remote sensing

MDPI

Article

Mapping Dwellings in IDP/Refugee Settlements Using Deep Learning

Omid Ghorbanzadeh [1,2,*], Alessandro Crivellari [3], Dirk Tiede [1], Pedram Ghamisi [2,4] and Stefan Lang [1]

[1] Christian Doppler Laboratory for Geospatial and EO-based Humanitarian Technologies GEOHUM, Department of Geoinformatics—Z-GIS, University of Salzburg, 5020 Salzburg, Austria
[2] Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5, 1030 Vienna, Austria
[3] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[4] Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitzer Str. 40, 09599 Freiberg, Germany
* Correspondence: omid.ghorbanzadeh@iarai.ac.at

Abstract: The improvement in computer vision, sensor quality, and remote sensing data availability makes satellite imagery increasingly useful for studying human settlements. Several challenges remain to be overcome for some types of settlements, particularly for internally displaced populations (IDPs) and refugee camps. Refugee-dwelling footprints and detailed information derived from satellite imagery are critical for a variety of applications, including humanitarian aid during disasters or conflicts. Nevertheless, extracting dwellings remains difficult due to their differing sizes, shapes, and location variations. In this study, we use U-Net and residual U-Net to deal with dwelling classification in a refugee camp in northern Cameroon, Africa. Specifically, two semantic segmentation networks are adapted and applied. A limited number of randomly divided sample patches is used to train and test the networks based on a single image of the WorldView-3 satellite. Our accuracy assessment was conducted using four different dwelling categories for classification purposes, using metrics such as Precision, Recall, *F*1, and Kappa coefficient. As a result, *F*1 ranges from 81% to over 99% and approximately 88.1% to 99.5% based on the U-Net and the residual U-Net, respectively.

Keywords: remote sensing; refugees; humanitarian operations; Africa

Citation: Ghorbanzadeh, O.; Crivellari, A.; Tiede, D.; Ghamisi, P.; Lang, S. Mapping Dwellings in IDP/Refugee Settlements Using Deep Learning. *Remote Sens.* **2022**, *14*, 6382. https://doi.org/10.3390/rs14246382

Academic Editor: Hossein M. Rizeei

Received: 11 November 2022
Accepted: 15 December 2022
Published: 16 December 2022

1. Introduction

The number of people who were forced to leave their homes as a result of natural disasters or human-made violence (by persecution or intimidation) around the world was estimated to be almost 80 million in 2019 and exceeded 84 million people by the middle of 2021 [1]. There are two types of populations who are displaced by conflict: those who remain within their home countries are known as internally displaced people (IDPs), while those who leave their home countries are considered refugees [2]. By late 2020, about 7 million refugees had been resettled in United Nations High Commissioner for Refugees (UNHCR) managed settlements in 132 countries across the globe [3].

The last two decades have seen an exponential increase in the availability, volume, and diversity of satellite imagery [4]. In the Remote Sensing (RS) community, however, there is an ongoing need for the development of automated image processing techniques for transforming satellite data into beneficial information [5]. The situation is not the same for all RS applications; some are better than others. For example, land cover semantic segmentation, scene classification, landslide detection, and local climate zone classification have received a lot of attention from machine learning and computer vision communities by developing popular labeled data sets of SEN12MS [6], BigEarthNet [7], L4S [8], and So2SatLCZ42 [9],

respectively. Such data sets can be utilized for consideration and pretraining of the machine and deep learning approaches that have now become the pioneers in constructing automated image processing techniques in the computer vision domain. More specifically, the RS data regarding human settlements, including GRID3 [10], WSF [11], and GHSL [12], provide foundational awareness about geo-referenced infrastructure and demographic data, the world settlement footprint, and the global human settlement layer, respectively [3]. Thus, such comprehensive labeled data sets facilitate the development of state-of-the-art automated image processing techniques for the respective RS applications and provide valuable information for population estimation [13], natural hazard mitigation [14], and sustainable development [15], etc.

However, this development is not significant for humanitarian operational challenges such as the monitoring of IDP and refugee camps [16,17]. RS data analysis to support humanitarian operations has been a research area for more than two decades, especially in conflict zones or refugee camps where population dynamics are high [18]. With the operational availability of very high-resolution (VHR) satellite imagery around the year 2000 and the ongoing technological developments, the uptake of RS technology [19,20] is steadily increasing in the humanitarian sector for camp analysis based on dwelling extraction and characterization [2]. However, in spite of current advancements in image processing techniques for RS data and the increased availability of imagery sources, extracting information from refugee camps has not yet been fully explored [21]. Several factors may contribute to the limited amount of research on refugee camps, including the high cost of acquiring and processing VHR satellite imagery and the absence of large labeled data sets for developing and training intended deep learning approaches [22]. It must be noted that refugee camps include small-scale dwellings and buildings that are often irregularly positioned and cannot be separated from background features such as bare ground and vegetation, since these features usually appear together in one coarse pixel of moderate-resolution images [23]. Thus, moderate-resolution satellite images (e.g., Sentinel-1/2) are normally not suitable for obtaining detailed and accurate classifications of dwellings and other structures, and as a result, this task is almost impossible without the use of VHR images [3]. Nevertheless, the main reason for the lack of large-scale data sets for refugee camps has to do with the security concerns associated with this field. [21,24,25]. The specific properties of dwellings such as varying sizes and a range of different building materials (e.g., wood and plastic tarp) are extra challenges for dwelling classification tasks within the remote sensing community [3].

Purely pixel-based approaches for dwelling extraction as an alternative to visual interpretation were rapidly replaced by spatially aware and context-sensitive ones such as object-based image analysis (OBIA) [22,23,26–29], template-matching [30], mathematical morphology-based algorithms [31,32], and, recently, deep learning approaches [21,33]. During the past decade, the standard deep learning (DL) approach to image classification has been chiefly convolutional neural network (CNN) which has continuously optimized and achieved cutting-edge performance [34–37]. CNN can extract features from input images using convolutional layers derived from these images and has achieved higher accuracies than traditional hand-crafted feature ones [38,39]. CNNs can be used for annotating earth surface objects, such as different kinds of dwellings, through learning the class-critical features. Increasing the depth of the CNN structure by adding more convolutional layers makes it possible to extract and learn more complex features. However, as the number of convolutional layers increases, the dimension of the image decreases, which complicates the calculation process. Therefore, Long et al. [40] introduced the fully convolutional network (FCN). The CNN and FCN algorithms are fast developing. They have already been applied widely in image classification [41,42] and semantic segmentation [40,43] tasks. The main difference between the structure of the FCN and the traditional CNNs is that there is no fully connected layer in FCNs, and fully connected layers are converted to convolutional and upsampling layers [44]. The input of CNN models is small square subsets of the image (image patches), including a specific single class [21], and CNN can provide a probability

of a single class for the centroid of each image patch [45]. Instead, FCNs (e.g., SegNet and U-net) are designed to represent an image-to-image mapping to predict per-pixel class labels [46]. In FCNs, the entire input image is convolved with the first layer, and the resulting feature maps are used for the second convolutional layer. This process continues until it results in a 2D matrix of class probabilities. The absence of fully connected layers enables FCNs to utilize two main properties of learning representations on local spatial input and use input image patches of arbitrary size [47]. U-net also adds skip-layer modules among the layers to refine the spatial accuracy and retain the semantic information of the introduced classes [43]. For both segmentation and classification tasks, some studies have compared the performance of CNNs and FCNs. For instance, Jiang et al. [48] compared their performances in different aspects such as resulting accuracy and cost. They concluded that the accuracy of these models was almost the same, but the FCN model was trained with a fewer number of parameters, which led to requiring less memory consumption. Lu et al. [49] proposed an FCN model and compared it with the classic spectral angle mapper technique, CNN, and the mask-region-based CNN (R-CNN) models to determine the distribution and number of dwellings along the Syria–Jordan border in the Rukban desert. Their experimental results indicated that FCN's semantic segmentation model was superior to CNNs, SAMs, and R-CNN masks in terms of overall accuracy by 4.49%, 3.54%, and 0.88%, respectively. Gao et al. [24] used the U-Net with the VGG16 as a backbone for their evaluations on the effectiveness of extracting refugee dwellings from the VHR imagery. Three types of pretrained weights were used in combination with the four bands. In terms of comparison, the U-Net model has been compared with some other standard models such as HRNet, FastSCNN, DeepLabV3, and SegFormer [50]. In a comprehensive work, extensive experiments have been conducted by [8] to compare U-Net with some other state-of-the-art DL segmentation models including ResU-Net, PSPNet, ContextNet, FCN-8s, LinkNet, FRRN-A, FRRN-B, DeepLab-v2, DeepLab-v3+, and SQNet, for landslide detection. The U-Net model was ranked as one of the top five most accurate models. Among various FCN architectures, this model has a simple and effective architecture to extract useful features by using fewer training sample patches [51]. Deep learning (DL) architectures, including both CNN and FCN, consist of many layers [52,53]. Thus, deeper models would lead to a higher accuracy over their shallow counterparts with lesser depth. However, deeper models reduce gradient signals to values close to zero within the backpropagation process, making the training more difficult [54]. He et al. [55] could alleviate this problem by employing a residual learning framework to avoid the vanishing of gradient signals and, consequently, assist in training deeper models. Such a residual shortcut route is already applied by [56] and played an important role in the preservation of valuable information. The application of residual learning has shown high performance in different fields such as semantic segmentation [57], image recognition [55], optical and Lidar data fusion [58], image fusion of Sentinel-2 data (Palsson et al., 2018), hyperspectral remote sensing image classification [59], and road extraction [47]; clouds classification [51].

The primary purpose of this paper is to evaluate two DL architectures, i.e., U-Net and residual U-Net, for WorldView-3 satellite image classification. The literature shows different CNN architectures being used for dwelling extraction. However, due to the limitations of CNNs (e.g., class partitioning issues and fuzzy boundaries) and lack of large data sets, some post-processing strategies were developed for improving the classification results [21]. Therefore, our contribution is to show the capabilities of DL models for dwelling classification tasks differentiating four categories of dwelling without requiring post-processing refinement. As we work on only one single refugee camp in a specific time frame, a limited number of training data is available.

2. Materials and Methods

2.1. Study Site and Data Descriptions

The study site in this paper is the refugee camp Minawao in the far northern region of Cameroon, at 10°33′30″N 13°51′30″E (see Figure 1), managed by the UNHCR and mainly

housing refugees from the neighboring Nigerian Borno State, where activities of the Boko Haram militia forced over 240,000 people to flee their homes [25]. The camp has existed since 2012 and shows a quite high dynamic in population number, increasing rapidly since early 2014. At the first image acquisition on 12 April 2015, the camp housed approximately 34,000 people. It has been used in other experiments as a reference test site for a typical refugee camp (e.g., [21]) since validated classifications for different time slots and different VHR images are available. The data and the validated classifications are obtained from humanitarian mapping products produced by the Department of Geoinformatics, University of Salzburg, and Spatial Services Ltd. in the context of operational humanitarian mapping services provided to Doctors without Borders (MSF). The existing reference classifications were derived from semi-automated methods, underwent visual checks, and were refined by trained operators prior to map production. MSF uses these mapping products to obtain actual population figures and plan for healthcare, water, and sanitation services and campaigns. Labeled dwelling structures as parts of these products were used in this study as training and test data.

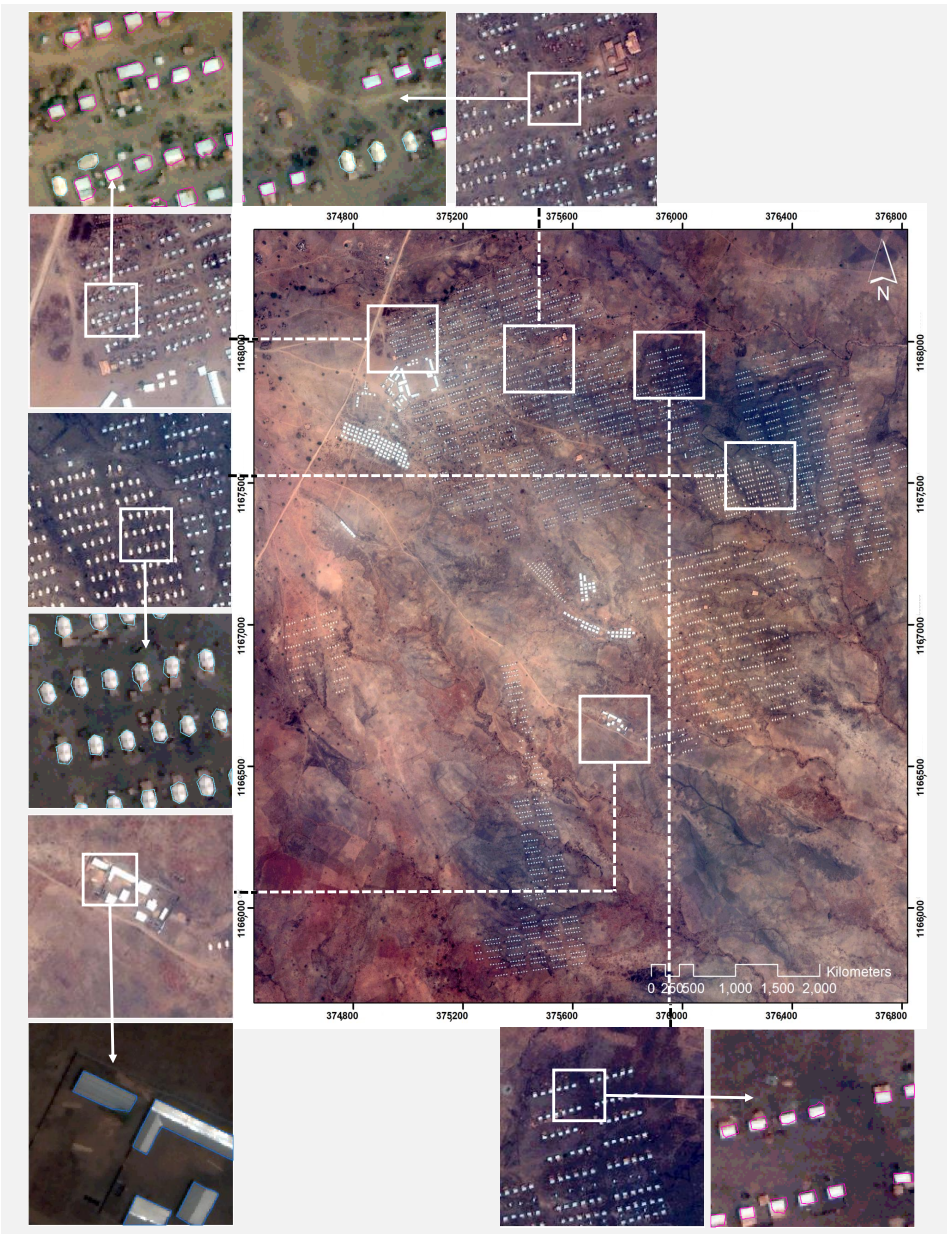

Figure 1. Refugee camp Minawao in the far North region of Cameron. The camp extent on 13 October 2015 and examples on the test set of image patches with ground truth labels.

The input images used in this study for training and testing are taken from WorldView-3 images captured on 12 April 2015, with four spectral bands: red (630–690 nm) green (510–580 nm), blue (450–510 nm), and near-infrared (770–895 nm). Bright dwellings, including Facility Buildings, Tunnel Shape, and Rectangular shapes, in this refugee camp are categorized as a proxy for estimating the population figures of this camp. We combined vegetation, dark structures, bare soil, and surrounding village houses into one class called Other Classes. The whole image was clipped into 128 × 128 pixel sample patches without any overlap in their X and Y directions. There are 525 image patches included of dwellings out of 882 patches that were created, which were then randomly divided into two sets, one half for training and one for testing.

2.2. U-Net

U-Net has shown an acceptable performance in the segmentation of high and very high-resolution satellite images and many other vision tasks such as medical image analysis (Abderrahim et al., 2020). U-Net was introduced by Ronneberger et al. [46], who identified an employed skip architecture to solve the cell tracking problem. Before [46], the skip architecture was proposed by Long et al. [40] to enhance the performance of object segmentation (PASCAL VOC) by the integration of the appearance representations obtained from shallow, deep encoder layers with the semantic information resulting from the decoder ones of an FCN. The overall architecture of the U-Net consists of a path of downsampling or encoding layers for low-level representations along with another path of upsampling or decoding layers for high-level ones [51]. The upsampling path is an asymmetrical part retaking the vanished information of the segmentation borders [60]. The downsampling path is similar to the common CNN architectures. It comprises convolution blocks consisting of two repeated convolutional layers with a filter size of 3 × 3 and an activation function of the rectified linear unit (ReLU) immediately after each layer. Four max-pooling layers with a filter size of 2 × 2 and a stride of 2 are used to downsample the results of the last convolutional layer of each block in the downsampling path. The number of feature maps in each convolution block is doubled. After the first convolution block, one layer of the input image goes to 16 feature maps and 256 in its highest number after the fifth block. The upsampling path is an inverted form of the downsampling one; thereby a copy of the feature maps is concatenated to the maps of the corresponding decoding blocks in the upsampling path. Moreover, the number of feature maps is halved after each block [46]. Concatenating the feature maps resulting from encoding layers to the decoding blocks can improve the object segmentation by retaining more information from the object boundaries [61]. This U-shaped architecture consists of 19 convolutional layers. The last convolutional layer maps each resulting 16 component feature vector to a semantic label.

2.3. Residual U-Net

The architecture of the residual U-Net is based on the U-Net with reformulated layers that learn residual functions. The residual learning strategy is applied to improve learning and the model performance as well as to ease the optimization of the U-Net by Zhang et al. [47]. The architecture of a residual neural network is constructed by stacking a series of fundamental structural elements, called residual unit. The general form of each residual unit can be represented by Equations (1) and (2)

$$y_i = h(x_i) + F(x_i, W_i) \tag{1}$$

$$x_{i+1} = f(y_i) \tag{2}$$

where the input and output of the ith residual unit are defined by x_i and x_{i+1}, and $f(y_i)$ and $F(\cdot)$ are defined as the activation and the residual functions, respectively. The function of $h(\cdot)$ simply computes identity mapping, which means $h(x_i) = x_i$. Therefore, $h(\cdot)$ has no parameter, and it can add the layer's output before the residual unit to the next layer. The convolution layer typically reduces the spatial resolution of the input image. Therefore,

the dimension of x_i may be higher than that of $F(x_i)$. For instance, a 3×3 convolution on an input image with a dimension of $a \times a$ results in a feature map with a dimension of $(a-2) \times (a-2)$. Then, a linear projection W_i is used to match the dimension of x_i and $F(x_i)$ to perform summation and be considered as the input for the layer ahead (see Figure 2).

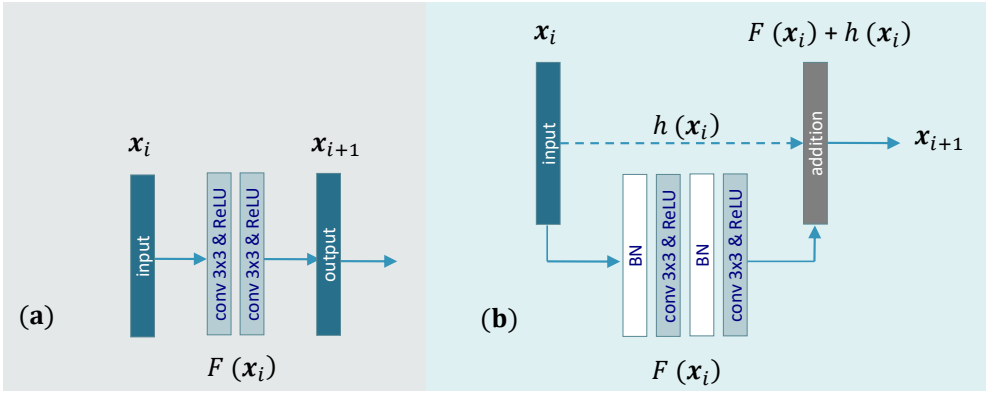

Figure 2. Illustration of the architecture of: (**a**) a typical convolutional process in a U-Net; and (**b**) a residual U-Net with an identity mapping.

There are two main differences between the U-Net and the residual U-Net. The first one is that the residual U-Net uses convolution layers with a stride of 2 instead of standard pooling layers to perform downsampling on the resulting feature maps. The second difference is that instead of the normal neural units used for structuring the U-Net, residual units are applied for building the residual U-Net [62]. Each building block of the residual U-Net is made by two 3×3 convolution layers followed by a ReLU activation layer and a batch normalization (BN) along with the convolution layers. Each block also includes an identity mapping of h (x_i) which is for transforming the input of a block to its corresponding output. The upsampling path consists of three residual blocks. An upsampling of the resulting feature maps is located before each residual block of this path. Finally, a convolution layer with a 1×1 filter size and a softmax activation function is used to convert the resulting probabilities to the defined semantic labels. The residual U-Net consists of 15 convolutional layers, which is less than the U-Net.

It must be mentioned that in a conventional residual U-Net, a sigmoid activation function (see Equation (3)) is used for normalizing the output to a probability distribution over only two predicted classes. The sigmoid activation function is used for binary classification in the logistic regression model and basic neural networks. The sum of the resulting probabilities will not be one necessarily as this function is applied to each element of the raw output data independently. Thus, it is useful for cases where the outputs are not mutually exclusive. However, in this study, we have a multi-class classification problem. That is why we replaced the sigmoid with the softmax activation function, which looks like Equation (4). The softmax takes into account all the output data, so probabilities are always interrelated, and the sum of them is one.

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{3}$$

$$softmax(z_j) = \frac{e_j^z}{\sum_{n=1}^{K} e^{z_k}} \quad for \quad j = 1 \cdots K \tag{4}$$

where z and K refer to the input vector and number of classes in the multi-class case, respectively. The similarity of these activation functions can be easily seen. The difference is the denominator that is a summation of all of the values. Turning to the applied architecture networks, Figure 3 shows the network structure of the U-Net and residual U-Net in a more comparable way.

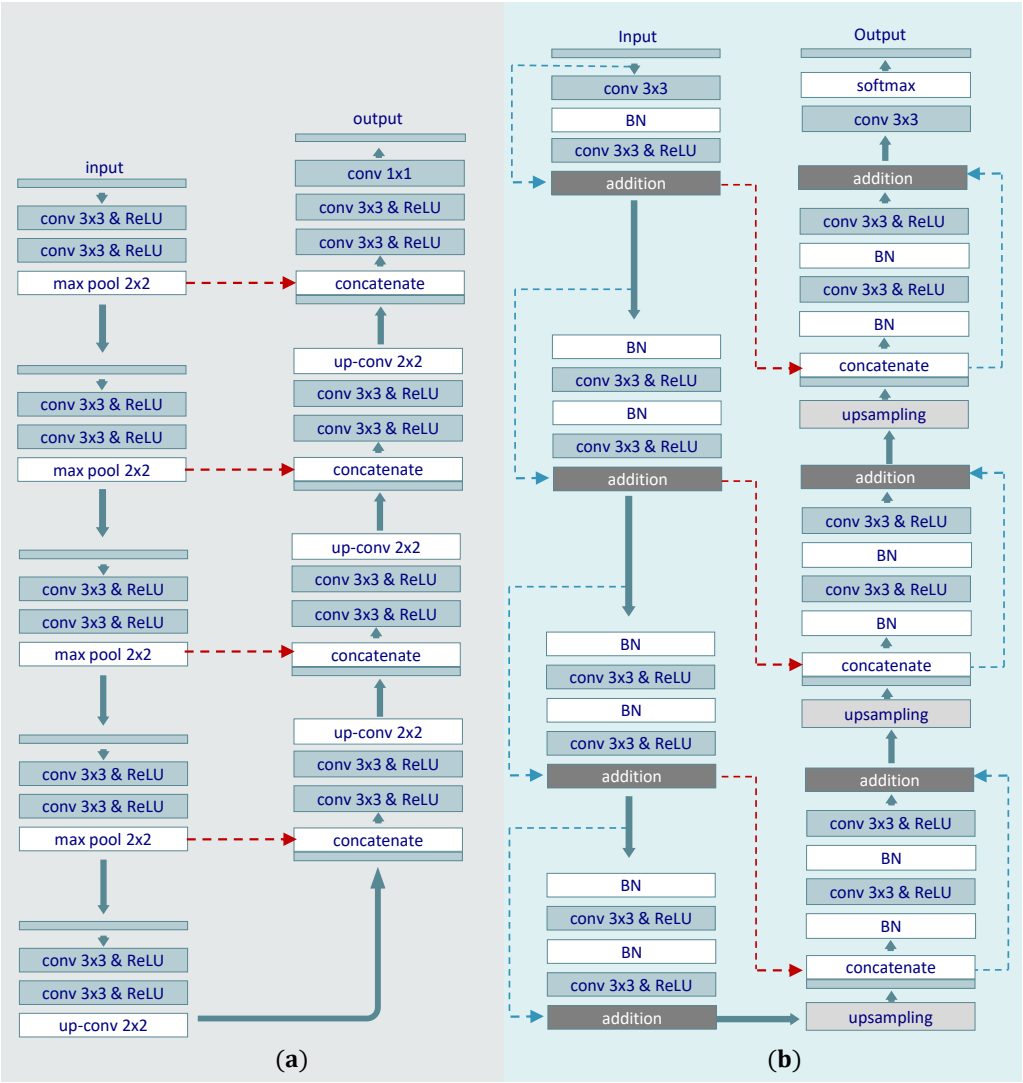

Figure 3. Illustration of the architecture of: (**a**) a typical convolutional process in a U-Net; and (**b**) a residual U-Net with an identity mapping. The dotted red and blue refer to the skip connections between low levels and high levels of the network and identity mapping, respectively.

2.4. Applied Loss Functions

The cross-entropy loss function was used for our applied semantic segmentation networks. Let $p_s(\cdot)$ stand for the mapping functions of the prediction in the model. Accordingly, the cross-entropy loss $\mathcal{L}_{ce}(\theta_s, x, y)$ has the following Formula (5):

$$\mathcal{L}_{ce}(\theta_s, x, y) = -\sum_{k=1}^{n_v} y^{(k)} \log p_s(x)^{(k)}, \tag{5}$$

where n_v indicates how many classes are involved in the classification task, and y is the ground truth of the input sample x [4].

2.5. Implementation Details

In this study, we specifically used the open-source packages of the U-Net and residual U-Net developed by [46,47]. Both algorithms were built on Python 3.6, TensorFlow, and the Keras functional API. To compare the algorithms in the same situation and to evaluate the models with a limited number of training data, they were trained from scratch, and no data augmentation was used in their training process. Theoretically, both architectures can work with the arbitrary size of image patches as training input data. Based on dividing our

WorldView-3 images into training and testing, around 270 training image patches with a size of 128 × 128 were randomly selected for the training of both architectures. The fixed size for the image patches was selected due to our available data, decreasing the randomness of adverse effects in the results, and our applied hardware. In this study, to train the U-net and the residual U-Net, we used Adam optimizer for optimizing both architectures due to its capabilities in convergence performance [63]. Both architectures were trained with the same default learning rate parameters of 0.001 with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The batch size was set as 4 during training; 60 epochs derived our best results from the U-Net and 32 epochs from the residual U-Net.

2.6. Evaluation Metrics

True positives (*TP*) indicate correctly classified dwellings; false positives (*FP*) indicate wrong classification pixels which belong to another class; false negatives (*FN*) refer to misclassified pixels based on the ground truth. Typical evaluation metrics of precision, recall, *F*1, and kappa coefficient were considered to quantitatively evaluate the performance of our applied models. Precision represents the proportion of pixels that are correctly detected for each class. The recall, also known as sensitivity, is the proportion of pixels in the labeled data allocated to the correct class [64,65]. *F*1 is a helpful quantitative metric to measure the balance between precision and recall (see Equations (6)–(8)). The kappa coefficient, calculated from the resulting confusion matrices, also corresponds to the agreement degree between the labeled data and the resulting classification.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

3. Results

The classification models we developed were finalized through the use of softmax, which provides a probability distribution over each of the four classes considered. To determine the class in which a pixel belongs, the maximum argument of the softmaxed output tensor is taken into account. According to each model, the output probability maps, or so-called heat maps, can vary based on how much of the details each model produces in regard to the class boundaries and the *TP*, *FP*, and *FN* detections. The following sections provide qualitative and quantitative assessments of the outputs from the applied models.

3.1. Qualitative Assessment

We demonstrate the effectiveness of U-Net and residual U-Net for dwelling detection from WorldView-3 satellite imagery. Both architectures were trained and tested with the same input data to compare the performance and the resulting accuracy. Within this study, we applied two different network architectures to observe the accuracy of the classification of the bright dwelling structures. The immediate result of each structure is the heat map of probabilities. These maps are created based on each architecture and are represented in Figure 4. For a subjective comparison, visually inspecting Figure 4b shows that the U-Net resulting in a probability of the class of Rectangular Shape on the corner of a Facility Building led to a separate object that wrongly classified the Rectangular Shape in this area (see Figure 5b), while based on Figure 6b, this matters less for the case of the residual U-Net so that the whole area of that Facility Building is correctly classified. Turning to the latter, Figures 5 and 6 illustrate some example image patches of the test set with the ground truth information (first column) and classification results of the U-Net and the residual U-Net (second column) along with comparison by overlapping the ground truth polygons and the classification results.

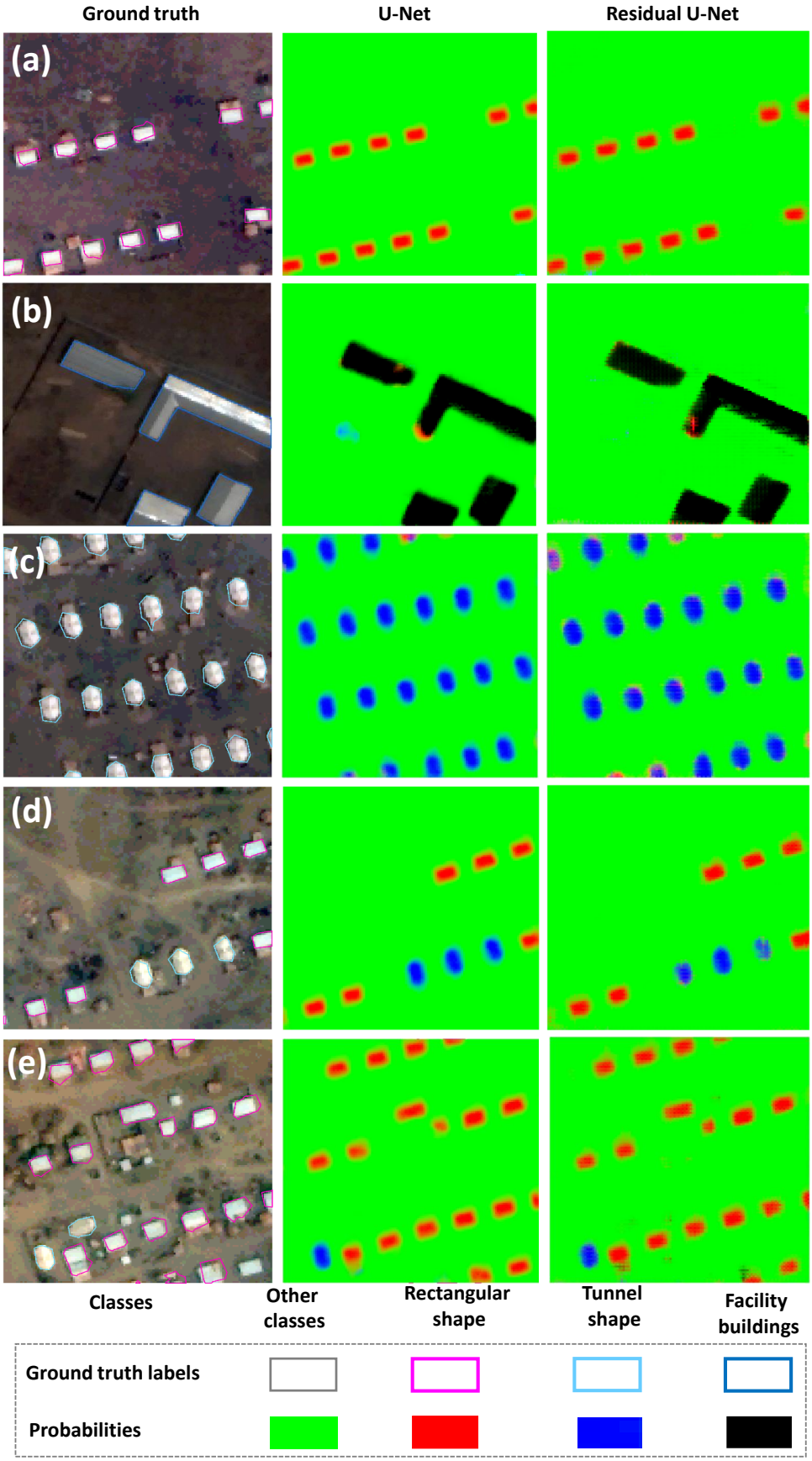

Figure 4. Heat maps demonstrate the probability distribution over different classes. The image patches (**a**–**e**) are selected from the test set.

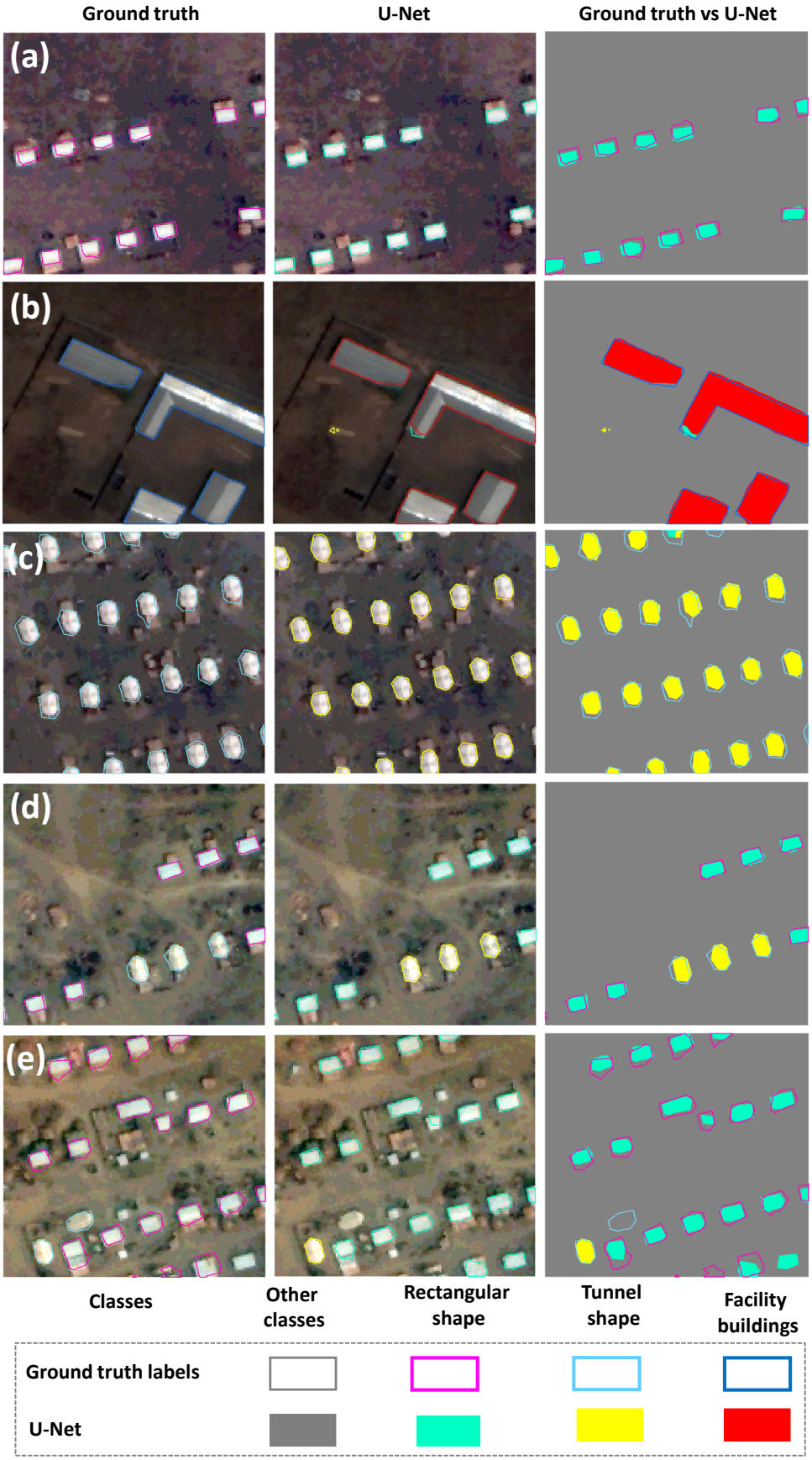

Figure 5. Example classification results on the test set of image patches resulting from U-net and comparison with the ground truth labels. The image patches (**a**–**e**) are selected from the test set.

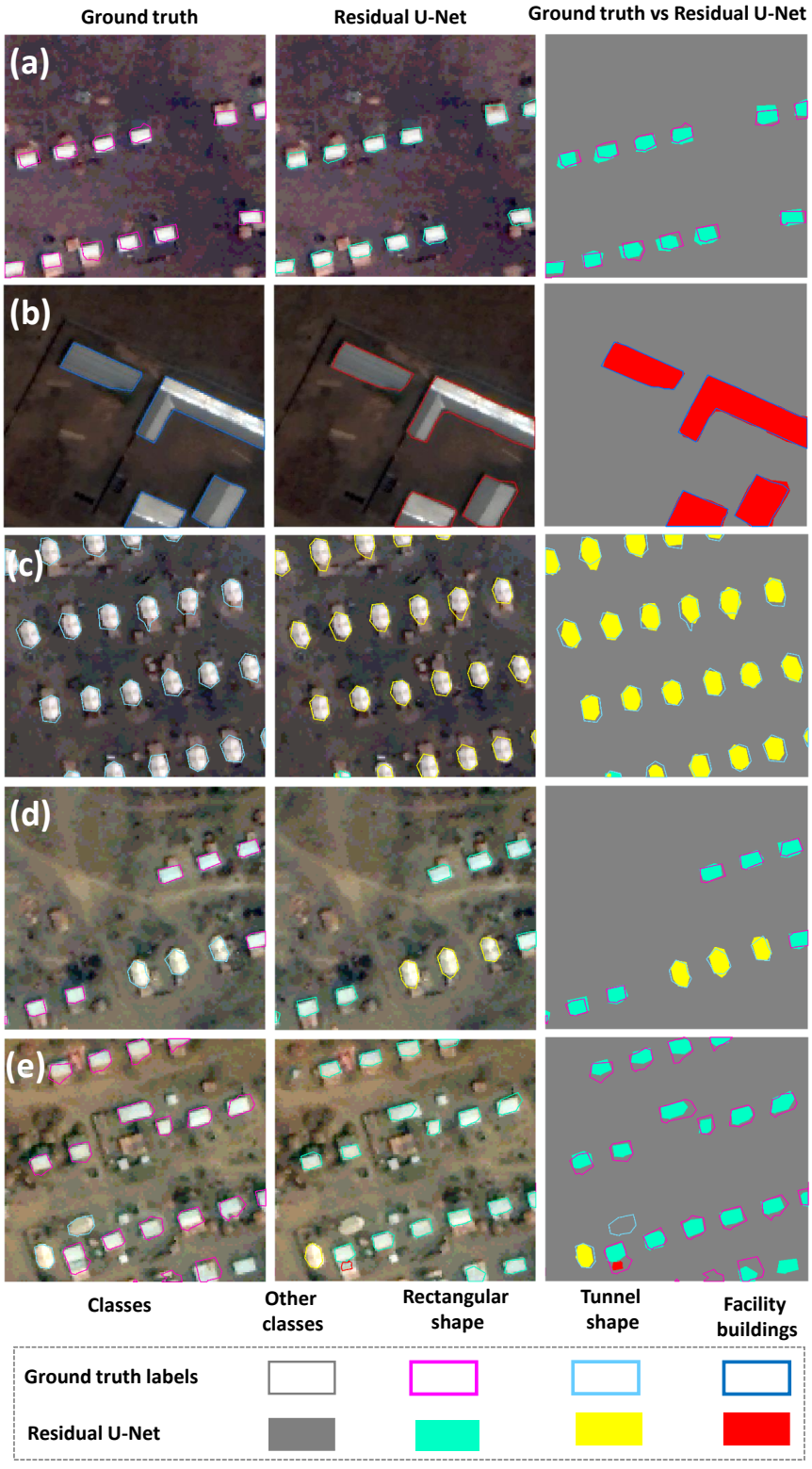

Figure 6. Example classification results on the test set of image patches resulting from residual U-net and comparison with the ground truth labels. The image patches (**a**–**e**) are selected from the test set.

3.2. Quantitative Assessment

Table 1 contains numerical results of the accuracy assessment of the trained U-Net and residual U-Net. The accuracy was assessed on the hold-out test subset of the image patches. As mentioned above, the focus of our trained architectures is set on the bright dwelling structures, including Facility Buildings, Tunnel Shape, and Rectangular shape in

a refugee camp, which is essential for population estimations. We calculated *TP*, *FP*, and *FN* values for each image patch from the test data set to obtain precision, recall, and *F*1. Moreover, the confusion matrix was created to evaluate the applied architectures' behavior on different classes and calculate the kappa coefficient (see Figure 7). Additionally, to evaluate the robustness and transferability of the trained U-Net and residual U-Net, we applied them on more than 250 randomly selected image patches, which were created without any overlap in between. Table 1 depicts the resulting precision, recall, *F*1, and kappa coefficient values for the Other Classes, Rectangular shape dwellings, Tunnel Shape dwellings, and Facility Buildings. The accuracy assessment was conducted for randomly selected image patches that were not used for the training process. Except for the precision of Facility Buildings, the resulting precision, recall, and *F*1 values from the residual U-Net for every class were higher than those resulting from the U-Net. The Other Classes had the lowest difference among the two architectures as for all evaluation metrics; the difference was less than one percent. Tunnel Shape dwellings obtained the highest difference of 6.47 percentage points in precision resulting from each architecture. However, the recall Rectangular shape evaluation metric had the highest improvement of 12.33 percentage points by using the residual U-Net. For the *F*1 measure, known as the weighted harmonic mean of precision and recall, the resulting values for Rectangular shape dwellings, Tunnel Shape dwellings, and Facility Buildings increased by 8.92, 5.9, and 2.66 percentage points, respectively. The kappa coefficients, which are considered a comprehensive summary of the confusion matrices of the classification results were 82.26% for the U-Net and 82.37% for the residual U-Net.

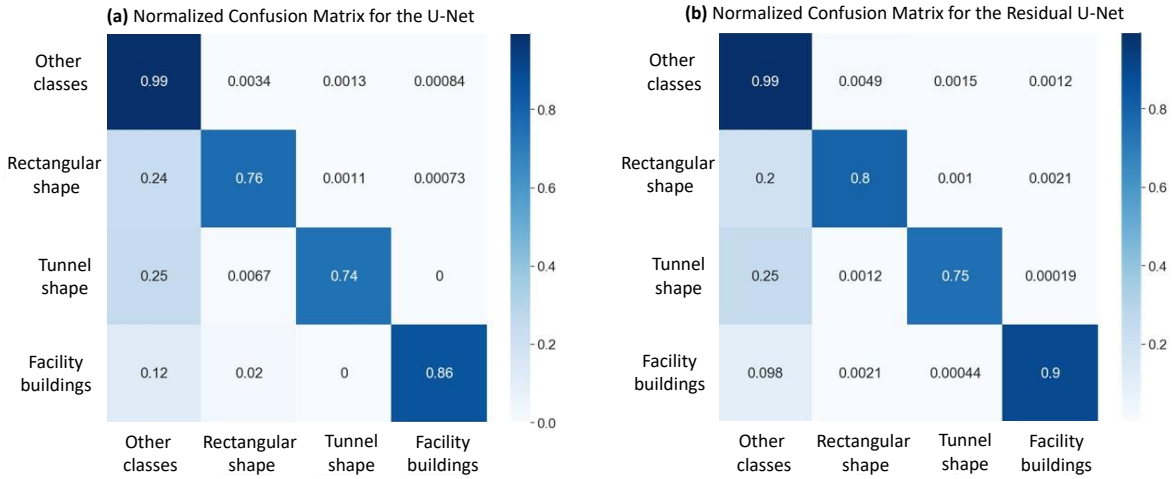

Figure 7. Resulting normalized confusion matrices for: (**a**) the U-Net; and (**b**) the Residual U-Net.

Table 1. Dwelling classification results for the testing image patches for both trained network architectures of the U-Net and the residual U-Net. Accuracies are stated as percentages of precision, recall, *F*1, and kappa coefficient.

Model	**U-Net**			**Residual U-Net**		
Evaluation metrics	Precision	Recall	*F*1	Precision	Recall	*F*1
Other classes	98.8	99.45	99.12	99.31	99.61	99.46
Rectangular shape	87.24	99.45	81.01	92.03	87.93	89.93
Tunnel shape	89.19	76.18	82.17	95.66	81.6	88.07
Facility buildings	90.85	87.54	89.17	90.06	93.67	91.83
Kappa coefficient		82.26			82.37	

4. Discussion

In this study, we applied two U-Net-based FCNs for dwelling classification on a single WorldView-3 image. Within the field of semantic segmentation (including in the context

of refugee camp analysis), several current studies have proposed that a deeper CNN or FCN architecture would have resulted in a classification with higher accuracy [47]. The downside is that dealing with deeper networks leads to problems such as vanishing the gradients and losing the information at the end of the network. Therefore, in this study, we applied residual U-Net and compared the results with the performance of standard U-Net. The architectures were trained and tested on the same image of a refugee camp. In order to demonstrate the transferability of the trained architectures, we did not consider overlaps between the image patches. The data set was randomly split into training and testing data. Although a limited amount of data was used in this study, and the division rate was in the proportion of almost 1:1 (255 patches for training and 270 for testing), both architectures obtained remarkable classification results. They suggest the high transferability of the applied architectures. Moreover, despite reaching an almost similar kappa coefficient on the refugee camp classification task, the residual U-Net has illustrated higher predictive power to produce segmentation and classification with higher accuracy in terms of precision, recall, and *F*1 than the standard U-Net. The results underscore that using a residual learning framework in a classification task improves results mainly in the boundary zones of the classified objects. Therefore, the experimental results demonstrated that the residual learning strategy is helpful to improve the learning capabilities of the U-Net. This increases performance in terms of resulting accuracies for dwelling classification. Moreover, the residual learning strategy boosts the convergence of the U-Net as the best results of the residual U-Net were obtained within 28 epochs less than the U-Net. This may be due to the defined skip connections in each residual unit as well as among downsampling and the upsampling paths of the U-Net, which could ease information transmissions in computation processes. A higher number of epochs was needed for the U-Net to obtain the best results as compared with the residual U-Net, illustrating the stronger generalization ability of the residual U-Net in this classification task compared to the U-Net.

Compared to previous CNN-based works on refugee camp segmentation and classification, we achieved convincing results even by using a single image and a pretty limited number of patches without any data augmentation scheme. Therefore, our classification results show higher accuracy compared to the current study of [21]. The mentioned study integrated a CNN model with an OBIA rule-based post-processing refinement process to classify dwellings in the same camp area. The *F*1 metric for all classes was more than 90% using an object-based accuracy assessment. In contrast, our results are based on and assessed per single pixel range from more than 88% for the Tunnel shape class to 99.46% for the other classes but without applying any post-processing refinement. Similar image segmentation and classification tasks were carried out by applying the U-Net and residual U-Net for classification of other geographical features, such as landslides, confirm our results in this study. Liu et al. [51] and Yi and Zhang [66], for instance, applied the U-Net and residual U-Net and compared their performance for landslide segmentation and classification from VHR imageries. All these studies obtained more accurate results using the residual U-Net than the standard U-Net.

Using AI for any application has always been fraught with ethical concerns [67]. The RS data-based DL applications are dependent on large data sets, some of which can be very sensitive, as in the present study. Gathering data from RS sources has long existed, but due to higher image resolutions, increasing sources of RS data, and because of the use cases that directly impact security and the lives of the people living in the study area, ethical issues are becoming increasingly important. In this study, the VHR images and labels of the training and test data sets are sourced by humanitarian mapping products generated by the Department of Geoinformatics, University of Salzburg, and Spatial Services Ltd. in support of operational humanitarian mapping services given to MSF. Hence, a better understanding of the ground situation in the IDP/refugee camps, such as the number of people still living there, requires more detailed and accurate information derived from the provided RS data. Due to ethical concerns and the safety of people residing in the case study area, we chose the least sensitive case that may no longer exist. As a result, the data

used in this study cannot be considered as part of an ongoing humanitarian aid project, but rather a test case to develop and practice DL models and to provide solutions for upcoming needs. Furthermore, we do not provide any information that may be considered sensitive, such as population estimation or any other type of statistic.

5. Conclusions

We applied U-Net and residual U-Net for refugee camp classification using high-resolution satellite imagery within the present study. The network of the residual U-Net could show the strengths of both U-Net architecture and residual learning by obtaining slightly higher accuracies in this study. Although both architectures achieved excellent training performances by effectively extracting the spectral and spatial characteristics of image patches, the residual U-Net revealed a more precise and more accurate object boundary representation than the U-Net, especially for the class of Facility Buildings. In the future, we will evaluate several common state-of-the-art algorithms using a larger training set including several camps around the world in order to benchmark our results and data sets.

Author Contributions: Conceptualization, O.G.; methodology, O.G. and A.C.; validation, O.G. and A.C.; investigation, O.G. and D.T.; resources, D.T. and S.L.; data curation, D.T. and S.L.; writing—original draft preparation, O.G.; writing—review and editing, O.G., D.T., S.L. and P.G.; visualization, O.G. and A.C.; supervision, D.T., S.L. and P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Institute of Advanced Research in Artificial Intelligence (IARAI) GmbH, the Austrian Federal Ministry for Digital and Economic Affairs, the Christian Doppler Research Association, and MSF Austria (Ärzte ohne Grenzen Austria).

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the anonymous referees for their valuable comments/suggestions that have helped us improve an earlier version of the manuscript.

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

References

1. 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]
2. 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]
3. 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]
4. Xu, Y.; Ghamisi, P. Universal Adversarial Examples in Remote Sensing: Methodology and Benchmark. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
5. Wang, W.; Chen, Y.; Ghamisi, P. Transferring CNN With Adaptive Learning for Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
6. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS–A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**, arXiv:1906.07789.
7. Sumbul, G.; Charfuelan, M.; Demir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan, 28 July–2 August 2019; pp. 5901–5904.
8. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4Sense: Reference Benchmark Data and Deep Learning Models for Landslide Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
9. Zhu, X.X.; Hu, J.; Qiu, C.; Shi, Y.; Kang, J.; Mou, L.; Bagheri, H.; Häberle, M.; Hua, Y.; Huang, R.; et al. So2Sat LCZ42: A benchmark dataset for global local climate zones classification. *arXiv* **2019**, arXiv:1912.12171.
10. Center for International Earth Science Information Network (CIESIN); Flowminder Foundation; United Nations Population Fund (UNFPA); WorldPop, University of Southampton. Mapping and Classifying Settlement Locations 2020. Available online: https://eprints.soton.ac.uk/469540/ (accessed on 10 November 2022).

11. Marconcini, M.; Metz-Marconcini, A.; Üreyen, S.; Palacios-Lopez, D.; Hanke, W.; Bachofer, F.; Zeidler, J.; Esch, T.; Gorelick, N.; Kakarla, A.; et al. Outlining where humans live, the World Settlement Footprint 2015. *Sci. Data* **2020**, *7*, 1–14. [CrossRef]
12. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [CrossRef]
13. Nations, U. *World Urbanization Prospects: The 2005 Revision*; United Nations Publications: New York, NY, USA, 2011.
14. Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Freire, S.; Haag, F.; Halkia, M.; Julea, A.; Kemper, T.; Soille, P. Global human settlement analysis for disaster risk reduction. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Berlin/Heidelberg, Germany, 11–15 May 2015.
15. Aguilar, R.; Kuffer, M. Cloud computation using high-resolution images for improving the SDG indicator on open spaces. *Remote Sens.* **2020**, *12*, 1144. [CrossRef]
16. Ghaffarian, S.; Roy, D.; Filatova, T.; Kerle, N. Agent-based modelling of post-disaster recovery with remote sensing data. *Int. J. Disaster Risk Reduct.* **2021**, *60*, 102285. [CrossRef]
17. Ghaffarian, S.; Kerle, N.; Pasolli, E.; Jokar Arsanjani, J. Post-disaster building database updating using automated deep learning: An integration of pre-disaster OpenStreetMap and multi-temporal satellite data. *Remote Sens.* **2019**, *11*, 2427. [CrossRef]
18. Witmer, F.D. Remote sensing of violent conflict: Eyes from above. *Int. J. Remote Sens.* **2015**, *36*, 2326–2352. [CrossRef]
19. Gruca, A.; Herruzo, P.; Rípodas, P.; Kucik, A.; Briese, C.; Kopp, M.K.; Hochreiter, S.; Ghamisi, P.; Kreil, D.P. CDCEO'21-First Workshop on Complex Data Challenges in Earth Observation. In Proceedings of the the 30th ACM International Conference on Information & Knowledge Management, Virtual, 1–5 November 2021; pp. 4878–4879.
20. Rizeei, H.M.; Pradhan, B. Urban mapping accuracy enhancement in high-rise built-up areas deployed by 3D-orthorectification correction from WorldView-3 and LiDAR imageries. *Remote Sens.* **2019**, *11*, 692. [CrossRef]
21. 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**, *54*, 127–140. [CrossRef]
22. 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. [CrossRef]
23. 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. *PFG Photogramm.* **2013**, *3*, 185–195. [CrossRef] [PubMed]
24. Gao, Y.; Gella, G.W.; Liu, N. Assessing the Influences of Band Selection and Pretrained Weights on Semantic-Segmentation-Based Refugee Dwelling Extraction from Satellite Imagery. *AGILE GISci. Ser.* **2022**, *3*, 1–6. [CrossRef]
25. Gella, G.W.; Wendt, L.; Lang, S.; Braun, A.; Tiede, D.; Hofer, B.; Gao, Y.; Riedler, B.; Alobaidi, A.; Schwendemann, G.M. Testing transferability of deep-learning-based dwelling extraction in refugee camps. *GI_Forum* **2021**, *9*, 220–227. [CrossRef]
26. Lang, S.; Tiede, D.; Hölbling, 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]
27. Lüthje, F.; Tiede, D.; Füreder, P. Don't see the dwellings for the trees: Quantifying the effect of tree growth on multi-temporal dwelling extraction in a refugee camp. In Proceedings of the GI_Forum, Salzburg, Austria, 7–10 July 2015
28. Tiede, D.; Lang, S.; Hölbling, D.; Füreder, P. Transferability of OBIA rulesets for IDP camp analysis in Darfur. In Proceedings of the GEOBIA, Ghent, Belgium, 29 June–2 July 2010.
29. Ghorbanzadeh, O.; Shahabi, H.; Crivellari, A.; Homayouni, S.; Blaschke, T.; Ghamisi, P. Landslide detection using deep learning and object-based image analysis. *Landslides* **2022**, *19*, 929–939. [CrossRef]
30. 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]
31. Kemper, T.; Jenerowicz, M.; Pesaresi, M.; Soille, P. Enumeration of dwellings in Darfur Camps from GeoEye-1 satellite images using mathematical morphology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *4*, 8–15. [CrossRef]
32. Laneve, G.; Santilli, G.; Lingenfelder, I. Development of automatic techniques for refugee camps monitoring using very high spatial resolution (VHSR) satellite imagery. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 31 July–4 August 2006; pp. 841–845.
33. 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] [PubMed]
34. 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]
35. Duan, Y.; Zhang, W.; Huang, P.; He, G.; Guo, H. A New Lightweight Convolutional Neural Network for Multi-Scale Land Surface Water Extraction from GaoFen-1D Satellite Images. *Remote Sens.* **2021**, *13*, 4576. [CrossRef]
36. Zheng, J.; Fu, H.; Li, W.; Wu, W.; Yu, L.; Yuan, S.; Tao, W.Y.W.; Pang, T.K.; Kanniah, K.D. Growing status observation for oil palm trees using Unmanned Aerial Vehicle (UAV) images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 95–121. [CrossRef]
37. Haq, M.A.; Ahmed, A.; Khan, I.; Gyani, J.; Mohamed, A.; Attia, E.A.; Mangan, P.; Pandi, D. Analysis of environmental factors using AI and ML methods. *Sci. Rep.* **2022**, *12*, 1–16. [CrossRef]

38. Shahabi, H.; Rahimzad, M.; Tavakkoli Piralilou, S.; Ghorbanzadeh, O.; Homayouni, S.; Blaschke, T.; Lim, S.; Ghamisi, P. Unsupervised deep learning for landslide detection from multispectral sentinel-2 imagery. *Remote Sens.* **2021**, *13*, 4698. [CrossRef]
39. Srivastava, A.; Yetemen, O.; Saco, P.M.; Rodriguez, J.F.; Kumari, N.; Chun, K.P. Influence of Orographic Precipitation on Coevolving Landforms and Vegetation in Semi-arid Ecosystems. *Earth Surf. Process. Landforms* **2022**, *47*, 2846–2862. [CrossRef]
40. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
41. Jozdani, S.E.; Johnson, B.A.; Chen, D. Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. *Remote Sens.* **2019**, *11*, 1713. [CrossRef]
42. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sens.* **2018**, *10*, 1119. [CrossRef]
43. Cui, B.; Fei, D.; Shao, G.; Lu, Y.; Chu, J. Extracting raft aquaculture areas from remote sensing images via an improved U-net with a PSE structure. *Remote Sens.* **2019**, *11*, 2053. [CrossRef]
44. Sherrah, J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv* **2016**, arXiv:1606.02585 .
45. DeLancey, E.R.; Simms, J.F.; Mahdianpari, M.; Brisco, B.; Mahoney, C.; Kariyeva, J. Comparing deep learning and shallow learning for large-scale wetland classification in Alberta, Canada. *Remote Sens.* **2019**, *12*, 2. [CrossRef]
46. 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.
47. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
48. Jiang, X.; Wang, Y.; Liu, W.; Li, S.; Liu, J. Capsnet, cnn, fcn: Comparative performance evaluation for image classification. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 840–848. [CrossRef]
49. Lu, Y.; Koperski, K.; Kwan, C.; Li, J. Deep Learning for Effective Refugee Tent Extraction Near Syria–Jordan Border. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1342–1346. [CrossRef]
50. Tang, X.; Tu, Z.; Wang, Y.; Liu, M.; Li, D.; Fan, X. Automatic Detection of Coseismic Landslides Using a New Transformer Method. *Remote Sens.* **2022**, *14*, 2884. [CrossRef]
51. Liu, P.; Wei, Y.; Wang, Q.; Chen, Y.; Xie, J. Research on post-earthquake landslide extraction algorithm based on improved U-Net model. *Remote Sens.* **2020**, *12*, 894. [CrossRef]
52. Kalantar, B.; Ueda, N.; Saeidi, V.; Janizadeh, S.; Shabani, F.; Ahmadi, K.; Shabani, F. Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane, Australia. *Remote Sens.* **2021**, *13*, 2638. [CrossRef]
53. Naderpour, M.; Rizeei, H.M.; Ramezani, F. Forest fire risk prediction: A spatial deep neural network-based framework. *Remote Sens.* **2021**, *13*, 2513. [CrossRef]
54. Meng, Z.; Li, L.; Tang, X.; Feng, Z.; Jiao, L.; Liang, M. Multipath residual network for spectral-spatial hyperspectral image classification. *Remote Sens.* **2019**, *11*, 1896. [CrossRef]
55. 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.
56. Zelikowsky, M.; Bissiere, S.; Hast, T.A.; Bennett, R.Z.; Abdipranoto, A.; Vissel, B.; Fanselow, M.S. Prefrontal microcircuit underlies contextual learning after hippocampal loss. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 9938–9943. [CrossRef]
57. Li, L. Deep residual autoencoder with multiscaling for semantic segmentation of land-use images. *Remote Sens.* **2019**, *11*, 2142. [CrossRef]
58. Seydi, S.T.; Rastiveis, H.; Kalantar, B.; Halin, A.A.; Ueda, N. BDD-Net: An End-to-End Multiscale Residual CNN for Earthquake-Induced Building Damage Detection. *Remote Sens.* **2022**, *14*, 2214. [CrossRef]
59. Wang, L.; Zhang, J.; Liu, P.; Choo, K.K.R.; Huang, F. Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification. *Soft Comput.* **2017**, *21*, 213–221. [CrossRef]
60. Khryashchev, V.; Larionov, R. Wildfire segmentation on satellite images using deep learning. In Proceedings of the 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), Moscow, Russia, 11–13 March 2020; pp. 1–5.
61. Abderrahim, N.Y.Q.; Abderrahim, S.; Rida, A. Road segmentation using u-net architecture. In Proceedings of the 2020 IEEE International conference of Moroccan Geomatics (Morgeo), Moscow, Russia, 11–13 March 2020; pp. 1–4.
62. Yang, Z.; Xu, C.; Li, L. Landslide Detection Based on ResU-Net with Transformer and CBAM Embedded: Two Examples with Geologically Different Environments. *Remote Sens.* **2022**, *14*, 2885. [CrossRef]
63. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
64. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. *Remote Sens.* **2020**, *12*, 1737. [CrossRef]
65. Rahmati, O.; Panahi, M.; Ghiasi, S.S.; Deo, R.C.; Tiefenbacher, J.P.; Pradhan, B.; Jahani, A.; Goshtasb, H.; Kornejady, A.; Shahabi, H.; et al. Hybridized neural fuzzy ensembles for dust source modeling and prediction. *Atmos. Environ.* **2020**, *224*, 117320. [CrossRef]

66. Yi, Y.; Zhang, W. A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal RapidEye satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6166–6176. [CrossRef]
67. Kochupillai, M.; Kahl, M.; Schmitt, M.; Taubenböck, H.; Zhu, X.X. Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities. *IEEE Geosci. Remote Sens. Mag.* **2022**, 2–36. [CrossRef]