et al., 2017).

3.2 Pretrained Model

In this research, the used pre-trained model is named (Building Footprint Extraction–USA) (ESRI, 2022a), trained by ESRI company. The model has been trained based on using satellite imagery data, and the model precision got 0.718. The data used to train the pre-trained model was based on satellite imagery with a resolution of 30 cm. Therefore, the orthomosaic that was obtained for the study areas was resampled to 30 cm resolution; otherwise, the obtained results were inaccurate.

3.3 Training data

Prior to using the selected deep learning model from section 3.2, it should be trained. In the training process, it is necessary to use similar data to the test area in the model training. Otherwise, the result will be erroneous; also, it is preferred to use a large amount of data. Occasionally the pre-trained model has been trained based on a sample from a specific city or one country.

Some buildings from both study areas are selected randomly for the model training purpose. Later, the feature labeling of the buildings was done manually in both study areas. The percentage of the selected training data was around %25 of the total features, as shown in Figure 3.

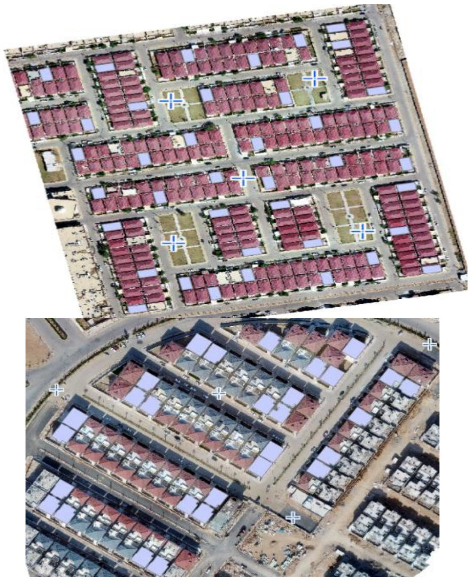

Figure 3: The selected buildings used in the training data for both study areas showing the location of GCPs places and the highlighted buildings refer to the buildings that used in the fine-tune process. (the up image is study area 1 and the down image is study area 2)

3.4 Experiments and results

In this stage, the trained Mask R-CNN model has been used in the building detection in both study areas. For obtaining optimum results for extracting building footprints from UAV imagery, three experiments have been achieved to identify the optimum pipeline:

3.4.1 Experiment One: Building extraction based on the Pretrained model only

For an initial test, the pre-trained model (Building Footprint Extraction – USA), which is mentioned in section 3.1, has been applied for both study areas. The used model was based on the trained data from the satellite imagery only. Based on the recommendation of the author, the optimal result is obtained by using images with 30 cm resolutions. For that purpose, the available orthophotos have been resampled to 30 cm using the Nearest interpolation method. The result of the building footprint extraction based on the pre-trained model is shown in Figure 5. The Results show that the buildings are clearly extracted from the first study area, but some of the buildings are margined as one building. In contrast, in the second study area, the buildings are partially detected, and a large number of buildings are merged together.

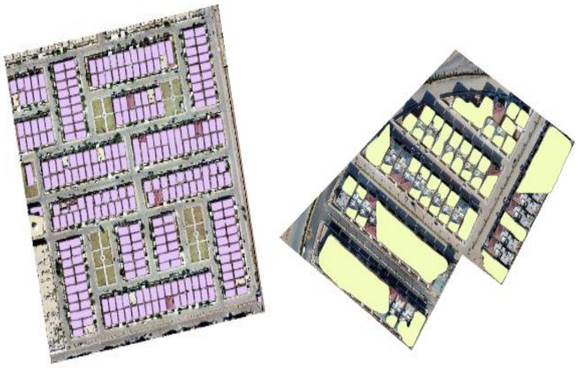

Figure 5: Results one of applying the model pretrained model only, in (Left) study area one most of the buildings detected, however, merged together. In (right) study area two detection is bad and merged as a big polygon.

3.4.2 Experiment Two: Building extraction based on training data only

In the second test, the Mask R-CNN model was trained based on using the training data from the study area only without using satellite imagery. Each study area has been resampled into 30cm then from each study area, 25% of the buildings have been selected as training data, as shown in Figure 3. Later the selected data are labeled and exported for training data. The Exported training data has been clipped into sub-images with size (128×128) because it is better to use small tiles with very high-resolution UAV imagery. Thus (25295) tiles have been obtained. Later, the trained data was applied to the study area to extract other buildings. Results are shown in Figure 4. The left image (study area 1) shows around half of the buildings have been detected. However, in the right image (study area 2) detection rate is higher than in study area 1.

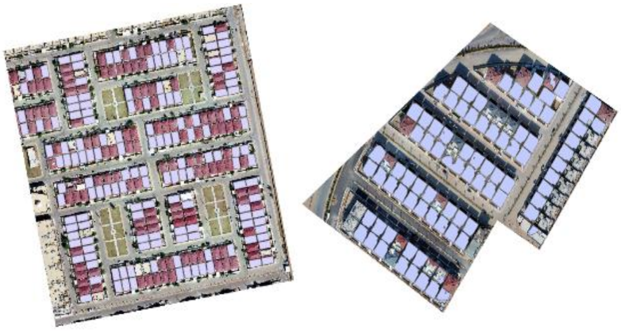

Figure 4 : Results for experiment two, some of the buildings have been detected in the left case study and the right case study most of the buildings were detected but merged some of them.

3.4.3 Experiment Three: Using training data and Pretrained model (Fine-tune model)

In the third test, a new model was obtained based on using the pre-trained model (Building Footprint Extraction – USA) and the same training data used in experiment two.

This process is called a fine-tuning pre-trained model. The Fine-tuned model has been used for extracting building footprints from UAV imagery. Similarly, based on the model requirement, both images have been resampled to 30 cm. In Figure 6, detection results are shown for both study areas (study area 1&2). It can be noticed that a higher rate of buildings has been detected in both study areas by using the fine-tuned model. Furthermore, the results show that high percentages of buildings have been extracted successfully, and the merging building is very low.

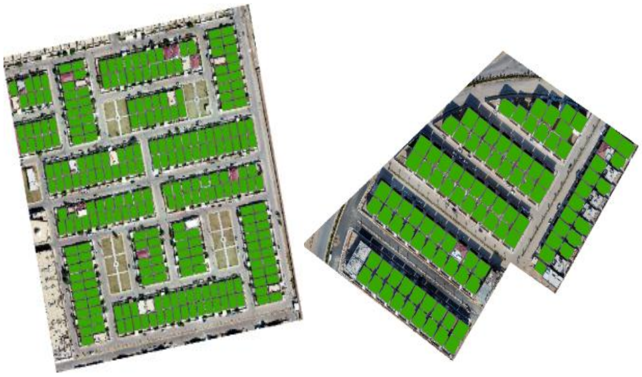

Figure 6: Results for experiment three, most of the buildings detected correctly in both case studies.

4. Accuracy Assessments

For accuracy assessment and to find the performance of the applied method, the buildings in both study areas have been digitized manually using ArcGIS Pro to get the ground truth layer. The accuracy metrics provided by the confusion matrix were used by comparing the locations of the predicted building polygons and the ground truth buildings. For the assessment purpose, three indices have been determined Precision, Recall, and F1-Score. The metric of Intersection over Union (IoU) has been calculated as follows:

$$IoU\ = \frac{Area(A\cap B)}{Area(A\cup B)} \qquad 4.1$$

Where A represents the predicted boundary and B ground truth building polygons, IoU is equal to the intersection area of A and B divided by their union area, which showed in Figure 7 :

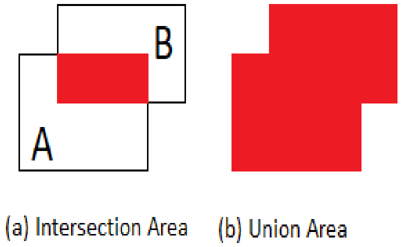

Figure 7: Definition of IoU areas, referred to as the red regions in the figures

When a value of IoU between the predicted polygon and ground truth is (≥ 0.5), it means extracted building polygon is True positive (TP). Otherwise, the extracted polygon was false positive (FP). The ground truth polygon not extracted or missed in the detections was denoted as a false negative (FN). N and M represent the number of ground truth buildings polygon and predicted building footprints, respectively. Based on counting (TP, FP, FN) numbers, scores for Precision, Recall, and F1-Score have been calculated as below:

Precision indicates the proportion of building footprints correctly identified by the proposed approach:

$$Precision\ = \frac{TP}{TP+FP}\ = \frac{TP}{M} \qquad 4.2$$

while recall indicates the proportion of building footprints in the validation data that were correctly detected by the approach:

$$Recall = \frac{TP}{TP+FN}\ = \frac{TP}{N} \qquad 4.3$$

F1 is used to balance precision and recall parameters:

$$F1score\ = 2 \times \frac{Precision \times Recall}{Precision+Recall}\ = 2 \times \frac{TP}{M+N} \qquad 4.4$$

All calculated scores for both case studies have been collected in below Table 1:

Table 1: Accuracy assessments scores for this research based on object detection

Cases	Exp.	Recall	F1 Score	Precision	TP	FP	FN
Study area 1	Exp. 1	0.8379	0.8773	0.7965	243	21	47
	Exp. 2	0.5207	0.6771	0.5132	151	5	139
	Exp. 3	**0.9207**	**0.9570**	**0.9186**	**267**	**1**	**23**
Study area 2	Exp. 1	0.3153	0.4375	0.2312	35	14	76
	Exp.2	0.7297	0.8020	0.7151	81	10	30
	Exp.3	**0.9099**	**0.9439**	**0.9072**	**101**	**2**	**10**

Considering the first test for the building extraction, it can be noticed that some of them are shown to be merged. This is because the applied extraction model was based on using training data sourced to satellite imagery only, in which the geometric resolution is different from UAV.

In the second test, the rate of the merged building is less due to using a very small number of training data during the model training. In the case of the third test, it is clear that the results are better than both tests. This improvement is achieved by the Fined-Tuned Model in which the Pretrained model and local training data were used for the model creation. It can be noticed from the table-1 the best method for building footprint extraction from UAV orthomosaics is fine-tuned model obtained from the third experiment. The F1-Score has the highest value in test three for both study areas, **0.9570** and **0.9439**.

5. Discusion

The current study, building extraction from UAV imagery is considered to be a very challenging task due to the various geomatic and spectral resolutions of images (Boonpook et al., 2018). The current study found that using the Fine-tuning Mask R-CNN method, performs effectively in identifying buildings in UAV images with a fairly high level of accuracy than Pre-training model or model training based on UAV imagery only. The analysis shows that the used training model is able to accurately detect most of the buildings successfully, including buildings with complex features. Furthermore, adjacent buildings to each other are also clearly has been extracted, even in cases where the buildings have multiple patterns. These findings further support the idea of using similar data in model training will enhance the result obtained from deep learning .

Although, there were still some errors in detecting buildings. The present findings appear to outperform other research. For instance,(Liu et al., 2021)achieved F1 scores of 0.9213, 0.8850, and 0.9019 for three different study areas by training a new model using U-Net and a large number of images. Similarly, (Kokeza et al., 2020) employed the Res-U-Net algorithm to extract buildings from UAV imagery by training orthophoto. After a lengthy processing procedure, they obtained F1 scores of 0.9393. To obtain better accuracy, and to generalize the model so it can be applied to different types of buildings, it is recommended use train the model based on using various types of buildings.

6. Conclusion

The experiments conducted in this research aimed to evaluate the performance of different deep learning models for building footprint extraction. The results showed a clear improvement in performance as the models became more sophisticated. The first experiment, which used only a pre-trained deep learning model, achieved an F1 score of **0.8773 & 0.4375**. This result indicates that pre-trained models cannot be effective for all different places in building footprint extraction tasks.

However, when a new model was trained using training data collected by the researchers', the F1 score dropped to 0.6771 & 0.8020. This suggests that, in cases where the training data is limited, it may not be sufficient to train a model from scratch and achieve good performance.

The final experiment, which involved fine-tuning a pre-trained model, produced the best result with an F1 score of **0.9570 & 0.9439**. This highlights the importance of fine-tuning pre-trained models, as it allows for better adaptation to the specific task and data. The results of this experiment demonstrate the potential of deep learning models for building footprint extraction and the value of using pre-trained models in this context.

In summary, these experiments have provided valuable insights into the use of deep learning models for building footprint extraction. The results support the conclusion that fine-tuning pre-trained models is a highly effective approach in cases where limited training data is available, and suggests that further research in this area has the potential to produce even better results.

References

ABDOLLAHI, A., PRADHAN, B., GITE, S. & ALAMRI, A. 2020. Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture. *IEEE Access,* 8**,** 209517-209527.

ALSABHAN, W. & ALOTAIBY, T. 2022. Automatic Building Extraction on Satellite Images Using Unet and ResNet50. *Computational Intelligence and Neuroscience,* 2022**,** 5008854.

ALSABHAN, W., ALOTAIBY, T. & DUDIN, B. 2022. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience,* 2022**,** 4831223.

AMMOUR, N., ALHICHRI, H., BAZI, Y., BENJDIRA, B., ALAJLAN, N. & ZUAIR, M. 2017. Deep learning approach for car detection in UAV imagery. *Remote Sensing,* 9**,** 312.

AUNG, H. T., PHA, S. H. & TAKEUCHI, W. 2022. Building footprint extraction in Yangon city from monocular optical satellite image using deep learning. *Geocarto International,* 37**,** 792-812.

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.

CHAFIQ, T., HACHIMI, H., RAJI, M. & ZERRAF, S. U-Net: deep learning for extracting building boundary collected by drone of Agadir's harbor. International Conference on Digital Technologies and Applications, 2021. Springer, 111-121.

CHEN, B., CHEN, Z., DENG, L., DUAN, Y. & ZHOU, J. 2016. Building change detection with RGB-D map generated from UAV images. *Neurocomputing,* 208**,** 350-364.

CHEN, J., WANG, G., LUO, L., GONG, W. & CHENG, Z. 2020. Building area estimation in drone aerial images based on mask R-CNN. *IEEE Geoscience and Remote Sensing Letters,* 18**,** 891-894.

CHITTURI, G. 2020. Building Detection in Deformed Satellite Images Using Mask R-CNN.

DARANAGAMA, S. & WITAYANGKURN, A. 2021. Automatic Building Detection with Polygonizing and Attribute Extraction from High-Resolution Images. *ISPRS International Journal of Geo-Information,* 10**,** 606.

ESRI. 2022a. *Deep learning model to extract building footprints from high-resolution aerial and satellite imagery.* [Online]. Available: https://www.arcgis.com/home/item.html?id=a6857359a1cd44839781a4f113cd5934 [Accessed].

ESRI. 2022b. *How Mask R-CNN Works* [Online]. Available: https://developers.arcgis.com/python/guide/how-maskrcnn-works/?rsource=https%3A%2F%2Flinks.esri.com%2FDevHelp_HowMaskRCNNWorks [Accessed].

GAVANKAR, N. L. & GHOSH, S. K. 2018. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing,* 51**,** 182-193.

HE, K., GKIOXARI, G., DOLLÁR, P. & GIRSHICK, R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision, 2017. 2961-2969.

IDRIS, I., MUSTAPHA, A., CALEB, O., ALIYU, M. B., OLUMIDE, M. A. & AHMAD, S. H. 2021. Application of Artificial Neural Network for Building Feature Extraction in Abuja. *Educational Research (IJMCER),* 3**,** 09-15.

KOKEZA, Z., VUJASINOVIC, M., GOVEDARICA, M., MILOJEVIC, B. & JAKOVLJEVIĆ, G. 2020. Automatic building footprint extraction from UAV images using neural networks. *Geodetski Vestnik,* 64**,** 545.

LAFARGE, F., DESCOMBES, X., ZERUBIA, J. & PIERROT-DESEILLIGNY, M. 2008. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of photogrammetry and remote sensing,* 63**,** 365-381.

LI, W., HE, C., FANG, J., ZHENG, J., FU, H. & YU, L. 2019. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing,* 11**,** 403.

LI, Y., XU, W., CHEN, H., JIANG, J. & LI, X. 2021. A novel framework based on mask R-CNN and Histogram thresholding for scalable segmentation of new and old rural buildings. *Remote Sensing,* 13**,** 1070.

LIU, P., LIU, X., LIU, M., SHI, Q., YANG, J., XU, X. & ZHANG, Y. 2019a. Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. *Remote Sensing,* 11**,** 830.

LIU, W., XU, J., GUO, Z., LI, E., LI, X., ZHANG, L. & LIU, W. 2021. Building footprint extraction from unmanned aerial vehicle images via PRU-Net: Application to change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 14**,** 2236-2248.

LIU, W., YANG, M., XIE, M., GUO, Z., LI, E., ZHANG, L., PEI, T. & WANG, D. 2019b. Accurate Building Extraction from Fused DSM and UAV Images Using a Chain Fully Convolutional Neural Network. *Remote Sensing,* 11**,** 2912.

RAGHAVAN, R., VERMA, D. C., PANDEY, D., ANAND, R., PANDEY, B. K. & SINGH, H. 2022. Optimized building extraction from high-resolution satellite imagery using deep learning. *Multimedia Tools and Applications***,** 1-15.

RAN, S., GAO, X., YANG, Y., LI, S., ZHANG, G. & WANG, P. 2021. Building multi-feature fusion refined network for building extraction from high-resolution remote sensing images. *Remote Sensing,* 13**,** 2794.

RASTOGI, K., BODANI, P. & SHARMA, S. A. 2022. Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International,* 37**,** 1501-1513.

RONNEBERGER, O., FISCHER, P. & BROX, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 2015. Springer, 234-241.

SADEQ, H. A. 2019. Accuracy assessment using different UAV image overlaps. *Journal of Unmanned Vehicle Systems,* 7**,** 175-193.

SHACKELFORD, A. K., DAVIS, C. H. & WANG, X. Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery. IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IEEE, 1996-1999.

SHEPPARD, C. & RAHNEMOONFAR, M. Real-time scene understanding for UAV imagery based on deep convolutional neural networks. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017. IEEE, 2243-2246.

STILLER, D., STARK, T., WURM, M., DECH, S. & TAUBENBÖCK, H. Large-scale building extraction in very high-resolution aerial imagery using Mask R-CNN. 2019 Joint Urban Remote Sensing Event (JURSE), 2019. IEEE, 1-4.

TEJESWARI, B., SHARMA, S. K., KUMAR, M. & GUPTA, K. 2022. BUILDING FOOTPRINT EXTRACTION FROM SPACE-BORNE IMAGERY USING DEEP NEURAL NETWORKS. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 43**,** 641-647.

TIEDE, D., SCHWENDEMANN, G., ALOBAIDI, A., WENDT, L. & LANG, S. 2021. Mask R- CNN-based building extraction from VHR satellite data in operational humanitarian action: An example related to Covid- 19 response in Khartoum, Sudan. *Transactions in GIS,* 25**,** 1213-1227.

WANG, O., LODHA, S. K. & HELMBOLD, D. P. A bayesian approach to building footprint extraction from aerial lidar data. Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), 2006. IEEE, 192-199.

WANG, Y., LI, S., TENG, F., LIN, Y., WANG, M. & CAI, H. 2022. Improved mask R-CNN for rural building roof type recognition from uav high-resolution images: a case study in hunan province, China. *Remote Sensing,* 14**,** 265.

WEI, S., JI, S. & LU, M. 2019. Toward automatic building footprint delineation from aerial images using CNN and regularization. *IEEE Transactions on Geoscience and Remote Sensing,* 58**,** 2178-2189.

WEN, Q., JIANG, K., WANG, W., LIU, Q., GUO, Q., LI, L. & WANG, P. 2019. Automatic building extraction from Google Earth images under complex backgrounds based on deep instance segmentation network. *Sensors,* 19**,** 333.

WOLF, P. R., DEWITT, B. A. & WILKINSON, B. E. 2014. *Elements of Photogrammetry with Applications in GIS,* McGraw-Hill Education.

XIONG, J., CHEN, T., WANG, M., HE, J., WANG, L. & WANG, Z. 2022. A Method for Fully Automatic Building Footprint Extraction From Remote Sensing Images. *Canadian Journal of Remote Sensing,* 48**,** 520-533.

ZHANG, L., WU, J., FAN, Y., GAO, H. & SHAO, Y. 2020. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. *Sensors,* 20**,** 1465.

ZHAO, K. 2019. Using Deep Neural Networks for Automatic Building Extraction with Boundary Regularization from Satellite Images.

ZHAO, K., KANG, J., JUNG, J. & SOHN, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018. 247-251.

ZHOU, J., LIU, Y., NIE, G., CHENG, H., YANG, X., CHEN, X. & GROSS, L. 2022. Building Extraction and Floor Area Estimation at the Village Level in Rural China Via a Comprehensive Method Integrating UAV Photogrammetry and the Novel EDSANet. *Remote Sensing,* 14**,** 5175.