



Article

Mapping of Dwellings in IDP/Refugee Settlements from Very High-Resolution Satellite Imagery Using a Mask Region-Based Convolutional Neural Network

Getachew Workineh Gella ^{1,*} , Lorenz Wendt ¹, Stefan Lang ¹, Dirk Tiede ¹ , Barbara Hofer ¹ , Yunya Gao ¹ and Andreas Braun ^{1,2}

¹ Christian Doppler Laboratory for Geospatial and EO-Based Humanitarian Technologies (GEOHUM), Department of Geoinformatics—Z_GIS, Paris Lodron University of Salzburg, Schillerstrasse 30, 5020 Salzburg, Austria; lorenz.wendt@plus.ac.at (L.W.); Stefan.Lang@plus.ac.at (S.L.); dirk.tiede@plus.ac.at (D.T.); barbara.hofer@plus.ac.at (B.H.); yunya.gao@plus.ac.at (Y.G.); andreas.braun@plus.ac.at (A.B.)

² Department of Geography, University of Tübingen, Rümelinstr. 19-23, D-72070 Tübingen, Germany

* Correspondence: getachewworkineh.gella@plus.ac.at

Abstract: Earth-observation-based mapping plays a critical role in humanitarian responses by providing timely and accurate information in inaccessible areas, or in situations where frequent updates and monitoring are required, such as in internally displaced population (IDP)/refugee settlements. Manual information extraction pipelines are slow and resource inefficient. Advances in deep learning, especially convolutional neural networks (CNNs), are providing state-of-the-art possibilities for automation in information extraction. This study investigates a deep convolutional neural network-based Mask R-CNN model for dwelling extractions in IDP/refugee settlements. The study uses a time series of very high-resolution satellite images from WorldView-2 and WorldView-3. The model was trained with transfer learning through domain adaptation from nonremote sensing tasks. The capability of a model trained on historical images to detect dwelling features on completely unseen newly obtained images through temporal transfer was investigated. The results show that transfer learning provides better performance than training the model from scratch, with an mIoU range of 4.5 to 15.3%, and a range of 18.6 to 25.6% for the overall quality of the extracted dwellings, which varied on the bases of the source of the pretrained weight and the input image. Once it was trained on historical images, the model achieved 62.9, 89.3, and 77% for the object-based mean intersection over union (mIoU), completeness, and quality metrics, respectively, on completely unseen images.

Keywords: deep learning; dwelling extraction; refugee settlement; IDPs; Mask R-CNN



Citation: Gella, G.W.; Wendt, L.; Lang, S.; Tiede, D.; Hofer, B.; Gao, Y.; Braun, A. Mapping of Dwellings in IDP/Refugee Settlements from Very High-Resolution Satellite Imagery Using a Mask Region-Based Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 689. <https://doi.org/10.3390/rs14030689>

Academic Editors: Bogdan Zagajewski and Klaus Komp

Received: 6 December 2021

Accepted: 29 January 2022

Published: 1 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As of 2019, there were around 79.5 million persons displaced from their homes as a result of either natural or human-made disasters globally, and, in mid-2021, this number surpassed 84 million [1]. These populations might either stay within their national borders (internally displaced populations (IDPs)), or they may cross international borders (refugees), and a considerable share of these people stay in refugee settlements [2]. The livelihoods of refugees are dependent on support from the host government and international aid organizations. Therefore, information related to the attributes of refugee settlements, such as the residing population, the number of dwellings, and the available infrastructure, has paramount importance for both emergency responses and the long-term socioeconomic planning of refugee camps [3,4].

This information can be collected by being onsite, through physical observation and measurement, and through other data collection approaches, which involve the presence of data collection analysts/experts. This strategy is not efficient in terms of resources, time,

and the safety of the data collectors, especially in camps with proximities to insecure and conflict-stricken areas. More importantly, in camps where frequent and regular monitoring with regular updates is needed, field data collection approaches are almost impractical. In such situations, the use of remote-sensing approaches becomes a viable alternative as an information source for the humanitarian response [2,5,6].

Remote sensing has been applied to monitor the various attributes of IDP/refugee camps, including the infrastructure evolution [7], the environment [8–11], refugee camp expansion [9,12–15], and natural hazard vulnerability analysis [16], towards an estimation of the residing populations within the camps [17–19], using various interpretation, prediction, classification, and modeling approaches. There are previous works that are specific to dwelling detection in IDP/refugee camp [4,20–24]. While these strategies can provide useful results, they always demand the partial or full intervention of experts in the information extraction workflow, as these methods rely on either the creation of an object-based ruleset, semiautomatic thresholding, or manual digitization.

IDPs and refugee camps change dynamically, and they are controlled by an influx of incoming and outgoing populations, which subsequently change the extent of the camp, the dwelling structures, and the population sizes [12,25]. Therefore, the close monitoring of these changes requires frequent information updates and rapid information appraisal for the humanitarian response [2]. Beyond immediate information retrieval, there is also a significant challenge in dwelling feature extraction in refugee camps using a manual approach, which is demonstrated by the inherent characteristics of the dwelling structures themselves. This includes extremely small sizes, but extremely large numbers, and the strong similarity of some of the dwelling features with the surrounding soil or background, as most structures are made from unconventional building materials, such as grass, wood, mud, and plastic. Both situations call for automatic information extraction strategies.

By leveraging advances in computer vision and the proliferation of very high-resolution satellite imagery, deep convolutional neural networks (CNN) play a major role in segmentation and object detection tasks [26]. There have been promising works that use segmentation and detection tasks for automatic building extraction [27–30]. Using the availability of high-resolution images, high-performance computing, and advanced deep learning models, there have also been mappings of buildings at the continental [31] and global scales [32]. According to an in-depth analysis by the authors of [33], the global datasets on built-up areas are less representative for IDP/refugee settlements, which are characterized by a low probability of detection, the underestimation of the built-up extent, and the complete overlooking of a few refugee settlements. A comprehensive assessment on the relevance of Open Street Maps (OSM) for monitoring the sustainable development goals in refugee settlements, by the authors of [34], also concludes that although OSM has relevant thematic layers for the intended goal, the information is mapped long after refugee settlements form, and it is not frequently updated. From the qualitative inspection made on the test site during the writing phase of this report, the underrepresentation of dwelling structures in IDP/refugee camps is also common in well-known generic open datasets, such as OpenStreetMap (OSM), Bing Maps, and Google Maps (Appendix A Figures A1 and A2 for details in specific study sites), and specific thematic layers hosted by entities working on IDP/refugee settlements [35]. All of this justifies the requirements for the detailed mapping of IDP/refugee settlements from very high-resolution satellite images.

Recently, there have been a few encouraging studies dedicated to the deep-learning-based dwelling extraction of refugee camps. The authors of [36] mapped dwelling features with a combination of object-based image analysis (OBIA) approaches with CNNs. Similar to this research, the authors of [37] used a CNN model to count the dwelling features in the Rukban area at the border between Syria and Jordan. To use a deep learning approach as a dwelling extraction pipeline in the operational setting of the humanitarian response, the authors of [38] tested a Mask region-based convolutional neural network (Mask R-CNN) to map IDP/refugee dwellings in Sudan, where they observed model performance variations based on the input image and the test site characteristics. Similarly, after investigating the

potential of CNN approaches for dwelling detection in comparison with OBIA strategies in Cameroon, the authors of [39] recommended a further study on the temporal and spatial transferability, with multisource and multitemporal imagery. While these findings have provided promising results, our research answers the following research questions, which still remained unanswered:

- Transfer learning through domain adaptation [40]: Given that deep learning models are data driven, can we gain leverage from openly available pretrained weights generated from non-remotely sensed datasets through domain adaptation for dwelling detection? Which datasets perform better with regard to the provision of weights? Is there a pronounced performance difference if we train the model from scratch?
- Temporal transferability: Given that camps are very dynamic (Figure 1), is it possible to extract dwelling features using a model trained on images obtained in the past, every time new imagery is obtained?

By evaluating the temporal transferability in combination with transfer learning through domain adaptation, we believe our study makes a profound contribution to the rapid mapping of IDP/refugee settlements to support operational Earth-observation-based humanitarian responses.

2. Materials and Methods

2.1. Study Site

Minawao is situated in the northern part of Cameroon. According to the [41], this study site hosted a large number of refugees. The camp consists of some facility buildings, and a large number of UNHCR-issued standard tents and tukuls. UNHCR-issued standard tents may vary in size on the basis of the intended family size and the phase of camp transition [42]. Tukuls are dwelling structures made from either wood or thatch, which have circular, rectangular, or square shapes, with either conical, pitched, or pyramid-shaped rooftops [43]. The camp is situated in a semiarid environment, where the dry season lasts from November to March. The rest months are grouped to the wet season, with a small amount of rainfall that peaks in August, where the monthly total reaches 228 milometers [44].

2.2. Data, Processing, and Sample Preparation

The study used very high-resolution WorldView-3 and WorldView-2 images, with spatial resolutions of 0.5 and 0.3 m, respectively, in the panchromatic band. Table 1 shows the specifications of the utilized images. The images of the Minawao refugee camp in Cameroon were taken at different times that accounted for both the dry and wet seasons of the test site. Though the images were obtained as part of a remote-sensing-based operational humanitarian response, the specific imaging years were selected on the basis of the data availability in our in-house image archives, and on the possibility of capturing the spatiotemporal dynamics of the dwellings in the test sites. The annotated shape files used for the model training and testing were obtained as part of the long-term engagement of our research group in generating dwelling features from very high-resolution satellite images for humanitarian responses.

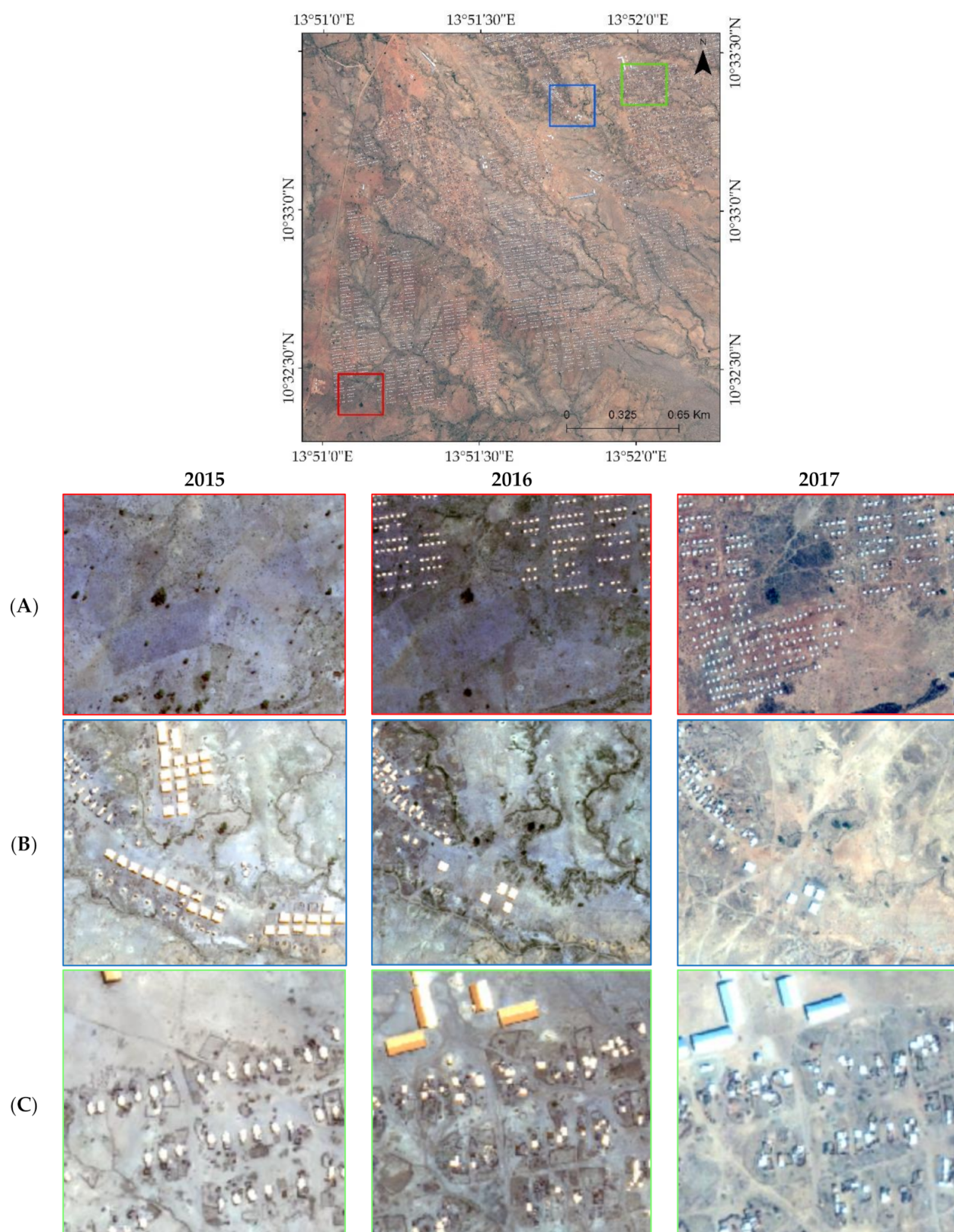


Figure 1. IDP/refugee camp dwelling dynamics through time: Top (full camp) and (A) indicate a rapid spatial expansion of dwelling structures; (B) indicates changes in presence and type of dwellings, as dwellings in 2015 images were dismantled, and tunnel-shaped dwellings were replaced by rectangular structures; and (C) indicates changes in dwelling type (structure) and number. Border colors correspond to the same-colored boxes on the full camp image, where the subsets were taken.

The labels were generated using both OBIA, manual digitalization, and a combination of OBIA with manual digitization for the postprocessing. Before using the annotated samples, quality assurance measures were undertaken by checking the congruency of the annotated dwelling feature boundary with the dwellings on the image, correction of the geometry, screening for the presence of systematic shifts, and the attribute correction of mislabelled objects. They were then converted to a binary raster file, with the same resolution as the input satellite images, for further operation.

Table 1. Summary of utilized images.

Date	Sensor	Ground Sampling Distance (m)	Bit Depth	Processing Level
10 March 2015	WorldView-3	0.5	16 bits	Ortho ready standard product
21 June 2015	WorldView-3	0.5	16 bits	Ortho ready standard product
3 June 2016	WorldView-3	0.3	16 bits	Ortho ready standard product
12 February 2017	WorldView-2	0.5	16 bits	Ortho ready standard product

As is indicated in Table 1, the WorldView-3 and WorldView-2 images were obtained as an ortho-ready standard product. Products with this processing level from DigitalGlobe are radiometrically corrected and projected to a plane surface using user-defined map projection and datum [45]. The radiometric correction includes the removal of sensor-related anomalies and a further conversion to absolute radiometric numbers, while the geometric operations include corrections for errors in the spacecraft orbit position, the attitude uncertainty, the effects of the earth curvature, and the panoramic distortion.

As the obtained images are radiometrically corrected, under the assumption that deep learning models are robust for the scene variation that comes from variations in the sun illumination, a further atmospheric correction was not applied on the images. Pan sharpening was produced using the Gram–Schmidt algorithm [46]. As DigitalGlobe ortho-ready standard products are projected to the standard plane without accounting for the local topography, the orthorectification was performed using the advanced spaceborne thermal emission and reflection (ASTER) digital elevation model (DEM), accessed from USGS Earth Explorer [47]. As the last step of the image pre-processing step, before ingesting to the model, all the images were linearly scaled to 8-bit RGB images, with the enhancement conducted by using histogram stretching, with 2.5 standard deviations. As indicated in Figure 2A–C, all of the image and label pairs, together with the annotated binary raster image, were partitioned into smaller image tiles of 256-by-256 pixels, in height and width, respectively.

Though CNNs, in general, and Mask R-CNNs, in particular, can handle smaller as well as larger image tiles than the image tile we used, the sizes of the tiles were chosen mainly by considering the computational efficiency and the provision of a model to better learn the spatial features [48,49]. Subsequently, individual object masks and bounding boxes were generated from the respective label tiles. Before training, these image tiles were randomly shuffled with a user-defined seed value (to reproduce the results in subsequent experiments), partitioned to the training, validation, and testing sets, with ratios of 60, 20, and 20%, respectively.

The proper train–test ratio is acknowledged in machine learning research, which [50,51] has investigated its impact on the model performance by randomly assigning split ratios, while the authors of [52] implemented algorithmic approaches, which further require data clustering. As in both cases, there is no hard rule, and we have allotted large samples to the training for two reasons. The first reason is to allow the model to learn the existing dwelling-feature-attribute heterogeneity from large datasets. The second reason is to reduce the chance of the overfitting of the very deep model during transfer learning through domain adaptation. The training set is used to adjust the model weights and biases during training, while the validation set is used for the hyperparameter tuning (monitoring model

performance during training and selecting optimal combinations) [53]. The test set is used to evaluate the generalization performance of the trained model on unseen data.

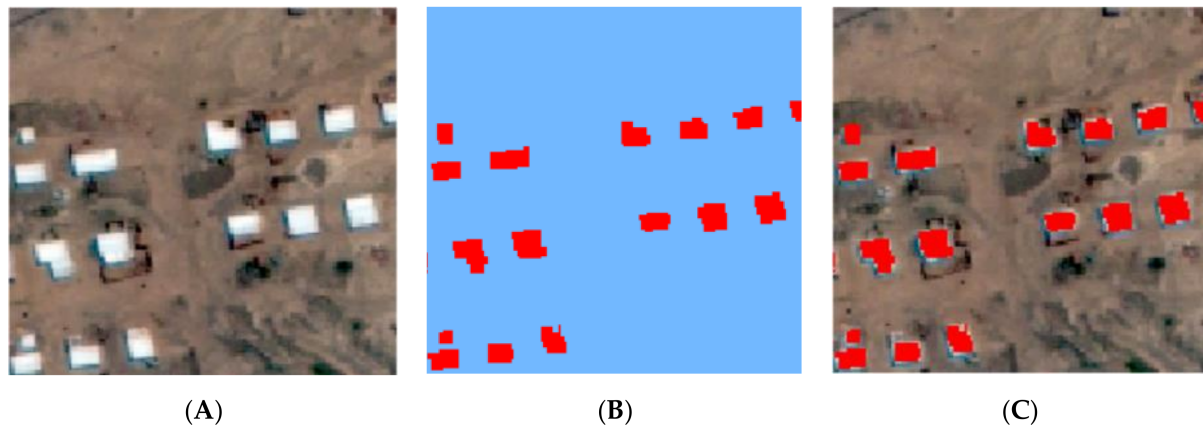


Figure 2. Sample images: (A) input image is shown in RGB; (B) label raster, where blue indicates background, and red indicates dwelling features; and (C) labeled raster on top of an input image with RGB color.

For transfer learning through domain adaptation, the study used two pretrained model weights, which were generated from a large number of non-remotely sensed datasets: Common Object Context (COCO) [54] and ImageNet [55]. The weights from COCO were generated from a large number of images with 81 object classes, while the ImageNet weights were generated from more than 50 million images, with more than 20000 categories. The ImageNet weights, contributed by the authors of [56], were obtained from an open repository [57], while the COCO weights were obtained from [58].

2.3. Model and Training Procedure

For the overall training and testing operations, an instance segmentation Mask R-CNN [59] was used. The model takes input images, object masks with class labels, and the respective bounding boxes of each object, and it produces individual object masks, object classes with bounding boxes, and detection probabilities. The model evolved from a combination of object-detection models [60–63] and semantic segmentation models [64,65].

The model architecture is mainly built from three main engines: feature extraction modules with ResNet architecture [56]; a regional proposal network (RPN) [60]; and a mask head. The feature extraction module generates a stack of abstract features with different levels of spatial detail on the basis of the provided anchor scale. The use of anchor ratios and anchor scales is mainly intended to accommodate the existing size and dimension variations of the objects within the study site. For example, the dwelling features range from small 2-m² huts, to large facility buildings (~400 m²). In order to accommodate this, the study used anchors of sizes 4, 8, 16, 32, and 64, which the feature extraction module uses to produce an abstract feature layer of $[B, A, C, W, H]$. To accommodate the length and width variations or the aspect ratios, the anchor sizes were further scaled with ratios of 0.5, 1, and 2, respectively. Thereby, B is the batch size, which depends on the parameter set of the number of image chips fed into the feature extraction module; A is the number of anchor scales; C is the number of channels, which is determined by the number of channels in the last convolutional layer; and H is the image height and width of the output feature map. Using generated features, the RPN layer generates candidate object proposals. These object proposals were further passed through a region of interest align (RoIA) module, which produced fixed-size feature maps. These fixed-size feature maps, aligned to the region proposals, were simultaneously fed into two independent branches. The first branch has two fully connected subnetworks: a bounding box regressor and a bounding box SoftMax, which further yielded bounding box coordinates and bounding box

classes, respectively. The second branch is a fully convolutional network [66] that produces object masks (Figure 3 for implementation detail).

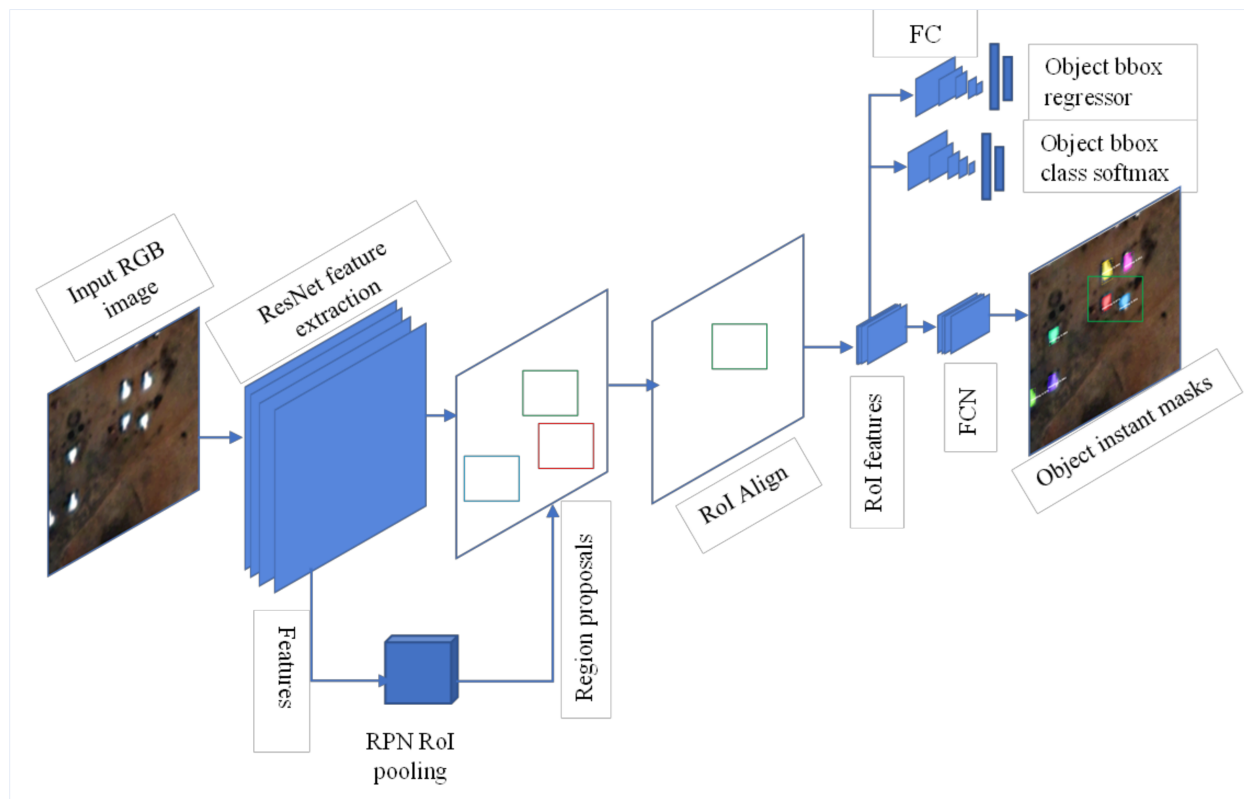


Figure 3. Mask R-CNN architecture.

During the training phase, the model is optimized with stochastic gradient descent (SGD) [67] by recursively adjusting the cost from the multitask loss function [59,62] provided in Equation (1):

$$L = L_{bbx} + L_{cls} + L_{msk} \quad (1)$$

where L is the multitask loss, and L_{bbx} and L_{cls} are the bounding box and class loss, respectively. For the bounding box loss, the $L1$ smooth loss [62] is implemented for both the regional proposal network (RPN) and the Mask R-CNN mask object bounding box regression tasks. The L_{msk} is the mask loss [59]. For both the bounding box and mask classes, a sparse categorical cross-entropy loss was implemented.

To achieve the objective of checking the performance of transfer learning through domain adaptation, during the training phase, the model was trained by two weight initialization approaches. The first approach starts the training from scratch, where the model weights are initialized from a random normal distribution. The second approach is domain adaptation through transfer learning, using pretrained weights generated from the COCO [54] and ImageNet [55] datasets. Both sources of weights are designed to segment objects for a large number of classes. As the current study has two classes (dwelling and background), the last layer weights of the pretrained weights were excluded from usage. This includes the weights for the bounding box regressor, the bounding box class logits, and the object mask. The final SoftMax layer is removed and replaced with SoftMax, which predicts the probabilities of the two classes. Then, the overall model weights are hierarchically fine-tuned. Firstly, the head layers are fine-tuned with a learning rate of 0.01, followed by the first four layers, with a learning rate of 0.001. During this phase, the weights of the other layers were set to freeze. While transferring the learned weights from a deep model trained with a substantially large-source dataset, freezing the first few layers boosts the accuracy [68,69]. Finally, all the layers were updated with a learning rate of

0.0001. This approach enables the model to first learn the task-specific abstract features, and then slowly transfer the model knowledge to learn the features from a new target task.

In both cases, though the model has many hyperparameters, the values of the ones that we consider to be the most sensitive, are optimized. These are the minimum confidence of detection, the minimum threshold for non-maximum suppression (NMS) [60,70], and the learning rate. The NMS values control the double detection of a single object as overlapping features, as well as the proliferation of false positives, especially the false inclusion of the background as a part of the dwelling features. The ideal way to select the optimal hyperparameters of the model would be through a grid search strategy; however, given that the model is computationally demanding, the optimization was performed by consulting the hyperparameters reported in previous studies [71,72], through trial and error.

To achieve the objective of the testing, which is the transferability of the trained model to learn across time, the model training, testing, and the transferring procedure, presented in Figure 4, were implemented. At first, the model was trained and tested on a one-time older image and was transferred to newly obtained images from both the same season and from different seasons. To understand the performance difference of the mixing images from both seasons, the model was first trained and tested on the combined images obtained from the wet and dry seasons of 2015, and then transferred to unseen images obtained in June 2016 and February 2017. Though both experiments can provide insights into the impacts of the seasonal variation on the temporal transferability, to account for the combined effects of the time gap and the seasonal variation, another model was trained, with mixed images from 2015 and 2016 in the same season, and from 2015 and 2016 from a different season, and all were transferred to 2017. In addition to these experiments, the temporal transferability was also tested with the further fine-tuning of the trained model with a few samples from the target images (20% of the samples), as implemented in [38].

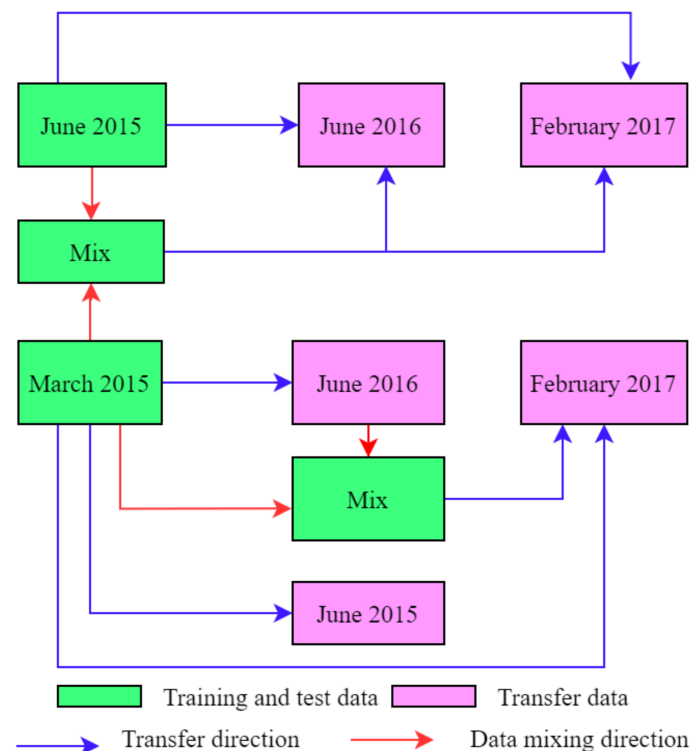


Figure 4. Model training, testing, and transferring procedure.

To evaluate the segmentation performance of the model, the pixel label precision, recall, and F-1 score [73], as well as the object-based mean intersection over union (*MIoU*) were used, which are indicated in Equations (2)–(6). The F-1 score, also called the “dice

score”, provides a measure for the holistic performance of the model by balancing the recall and precision scores. The *MIoU*, which is also called the “Jaccard index”, provides the geometric fit of the segmented and reference dwelling features [74]. The evaluation metrics were computed as:

$$\text{Precision } (P_r) = \frac{T_p}{T_p + F_p} \quad (2)$$

$$\text{Recall } (R_c) = \frac{T_p}{T_p + F_n} \quad (3)$$

$$F-1 \text{ score } (F1) = 2 * \frac{(P_r * R_c)}{(P_r + R_c)} \quad (4)$$

$$\text{Mean Intersection over union (MIoU)} = \frac{1}{N} \sum_{i=1}^N IoU \quad (5)$$

$$\text{Intersection over Union } (IoU) = \frac{(P \cap R)}{(P \cup R) + e} \quad (6)$$

where T_p , F_p , and T_n are the true positive, false positive, and false negative pixels, respectively, while N is the number of image tiles used in the test; the P and R are predicted, and the reference dwelling features or object masks, and e is a constant for a safe division, which is 10^{-6} .

A large number of extremely small-size buildings and a few larger buildings usually dominate IDP/refugee settlements. If the model performs well to detect larger buildings, and fails to accurately localize and detect smaller ones, the object-based and pixel-based accuracy metrics could be misleading. Hence, accuracy metrics that account for the number of detected objects, and their spatial fit with the reference objects, can provide a full picture of the model performance. In humanitarian responses, the completeness and correctness of the generated dwellings are more important than the pixel accuracy, as every further subsequent decision is based on the numbers and types (size) of the dwellings. To serve this purpose, the completeness, correctness, and quality metrics were computed as in [75] and [76], and are provided in Equations (7)–(10):

$$\text{Correctness } (C_r) = \frac{T_{po}}{T_{po} + F_{po}} \quad (7)$$

$$\text{Completeness } (C_m) = \frac{T_{po}}{T_{po} + F_{no}} \quad (8)$$

$$\text{Quality } (Q_l) = \frac{T_{po}}{T_{po} + F_{po} + F_{no}} \quad (9)$$

$$\text{Object-based F-score } (F1_o) = 2 * \frac{C_r * C_m}{C_r + C_m} \quad (10)$$

where T_{po} are the true positive objects that are detected and the overlap with the reference objects; F_{po} are the false positive objects, which are detected objects that are not actual dwellings; and F_{no} are the false-negative detections, which are the objects that are not detected by the model. The overall workflow was implemented using Python. The core TensorFlow [77] based implementation of the Mask R-CNN model from [58] was integrated, with custom-made input–output, image pre-processing, accuracy, and visualization scripts.

3. Results

The results clearly show that training the model with weight initializations from the weights generated from large-image sets through domain adaptation has profoundly better results than training the model from scratch using the values from the random latent space. It was also learned that, although the results obtained from both the ImageNet and COCO

weight initializations outweigh the results obtained from the random initializations, there are also performance and consistency variations among the pretrained weights. Similarly, transferring the model across time has good potential for predicting the dwelling features from completely unseen new images.

As is indicated in Table 2, when the model was trained by different weight initialization approaches, it showed strong variations in the dwelling detection performance. When the model was initialized with pretrained weights, it performed better than a model initialized with random values from normal distributions. During the testing phase, a model initialized with random values has almost 10% lower MIOU values than a model initialized with ImageNet weights, and 12% lower than a model initialized with weights from the COCO dataset. In terms of the number of detected objects, when the model is initialized with randomly initiated weights, it strongly underestimates the number of dwellings. Additionally, the detected numbers consist of contributions from a large number of false-positive rates (2594 out of 9669 detected dwellings). The overall quality is also almost 20% lower than in the models trained with pretrained weights.

When all the trained models were exposed for completely unseen images from June 2016 and February 2017, almost the same performance scenario was observed. The model trained from scratch achieved lower performance than all the models trained with pre-trained weights. This is evident from both the pixel-based and object-based metrics presented in Table 2. It is also clear that, when we look at it in terms of time, the performance of all the June 2016 model images is better than the observed performance in February 2017, which is almost 10% for the overall quality and the *MIOU* values. Though a model initialized with ImageNet almost performed better in terms of the overall completeness and correctness of the detected objects in the testing phase, in all the transferability tests, a model initialized with COCO weights showed consistently better performance. Because of this, further transferability test experiments were performed using a model initialized with COCO weights.

Similar to the quantitative measures presented in Table 2, the detection results, presented in Figure 5, also clearly show the better detection performance of the model initialized with pretrained weights. The model trained from scratch has merged nearby objects and the double prediction of a single object, which is an indication of the inability of the RPN module to properly localize the dwelling objects. It also failed to predict even bright objects, which contrast well with the background.

Table 2. Training and transfer performances with different model weight initializations trained on an image from the wet season.

June 2015 trained and tested on samples from the same image													
Model weights	Pixel-based metrics					Object-based metrics							
	R_c	P_r	F1	MIOU	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
Scratch	0.845	0.809	0.804	0.493	10,383	9669	7075	2594	3296	0.682	0.732	0.706	0.546
ImageNet	0.865	0.871	0.857	0.593	10,383	10,015	8843	1172	1513	0.854	0.883	0.868	0.767
COCO	0.856	0.894	0.866	0.613	10,383	10,676	8914	1762	1484	0.857	0.835	0.846	0.733
Trained on June 2015 image and transferred to June 2016 image													
Model weights	Pixel-based metrics					Object-based metrics							
	R_c	P_r	F1	MIOU	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
Scratch	0.792	0.850	0.793	0.476	17553	21601	13087	8514	3864	0.772	0.606	0.679	0.514
ImageNet	0.851	0.886	0.857	0.595	17553	16158	14428	1730	2949	0.830	0.893	0.860	0.755
COCO	0.863	0.901	0.872	0.629	17553	16655	14720	1935	2450	0.857	0.884	0.870	0.770

Table 2. Cont.

Model weights	Trained on June 2015 image and transferred to February 2017 image												
	Pixel-based metrics					Object-based metrics							
	R_c	P_r	F1	MIoU	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
Scratch	0.842	0.730	0.753	0.387	20912	16649	11818	4831	9088	0.565	0.710	0.629	0.459
ImageNet	0.847	0.760	0.784	0.432	20912	17983	15403	2580	5895	0.723	0.857	0.784	0.645
COCO	0.850	0.812	0.820	0.507	20912	18239	15743	2496	5309	0.748	0.863	0.801	0.669

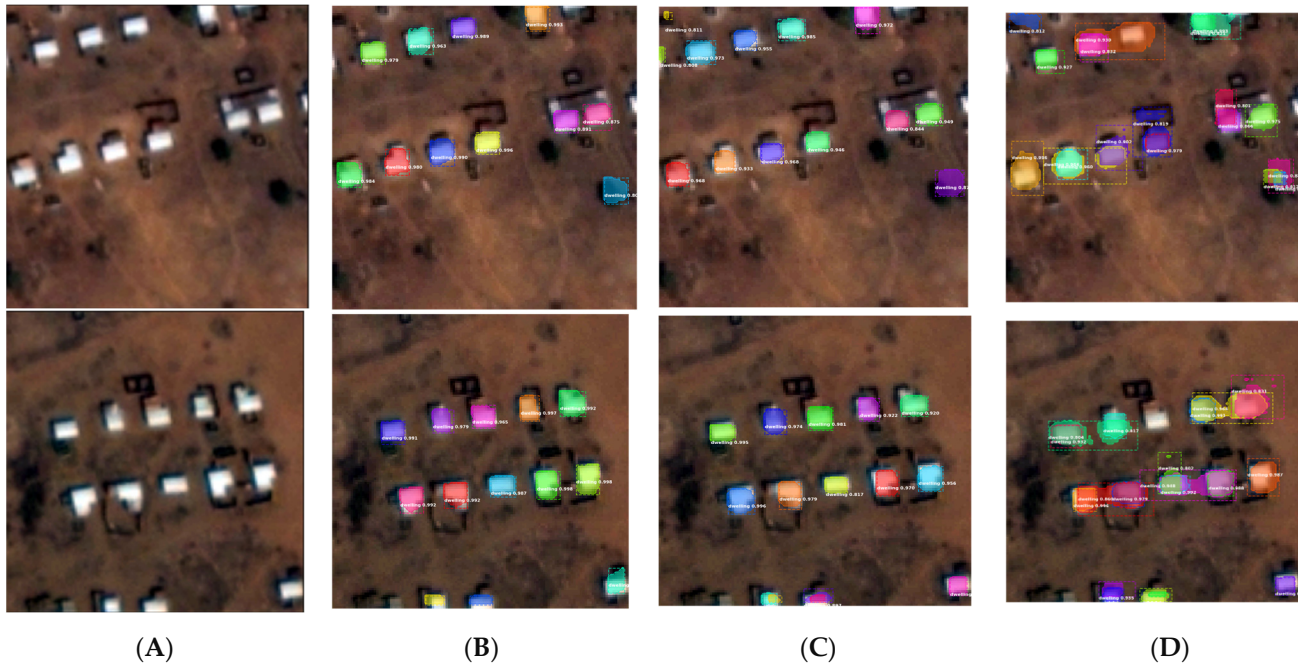


Figure 5. Sample model prediction results of the test dataset using different weight initializations during a training phase for the image taken in June 2015. (A) Reference image, (B,C) are predictions from Mask R-CNN trained with weight initializations from COCO and ImageNet weights respectively, (D) predictions from Mask R-CNN weight initializations with random weights and trained from scratch and rows indicate patches from different locations in the test site.

The results presented in Table 2 were obtained from models trained in June 2015 (wet season), and where the model performance was tested on wet- and dry-season images from June 2016 and February 2017, respectively. To acquire further insight, a model was also trained on the dry-season image obtained in March 2015, and its transferability was tested on the same images from 2016 and 2017, and June 2015 (same year, different season). As indicated in Table 3, except for the *MIoU* value, in all the performance metrics, the transferability test achieved higher performances for the target images from June 2015 (smaller time gap with source images used to train the model), followed by an image from June 2016.

Table 3. Model performance variations with a model trained on dry-season image (March 2015).

Test March 2015	MIoU	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
	0.669	19,002	14,077	13,902	175	4844	0.742	0.988	0.847	0.735
Transfer										
June 2015	0.460	8146	6194	5988	206	2017	0.748	0.967	0.843	0.729
June 2016	0.546	17,553	13,826	12,728	1098	4579	0.735	0.921	0.818	0.692
February 2017	0.465	20,912	17,086	14,845	2241	5992	0.712	0.869	0.783	0.643

If we compare the results presented in Tables 2 and 3, there is performance variation when the season of training and the transfer images are different. The performance is better when the source and target images are taken from the same season. When the temporal gap between the source and the target images widens, this does not always hold (for example, for the year 2017).

As is indicated in Table 4, the combination of images from different seasons provides better results than training the model using a one-time image. Compared to the results presented in Table 3, using a one-time image for the model training, mixing images from two seasons provided an improvement of 6% and 4.1% for the overall quality of the detected objects for June 2016 and February 2017, respectively. Similarly, it also achieved 8.6% and 5.3% improvements for the *MIoUs* for June 2016 and February 2017, respectively. Though the mixing provided better results than using one-time images for training the model, the rate of improvement for newly obtained images with different times was not the same. Compared to the performance of its transferability in the image from June 2016, the same model shows a 6.6% degradation in the quality of the detected objects when transferred to February 2017. Given that the mixed images contain both seasons, which is the common denominator for both target images, this performance variation could be attributed to the temporal gap.

Table 4. Performance of the model trained with images mixed from two seasons of the same year.

Test, June and March 2015	<i>MIoU</i>	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
	0.625	28,656	24,782	22,693	2089	5664	0.800	0.916	0.854	0.745
Transfer										
June 2016	0.626	17,553	15,928	14,247	1681	3059	0.823	0.895	0.857	0.750
February 2017	0.518	20,912	19,569	16,532	3037	4580	0.783	0.845	0.812	0.684

Training the model on combined images from two seasons in 2015 and comparing its transferability in 2016 and 2017 with different seasons may not indicate the impact of a temporal gap on the model transferability. The transferability was also tested by training the model on images from the dry season of 2015 and the wet season of 2016, and, after transferring them to 2017, it yielded an improved performance (Table 5). As the temporal gap of the images decreased, the quality of the detected objects increased, from 68.4% to 71.2%. Similarly, the *MIoU* values show an increment from 51.8 to 53.2%.

Table 5. Model performance trained with mixed images from two seasons with different years.

Test, March 2015 and June 2016	<i>MIoU</i>	Ref count	Pred count	T_{po}	F_{po}	F_{no}	C_r	C_m	$F1_o$	Q_l
	0.466	8146	6658	6326	332	1614	0.797	0.950	0.867	0.765
Transfer										
February 2017	0.532	20,912	17,104	15,440	1664	4572	0.772	0.903	0.832	0.712

All of the temporal transfer results presented in Tables 2–5 were without any further fine-tuning of the trained model. As can be seen in Table 6, in all the experiments, the further fine-tuning of the trained model, with little additional data, produced improvements in the *MIoU* values ranging from almost 1% to 6%. Despite its consistent improvement in the *MIoU* values, the impacts on the overall quality of the detected objects are mixed. For the transfers with wider temporal gaps and a cross-season between the source and the target images (from 2015 to 2017), there was an improvement in the overall quality of the detected dwellings. When the source and the target images have the same season and a relatively small temporal gap, the performance improvement is either very small, or there is a degradation in the overall quality of the detected dwellings. This is evidenced by

the transfer test results from June 2015 to June 2016, and from March 2015 to June 2015 (Table 6).

As a general observation, from all the transferability tests obtained from the model trained with weight initializations from the COCO dataset, there is an underestimation of the absolute dwelling counts, which is characterized by the prevalence of false-negative dwellings rather than false positives (wrongly detected non-dwelling features). For all the transferability test sites, it took the model less than one hour to finish the inference for the whole study site (Figure 6), when executed on a desktop computer equipped with an Intel(R) Core(TM) i5-8500 CPU @ 3.00GHz.

Table 6. Dwelling detection improvement after further fine-tuning with a few samples from the target image.

Training (Source)	Transfer (Target)	MIoU			Q_I		
		Without Fine-Tune	With Fine-Tune	% Δ	Without Fine-Tune	With Fine-Tune	% Δ
June 2015	June 2016	0.629	0.638	0.93	0.770	0.762	−0.77
June 2015	February 2017	0.507	0.551	4.39	0.669	0.720	5.05
March 2015	June 2015	0.460	0.508	4.83	0.729	0.734	0.52
March 2015	February 2017	0.465	0.525	5.98	0.643	0.705	6.18

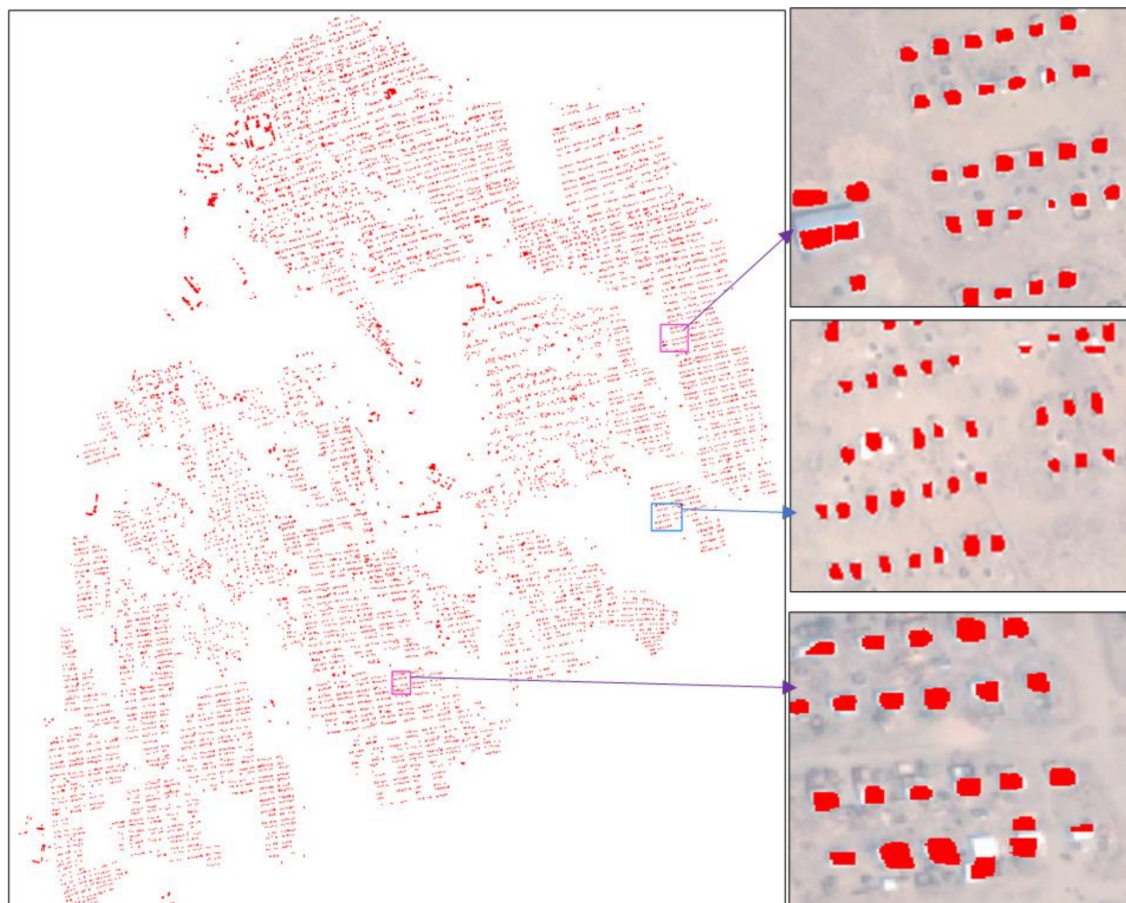


Figure 6. Model (initialized with COCO weights) prediction on unseen image from February 2017 for the full camp extent, with detailed subset RGB images overlaid with detected dwellings.

4. Discussion

As is indicated in Table 2, all of the test results obtained from the model with the weight initialization from the pretrained weights performed much better than the results

obtained from training the model from the scratch (almost 18.6–25.6%, in Q_I and 3.4 to 12% in MIoU, based on the images of different sites). As reported in [78], whether the model is initialized with pretrained weights through transfer learning, from either related or distant domains in a specific task, the results are better than starting the model training from random weight initializations. An extended experiment in [79] for medical imaging segmentation also reports that pretraining through domain adaptation improves the segmentation tasks, but the transfer learning was significantly better if it was pretrained with the same domain images. Contrary to these, the authors of [80] present counterarguments with their experiments, which state that, with the exception of speeding up the convergence time, pretraining from ImageNet does not improve the accuracy performance, which is contrary to our findings.

The results reported in the current study are better in both the testing and transferability phases than the model performance reported by [28,81,82] on the Mask R-CNN based on building detection, while the authors of [83] report that the model performance results are slightly better than the results in the current study. The performance variation could be attributed to the variability in the input data in terms of the size and the proper selection of the model hyperparameters. It should also be noted that, except for [28], which reports object-based accuracy metrics, pixel-based metrics are not always comparable. In the tent detection task similar to the current study, a study by the authors of [37], using a modified VGG16 image segmentation model, achieved an *MIoU* of 67%, and dwelling count errors ranging from 1.55 to 15.88%, based on the input images in the Rukban refugee area, near the Syria–Jordan border. It should be noted that the dwelling counts reported in their study are just gross counts, which do not indicate the percentage of true positives, number of falsely detected non-tent objects, and number of tents their model failed to predict (false negatives), it is not easy to make a fair comparison. Similar to this, the authors of [84], in their study conducted in Darfur using mathematical morphology, report predicted dwelling counts that are in good agreement with the manual counts, but without the disaggregation of the correctly and wrongly predicted dwellings. In a study that combines a self-designed CNN with an object-based image analysis approach in a refugee camp, the authors of [36] report an F-1 score that ranges to 91% in the same study site. This inflated accuracy report is mainly attributed to the inclusion in their study of only bright dwellings with different sizes, without accounting for the small grey and dark dwelling structures, which is in contrast to our study.

Building extraction studies that utilize Mask R-CNN have hybrid findings, which are governed by the complexities of the building structures. For example, a study undertaken in Las Vegas, Paris, Shanghai, and Khartoum [72] reports F-1 scores of 0.881, 0.760, 0.646, and 0.578, respectively, which are lower than our findings obtained by transfer learning through domain adaptation (Table 2). A study with an improved version of Mask R-CNN, implemented in a well-developed urban setting, achieved a pixel-based F-1 score of 0.887, which is in fair agreement with the results we report for the test phase (Table 2). Their study also acknowledges the complexity posed by the extreme variation in the building sizes, shapes, structures, and aspect ratios. As is shown in the comparisons provided in Table 7, the Mask R-CNN yielded comparable, or even better, results than the studies applied in urban settings with less complex building structures than IDP/refugee camps.

Table 7. Comparison of past study results using commonly used pixel- and object-based accuracy metrics.

Studies	R_c	P_r	$F1$	$MIoU$	$F1_o$	Model
[28]	-	-	-	-	0.78	Mask RCNN
[72]	-	-	0.578–0.881 *	-	-	Mask RCNN
[72]	-	-	0.568–0.879 *	-	-	Mask R-CNN with boundary regularization
[38]	-	-	0.390–0.930 *	-	-	Mask R-CNN with point rendering in mask head
[71]	-	-	-	0.48–0.70 *	-	Mask R-CNN
[83]	0.863	0.895	0.878	-	-	Mask RCNN (ResNet101 backbone)
[83]	0.872	0.904	0.887	-	-	Mask RCNN with rotation anchors and replacement of 2D convolution with receptive field block (RFB) in mask head
[76]	-	-	0.923	-	0.936	Multiconstraint graph segmentation
Ours (test) **	0.856	0.894	0.866	0.613	0.846	Mask R-CNN (ResNet101 backbone)
Ours (Transfer 1) **	0.863	0.901	0.872	0.629	0.870	Mask R-CNN (ResNet101 backbone)
Ours (Transfer 2) **	0.850	0.812	0.820	0.507	0.801	Mask R-CNN (ResNet101 backbone)

* results from model initialized with COCO weights, ** based on complexities of different test sites included in the study.

One of the general observations in all the temporal transferability tests is the underestimation of the dwelling counts, which is characterized by the prevalence of the false negatives over the false positives (Figure 7). This is because of the nature and evolution of some dwellings within the period between the source and the target image sensing dates [36], for example, tukuls. As tukuls are dynamic, there is no sufficient number of tukuls in the 2015 images and that were proliferated in the testing phase, which creates a sample imbalance during the model training and a distribution problem in the transferability phase. This is most evident in the June 2016 and February 2017 images. The authors of [23] also note the challenges posed from the variations in the dwelling types and the changing geographical variables because of the seasons, for effective model transferability.

As is indicated in the results section, in the transferability test, there are occasions where the model has a lower $MIoU$, but performs better on the overall completeness, correctness, and quality of the detected objects than other tests that have relatively better $MIoUs$. This could be attributed to the detection of dwelling features with inaccurate localizations of the object boundaries, where some dwelling structures are detected, but their boundaries are not precise, especially for dwellings that have half-bright or half-brown or half-grey rooftops (Figure 8A).

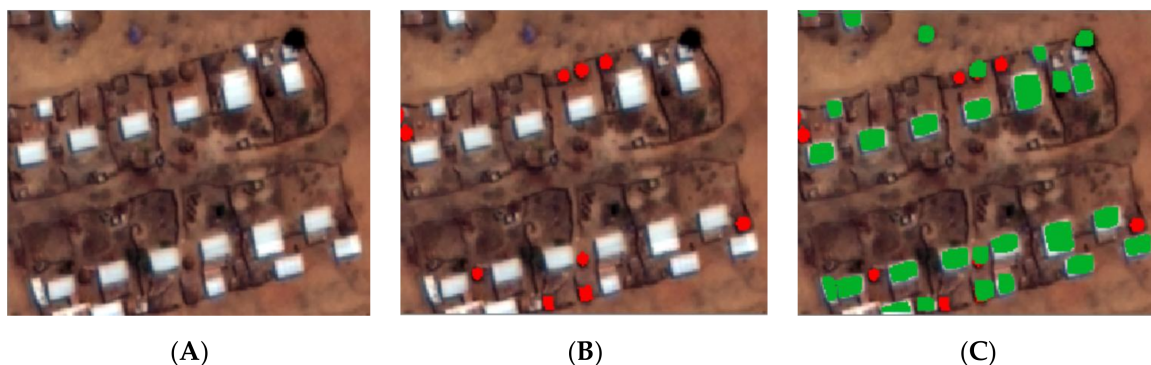
**Figure 7.** Model performance on tukuls: (A) input image from June 2016 shown in RGB; (B) tukul labels on top of RGB; and (C) predicted green and false negatives (red) on top of input RGB image.



Figure 8. Imperfect predictions of the model: (A) prediction with inaccurate object-boundary localization; and (B) missing dwelling features (false negatives).

In humanitarian responses, the prediction of dwelling features with precise boundaries, as is demanded in the cadastral application, may be less important than an accurate number of the dwellings and their locations [28]. These kinds of predictions are still valid for estimating dwelling counts and for mapping the distribution of dwellings in IDP/refugee settlements. More importantly, the tests or transfer experiments with relatively lower *MIoUs*, but with better completeness and quality metrics, indicate the use these metrics with caution in dwelling detection tasks for the humanitarian application. If the model detects relatively few but large-size dwellings better, and fails to properly detect many smaller ones, the model could still have a good *MIoU* value.

As can be observed from the analysis results, the performance of the transferring trained model to detect dwelling objects on completely unseen images is synchronized with two main factors. The first factor is the seasonal similarity of the images taken. As can be easily compared in the results from Figure 9, in the experiments where the training and transfer images were taken in the same season (either wet or dry), the performances are relatively better than those in the cross-season transfer. As can be seen in Figure 10, this can be somewhat overcome by mixing the images from different seasons, depending on the availability of the images and their respective labels. The second factor is the temporal gap between the training and transfer images. The wider the gap between the training and transfer images, the more evident are the changes in the dwelling spectral characteristics and dwelling types. In this study, the individual contributions of the seasonal difference and the temporal gap are not quantified separately.

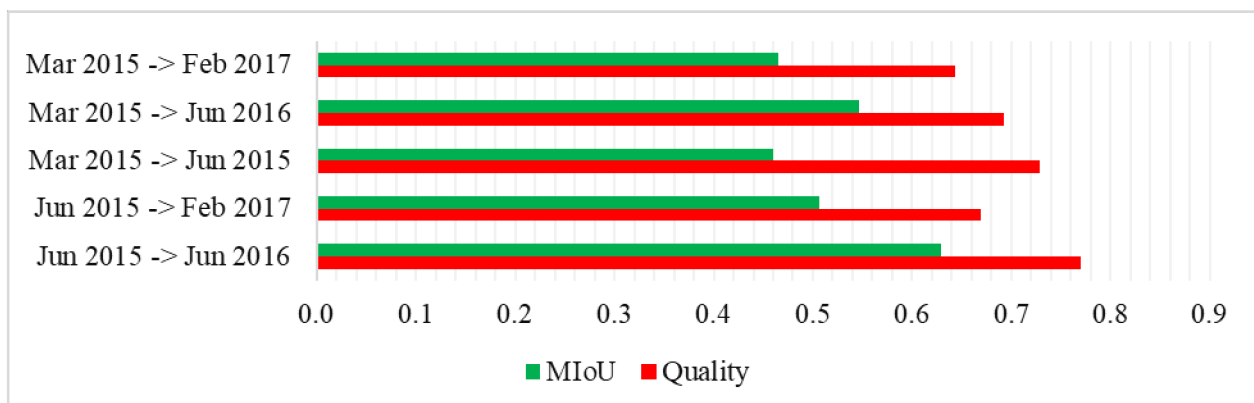


Figure 9. Model transfer performance with the varying seasonal and temporal gaps between the training and transfer images. Initial dates are imaging dates of input images used for training and testing, while end dates are imaging dates of images used for the transfer.

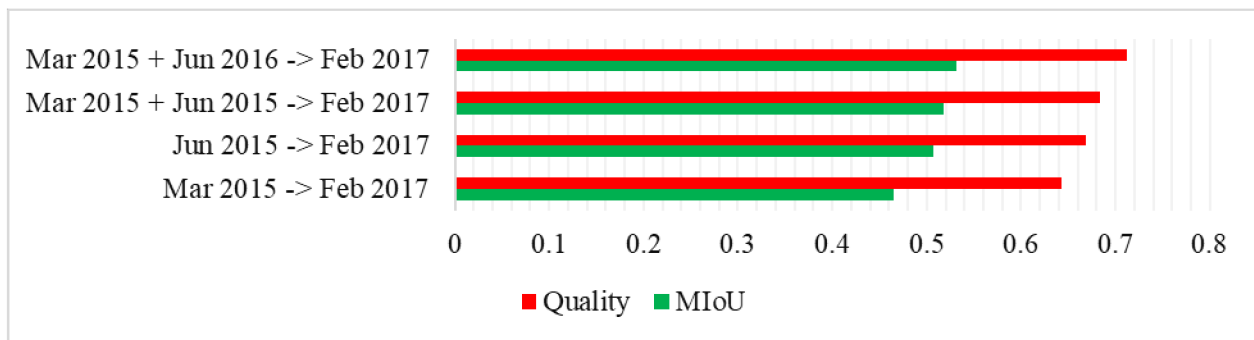


Figure 10. Impacts of combining training samples during training on transfer performance.

Further fine-tuning of the trained model, with few images and respective labels for the temporal transferability, did not provide a consistent result, which is contrary to the findings of [38]. The further fine-tuning of a model trained with the same season as the target image, or with a small temporal gap, did not improve the performance, and even degraded the results. This is mainly related to the overfitting of the model with few datasets, and the quick convergence, without learning additional features in the wider target image, where the learned knowledge is going to be transferred. Further fine-tuning improves the results of transfer learning if there is cross-seasonal variation between the source and target images, or if there is a wider temporal gap. This is because it enables the model to learn new features from the new image, added as a result of the spectral variation by season and dwelling evolution, within the source- and target-image sensing dates.

As is stated in [78], while the use of a pretrained network to boost the model performance has been acknowledged, when the source input and the target input datasets are distant or not related, or the adaptation data are very few compared to the source dataset and the network is deep, transfer learning leads to the overfitting of the model. In general, although Mask R-CNN with pretrained weights has been reported to achieve high accuracy performance in tasks that are not taking overhead satellite images [85], it should be noted that transfer learning through the domain adaptation to target tasks with resembling objects in the source task can produce a better performance [78]. Even pretraining with the same domain is reported to produce a significant performance boost [79].

If we assume that some parts of the false negatives were offset by false positives, the model is providing reasonable dwelling counts (e.g., 6.5% deviation from reference counts on 2017 image (Table 4)). If this is coupled with proper postprocessing, it could be a good way to shorten the information delivery time in an operational setting. As can be learned from [28], in humanitarian responses, accuracy and timeliness are tradeoffs when urgent information retrieval is at stake, which suggests that Mask R-CNN-based dwelling extraction, combined with postprocessing operations, is a relief from the strain of manual digitization. The potential of deep-learning-based dwelling detection to boost the efficiency and accuracy in operational settings specific to the humanitarian response on IDP/refugee camps is also noted in [38].

5. Conclusions

On the basis of these analyses and previous studies, we draw the following conclusions. Transfer learning with domain adaptation from different tasks is much better than training the model from scratch with initializations from a random normal distribution. The temporal transferability of a model is affected by the seasonal similarity and temporal distance between the source and target images, where a trained model is supposed to undertake the dwelling extraction. The model performs better in same-season images than in cross-season images. Similarly, the model performs better when the time gap between the source and target images is smaller. The combined effect of the seasonal variation and the time gap can be achieved by training a model with combined cross-

season images from different years. If constrained by the unavailability of cross-season images, the transferability of a model trained with one-time images could be improved by further fine-tuning the model with a few samples from newly obtained images. As can be observed from all the test and transfer results, the model can detect dwellings with good completeness and quality, ranging from 86–96% and 69–77%, respectively. With this level of accuracy, the model detects all the dwellings in the test site in less than one hour. Therefore, if the model is implemented in an operational setting where there is a request for an immediate information appraisal for an emergency response, deep convolutional neural network-based Mask R-CNN outputs, with minimum postprocessing, generate the relevant information with less time and effort. As the current study is mainly focused on dwelling extraction, transfer learning by domain adaptation, and temporal transferability, we are envisioning further innovations on the model side, which will evaluate the scalability of the model across geographies, and more hybrid sensors.

Author Contributions: Conceptualization, G.W.G., L.W., S.L., D.T. and B.H.; methodology, formal analysis, G.W.G.; investigation, G.W.G.; writing—original draft preparation, G.W.G.; writing—review and editing, G.W.G., L.W., S.L., D.T., B.H., Y.G. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, the Christian Doppler Research Association (CDG), and Médecins Sans Frontières (MSF) Austria.

Data Availability Statement: Data will be provided upon request from authors.

Acknowledgments: We would like to acknowledge the help through Christian Doppler Laboratory for Geospatial and EO-based Humanitarian Technologies (GEOHUM). We are also grateful for the anonymous reviews.

Conflicts of Interest: Authors have declared no conflicts of interest.

Appendix A

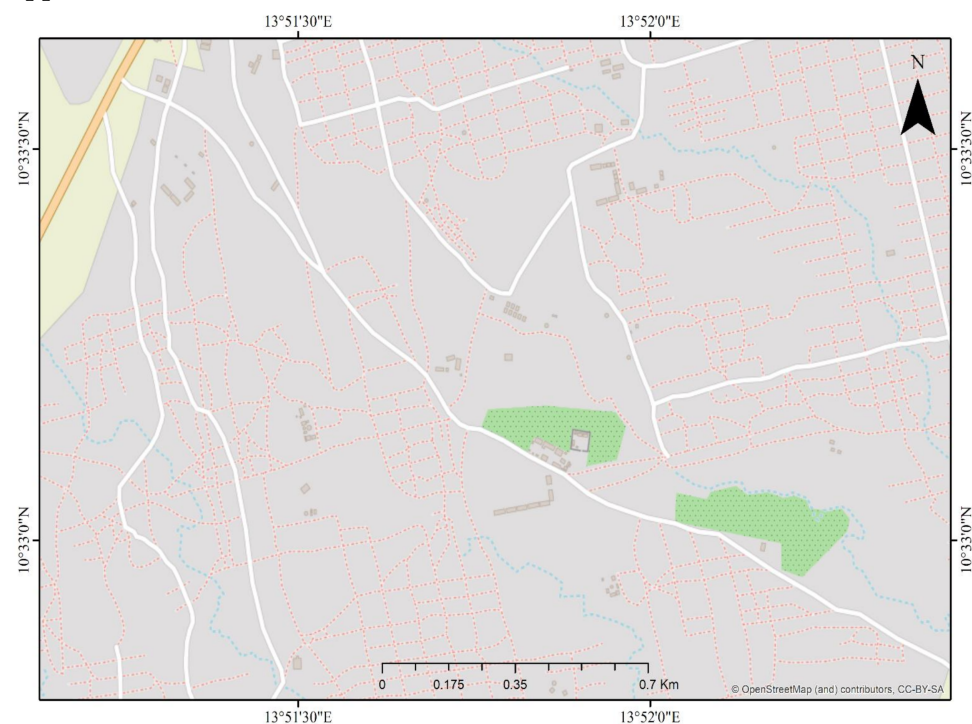


Figure A1. OSM maps for some part of a test site. As indicated on the map, except block outlines and few grey polygons, there are no sufficient building footprint on the map that can help for camp planning and monitoring.

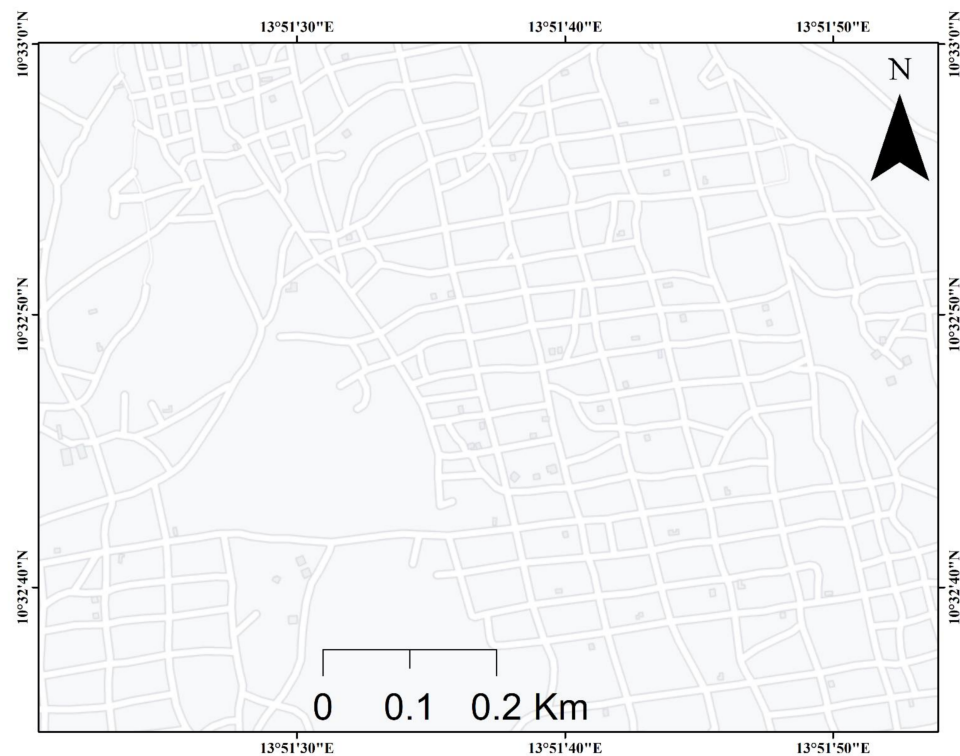


Figure A2. Print screen of Google Maps for some part of a test site. As indicated on the figure, except few leapforningly digitized building footprints, there is no detailed building features in the test site.

References

1. UNHCR Refugee Data Finder. Available online: <https://www.unhcr.org/refugee-statistics/> (accessed on 15 June 2021).
2. Kemper, T.; Heinzel, J. Mapping and monitoring of refugees and internally displaced people using EO data. In *Global Urban Monitoring and Assessment: Through Earth Observation*; Weng, Q., Ed.; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2014; pp. 195–216. ISBN 9781466564503.
3. Kranz, O.; Zeug, G.; Tiede, D.; Clandillon, S.; Bruckert, D.; Kemper, T.; Lang, S.; Caspard, M. Monitoring Refugee/IDP camps to Support International Relief Action. In *Geoinformation for Disaster and Risk Management—Examples and Best Practices*; Altan, O., Backhaus, R., Boccardo, P., Zlatanova, S., Eds.; Joint Board of Geospatial Information Societies (JB GIS): Copenhagen, Denmark, 2010; pp. 51–56. ISBN 978-87-90907-88-4.
4. Wang, S.; So, E.; Smith, P. Detecting tents to estimate the displaced populations for post-disaster relief using high resolution satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *36*, 87–93. [[CrossRef](#)]
5. Lang, S.; Füreder, P.; Riedler, B.; Wendt, L.; Braun, A.; Tiede, D.; Schoepfer, E.; Zeil, P.; Spröhnle, K.; Kulesa, K.; et al. Earth observation tools and services to increase the effectiveness of humanitarian assistance. *Eur. J. Remote Sens.* **2020**, *53*, 67–85. [[CrossRef](#)]
6. Jenerowicz, M.; Wawrzaszek, A.; Drzewiecki, W.; Krupinski, M.; Aleksandrowicz, S. Multifractality in Humanitarian Applications: A Case Study of Internally Displaced Persons/Refugee Camps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4438–4445. [[CrossRef](#)]
7. Tomaszewski, B.; Tibbets, S.; Hamad, Y.; Al-Najdawi, N. Infrastructure evolution analysis via remote sensing in an urban refugee camp—Evidence from Za’atari. *Procedia Eng.* **2016**, *159*, 118–123. [[CrossRef](#)]
8. Braun, A.; Lang, S.; Hochschild, V. Impact of Refugee Camps on Their Environment A Case Study Using Multi-Temporal SAR Data. *J. Geogr. Environ. Earth Sci. Int.* **2016**, *4*, 1–17. [[CrossRef](#)]
9. Braun, A.; Fakhri, F.; Hochschild, V. Refugee camp monitoring and environmental change assessment of Kutupalong, Bangladesh, based on radar imagery of Sentinel-1 and ALOS-2. *Remote Sens.* **2019**, *11*, 2047. [[CrossRef](#)]
10. Leiterer, R.; Bloesch, U.; Wulf, H.; Eugster, S.; Joerg, P.C. Vegetation monitoring in refugee-hosting areas in South Sudan. *Appl. Geogr.* **2018**, *93*, 1–15. [[CrossRef](#)]
11. Hagenlocher, M.; Lang, S.; Tiede, D. Integrated assessment of the environmental impact of an IDP camp in Sudan based on very high resolution multi-temporal satellite imagery. *Remote Sens. Environ.* **2012**, *126*, 27–38. [[CrossRef](#)]
12. Benz, S.; Park, H.; Li, J.; Crawl, D.; Block, J.; Nguyen, M.; Altintas, I. Understanding a rapidly expanding refugee camp using convolutional neural networks and satellite imagery. In Proceedings of the IEEE 15th International Conference on eScience (eScience), San Diego, CA, USA, 24–27 September 2019.

13. Lang, S.; Tiede, D.; Höbling, D.; Füreder, P.; Zeil, P. Earth observation (EO)-based ex post assessment of internally displaced person (IDP) camp evolution and population dynamics in Zam Zam, Darfur. *Int. J. Remote Sens.* **2010**, *31*, 5709–5731. [\[CrossRef\]](#)
14. Lang, S.; Tiede, D.; Hofer, F. Modeling ephemeral settlements using VHSR image data and 3D visualization—The example of Goz Amer refugee camp in Chad. *PFG-Photogramm. Fernerkundung. Geoinf.* **2006**, *4*, 327–337.
15. Tiede, D.; Füreder, P.; Lang, S.; Höbling, D.; Zeil, P. Automated analysis of satellite imagery to provide information products for humanitarian relief operations in refugee camps -from scientific development towards operational services. *Photogramm. Fernerkundung Geoinf.* **2013**, *2013*, 185–195. [\[CrossRef\]](#)
16. Ahmed, N.; Firoze, A.; Rahman, R.M. Machine learning for predicting landslide risk of rohingya refugee camp infrastructure. *J. Inf. Telecommun.* **2020**, *4*, 175–198. [\[CrossRef\]](#)
17. Hadzic, A.; Christie, G.; Freeman, J.; Dimer, A.; Bullard, S.; Greiner, A.; Jacobs, N.; Mukherjee, R. Estimating Displaced Populations from Overhead. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020.
18. Ahmed, N.; Dlpitu, N.A.; Shadhin, M.S.K.; Jaki, M.A.F.; Hasan, M.F.; Islam, M.N.; Rahman, R.M. Artificial Neural Network and Machine Learning Based Methods for Population Estimation of Rohingya Refugees: Comparing Data-Driven and Satellite Image-Driven Approaches. *Vietnam J. Comput. Sci.* **2019**, *6*, 439–455. [\[CrossRef\]](#)
19. Green, B.; Blanford, J.I. Refugee camp population estimates using automated feature extraction. In Proceedings of the Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 10 January 2002.
20. Spröhnle, K.; Tiede, D.; Schoepfer, E.; Füreder, P.; Svanberg, A.; Rost, T. Earth observation-based dwelling detection approaches in a highly complex refugee camp environment—A comparative study. *Remote Sens.* **2014**, *6*, 9277–9297. [\[CrossRef\]](#)
21. Kahraman, F.; Ates, H.F.; Kucur Ergunay, S.S. Automated Detection Of Refugee/IDP TENTS FROM Satellite Imagery Using Two-Level Graph Cut Segmentation. In Proceedings of the CaGIS/ASPRS Fall Conference, San Antonio, TX, USA, 27–30 October 2013.
22. Spröhnle, K.; Fuchs, E.M.; Aravena Pelizari, P. Object-Based Analysis and Fusion of Optical and SAR Satellite Data for Dwelling Detection in Refugee Camps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1780–1791. [\[CrossRef\]](#)
23. Tiede, D.; Lang, S.; Höbling, D.; Füreder, P. Transferability of obia rulesets for idp camp analysis in darfur. In Proceedings of the GEOBIA 2010: Geographic Object-Based Image Analysis, Ghent, Belgium, 29 June–2 July 2010; Addink, E.A., van Coillie, F.M.B., Eds.; ISPRS: Ghent, Belgium, 2010; XXXVIII-4/C7.
24. Aravena Pelizari, P.; Spröhnle, K.; Geiß, C.; Schoepfer, E.; Plank, S.; Taubenböck, H. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens. Environ.* **2018**, *209*, 793–807. [\[CrossRef\]](#)
25. UNHCR Bangladesh Refugee Emergency: Population Factsheet (as of 30 September 2019). Available online: <https://reliefweb.int/sites/reliefweb.int/files/resources/68229.pdf> (accessed on 15 May 2021).
26. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*. [\[CrossRef\]](#)
27. Lu, Z.; Xu, T.; Liu, K.; Liu, Z.; Zhou, F.; Liu, Q. 5M-Building: A large-scale high-resolution building dataset with CNN based detection analysis. In Proceedings of the International Conference on Tools with Artificial Intelligence, ICTAI, Portland, OR, USA, 4–6 November 2019.
28. Tiede, D.; Schwendemann, G.; Alobaidi, A.; Wendt, L.; Lang, S. Mask R-CNN-based building extraction from VHR satellite data in operational humanitarian action: An example related to Covid-19 response in Khartoum, Sudan. *Trans. GIS* **2021**, *25*, 1213–1227. [\[CrossRef\]](#)
29. Pasquali, G.; Iannelli, G.C.; Dell’Acqua, F. Building footprint extraction from multispectral, spaceborne earth observation datasets using a structurally optimized U-Net convolutional neural network. *Remote Sens.* **2019**, *11*, 2803. [\[CrossRef\]](#)
30. Zhu, Q.; Liao, C.; Hu, H.; Mei, X.; Li, H. MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction from Remote Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6169–6181. [\[CrossRef\]](#)
31. Sirko, W.; Kashubin, S.; Ritter, M.; Annkah, A.; Bouchareb, Y.S.E.; Dauphin, Y.; Keyzers, D.; Neumann, M.; Cisse, M.; Quinn, J. Continental-Scale Building Detection from High Resolution Satellite Imagery. *arXiv* **2021**, arXiv:2107.12283.
32. Tiecke, T.G.; Liu, X.; Zhang, A.; Gros, A.; Li, N.; Yetman, G.; Kilic, T.; Murray, S.; Blankespoor, B.; Prydz, E.B.; et al. Mapping the world population one building at a time. *arXiv* **2017**, arXiv:1712.05839. [\[CrossRef\]](#)
33. Van Den Hoek, J.; Friedrich, H.K. Satellite-based human settlement datasets inadequately detect refugee settlements: A critical assessment at thirty refugee settlements in uganda. *Remote Sens.* **2021**, *13*, 3574. [\[CrossRef\]](#)
34. Van Den Hoek, J.; Friedrich, H.K.; Ballasiotes, A.; Peters, L.E.R.; Wrathall, D. Development after displacement: Evaluating the utility of openstreetmap data for monitoring sustainable development goal progress in refugee settlements. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 153. [\[CrossRef\]](#)
35. UNHCR SITE MAP Minawao/Gawar Cameroon. Available online: <https://im.unhcr.org/apps/campmapping/?site=CMRs004589> (accessed on 20 March 2021).
36. Ghorbanzadeh, O.; Tiede, D.; Wendt, L.; Sudmanns, M.; Lang, S. Transferable instance segmentation of dwellings in a refugee camp-integrating CNN and OBIA. *Eur. J. Remote Sens.* **2021**. [\[CrossRef\]](#)
37. Lu, Y.; Koperski, K.; Kwan, C.; Li, J. Deep Learning for Effective Refugee Tent Extraction near Syria-Jordan Border. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1342–1346. [\[CrossRef\]](#)

38. Quinn, J.A.; Nyhan, M.M.; Navarro, C.; Coluccia, D.; Bromley, L.; Luengo-Oroz, M. Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*. [CrossRef]
39. Ghorbanzadeh, O.; Tiede, D.; Dabiri, Z.; Sudmanns, M.; Lang, S. Dwelling extraction in refugee camps using CNN-First experiences and lessons learnt. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2018**, *42*, 161–166. [CrossRef]
40. Long, M.; Wang, J.; Cao, Y.; Sun, J.; Yu, P.S. Deep learning of transferable representation for scalable domain adaptation. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 2027–2040. [CrossRef]
41. UNHCR Cameroun–Extrême Nord Profil de la Population Réfugiée du Camp de Minawao Cameroun–Extrême Nord Profil de la Population Réfugiée du Camp de Minawao. Available online: <https://www.refworld.org/pdfid/553f384e4.pdf> (accessed on 12 April 2021).
42. *UNHCR Shelter Design Catalogue*; Shelter and Settlement Section Division of Programme Support and Management United Nations High Commissioner for Refugees: Geneva, Switzerland, 2016.
43. HHI Satellite Imagery Interpretation Guide Intentional Burning of Tukul. Available online: https://hhi.harvard.edu/files/humanitarianinitiative/files/siig_ii_burned_tukuls_3.pdf?m=1610658910 (accessed on 3 April 2021).
44. World Bank Climate Change Knowledge Portal: For Development Practitioners and Policy Makers. Available online: <https://climateknowledgeportal.worldbank.org/country/cameroon/climate-data-historical> (accessed on 12 February 2021).
45. Digital Globe DigitalGlobe Core Imagery Products Guide. Available online: <https://www.geosoluciones.cl/documentos/worldview/DigitalGlobe-Core-Imagery-Products-Guide.pdf> (accessed on 17 March 2021).
46. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpener. U.S. Patent 6,011,875, 4 January 2000.
47. USGS Surface Radar Topographic Mission (SRTM) Digital Elevation Model. Available online: <https://earthexplorer.usgs.gov/> (accessed on 15 September 2021).
48. Sabottke, C.F.; Spieler, B.M. The Effect of Image Resolution on Deep Learning in Radiography. *Radiol. Artif. Intell.* **2020**, *2*, e190015. [CrossRef] [PubMed]
49. Rukundo, O. Effects of Image Size on Deep Learning. Available online: <http://arxiv.org/abs/2101.11508> (accessed on 25 March 2021).
50. Pawluszek-Filipiak, K.; Borkowski, A. On the importance of train-test split ratio of datasets in automatic landslide detection by supervised classification. *Remote Sens.* **2020**, *12*, 3054. [CrossRef]
51. Nguyen, Q.H.; Ly, H.B.; Ho, L.S.; Al-Ansari, N.; Van Le, H.; Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Math. Probl. Eng.* **2021**, *2021*. [CrossRef]
52. Kahloot, K.M.; Ekler, P. Algorithmic Splitting: A Method for Dataset Preparation. *IEEE Access* **2021**, *9*, 125229–125237. [CrossRef]
53. Xu, Y.; Goodacre, R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J. Anal. Test.* **2018**, *2*, 249–262. [CrossRef] [PubMed]
54. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2015**.
55. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
56. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
57. Francois, C.; Joe, Y.-H.; De Thibault Main, B. Keras Code and Weights Files for Popular Deep Learning Models. Available online: <https://github.com/fchollet/deep-learning-models> (accessed on 1 January 2021).
58. Waleed, A. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. Available online: https://github.com/matterport/Mask_RCNN (accessed on 7 January 2021).
59. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22 October–29 October 2017.
60. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**. [CrossRef] [PubMed]
61. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
62. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 1440–1448.
63. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot Multibox Detector. Available online: <https://arxiv.org/abs/1512.02325> (accessed on 21 January 2021).
64. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
65. Hao, S.; Zhou, Y.; Guo, Y. A Brief Survey on Semantic Segmentation with Deep Learning. *Neurocomputing* **2020**. [CrossRef]

66. Zhang, K.; Han, Y.; Chen, J.; Zhang, Z.; Wang, S. Semantic Segmentation for Remote Sensing based on RGB Images and Lidar Data using Model-Agnostic Meta-Learning and Partial Swarm Optimization. *IFAC-Pap.* **2020**, *53*, 397–402. [\[CrossRef\]](#)
67. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. Available online: <http://arxiv.org/abs/1609.04747> (accessed on 3 January 2021).
68. Soekhoe, D.; Van der Putten, P.; Plaat, A. On the impact of data set size in transfer learning using deep neural networks. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Redondo Beach, CA, USA, 8–11 July 2007; Volume 9897 LNCS, pp. 50–60.
69. Ghafoorian, M.; Mehrtash, A.; Kapur, T.; Karssemeijer, N.; Marchiori, E.; Pesteie, M.; Guttmann, C.R.G.; de Leeuw, F.E.; Tempny, C.M.; van Ginneken, B.; et al. Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; Volume 10435 LNCS, pp. 516–524.
70. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection with One Line of Code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
71. Sebastian, O. Building Segmentation on Satellite Images. Available online: <http://arxiv.org/abs/1703.06870> (accessed on 27 April 2021).
72. Zhao, K.; Kang, J.; Jung, J.; Sohn, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 242–246.
73. Guyon, I.; Bennett, K.; Cawley, G.; Escalante, H.J.; Escalera, S.; Ho, T.K.; Macià, N.; Ray, B.; Saeed, M.; Statnikov, A.; et al. Design of the 2015 ChaLearn AutoML challenge. In Proceedings of the International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015.
74. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*. [\[CrossRef\]](#)
75. Vakalopoulou, M.; Karantzalos, K.; Komodakis, N.; Paragios, N. Building detection in very high resolution multispectral data with deep learning features. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1873–1876.
76. Hui, Z.; Li, Z.; Cheng, P.; Ziggah, Y.Y.; Fan, J. Building extraction from airborne lidar data based on multi-constraints graph segmentation. *Remote Sens.* **2021**, *13*, 3766. [\[CrossRef\]](#)
77. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
78. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 4, pp. 3320–3328.
79. Wen, Y.; Chen, L.; Deng, Y.; Zhou, C. Rethinking pre-training on medical imaging. *J. Vis. Commun. Image Represent.* **2021**, *78*. [\[CrossRef\]](#)
80. He, K.; Girshick, R.; Dollar, P. Rethinking imageNet pre-training. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4917–4926.
81. Zhao, K.; Kang, J.; Jung, J.; Sohn, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* **2018**, *2018*, 242–246. [\[CrossRef\]](#)
82. Hu, Y.; Guo, F. Building Extraction Using Mask Scoring R-CNN Network. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya, China, 22–24 October 2019; pp. 1–5.
83. Wen, Q.; Jiang, K.; Wang, W.; Liu, Q.; Guo, Q.; Li, L.; Wang, P. Automatic building extraction from google earth images under complex backgrounds based on deep instance segmentation network. *Sensors* **2019**, *19*, 333. [\[CrossRef\]](#)
84. Kemper, T.; Pesaresi, M.; Soille, P.; Jenerowicz, M. Enumeration of Dwellings in Darfur Camps from GeoEye-1 Satellite Images Using Mathematical Morphology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 8–15. [\[CrossRef\]](#)
85. Xu, B.; Wang, W.; Falzon, G.; Kwan, P.; Guo, L.; Chen, G.; Tait, A.; Schneider, D. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Comput. Electron. Agric.* **2020**, *171*. [\[CrossRef\]](#)