applied sciences

MDPI

Article

Comparing OBIA-Generated Labels and Manually Annotated Labels for Semantic Segmentation in Extracting Refugee-Dwelling Footprints

Yunya Gao *, Stefan Lang *, Dirk Tiede, Getachew Workineh Gella and Lorenz Wendt

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

* Correspondence: yunya.gao@plus.ac.at (Y.G.); stefan.lang@plus.ac.at (S.L.)

Abstract: Refugee-dwelling footprints derived from satellite imagery are beneficial for humanitarian operations. Recently, deep learning approaches have attracted much attention in this domain. However, most refugees are hosted by low- and middle-income countries where accurate label data are often unavailable. The Object-Based Image Analysis (OBIA) approach has been widely applied to this task for humanitarian operations over the last decade. However, the footprints were usually produced urgently, and thus, include delineation errors. Thus far, no research discusses whether these footprints generated by the OBIA approach (OBIA labels) can replace manually annotated labels (Manual labels) for this task. This research compares the performance of OBIA labels and Manual labels under multiple strategies by semantic segmentation. The results reveal that the OBIA labels can produce IoU values greater than 0.5, which can produce applicable results for humanitarian operations. Most falsely predicted pixels source from the boundary of the built-up structures, the occlusion of trees, and the structures with complicated ontology. In addition, we found that using a small number of Manual labels to fine-tune models initially trained with OBIA labels can outperform models trained with purely Manual labels. These findings show high values of the OBIA labels for deep-learning-based refugee-dwelling extraction tasks for future humanitarian operations.

Keywords: remote sensing; deep learning; semantic segmentation; label noise; OBIA; refugees

Citation: Gao, Y.; Lang, S.; Tiede, D.; Gella, G.W.; Wendt, L. Comparing OBIA-Generated Labels and Manually Annotated Labels for Semantic Segmentation in Extracting Refugee-Dwelling Footprints. *Appl. Sci.* **2022**, *12*, 11226. https://doi.org/10.3390/app122111226

Academic Editor: Anselme Muzirafuti

Received: 24 September 2022
Accepted: 3 November 2022
Published: 5 November 2022

1. Introduction

Forcibly displaced people (FDP) refers to a group of people who leave their home or home region involuntarily due to conflicts, generalized violence, persecution, or human rights violations [1]. The global FDP population surpassed 100 million in mid-2022, which represents more than 1% of the global population [2]. Refugees, as an important category of FDP, refer to a group of people who are outside their home country and have received recognized mandates from the United Nations (UN) [3]. Approximately 86% of global refugees are hosted by low- and middle-income countries which usually cannot provide sufficient living space, medical services, food, water, and sanitation [4]. United Nations High Commissioner for Refugees (UNHCR) [5] together with many non-governmental organizations such as Médecins Sans Frontières (MSF, Doctors Without Borders in English) [6] have provided multiple humanitarian operations, including allocation and distribution of living resources and medical services to protect refugees.

Estimating the refugee population in need is essential for the logistics planning of humanitarian aid [7]. However, in most cases, it is difficult to collect such information in the field due to security and access reasons [8,9]. With the availability of very high spatial resolution satellite imagery, remote sensing has become one of the most powerful tools for population estimation for humanitarian operations over the last decade [8,10–14]. Refugee-dwelling footprints are useful in population estimation [15,16]. Therefore, producing high-

quality refugee dwelling footprints efficiently can save much time for urgent humanitarian operations [17].

Multiple remote sensing image analysis techniques have been applied for this dwelling-extraction task during the last decade. Visual interpretation is thought to be capable of producing the refugee-dwelling footprints with the highest accuracy even though it still includes errors from inter-variability of observations [8]. However, it is labour-intensive and time-consuming [8]. Over the last decade, the OBIA approach has been widely applied to this task. Researchers have developed mature processing routines for producing refugee-dwelling footprints [18–20]. Nevertheless, due to the fast dynamics and diverse ontology of dwellings, the complexity and transferability of OBIA rulesets remain challenging [21].

In the past ten years, the development of deep learning in computer vision [22,23] has boosted a large number of applications in remote sensing [24–26]. However, related research for this extraction task is still limited. One main reason is the unavailability of adequate accurate label data of refugee dwellings [27] because most refugees are hosted by the countries where accurate label data of the footprints are not available in open sources, such as Open Street Map.

With the efforts of researchers and staff from humanitarian services at the Department of Geoinformatics Z_GIS, a large amount of label data for FDP dwellings has been produced by semi-automatic approaches (e.g., OBIA), with limited manual correction as post-processing [8]. Nevertheless, due to the urgency of the humanitarian operations, the produced OBIA labels may exhibit several types of annotation errors. Therefore, the OBIA label data are considered "noisy label data" compared with manually annotated label data (Manual labels). To the best of our knowledge, no research thus far has discussed whether we can make use of the generated tremendous OBIA labels to replace the Manual labels for this extraction task by deep learning approaches.

Currently, there are numerous deep learning models for remote-sensing-based building extraction tasks [28]. However, up to now, very few of them have been applied for this task. O. Ghorbanzadeh et al. [29] designed shallow convolutional neural networks (CNNs) to extract refugee dwellings in the Minawao refugee camp in northern Cameroon. The prediction results prove that CNNs have a high potential in this task. J. A. Quinn et al. [30] selected Mask-RCNN with ResNet-101 as a backbone to extract dwellings in multiple refugee camps. L. Wickert et al. [31] chose a Faster-RCNN model with ResNet-50 as a backbone to count dwellings in refugee camps. D. Tiede et al. [32] selected a Mask-RCNN model with ResNet-50 as a backbone to extract built-up structures in Khartoum in Sudan for fast humanitarian operations. G. W. Gella et al. [33] tested the spatial transferability of U-Net in this extraction task. G. W. Gella et al. [34] selected a Mask-RCNN model with ResNet-101 as a backbone to extract refugee dwellings in Cameroon. Y. Lu and C. Kwan [35] tested the performance of multiple neural networks and found that a fully neural network (FCN) outperforms Mask-RCNN in detecting refugee tents near the Syria–Jordan border.

On the basis of the findings from Y. Lu and C. Kwan [35], this research selected an FCN model for semantic segmentation for all experiments. Semantic segmentation can assign a class label to each pixel in an image to produce a fine-grained delineation of target objects with embedded spatial information [36] and has been proven to be robust in labelling noise in image analysis applications [37,38].

Up to now, many techniques have been developed to handle label noise in deep learning. Karimi et al. [39] reviewed related techniques and categorized them into six major classes, as shown in Figure 1. Recent progress in weakly supervised semantic segmentation (WSSS) enables CNNs to learn from weakly labelled data such as image-level annotations for semantic segmentation tasks. Related techniques such as self-supervised equivariant attention mechanism [40], group-wise learning [41], regional semantic contrast and aggregation [42], and adaptive early-learning correction [43] for WSSS also bring high value to make the best of the OBIA label data for refugee dwelling extraction tasks.

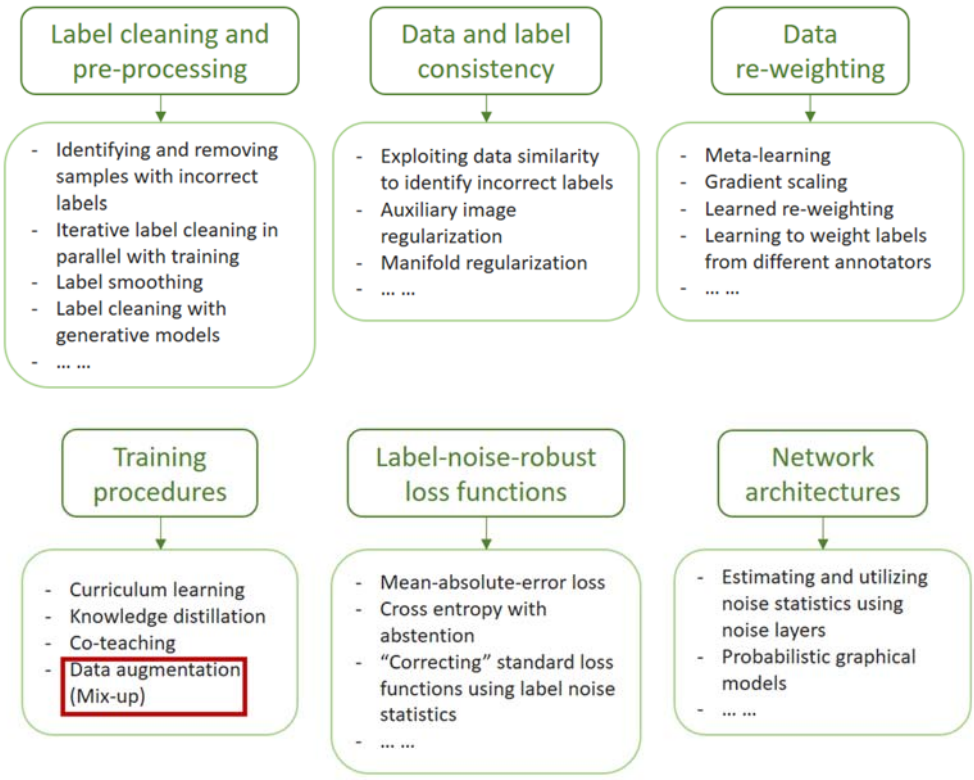

Figure 1. The overview of the state-of-the-art techniques of handling label noise in deep-learning-based approaches. The content was summarized based on Karimi et al. [39]. In this research, we chose "Data augmentation" by intermixing multi-source data to test whether the OBIA labels have potential to replace the Manual labels. The designation of experiments can be found in Section 2.5.

Among all techniques, intermixing training data from multiple sources proved to be a less intuitive but effective approach to increase the robustness of deep learning models for image segmentation [39]. Touzani and Granderson [44] intermixed image data from multiple study sites, various dates, and/or different sensors and trained models with label data from the Microsoft building footprint dataset to extract building footprints from satellite imagery. The results reveal that intermixing training data from multiple sources can produce promising results despite label noise. Inspired by these findings, this research selected intermixing as the strategy to handle the label noise issue as a starting point.

Inspired by this idea, we selected data from two refugee camps (Kule and Nguenyyiel in Ethiopia) from a dry season and a wet season (seen in Section 2.1) and designed multiple intermixing strategies (seen in Section 2.5) to make the comparison between the OBIA labels and the Manual labels in extracting refugee-dwelling footprints from very high-spatial-resolution satellite imagery (Pléiades-1, seen in Section 2.2) by semantic segmentation (U-Net with ResNet-34 as a backbone, seen in Section 2.3). In addition, we used a small amount (i.e., 10%, 20%, 30%, 40%, and 50%) of the Manual labels to fine-tune models trained with the OBIA labels to verify whether we can use a smaller amount of the Manual labels and existing OBIA labels to achieve similar performance as that of the models purely trained with a large amount of the Manual labels (seen in Section 2.5).

The main contribution of this research is that we first prove the OBIA labels, even though they include delineation errors, have very high potential to replace the manual labels, to a certain extent, in extracting refugee-dwelling footprints. Additionally, fine-tuning models trained with the OBIA labels with a small amount of the Manual labels can outperform models trained with purely Manual labels. The currently available OBIA labels from the last decade's humanitarian operations could be fuel for future deep-learning-based refugee-dwelling extraction tasks.

2. Materials and Methods

2.1. Study Sites

Kule and Nguenyyiel refugee camps are located in the Gambella region, Ethiopia. The Kule refugee camp was opened in 2014 in response to the major refugee influx from South Sudan and was fully occupied in 2016 [45]. The Nguenyyiel refugee camp was subsequently opened in 2016 to accommodate the influx of refugees from South Sudan due to the escalation of conflicts in July 2016 [46]. The location of the two camps together with the typical roof ontology of built-up structures can be found in Figure 2.

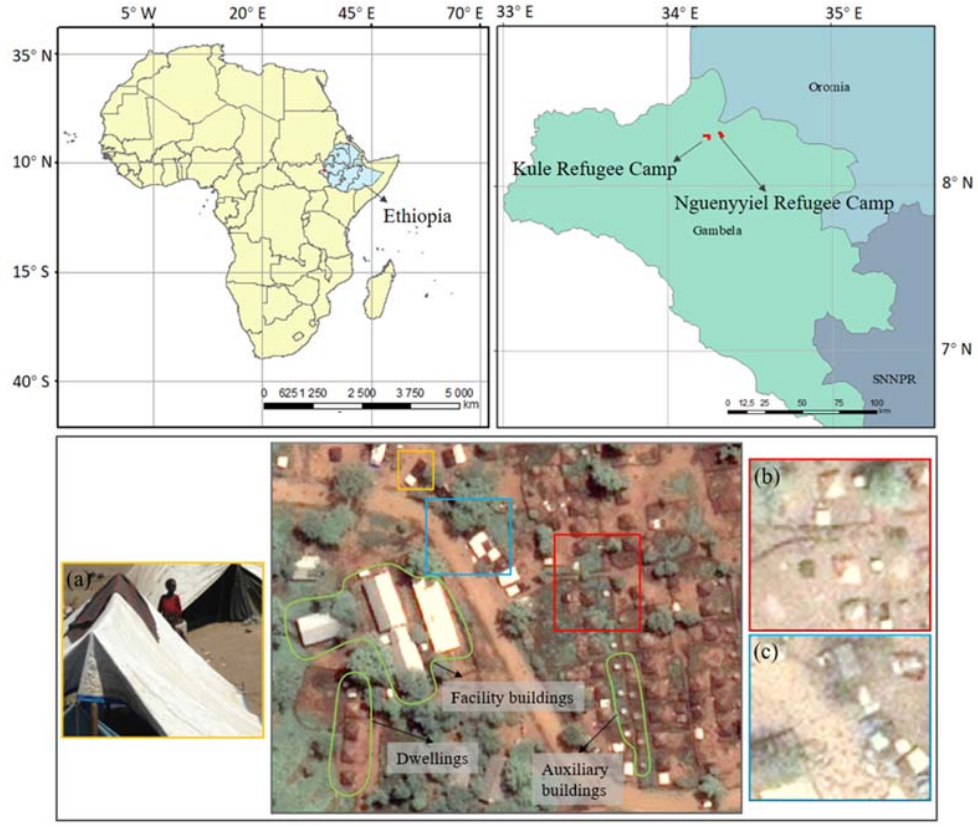

Figure 2. The location of Kule and Nguenyyiel refugee camps together with their typical roof ontology of built-up structures. The subfigure (**a**) is the photo provided by MSF to reveal ground situations. The satellite imagery in the middle was retrieved from the wet season (22 June 2018), while (**b**,**c**) were retrieved from the dry season (24 March 2017). Yellow boundary: an example showing residents putting a brown cloth on the top of refugee tents. Red boundary: an example showing the dwelling roof turned from bright-colored to brownish due to rusting and/or sand. Green boundary: examples of three types of built-up structures. Blue boundary: an example showing trees can occlude built-up structures in satellite imagery.

There are three main types of built-up structures in the two camps. Figure 2 shows several examples of the three types of built-up structures within the green boundary. The first type is for residential purposes, named dwellings. They are usually represented as white rectangles or squares with an area ranging from 20 to 40 m^2 from satellite imagery. The roof ontology of refugee dwellings may be subject to change by local refugee behaviors. For example, they covered dark brown cloth on the rooftop, as shown in Figure 2a (yellow boundary). In addition, the ontology can be changed by naturally rusting or sand. Thus, we can observe many dwellings that have turned to brown from white, as seen in Figure 2b (red boundary). The second type is auxiliary buildings that are generally used for storage or latrines. Their areas are usually less than 10 m^2. The last type is facility buildings, which

are usually white, bluish, brownish, or greyish rectangles whose area is usually larger than 50 m^2.

In Gambella, the wet season usually lasts from April to November, and the dry season spans from December to March [47]. The vegetation, especially trees, can occlude part of built-up structures on the ground, which brings difficulty to the refugee-dwelling extraction tasks. Figure 2c (blue boundary) shows an example where trees occluded built-up structures during the wet season.

2.2. Data Preparation

We chose satellite imagery from the Pléiades-1 sensor with a spatial resolution of 2 m. Considering the common size of dwellings (20 to 40 m^2) in the selected sites, we pan-sharpened the imagery with the panchromatic band from the same source and resampled it to 0.5 m in GeoTIFF format to help the models learn more geometric features. Two retrieval dates represent the dry season (24 March 2017) and the wet season (22 June 2018), respectively. The pixel depth is 16-bit. RGB bands were selected because we found that three input bands with pretrained weights from ImageNet datasets outperform four bands during the preliminary phase. All satellite images use WGS1984 as the geographic coordinate system and UTM 36N as the projected coordinate system.

The OBIA labels of the two camps include only a small part of auxiliary buildings. To make a comparable and systematic comparison, we detect only facility buildings and dwellings in this research for both types of labels. While the footprints of dwellings can be used for population estimation [7], the footprints of facility buildings could be auxiliary information for humanitarian operators. Due to the limited number of facility buildings, they can be easily deleted as well, if such information is not useful in operations. For these reasons, we use binary classes in this research, which are *Target built-up structures*, including both refugee dwellings and facility buildings, and *Background*.

The OBIA labels were provided by the Department of Geoinformatics—Z_GIS of Paris Lodron University of Salzburg. Manual labels were manually annotated by one expert and crosschecked by another expert using a GIS software package. The annotated polygon data were converted to GeoTIFF format with a spatial resolution of 0.5 m. Figure 3 presents the built-up structures in two camps during two seasons together with the OBIA-generated labels and the Manual labels. We can observe the OBIA labels have high quality, in general, but include some annotation errors, such as missing some built-up structures.

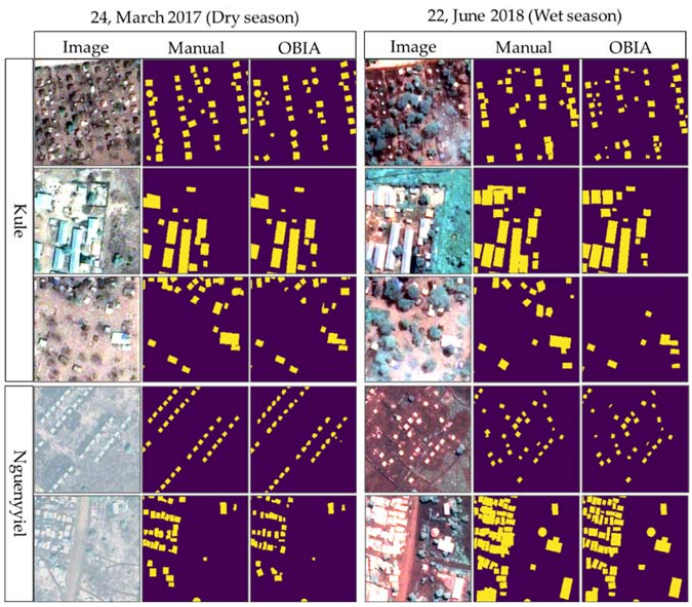

Figure 3. The examples of the OBIA labels and the Manual labels of built-up structures in the Kule and Nguenyyiel refugee camps during the dry and the wet seasons.

During data preparation, we firstly selected areas for training (and validation) and testing. The dimensions of the selected areas were kept the same for each camp across two seasons. Secondly, the selected training (and validation) data were converted into patches with an overlap of 32 pixels in a shape of (128, 128) pixels. This patch size has been applied in related research and proved to be effective [33]. Thirdly, we deleted patches without any target built-up structures to reduce the redundancy of the class *Background*, in turn, to improve training speed and reduce the imbalance between the class *Target built-up structures* and the class *Background*. Following that, the produced patches were divided into training patches and validation patches at a ratio of 9:1. The data preparation process is briefly shown in Figure 4 below. Table 1 shows the number of training and validation patches for the OBIA labels and the Manual labels.

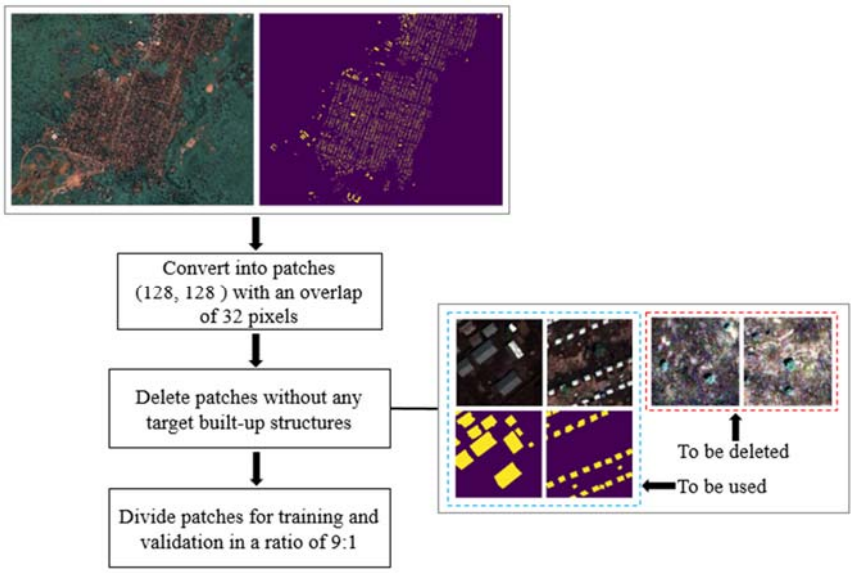

Figure 4. The brief process of data preparation.

Table 1. The number of training and validation patches for the OBIA labels and Manual labels.

Refugee Camp	Season	Training		Validation	
		OBIA	Manual	OBIA	Manual
Kule	Dry	8253	8286	917	921
	Wet	7930	8109	881	901
Nguenyyiel	Dry	6806	6745	756	749
	Wet	8106	8145	901	905

2.3. Model Structures and Hyperparameters

We selected U-Net [48] as an encoder–decoder structure with ResNet-34 [49] as a backbone (ResNet34-UNet) for all experiments. This architecture has been proven to work well in satellite image segmentation tasks [50].

Figure 5 presents its general structure. Overall, there are five encoder units (Encoder 1, 2, 3, 4, and 5) and five decoder units (Decoder 1, 2, 3, 4, and 5) in ResNet34-UNet. Encoder 1 (seen in Figure 6) converts an input image with a shape of (128, 128, 3) to a set of feature maps with a shape of (32, 32, 64) after a batch normalization (BN) layer, an activation layer, a zero-padding (ZP) layer, a convolution (Conv) layer, a second BN layer, and an activation layer together with a max-pooling layer. The output of the second ReLU layer will be connected to the first layer of Decoder 4. Here, we used rectified linear unit (ReLU) as the activation function. Encoders 2, 3, 4, and 5 are residual units, and they share the same basic structure as Encoder 1.

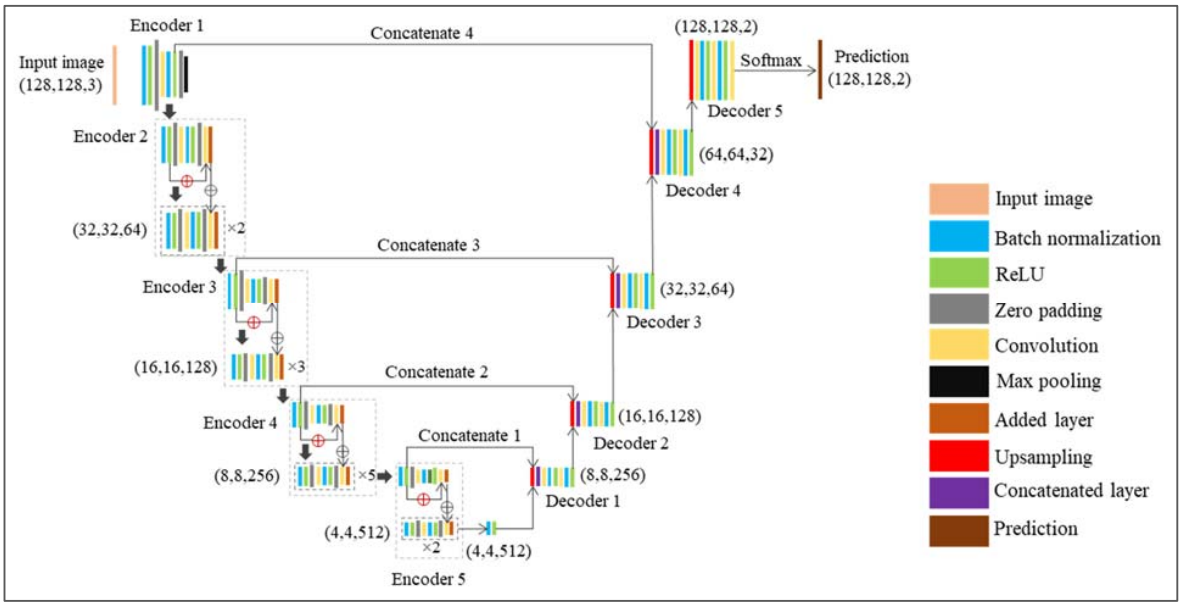

Figure 5. The general structure of ResNet34-UNet.

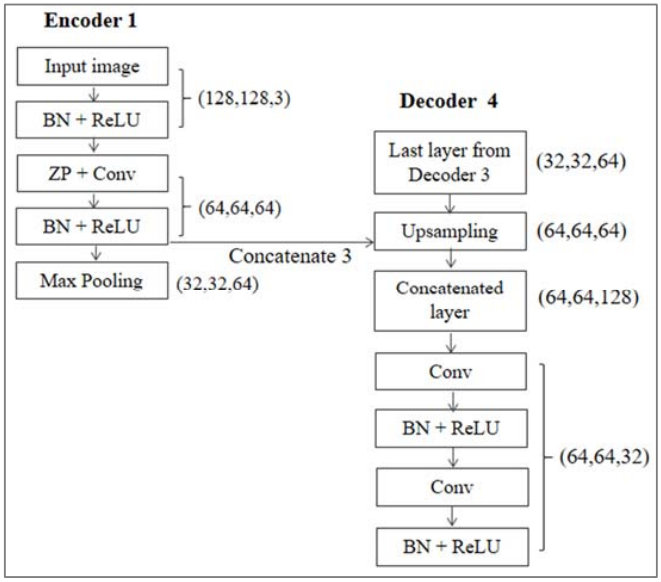

Figure 6. The basic structure of Encoder 1 and Decoder 4 of the ResNet34-UNet.

Decoders 1, 2, 3, and 4 share similar structures. Figure 6 shows the structure of Decoder 4. It starts from an upsampling layer converting the shape of the input layer from (32, 32, 64) to (64, 64, 64). Then, it will be concatenated with the output of the first ReLU layer. After that, the first Conv layer reduces the number of features from 128 to 32. Encoder 5 uses a convolution layer to replace a concatenation layer and then connects to a SoftMax operation in the end to produce prediction results.

2.4. Accuracy Metrics

The performance of the proposed model was evaluated by Precision, Recall, and Intersection over Union (IoU) of target built-up structures, which have been widely applied in image analysis domains [51]. The calculation of the three metrics can be found in Equations (1)–(3) where TP, FP, and FN refer to the number of the True Positive, the False Positive, and the False Negative pixels for the semantic class. Additionally, we attached the

results of F1-score and mean IoU in Appendix A, which can be helpful in comparison with outcomes from other related research in the future.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{3}$$

2.5. The Designation of Experiments

We designed two scenarios to compare the OBIA labels and the Manual labels. The designation of experiments for the two scenarios can be found in Figure 7.

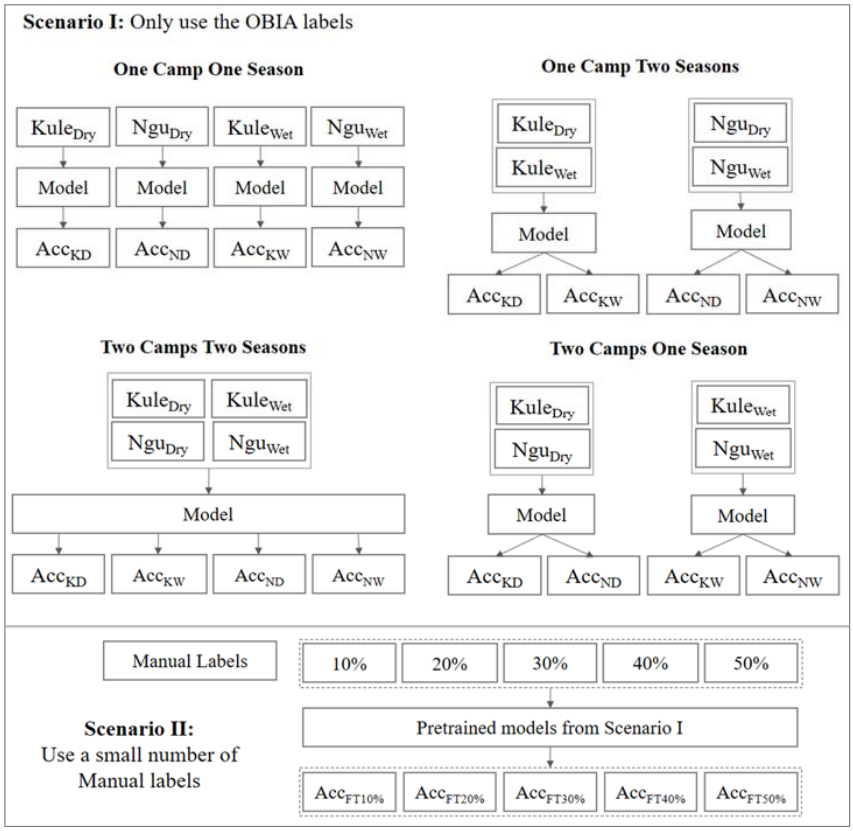

Figure 7. The designation of experiments for Scenario I and Scenario II. KuleDry means data from the Kule refugee camp during the dry season. AccKD means the accuracy of the model tested by data from KuleDry. Ngu denotes the Nguenyyiel refugee camp. Other abbreviations in Scenario I follow analogous rules. The FT in Scenario II denotes fine-tuning.

In Scenario I, we designed four strategies to train models with the OBIA labels from past humanitarian operations. The first strategy is to train models with data from one camp and one season (OCOS). The other three training strategies include intermixing data from (1) one camp during two seasons (OCTS), (2) two camps during the same season (TCOS), and (3) two camps during two seasons (TCTS). The output results of Scenario I are expected to illustrate whether we can use existing OBIA-generated labels to replace the Manual labels. The three intermixing strategies may help reveal the influences of combining different data sources on the performance of the selected model.

In Scenario II, we randomly selected different percentages (10%, 20%, 30%, 40%, and 50%) of the total training patches of the Manual labels to fine-tune models pretrained by

the four strategies in Scenario I. The reason that we made a random selection rather than a manual selection of patches with a high density of target built-up structures was to keep the data distribution of fine-tuning data similar to that of the training and testing data. The output results from Scenario II are expected to reveal whether we can use a smaller amount of the Manual labels based on models pretrained by the OBIA labels to replace a large number of the Manual labels.

Additionally, we designed baseline experiments where models were trained with purely Manual labels by the same strategies in Scenario I. The results from the two scenarios for the OBIA labels were compared with the performance of the baseline models for each camp during each season.

2.6. Set-Up

In this research, the ResNet34-UNet was implemented based on the Segmentation Model Python library [52]. For all experiments, we selected a batch size of 32, the Adam optimizer [53], and the Binary Cross Entropy as a loss function. We chose pretrained weights from ImageNet as initial weights. We used NVIDIA RTX3090 GPU to train and test models for all experiments in TensorFlow 2.7.0 environment. In terms of learning rates, we trained models with 200 epochs and chose 4×10^{-4} as an initial learning rate with 2×10^{-6} as a decay rate in Scenario I. In Scenario II, each model was fine-tuned with 10 epochs with a fixed learning rate of 1×10^{-4}. The selection of the learning rates was based on multiple tests during the primary phase of this research.

3. Results

Tables 2–4 show the IoU, Precision, and Recall values of all implemented experiments for the two scenarios, as shown in Figure 7. To simplify the text, we use *Kule* to replace the Kule refugee camp and *Kule Dry* (or *Wet*) to replace the Kule refugee camp during the dry (or wet) season. *Nguenyyiel* and *Nguenyyiel Dry* (or *Wet*) follow the same nomenclature. In addition, we use *OBIA-OCOS models*, and *Manual-OCOS models* to represent models trained with the OBIA labels and the Manual labels by the OCOS strategy, respectively. The other three strategies (OCTS, TCOS, and TCTS) follow the same nomenclature. The highest values of metrics for each camp and season by each training strategy are highlighted in blue. The highest values for each camp and season among all training strategies are highlighted in red.

Table 2. The IoU values of built-up structures of all implemented experiments (OCOS: One Camp One Season; OCTS: One Camp Two Seasons; TCOS: Two Camps One Season; TCTS: Two Camps Two Seasons). The mean IoU values can be found in Table A2 in Appendix A.

Refugee Camp	Season	Training Strategy	Baseline	Scenario I	Scenario II				
			Manual	OBIA	10%	20%	30%	40%	50%
Kule	Dry	OCOS	0.5857	0.5113	0.5829	0.6033	0.6019	0.5969	0.5999
		OCTS	0.5828	0.5202	0.5836	0.5927	0.5936	0.5958	0.5917
		TCOS	0.5887	0.5277	0.5916	0.5927	0.5927	0.5934	0.5927
		TCTS	0.5884	0.5276	0.5696	0.5847	0.5836	0.5866	0.5830
Kule	Wet	OCOS	0.6400	0.5539	0.6386	0.6541	0.6569	0.6412	0.6425
		OCTS	0.6459	0.5434	0.6188	0.6389	0.6446	0.6325	0.6413
		TCOS	0.6387	0.5273	0.6415	0.6470	0.6480	0.6493	0.6451
		TCTS	0.6339	0.5573	0.6296	0.6327	0.6480	0.6404	0.6364
Nguenyyiel	Dry	OCOS	0.6377	0.6051	0.6359	0.6393	0.6394	0.6389	0.6353
		OCTS	0.6388	0.6074	0.6129	0.6302	0.6375	0.6369	0.6366
		TCOS	0.6379	0.6051	0.6333	0.6371	0.6380	0.6385	0.6375
		TCTS	0.6391	0.6107	0.6399	0.6493	0.6471	0.6481	0.6440
Nguenyyiel	Wet	OCOS	0.6472	0.6196	0.6491	0.6538	0.6522	0.6493	0.6477
		OCTS	0.6465	0.6168	0.6486	0.6517	0.6518	0.6480	0.6452
		TCOS	0.6507	0.6167	0.6608	0.6636	0.6630	0.6611	0.6653
		TCTS	0.6634	0.6302	0.6330	0.6152	0.6624	0.6353	0.6604

Table 3. The Recall values of built-up structures of all implemented experiments.

Refugee Camp	Season	Training Strategy	Baseline	Scenario I	Scenario II				
			Manual	OBIA	10%	20%	30%	40%	50%
Kule	Dry	OCOS	0.6840	0.5912	0.7370	0.7313	0.7335	0.7299	0.7377
		OCTS	0.6910	0.5790	0.6798	0.7066	0.7071	0.7180	0.7135
		TCOS	0.6940	0.5879	0.7002	0.7035	0.7076	0.7135	0.7180
		TCTS	0.6959	0.5950	0.6553	0.6912	0.6922	0.7007	0.6960
Kule	Wet	OCOS	0.7116	0.6162	0.7470	0.7551	0.7324	0.7324	0.7349
		OCTS	0.7245	0.6005	0.6766	0.7120	0.7225	0.7228	0.7298
		TCOS	0.7148	0.5755	0.7209	0.7243	0.7307	0.7422	0.7377
		TCTS	0.7124	0.6234	0.6966	0.7051	0.7338	0.7281	0.7249
Nguenyyiel	Dry	OCOS	0.7760	0.7093	0.7536	0.7628	0.7623	0.7668	0.7679
		OCTS	0.7657	0.7110	0.7493	0.7552	0.7683	0.7635	0.7651
		TCOS	0.7698	0.7043	0.7863	0.7659	0.7667	0.7688	0.7677
		TCTS	0.7713	0.7186	0.7562	0.7779	0.7782	0.7825	0.7784
Nguenyyiel	Wet	OCOS	0.7553	0.7042	0.7474	0.7595	0.7694	0.7628	0.7553
		OCTS	0.7608	0.7045	0.7487	0.7592	0.7621	0.7595	0.7552
		TCOS	0.7638	0.6890	0.7627	0.7712	0.7717	0.7724	0.7747
		TCTS	0.7712	0.7138	0.7245	0.7626	0.7754	0.7667	0.7703

Table 4. The Precision values of built-up structures of all implemented experiments.

Refugee Camp	Season	Training Strategy	Baseline	Scenario I	Scenario II				
			Manual	OBIA	10%	20%	30%	40%	50%
Kule	Dry	OCOS	0.7909	0.8029	0.7360	0.7752	0.7704	0.7661	0.7625
		OCTS	0.7882	0.8367	0.8047	0.7862	0.7871	0.7777	0.7760
		TCOS	0.7950	0.8374	0.7923	0.7901	0.7849	0.7791	0.7726
		TCTS	0.7921	0.8232	0.8132	0.7914	0.7882	0.7827	0.7822
Kule	Wet	OCOS	0.8642	0.8455	0.8402	0.8346	0.8374	0.8374	0.8364
		OCTS	0.8562	0.8511	0.8787	0.8616	0.8568	0.8352	0.8410
		TCOS	0.8573	0.8629	0.8535	0.8583	0.8512	0.8385	0.8372
		TCTS	0.8518	0.8402	0.8674	0.8603	0.8471	0.8418	0.8389
Nguenyyiel	Dry	OCOS	0.7815	0.8048	0.8028	0.7979	0.7987	0.7930	0.7863
		OCTS	0.7940	0.8066	0.7710	0.7921	0.7891	0.7934	0.7912
		TCOS	0.7883	0.8111	0.7650	0.7911	0.7918	0.7902	0.7899
		TCTS	0.7886	0.8026	0.8063	0.7971	0.7935	0.7905	0.7886
Nguenyyiel	Wet	OCOS	0.8189	0.8376	0.8316	0.8245	0.8106	0.8135	0.8198
		OCTS	0.8114	0.8320	0.8291	0.8215	0.8183	0.8154	0.8159
		TCOS	0.8147	0.8545	0.8318	0.8263	0.8247	0.8210	0.8249
		TCTS	0.8260	0.8432	0.8337	0.7610	0.8196	0.7876	0.8224

3.1. Scenario I

In terms of IoU values, we can observe four main findings. Firstly, models trained with the Manual labels outperform models trained with the OBIA labels in both camps during both seasons under all four training strategies. Secondly, the accuracy differences in Kule cases are higher than those in the Nguenyyiel cases. For example, in the Kule Dry case, the OBIA-OCOS model can produce an IoU value of 0.5113, while it is 0.5857 for the Manual-OCOS model. However, for the Nguenyyiel Dry case, the IoU values of the OBIA-OCOS model and the Manual-OCOS model are 0.6051 and 0.6377, respectively. Thirdly, the IoU values from the wet season are higher than those from the dry season. For instance, the IoU values of the OBIA-OCOS model and the Manual-OCOS model in the Kule Wet case are 0.5539 and 0.6400, respectively, which are higher than the values (0.5113, 0.5857) from the dry season. Last but not least, we can observe that the changes in IoU

values (within 0.01, in general) brought by the three intermixing strategies are insignificant for the Manual labels. For the OBIA labels, the TCTS intermixing strategy shows a stable but limited improvement (approximately 0.01). The other two intermixing strategies cannot consistently improve the model performance. Instead, they decrease the IoU values in most cases.

The Recall values share similar findings as the IoU values, as shown in Table 3. We can observe the TCTS intermixing strategy brings a consistent improvement (approximately 0.01) for the models trained with the OBIA labels, but the other two strategies harm the model performance in most cases.

For Precision values as shown in Table 4, we can observe that the OBIA labels outperform the Manual labels in Precision values in all implemented experiments. On the contrary, the Manual labels outperform the OBIA labels in terms of Recall. For instance, in the Nguenyyiel Wet case, the Precision values of the OBIA-OCOS model and the Manual-OCOS model are 0.8545 and 0.8147, respectively. However, the Recall values of the same two models are 0.6890 and 0.7638, respectively. This finding indicates that the annotation errors in the OBIA labels cause models to miss much of TP in prediction.

In the following content, we used Kule Wet as an example to explain the above findings because Kule Wet includes the most typical examples for an explanation. Figure 8 displays the spatial distribution of predicted FP (blue) and FN (red) pixels of the OBIA-OCOS model (left) and the Manual-OCOS model (right).

Overall, we can observe the two models share a similar spatial distribution of FP and FN pixels, but the OBIA-OCOS model produced more falsely predicted pixels. We selected five typical examples (A)–(E) to further show the prediction differences between the two models, as shown in Figure 9. In general, we found four main sources of errors coming from facility buildings, trees, boundaries, and brown cloth on the rooftop of dwellings.

There are two main types of facility buildings in the selected camps. Most facility buildings are white and grey in terms of spectral characteristics. Only a few of them are brownish. The OBIA labels missed some of the white and grey facility buildings in training data as shown in Figure 9(1A). This can be the reason that the Manual-OCOS models perform better than the OBIA-OCOS models for this type. For brownish facility buildings, both models cannot perform well, as shown in Figure 9(2E). This may be due to the very limited number of such structures in training data and the high similarity to the surrounding environment.

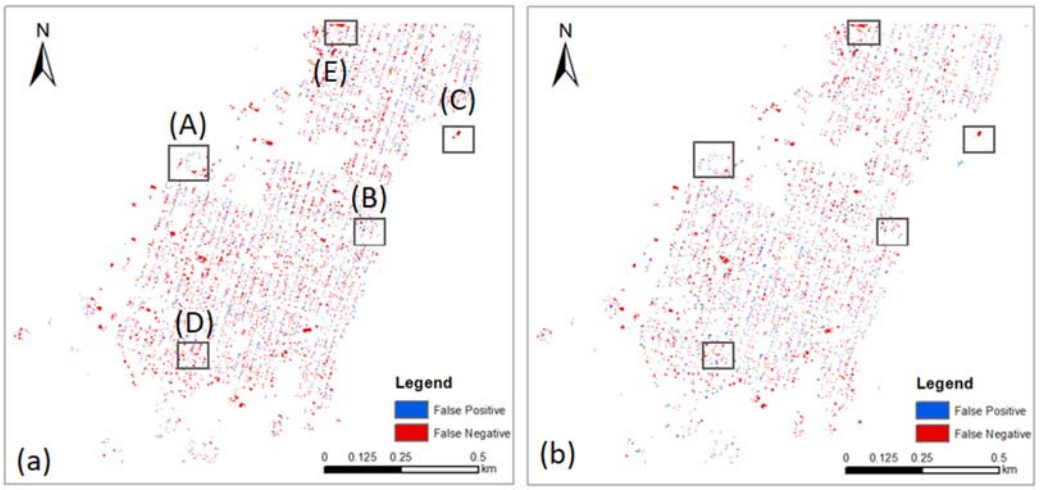

Figure 8. The spatial distribution of predicted FP (blue) and FN (red) pixels of the OBIA-OCOS model (**a**) and the Manual-OCOS model (**b**) from the Kule Wet case.

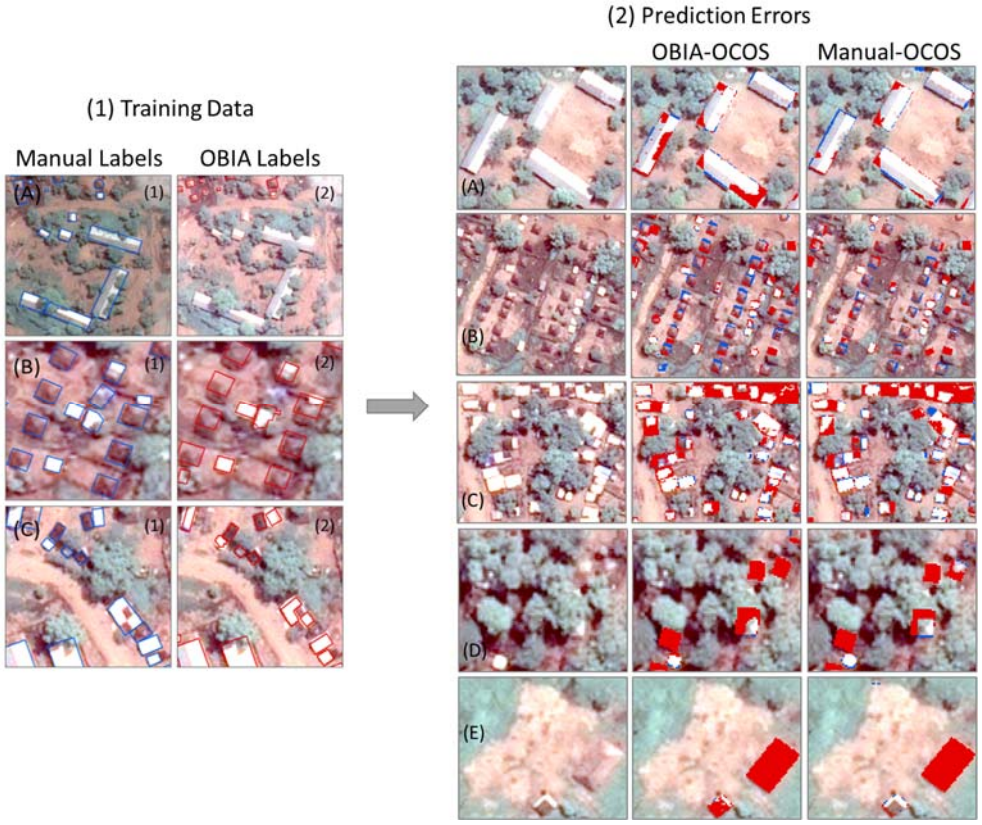

Figure 9. (**1**) The annotated errors in the original OBIA labels and (**2**) examples (**A**–**E**) showing the differences in FP (blue) and FN (red) pixels between the OBIA-OCOS model (middle) and the Manual-OCOS model (right) from the Kule Wet case.

The occlusion of trees increases the difficulty of detecting target built-up structures. Both types of models cannot detect most brownish dwellings partially covered by trees, as shown in Figure 9(2D). However, due to the limited number of tree-occluded brownish dwellings, this type of error may not play a significant role in the results. Additionally, we can observe the IoU values from the wet season are higher than those from the dry season. This indicates that the similarity between target-up structures and the surrounding environment may play a more important role in this extraction task.

The prediction errors around the boundaries of target built-up structures widely spread the whole testing area for both models, as shown in Figure 8. Figure 9(1B) presents an example showing the difference in the annotation of dwellings between the OBIA labels and the Manual labels. We can observe an obvious shift of the OBIA labels compared with the Manual labels, which may cause more predicted FN and FP pixels. In addition, due to the difficulty of determining the exact boundaries of built-up structures, the Manual-OCOS models also unavoidably have FN and FP pixels around the boundaries. These incorrectly predicted pixels around the boundaries are the major source of errors at pixel levels for both types of labels. This type of error is also the main reason that the Manual labels outperform the OBIA labels in terms of IoU values.

Last but not least, models trained with either type of label are unable to detect brown cloth on the top of large dwellings that are densely distributed, as shown in Figure 9(2C). This is likely because both spatial distribution patterns (too dense) and the size (much larger) are different from most dwellings. However, for smaller dwellings covered by brown cloth, the Manual-OCOS models perform better than OBIA-OCOS models. This is mainly because the OBIA labels usually ignore brown parts within a dwelling and separate them into several individual dwellings, as shown in Figure 9(1C).

3.2. Scenario II

In Scenario II, we used different shares (10%, 20%, 30%, 40%, and 50%) of the Manual labels to fine-tune the models pretrained with the OBIA labels from Scenario I. The results reveal that the fine-tuning approach can produce a comparable or even better performance than the models trained with purely Manual labels. For example, the IoU value of the Manual-OCOS model for the Kule Dry is 0.5857. Fine-tuning the OBIA-OCOS model with 20% of the Manual labels improves the IoU value to 0.6033. The fine-tuned models produced all of the highest IoU values (highlighted in red) for each camp during each season. Based on the existing results, it is not found that intermixing training strategies are influential in the fine-tuning results. Additionally, most of the highest IoU values occur in between 20% and 40% of cleaned labels. However, no such stable percentage was found to perform the best in all cases. More research is required to dissect the further relations between these factors.

4. Discussion

This research aims to verify whether we can use the OBIA labels to replace the Manual labels with intermixing strategies (Scenario I), and, if the intermixing strategies fail, whether we can use a smaller number of Manual labels to reduce the manual annotation work (Scenario II) to achieve similar performance to a large number of Manual labels.

Touzani and Granderson [44] collected building footprints from open data web portals from 15 cities in America as label data for a semantic segmentation model (DeepLab-v3+). The collected labels, similar to the OBIA labels, include many annotation errors. The models trained with mixed datasets from 15 cities can produce a mean IoU value of 0.89. The finding shows the high value of using mixed open datasets from multiple sources to replace manually annotated labels for extracting building footprints. However, the performance of the OBIA labels was slightly improved only by the TCTS intermixing strategy. The other two intermixing strategies even slightly harmed the models. This is perhaps because we selected only two refugee camps. It may be worthwhile to intermix data from more refugee camps to verify whether the intermixing strategy can help improve this extraction task.

As mentioned in the Section 1, many techniques have been applied to handle label noise in deep learning. In this research, we found that the improvement of intermixing data from multiple sources is limited under the initial set up, but fine-tuning based on the models trained with intermixing data has very high potential.

In terms of the selection of networks, even though it is not the main focus of this research, we can nevertheless find that the selected model performs quite well compared with other models in previous research. G. W. Gella et al. [34] selected a Mask-RCNN model with ResNet-101 as a backbone to extract refugee dwellings in Cameroon. They used WorldView-3 and WorldView-2 images with a spatial resolution of 0.5 m, which is the same as that used in this research. The Mask-RCNN model produced the mean IoU value up to 0.669 which is lower than the outputs of all experiments here, as shown in Table A2 in Appendix A. These comparison results are consistent with the results from Y. Lu and C. Kwan [35]. In addition, Ref. [54] compared multiple semantic segmentation models, including three classical deep learning architectures (U-Net, LinkNet, and Feature Pyramid Network (FPN)) and twelve backbones (VGG16, VGG19, ResNet-18, ResNet-34, DenseNet-121, DenseNet-169, Inception-v3, InceptionResnet-v2, MobileNet-v1, MobileNet-v2, EfficientNet-B0, and EfficientNet-B1) to test the validity of the OBIA labels compared with the Manual labels by using the same training and testing data of the Kule refugee camp during the dry and the wet seasons. The results show that these models perform similarly, in general. Therefore, it is necessary to undertake more state-of-the-art techniques in the future to discover more suitable workflows for this task.

5. Conclusions

From this research, we found that the OBIA labels can help the models produce satisfying predicted labels with IoU values of target built-up structures greater than 0.5—

even though the Manual labels still outperform the OBIA labels for each camp during each season. Intermixing the data from two camps and two seasons (TCTS) can slightly improve the performance of models trained with the OBIA labels. However, the other two strategies (OCTS and TCOS) cannot bring stable improvement and even sometimes harm the models. In addition, we found that both types of labels cannot properly detect brownish facility buildings, brownish dwellings partially occluded by trees, nor densely distributed large dwellings covered by brown cloth. The falsely predicted pixels around the boundary of target built-up structures may be the main source of prediction errors compared with other sources. Additionally, the reason that the Manual labels outperform the OBIA labels is mainly because the models trained with the Manual labels can produce fewer errors around the boundaries of structures. Additionally, it can be observed that the IoU values are higher during the wet season than during the dry season. This is likely because of the higher similarity between built-up structures and the surrounding environment during the dry season. Furthermore, we found fine-tuning models pretrained with the OBIA labels from Scenario I with smaller shares (10%, 20%, 30%, 40%, and 50%) of the Manual labels can enable the models to produce comparable or even better IoU values than the models trained with purely Manual labels.

These outcomes prove the remarkable potential of the OBIA labels to replace the Manual labels in future research. More research is required to make better use of the produced OBIA labels for future humanitarian operations.

Author Contributions: Conceptualization, Y.G., S.L. and D.T.; methodology and formal analysis, Y.G.; investigation, Y.G.; writing—original draft preparation, Y.G.; writing—review and editing, Y.G., S.L., D.T., G.W.G. and L.W. 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.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Those who are interested in the data can send data requests to yunya.gao@plus.ac.at. The data will be provided pending approval of the request by Médecins Sans Frontières (MSF) Austria.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Tables A1 and A2 present the results of F1-score and mean IoU values of all implemented experiments. The formula of the two metrics are shown in Equations (A1) and (A2).

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{A1}$$

$$\text{mIoU} = \frac{1}{n} \sum_{k=1}^{n} \frac{\text{TP}}{(\text{TP} + \text{FP} + \text{FN})} \ (n = 2) \tag{A2}$$

Table A1. The F1-score values of built-up structures of all implemented experiments.

Refugee Camp	Season	Training Strategy	Baseline	Scenario I	Scenario II				
			Manual	OBIA	10%	20%	30%	40%	50%
Kule	Dry	OCOS	0.7387	0.6766	0.7365	0.7526	0.7515	0.7476	0.7499
		OCTS	0.7365	0.6844	0.7370	0.7442	0.7449	0.7467	0.7435
		TCOS	0.7411	0.6908	0.7434	0.7443	0.7442	0.7449	0.7443
		TCTS	0.7409	0.6907	0.7258	0.7379	0.7371	0.7394	0.7366
Kule	Wet	OCOS	0.7805	0.7129	0.7794	0.7908	0.7929	0.7814	0.7824
		OCTS	0.7849	0.7041	0.7645	0.7797	0.7839	0.7749	0.7814
		TCOS	0.7795	0.6905	0.7816	0.7856	0.7864	0.7874	0.7843
		TCTS	0.7759	0.7157	0.7727	0.7750	0.7864	0.7808	0.7778
Nguenyyiel	Dry	OCOS	0.7787	0.7540	0.7774	0.7799	0.7801	0.7797	0.7770
		OCTS	0.7796	0.7558	0.7600	0.7732	0.7786	0.7781	0.7779
		TCOS	0.7789	0.7539	0.7755	0.7783	0.7790	0.7793	0.7786
		TCTS	0.7799	0.7583	0.7804	0.7874	0.7858	0.7865	0.7835
Nguenyyiel	Wet	OCOS	0.7858	0.7651	0.7873	0.7907	0.7895	0.7873	0.7862
		OCTS	0.7853	0.7630	0.7868	0.7892	0.7892	0.7864	0.7844
		TCOS	0.7884	0.7629	0.7957	0.7978	0.7973	0.7960	0.7990
		TCTS	0.7977	0.7731	0.7753	0.7618	0.7969	0.7770	0.7955

Table A2. The mean IoU values of built-up structures of all implemented experiments.

Refugee Camp	Season	Training Strategy	Baseline	Scenario I	Scenario II				
			Manual	OBIA	10%	20%	30%	40%	50%
Kule	Dry	OCOS	0.7819	0.7429	0.7795	0.7908	0.7900	0.7873	0.7888
		OCTS	0.7803	0.7481	0.7808	0.7853	0.7858	0.7869	0.7847
		TCOS	0.7834	0.7520	0.7849	0.7854	0.7853	0.7857	0.7852
		TCTS	0.7832	0.7518	0.7736	0.7813	0.7807	0.7821	0.7803
Kule	Wet	OCOS	0.8115	0.7664	0.8108	0.8186	0.8200	0.8119	0.8126
		OCTS	0.8145	0.7610	0.8005	0.8109	0.8139	0.8074	0.8120
		TCOS	0.8108	0.7528	0.8122	0.8151	0.8156	0.8162	0.8139
		TCTS	0.8082	0.7682	0.8061	0.8077	0.8155	0.8115	0.8094
Nguenyyiel	Dry	OCOS	0.8157	0.7993	0.8149	0.8165	0.8166	0.8163	0.8145
		OCTS	0.8163	0.8004	0.8030	0.8119	0.8156	0.8153	0.8151
		TCOS	0.8158	0.7992	0.8134	0.8154	0.8159	0.8161	0.8156
		TCTS	0.8164	0.8021	0.8169	0.8216	0.8205	0.8210	0.8189
Nguenyyiel	Wet	OCOS	0.8174	0.8033	0.8185	0.8208	0.8199	0.8184	0.8177
		OCTS	0.8170	0.8018	0.8182	0.8197	0.8198	0.8178	0.8163
		TCOS	0.8192	0.8019	0.8245	0.8259	0.8256	0.8246	0.8268
		TCTS	0.8258	0.8088	0.8102	0.8004	0.8252	0.8110	0.8242

References

1. UNHCR. *Global Trends: Forced Displacement in 2020*; UNHCR Global Trends: Copenhagen, Denmark, 2021; p. 72.
2. UNHCR. UNHCR: A Record 100 Million People Forcibly Displaced Worldwide. *UN News*, 23 May 2022.
3. UNHCR. *Persons Who are Forcibly Displaced, Stateless and Others of Concern to UNHCR, UNHCR The UN Refugee Agency*. 2021. Available online: https://www.unhcr.org/refugee-statistics/methodology/definition/ (accessed on 28 November 2021).
4. UNHCR. *Mid-Year Trends 2021*; UNCHR: Copenhagen, Denmark, 2021.
5. Elizabeth, U.; Stuart, M. *Report on UNHCR's Response to COVID-19*; UNCHR: Geneva, Switzerland, 2020.
6. Médecins Sans Frontières. Global Migration and Refugee Crisis—Record Numbers of People Have Been Forced from Home and Struggle to Find Safety. 2021. Available online: https://www.doctorswithoutborders.org/refugees (accessed on 28 November 2021).
7. Çelik, M.; Ergun, Ö.; Johnson, B.; Keskinocak, P.; Lorca, Á.; Pekgün, P.; Swann, J. Humanitarian logistics. In *New Directions in Informatics, Optimization, Logistics, and Production*; INFORMS: Catonsville, MD, USA, 2012; pp. 18–49.

8. Lang, S.; Füreder, P.; Riedler, B.; Wendt, L.; Braun, A.; Tiede, D.; Schoepfer, E.; Zeil, P.; Spröhnle, K.; Kulessa, K.; et al. Earth observation tools and services to increase the effectiveness of humanitarian assistance. *Eur. J. Remote Sens.* **2020**, *53*, 67–85. [CrossRef]
9. Kunz, N.; Van Wassenhove, L.N.; Besiou, M.; Hambye, C.; Kovacs, G. Relevance of humanitarian logistics research: Best practices and way forward. *Int. J. Oper. Prod. Manag.* **2017**, *37*, 1585–1599. [CrossRef]
10. Witharana, C. Who Does What Where? Advanced Earth Observation for Humanitarian Crisis Management. In Proceedings of the 2012 IEEE 6th International Conference on Information and Automation for Sustainability, Beijing, China, 27–29 September 2012; pp. 1–6.
11. Füreder, P.; Lang, S.; Hagenlocher, M.; Tiede, D.; Wendt, L.; Rogenhofer, E. Earth Observation and GIS to Support Humanitarian Operations in Refugee/IDP Camps. In Proceedings of the ISCRAM 2015—The 12th International Conference on Information Systems for Crisis Response and Management, Kristiansand, Norway, 24–27 May 2015.
12. Lang, S.; Schoepfer, E.; Zeil, P.; Riedler, B. Earth observation for humanitarian assistance. *GI Forum* **2017**, *1*, 157–165. [CrossRef]
13. Braun, A.; Hochschild, V. Potential and limitations of radar remote sensing for humanitarian operations. *GI Forum* **2017**, *1*, 1228–1243. [CrossRef]
14. Lang, S.; Füreder, P.; Rogenhofer, E. Earth observation for humanitarian operations. In *Yearbook on Space Policy 2016*; Springer: Cham, Switzerland, 2018; pp. 217–229.
15. Checchi, F.; Stewart, B.T.; Palmer, J.J.; Grundy, C. Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *Int. J. Health Geogr.* **2013**, *12*, 4. [CrossRef]
16. 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]
17. Füreder, P.; Tiede, D.; Lüthje, F.; Lang, S. Object-based dwelling extraction in refugee/IDP camps–challenges in an operational mode. *South-East. Eur. J. Earth Obs. Geomat.* **2014**, *3*, 539–544.
18. Kraffi, P.; Tiede, D.; Fureder, P. Template Matching to Support Earth Observation Based Refugee Camp Analysis in Obia Workflows—Creation and Evaluation of a Dwelling Template Library for Improving Dwelling Extraction Within an Object-Based Framework. In Proceedings of the GEOBIA 2016: Solutions and Synergies, Enschede, The Netherlands, 14–16 September 2016.
19. Tiede, D.; Krafft, P.; Füreder, P.; Lang, S. Stratified template matching to support refugee camp analysis in OBIA workflows. *Remote Sens.* **2017**, *9*, 326. [CrossRef]
20. Tiede, D.; Füreder, P.; Lang, S.; Hölbling, 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. Fernerkund. Geoinf.* **2013**, *2013*, 185–195. [CrossRef]
21. Tiede, D.; Lang, S.; Hölbling, 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 2021.
22. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef]
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
24. Zhu, M.; He, Y.; He, Q. A Review of Researches on Deep Learning in Remote Sensing Application. *Int. J. Geosci.* **2019**, *10*, 1–11. [CrossRef]
25. Zhang, L.; Zhang, L.; Kumar, V. Deep learning for Remote Sensing Data. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
26. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A review. *arXiv* **2017**, arXiv:1710.03959.
27. Fisher, A.; Mohammed, E.A.; Mago, V. TentNet: Deep Learning Tent Detection Algorithm Using A Synthetic Training Approach. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *2020*, 860–867.
28. Luo, L.; Li, P.; Yan, X. Deep learning-based building extraction from remote sensing images: A comprehensive review. *Energies* **2021**, *14*, 7982. [CrossRef]
29. 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.* **2018**, *42*, 161–166. [CrossRef]
30. 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*, 20170363. [CrossRef]
31. Wickert, L.; Bogen, M.; Richter, M. Lessons Learned on Conducting Dwelling Detection on VHR Satellite Imagery for the Management of Humanitarian Operations. *Sens. Transducers* **2021**, *249*, 45–53.
32. 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] [PubMed]
33. Gella, G.W.; Wendt, L.; Lang, S.; Braun, A. Testing Transferability of Deep-Learning-Based Dwelling Extraction in Refugee Camps. *GI Forum* **2021**, *9*, 220–227.

34. 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. [CrossRef]
35. Lu, Y.; Kwan, C. Deep Learning for Effective Refugee Tent. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 16–20.
36. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [CrossRef]
37. Lu, Z.; Fu, Z.; Xiang, T.; Han, P.; Wang, L.; Gao, X. Learning from weak and noisy labels for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 486–500. [CrossRef]
38. Kamann, C.; Rother, C. Benchmarking the Robustness of Semantic Segmentation Models with Respect to Common Corruptions. *Int. J. Comput. Vis.* **2021**, *129*, 462–483. [CrossRef]
39. Karimi, D.; Dou, H.; Warfield, S.K.; Gholipour, A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med. Image Anal.* **2020**, *65*, 101759. [CrossRef]
40. Wang, Y.; Zhang, J.; Kan, M.; Shan, S.; Chen, X. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2022; pp. 12272–12281.
41. Zhou, T.; Li, L.; Li, X.; Feng, C.M.; Li, J.; Shao, L. Group-Wise Learning for Weakly Supervised Semantic Segmentation. *IEEE Trans. Image Process.* **2021**, *31*, 799–811. [CrossRef]
42. Zhou, T.; Zhang, M.; Zhao, F.; Li, J. Regional Semantic Contrast and Aggregation for Weakly Supervised Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4289–4299.
43. Liu, S.; Liu, K.; Zhu, W.; Shen, Y.; Fernandez-Granda, C. Adaptive Early-Learning Correction for Segmentation from Noisy Annotations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2606–2616.
44. Touzani, S.; Granderson, J. Open data and deep semantic segmentation for automated extraction of building footprints. *Remote Sens.* **2021**, *13*, 2578. [CrossRef]
45. UNHCR. *Kule Camp Profile: Who Does What Where*; UNCHR: Copenhagen, Denmark, 2016.
46. UNHCR. *Nguenyyiel Refugee Camp*; UNCHR: Copenhagen, Denmark, 2020.
47. Weather Spark. *Climate and Average Weather Year Round in Gambēla*; Cedar Lake Ventures, Inc.: Excelsior, MN, USA, 2021.
48. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ISPRS Journal of Photogrammetry and Remote Sensing ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
51. Van Beers, F.; Lindström, A.; Okafor, E.; Wiering, M.A. Deep Neural Networks with Intersection Over Union Loss for Binary Image Segmentation. In Proceedings of the ICPRAM 2019—8th International Conference Pattern Recognition Applications and Methods, Prague, Czech Republic, 19–21 February 2019; pp. 438–445.
52. Yakubovskiy, P. *Segmentation Models*; GitHub Repository: San Francisco, CA, USA, 2019.
53. Bock, S.; Goppold, J.; Weiß, M. An improvement of the convergence proof of the ADAM-Optimizer. *arXiv* **2018**, arXiv:1804.10587.
54. Gao, Y.; Lang, S.; Tiede, D.; Gella, G.W.; Wendt, L. Comparing the Robustness of U-Net, LinkNet, and FPN Towards Label Noise for Refugee Dwelling Extraction from Satellite Imagery. In Proceedings of the 2022 IEEE Global Humanitarian Technology Conference (GHTC), Santa Clara, CA, USA, 8–11 September 2022; pp. 88–94.