Article

Remote Sensing and Deep Learning to Understand Noisy OpenStreetMap †

Munazza Usmani [1,2,‡], Francesca Bovolo [2,*,‡] and Maurizio Napolitano [2,‡]

1 Department of Information Engineering and Computer Science (DISI), University of Trento, Via Sommarive 9, 38123 Trento, Italy; musmani@fbk.eu
2 Center for Digital Society (DIGIS), Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy; napo@fbk.eu
* Correspondence: bovolo@fbk.eu; Tel.: +39-3316-294-549
† This paper is an extended version of our work published in the 4th International Electronic Conference on Remote Sensing (ECRS-2022), 25–27 January 2022.
‡ These authors contributed equally to this work.

Abstract: The OpenStreetMap (OSM) project is an open-source, community-based, user-generated street map/data service. It is the most popular project within the state of the art for crowdsourcing. Although geometrical features and tags of annotations in OSM are usually precise (particularly in metropolitan areas), there are instances where volunteer mapping is inaccurate. Despite the appeal of using OSM semantic information with remote sensing images, to train deep learning models, the crowdsourced data quality is inconsistent. High-resolution remote sensing image segmentation is a mature application in many fields, such as urban planning, updated mapping, city sensing, and others. Typically, supervised methods trained with annotated data may learn to anticipate the object location, but misclassification may occur due to noise in training data. This article combines Very High Resolution (VHR) remote sensing data with computer vision methods to deal with noisy OSM. This work deals with OSM misalignment ambiguity (positional inaccuracy) concerning satellite imagery and uses a Convolutional Neural Network (CNN) approach to detect missing buildings in OSM. We propose a translating method to align the OSM vector data with the satellite data. This strategy increases the correlation between the imagery and the building vector data to reduce the noise in OSM data. A series of experiments demonstrate that our approach plays a significant role in (1) resolving the misalignment issue, (2) instance-semantic segmentation of buildings with missing building information in OSM (never labeled or constructed in between image acquisitions), and (3) change detection mapping. The good results of precision (0.96) and recall (0.96) demonstrate the viability of high-resolution satellite imagery and OSM for building detection/change detection using a deep learning approach.

Keywords: remote sensing; crowdsourcing; data reliability; deep learning; change detection

1. Introduction

Geographic information gathered by people known as Volunteer and Technical Communities (VTCs) or digital humanitarians [1], frequently on a volunteer basis [2], is referred to as Volunteered Geographic Information (VGI). VGI enables the quick collection of accurate information before, during, and after a catastrophe, making this information open and publicly available [3] and addressing the shortcomings of existing mapping technologies and data sources [4,5]. The amount of crowdsourcing mapping information via web applications such as Google Maps and OpenStreetMap (OSM) is substantial, encompassing a considerable portion of the world's current human settlements. Among them, OSM is a collaborative effort founded in 2004 by Steve Coast with the goal of "creating a free editable map of the globe". The majority of mapping efforts during the first year were concentrated on road and transportation networks [6]. From then on, OSM has regularly been updated

with new geographical data, such as buildings and their functioning, land use, public transit information, etc. This type of information enables local governments and societies to undertake better risk management and handle emergencies, and it is frequently used in disaster management programs. The number of users/contributors involved in OSM is always growing. In the OSM community, there are around 5.5 million users, with 1 million contributors that are supposed to make over 3 million modifications each day, as well as specialist groups such as the Humanitarian OSM Team (HOT) that work on improving OSM data to help in need. Even though most OSM road network data are projected and comprehensive in comparison to other elements [7], building footprints are one example of an OSM feature whose coverage and completeness varies substantially not only between but also within nations [8]. Remote places, for example, have less coverage than densely populated metropolitan areas. These differences are a result of socioeconomic factors, including, but not limited to, population density and distribution, accessibility to major cities, and the location of contributing users. Although the information on LULC is widely available for metropolitan regions, several rural structures are still not mapped. Up-to-date building maps are critical for demographic studies and assisting different organizations in crisis response planning or change detection analysis after a natural disaster. So, it is necessary to create and update the building footprint database and highlight the missing data. A large volume of building annotation data was acquired and made publicly available via OSM. However, in OSM, volunteer annotations suffer from some major flaws, most of which are caused by rare up-to-date satellite imagery and incomplete/incorrect annotation information by volunteers [9]. The noises in OSM are outlined below, and they will be addressed in the suggested solution.

1. Building footprints are regularly labeled but are up to 9 m off-center on the image plane [10]. The reason for this shift is that images used to digitize the footprint are different from the images used for further analysis. The misalignment issue is referred as registration noise, as shown in Figure 1.

Figure 1. Registration noise: Misalignment.

2. Missing objects in the annotation dataset: occurs because they have not been noticed by the volunteers, or may have been constructed in the time between the data acquisition. This ambiguity in the labels is known as omission noise (see Figure 2).
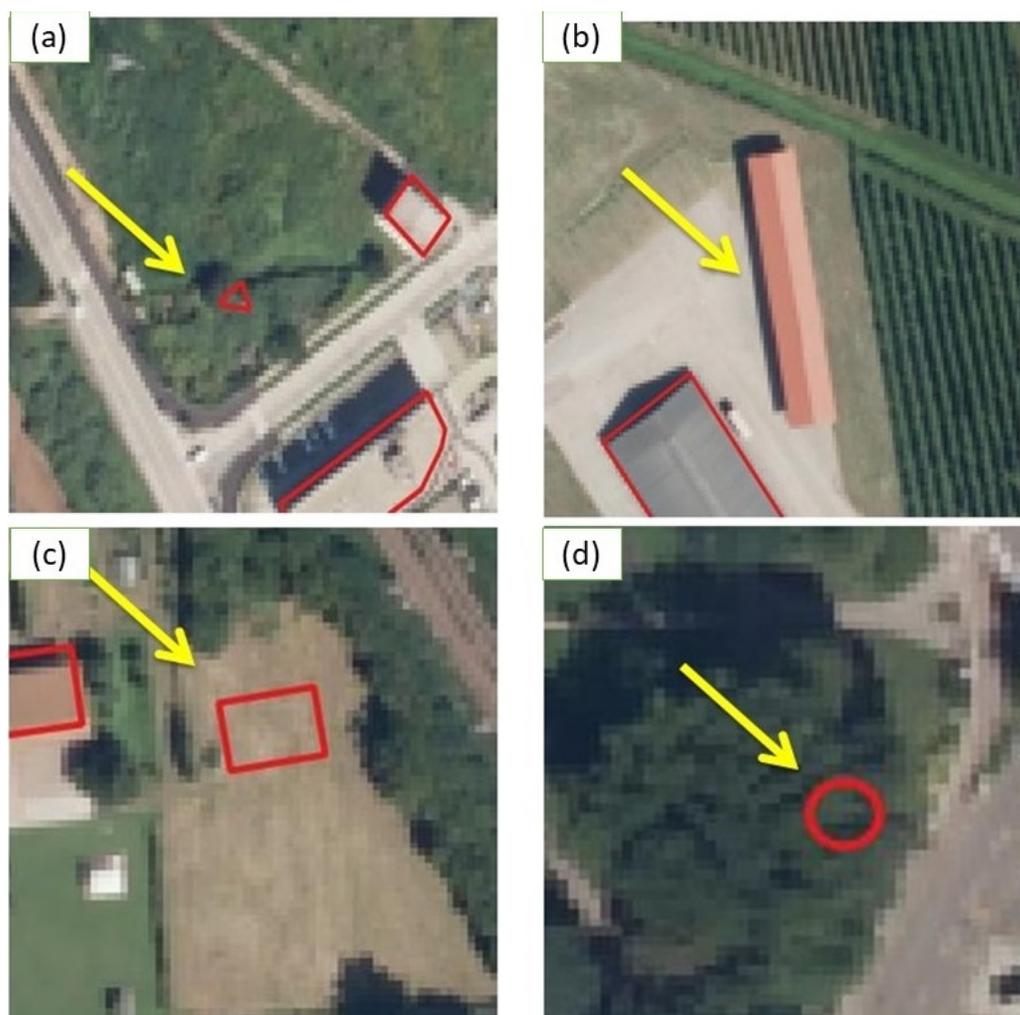
Figure 2. Each red box represents OSM data. Omission noise: (**a**) misclassification, (**b**) missing buildings, (**c**) misclassification, and (**d**) misclassification.

3. Objects may also remain in annotation when they do not exist at present, like buildings destroyed in natural disasters.

   With the increased use of VGI for disaster response and preparedness, different methodologies for assessing the quality and accuracy of VGI have been proposed, for example, in terms of data completeness, logical consistency, positional, thematic, semantic, spatial accuracy, temporal quality, and usability [11–13]. In this work, we addressed and handled three OSM-mentioned noises: misalignment, missing buildings, and updated data. We offer a methodology for improving misalignment with satellite imagery, predicting OSM missing coverage, and identifying gaps in OSM building footprints using regularly acquired remote sensing observations. Remote sensing data availability has increased dramatically, allowing a far better picture of our globe. With the recent evolution in deep learning and low-priced high-performance GPUs, object detection in satellite images is gaining popularity. It is very supportive in a wide range of applications to be able to correctly identify various sorts of objects in aerial imagery, such as building structures, roadways, and vegetation, as well as other categories, including map creation and maintenance, urban planning, environmental monitoring, and disaster relief. Building footprint extraction is a hot topic in the domain of remote sensing. However, due to differences in the layout of structures, complex environmental interference, shadows, and lighting circumstances, the automatic segmentation of building footprints from high-resolution images is still challenging. The most frequently used strategies in this situation are classification-based algorithms that use spectral, structural, and context information. Ref. [14], for example,

employs a Support Vector Machine (SVM) classifier to define the Pixel Shape Index (PSI), a shape that estimates the distance between adjacent gray pixels in each direction and combines it with spectral properties to extract buildings. However, the fundamental issue with these classification-based techniques is that they require several semantic labels to construct a classifier. The process is expensive and tends to limit large-scale implementation. Labeling satellite image data is a frequent problem for creating a variety of semantic maps for vast areas, whether for urban mapping or identifying land use/cover at a large scale [15]. Fully Convolutional Networks (FCNs) for the semantic labeling of urban areas or CNNs for land use classification are two methods that came into the community of computer vision for semantic segmentation and have been used effectively with cutting-edge outcomes on RGB remote sensing data [16]. Many works have studied the potential of crowd-sourcing to gather data and supplement the information generated from satellite imagery. However, there are still significant issues with quality and dependability, which are widely discussed in the research [17]. The OSM data, especially for humanitarian or disaster management purposes, continue to be limited by spatially variable data quality and the absence of suitable reference data. These issues of OSM can affect the performance of classification. OSM noise reduction and reliability improvement may benefit it. In this paper, we propose a composite fusion architecture that combines information from remote sensing images and OSM throughout the network and can handle the noise in OSM data. We experimentally show that the proposed method can perform better segmentation with quality-improved OSM (automatic pre-processing) than with noisy OSM. For this sake, we propose a workflow that incorporates the following techniques: (1) co-registration to handle misalignment noise, (2) missing object recognition, and, (3) updated data using advanced deep learning algorithms.

The related work for this research is described in Section 2, followed by the proposed methodology for removing noise in OSM building annotations (Section 3). Section 4 depicts the dataset and experimental setup. In addition to case studies, we contrast the findings of our proposed method with noisy and processed OSM layers in Section 5. After discussion in Section 6, the paper concludes with Section 7.

2. Background: Reducing OSM Noise Using Remote Sensing or Deep Learning

Despite being the most well-known and commonly utilized VGI platform, the OSM geographic heterogeneity data quality and availability continue to be an ongoing issue because of the incredibly varied volunteer mapping behaviors [9]. OSM suffers poor quality control due to enormous data like spatial variabilities, including the completeness of building footprints, which continues to be a significant difficulty on many regional scales [18]. Identifying missing regions in OSM is an important step for its reliability and effective voluntary contribution management. However, these issues are somehow improved because of the many tools created by communities and companies for system evaluation. For example, OSMCha is developed by the Mapbox company and used for data verification or to identify vandalism on data; JOSM is a Java editor for better and more efficient mapping; and more beneficial is the HOT, which was initialized for a quick response of mapping in any natural disaster situation and is used to update the map [19]. Along with the OSM-related research mentioned above, several researchers have explored how to apply deep learning technologies to a range of satellite image processing tasks, including LULC mapping. Deep learning architectures, such as CNNs have several advantages, including their independence from past information and hand-crafted features, which have aided their capacity to generalize more effectively. With encouraging outcomes, many CNN models have been proposed and used for semantic segmentation [20].

The task of generating building footprints falls under the semantic segmentation branch. Recent research in the remote sensing community has also attempted to generate building footprints accurately through the use of CNN models. Deep neural networks (remote sensing review in [16]) have recently been utilized in conjunction with other image processing approaches to successfully recognize and outline buildings in metropolitan

settings [21]. In the post-processing stage, the pixel (or region) level detection is usually combined into vector graphics. In [22], the authors utilized a CNN approach by skipping the post-processing step: vector footprints of buildings are learned automatically by specifying the building outline characterization as an active contour model and learning the parameters with a CNN approach. The fundamental disadvantage of employing CNN methods in remote sensing, regardless of the algorithm, is the requirement for a significant amount of labeled data for supervised classification. OSM semantic information has been employed as repositories of annotated data collections in recent studies [23–25]. However, because CNN structures are inherently invariant to spatial transformations, there are frequently non-sharp borders and visibly inferior outcomes in CNN-based semantic segmentation methods. In this situation, Mask R-CNN has been utilized to improve pixel-level segmentation. After training based on FCN in [26], Mask R-CNN is used to properly localize instance borders and give the most likely label to every pixel.

Considering the potential of using OSM data to train deep learning models, a data quality inconsistency issue is emerging. CNNs trained on this kind of reference data typically learn to predict the location of the item but not its precise extent [27]. The authors in [27] proposed a loss function to deal with noisy data, while [28] used a Recurrent Neural Network (RNN) to obtain accurate classification maps with a small dataset that is manually annotated. Others, such as [29], suggested pre-training with the entire dataset on a large scale and applying a domain adaptation with hand-labeled data. In recent years, from basic to advanced classification methods, remote sensing fusion with OSM has become of high relevance. Follow-up research [30] verified the suitability of ML-based building identification models for this application area.

Here, we focus on polygon annotations (buildings) in OSM because the noises described in Section 1 are mostly visible in the building data, and most rural regions are not entirely mapped in OSM. However, the concept can be extended to other OSM elements. The difficulty of applying a highly trained building detection model to geographically remote areas where building appearance may be significantly distinct and varied is one of the challenges we identified in the machine-assisted humanitarian mapping application. Our work addresses this issue through a case of fine-tuning with a small number of labels.

3. Proposed Methodology

In this work, a method is proposed to handle three noises of OSM using high-resolution satellite imagery based on a deep-learning segmentation model. We go over the steps for handling data quality and the follow-up method to map the missing buildings in OSM in the final output. The proposed method is then employed for change detection analysis. The suggested approach for data reliability issues and building a segmentation map is depicted in Figure 3. To produce a suitable training dataset for deep convolutional networks, data pre-processing is a key requirement in dealing with issues like misalignment and high dimensions. The dataset pre-processing in (Section 3.1), and the deep learning approach (Section 3.2) were utilized to train a binary classifier. Each step of the proposed technique is described below.
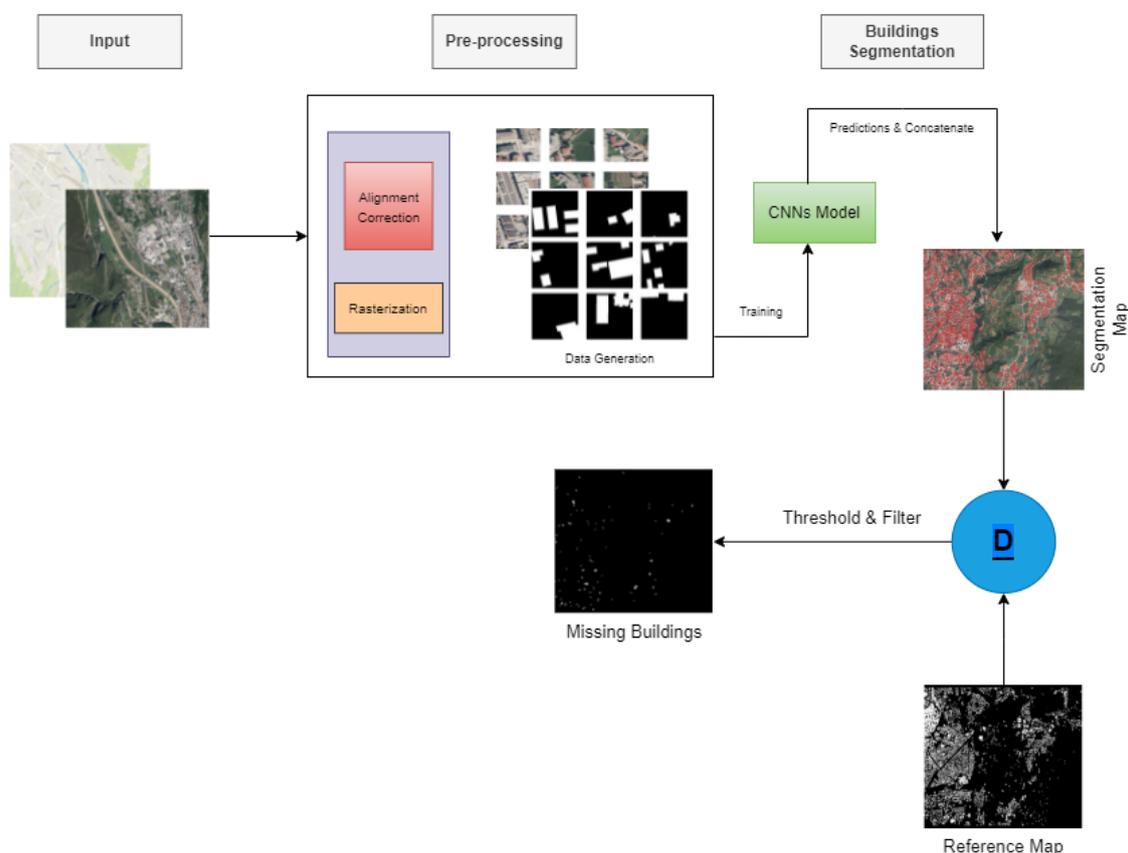
Figure 3. Proposed approach.

3.1. Pre-Processing

The difficulty in aligning OSM vector with imagery is referred to as a building registration problem (Figure 1). Image registration is crucial in combining images captured from various perspectives, at different times, or with different sensors. It is a method for calculating the point-to-point correlation between two scenes captured through a different source. The suggested image registration method employs an area-based approach, to achieve similarity using the Cross-Correlation (CCR) measure. We use a CCR-based technique for resolving the building misalignment error [31]. CCR has been used to solve registration issues in a variety of domains with great success [32,33]. The CCR is measured between a master (remote sensing image) and slave (OSM polygons) object, and we refer to these as a target and a source object, respectively. In this instance, it is possible to calculate the offsets in both directions (row and column), which correspond to the translation coefficients. Building structures usually exist in small groupings, each with the same alignment flaws. So, we align groups of buildings rather than individual structures. We have about 15 high-dimension (17,500 × 15,350/pixel) RS images with high spatial resolution. After extracting OSM data (3000 OSM buildings), we process and apply the CCR technique to each image separately. Handling each image separately and considering clusters of buildings significantly reduces the computational effort and improves numerical efficiency. Furthermore, using groups of buildings rather than single structures reduces the reliance on the building probability map quality.

Here, the CCR-based strategy maximizes/increases the correlation between a source and a target object while ensuring that buildings close to each other have similar shift correction vectors. We calculate the image gradient and cross correlation between polygons and gradient magnitude for an image window containing a cluster of buildings. The correlation coefficient will be at its maximum when the source polygons and target imagery are aligned. Alignment is estimated by shifting every source polygon to a high-building

probability. The process for the alignment of images is shown in Figure 4, where s and t represent source and target images, respectively. A function $g_\gamma(s, t) = u$ is calculated between the images with the CNN approach, where γ is the parameters of the convolutional layers, and u is the displacement between the images. The CNN architecture used in the proposed method is the Bayesian fully convolutional network. After defining the parameters, the defined algorithm takes the s and t inputs and calculates the maximum correlation (ϕ).
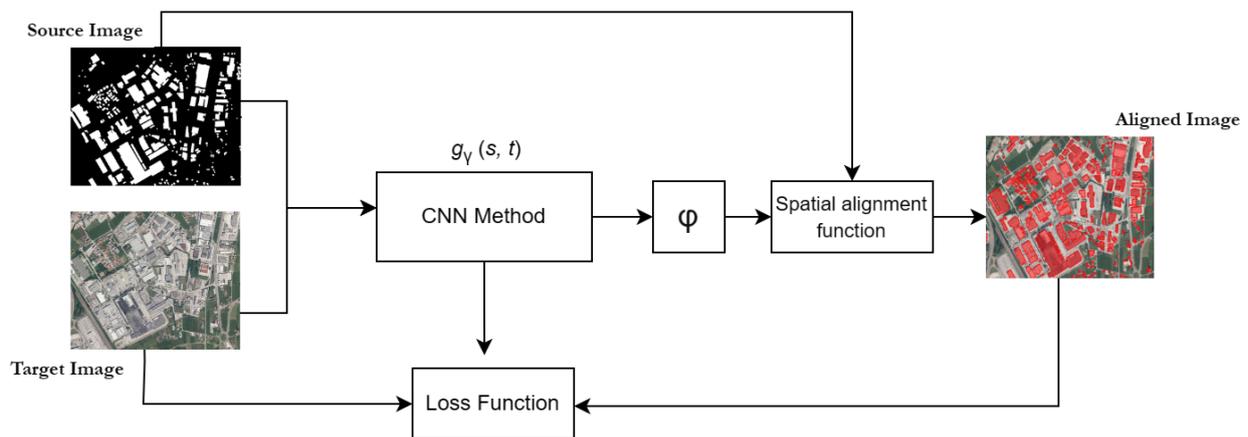
Figure 4. CCR-based approach.

The spatial transformation function uses the ϕ generated by the CNN method to re-sample s and obtain the warped image $s \circ \phi$. The proposed method learns the optimal parameter values by minimizing the difference between $s \circ \phi$ and t. The final loss can be calculated by the sum of the image dissimilarity $L_{image}(s, t)$ and the regularizing function (R) as:

$$Loss_{total} = L_{image}(s, t) + \alpha R(u) \tag{1}$$

It specifies the use of Cross Correlation (CC) as the main loss function for evaluating the similarity between the distorted s and t images. The definition of CC is as below:

$$CC_{(t, s \circ \phi)} = \frac{\left(\sum_{x \in \Omega}(t(x) - \bar{t}(x))(s \circ \phi(x) - \bar{s} \circ \phi(x))\right)^2}{\left(\sum_{x \in \Omega}(t(x) - \bar{t}(x))\right)^2 (s \circ \phi(x) - \bar{s} \circ \phi(x))^2} \tag{2}$$

where $t(x)$ is the grey value of the target image, $\bar{t}(x)$ is the average grey value of the target image, $s \circ \phi(x)$ is the grey value of the warped image, and $\bar{s} \circ \phi(x)$ is the average grey value of the warped image. The regularization term R regularizes the overall smoothness of the predicted displacements. The parameter of the regular term is α, and its value varies between 0 and 1. Finally, by the evaluation of the function, we can find the predicted registration field.

After alignment, in the second step of pre-processing, OSM vector data have to be rasterized on the same resolution as RS imagery. The large dimension images split into small equal-size patches. These are input patches for the training of CNNs, which is Mask R-CNN.

3.2. Deep Learning Approach

Mask R-CNN is a two-stage instance/object identification network used to segment a LULC class (buildings). Single object identification systems, such as Single-Shot Detector (SSD), merely learn bounding box regression and associated class probabilities. Although they have higher inference speeds, Mask R-CNN routinely outperforms them in terms of accuracy and incorporates semantic output. Mask R-CNN is an improved version of Faster R-CNN [34]. The first-stage regional proposal network (RPN) generates regions of interest

from a pre-defined set of anchors and feature maps using the Resnet-101 backbone with a Feature Pyramid Network (FPN). We employ a mix of binary cross entropy and a soft Jaccard loss as a loss function. The mechanism proposed by [35] is to generalize the discrete Jaccard index into a differentiable version. As a result, the network may directly optimize the loss throughout the training phase. The Jaccard index may be considered a measure of similarity between a limited number of sets. It may be defined as follows for two sets of ground truth (*A*) and segmentation results (*B*):

$$J(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cup B|} \tag{3}$$

A pixel classification challenge may be thought of as an image segmentation task. Thus, we apply a common classification loss function for a binary cross entropy, designated as *H*, to each output channel separately. Combining *J* and *H* gives the following final formula for the loss function:

$$L = \alpha H + (1 - \alpha)(1 - J) \tag{4}$$

By reducing (4), we maximize the estimated probability for the proper class for each pixel and the Intersection Over Union (IOU) between the OSM masks and related predictions. To train the Mask R-CNN, we modify its parameters according to our problem. We use randomly initiated weights and a binary cost loss function. The model is built using pre-trained weights from ImageNet [36]. This architecture enables the internal network representation for the semantic segmentation map to combine both data streams. Moreover, to explore even further from this segmentation map, we use a differential algorithm in the final layer that learns how to highlight the missing information in OSM using the related information from the OSM and the segmentation results.

4. Experimental Analysis

4.1. Dataset

We used Trento, Italy airborne data with a 100cm resolution for this study. We utilized RGB images and randomly chose 10,560 images for training and 300 images for evaluation. The patches of size 512 × 512 were cut from the given high-resolution satellite images. We obtained the semantic information in Shapefile format from Geofabrik to analyze the coverage of OSM building footprints in the region (data downloaded in July 2021). We used the 2021 data to check the goodness of the method for finding missing objects and we validated the findings in 2021 with the most recent OSM data in 2023. We used about 8k pre-processed OSM semi-urban buildings annotations to train the CNN model. We also tested the CNN-trained model on unseen building annotations (Beirut case I) that were spatially disconnected from training data.

4.2. Experimental Setup

We used Google Colab GPU to train the model for a total of 120 iterations with a 0.0001 learning rate, and Adam was used for optimization. With each training repetition, the learning rate decreased linearly. We trained the network in three steps: the first phase, training the head with 45 epochs; the second phase, training the backbone stage of ResNet-101 with 30 epochs; and the third phase, training all layers with 45 epochs. These were all inspired by the training methodology on the COCO dataset in [37]. Anchor scales of 8, 16, 32, 64, and 128 were investigated for instance segmentation on satellite images. Because the dataset had a substantial proportion of small structures and the input patch had a maximum size of 512 × 512 pixels, we investigated lower anchor sizes. We exclusively utilized RGB images since they have high resolution, sharpened characteristics, and the smallest memory capacity among the four channels of imagery. However, we expanded the images from small to large in augmentation as the network input to accommodate structures of varied scales. We used the majority of the hyper-parameters that were used to

train the COCO dataset, and we implemented the Mask R-CNN by [37] publicly available implementation. However, it is demonstrated that the training speed-improving hyper-parameters like Mini mask shape (56, 56) and effective mini-batch size 4, which were used, have a considerable effect on total detection performance. For loss calculation, we utilized $\alpha = 0.7$, randomly chosen in the range [0, 1]. The first experiment was conducted to evaluate the classifier segmentation performance using noisy OSM semantic information, and in the second one, we processed the noisy OSM annotations and used them as an input layer with high-resolution imagery to verify the effectiveness of using the quality-improved OSM information. Two further experiments on missing building prediction and change detection analyses are also presented in Sections 5.1 and 5.2, respectively. For the missing building case, the Ohsome Quality Analyst (OQT) indicator was used (in Appendix A) to evaluate the results.

5. Results and Comparison

Quantitative Results: This section contains the quantitative results for the proposed approach showing the Mask R-CNN accurate specificity and reliable building information when it deals with support to the OSM community. Table 1 also shows the comparison of the F1 score, average precision, and overall accuracy of the applied methods. The F1 score combines accuracy from the precision measure and completeness from the recall measure to create a harmonic average. Quantitatively, the proposed method achieves an F1 score of 0.96, whereas the Mask R-CNN with limited labeled data in [36] and FCN without pre-processing in [29] achieve 0.71 and 0.83 F1 scores, respectively. The overall accuracy of the proposed method is improved by 0.21 and 0.6, compared to the Mask R-CNN with noisy OSM, and FCN, respectively. Considering the evaluation values of the F1 score and overall accuracy, the proposed method with pre-processing shows competitive results. It shows that the dataset fusion (remote sensing and OSM), using the proposed experimental setup could facilitate better model fitting and improve performance. With OSM quality improved data, the score achieved by the proposed approach demonstrates that this strategy enhances overall accuracy and precision. The confusion matrix for the applied methodology is shown in Figure 5.
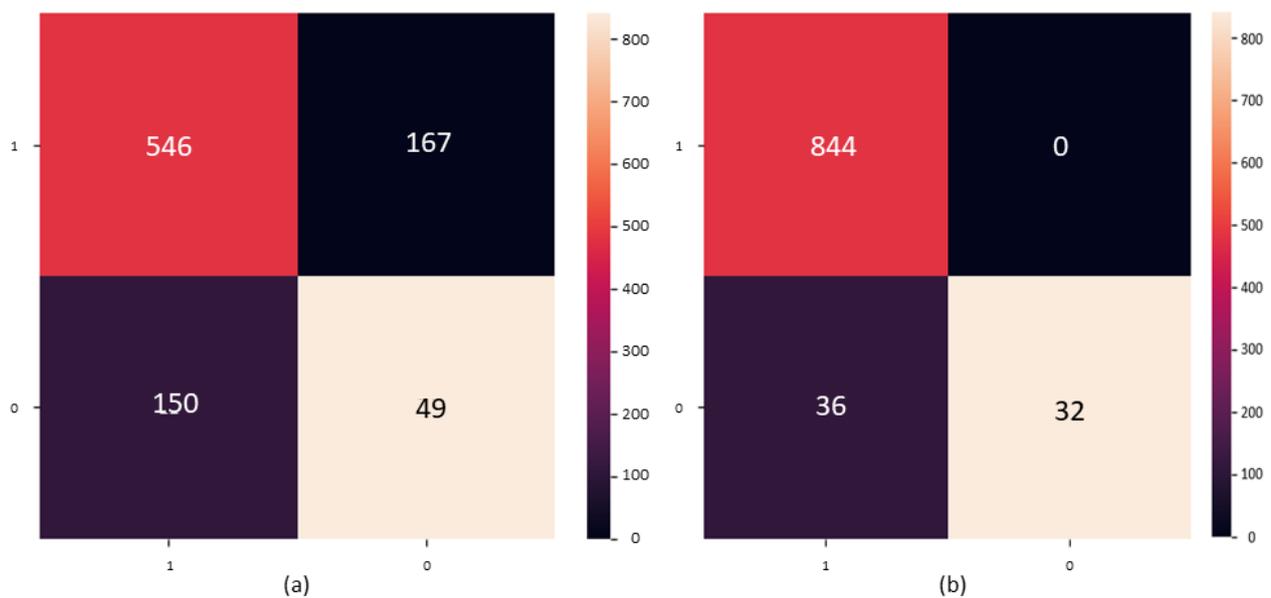
Figure 5. Confusion matrix for the experiments: (**a**) with noisy OSM, (**b**) with processed OSM.

Table 1. Performance evaluation.

Mask R-CNN (Proposed)	Precision	Recall	F1 Score	mAP	OA
(a) With Noisy OSM	0.76	0.77	0.76	0.60	0.62
(b) With Processed OSM	0.96	0.96	0.96	0.76	0.93
FCN Approach (Comparison)	0.85	0.77	0.83	0.80	0.87

Qualitative Results: To compare the qualitative performance of the two methods, we show how noise in OSM affects the final results, and how pre-processed annotations improve the network performance. Predictions for the chosen model are presented in Figure 6, showing output with noisy OSM and with quality-improved OSM data. The different colors highlight each building (instance segmentation). During learning with noisy OSM, the classifier shows misclassification, such as vegetation pixels mixed with building predictions. Because of the shifting in the coordinates, the classifier gets confused between pixels of vegetation and buildings. Compared to the network trained on noisy OSM, our method gives a better prediction of the building footprints with precise boundaries. Apart from performance with noisy and processed OSM, we also discuss the multi-performance of the proposed approach (with processed OSM) in change detection and missing building analysis in the two case studies below.

5.1. Missing Building Case

In this experiment, we perform segmentation with the CNN approach to see how well deep learning and remote sensing can predict not only buildings in that area but also missing OSM construction footprints. The predicted missing buildings for the Trento area are depicted in Figure 7. According to the Ohsome Quality Analyst (OQT) tool [38], the Trento area in 2021 is well mapped. The OQT is described as a functioning OSM data quality analysis software, accessible via a web interface (OQT Website). Appendix A provides additional details regarding the OQT indicators. The OQT indices show a mapping saturation between 97% and 100% and a contribution of about 99.16% edits (buildings) in the past three years, indicating that Trento is well-mapped in 2021. However, a visual assessment of the OSM building footprint's completeness indicates that some areas are still unmapped. We applied a differentiating algorithm at the last layer of the network, which takes the segmented output and OSM information to check the presence of buildings and shows buildings detected in the high-resolution imagery (segment output) and missing in OSM. The method successfully detected 63 missing buildings in 2021 OSM data in an area of about 19 km^2. We validated the results computed in 2021 by taking the most recent OSM data (2023): 21 buildings among 63 were added later on by the OSM contributors, and about 42 missing buildings still need to be mapped (wherein 31 are true positive and 11 are false positive). Accordingly, the proposed method may become available as support for the OSM community to faster and improve the information.

In Figure 8, yellow color polygons are from 2021, while green (correct) and red (incorrect) polygons are the ones detected by the proposed method. The last column is about the updated OSM (2023). There are still some limitations during mapping missing buildings, such as in the football court. Because of texture/color, it is detected as a missing building. These results show that remote sensing data are reliable and up to date as compared to the OSM, and also indicate that OSM is being updated with time.

Figure 6. (a) Left column: image instance segmentation with noisy and, (**b**) right column: quality improved (right) OSM annotations. Each bounding box/color shows a building.
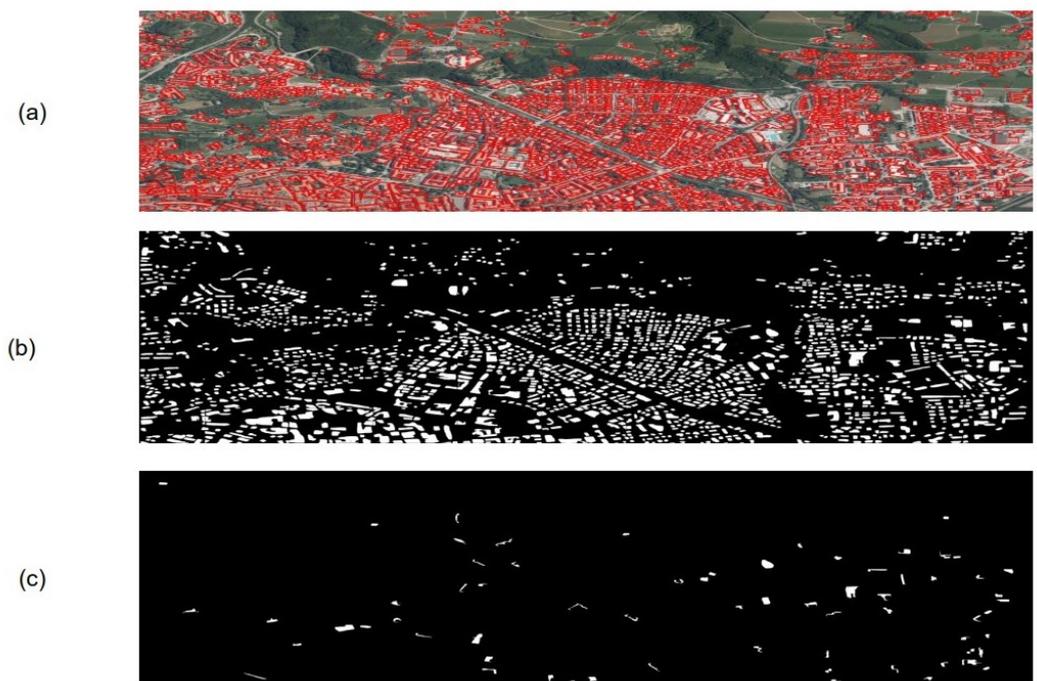
Figure 7. Missing building case on Trento: (**a**) RS and OSM overview, (**b**) buildings prediction, (**c**) missing buildings in OSM.
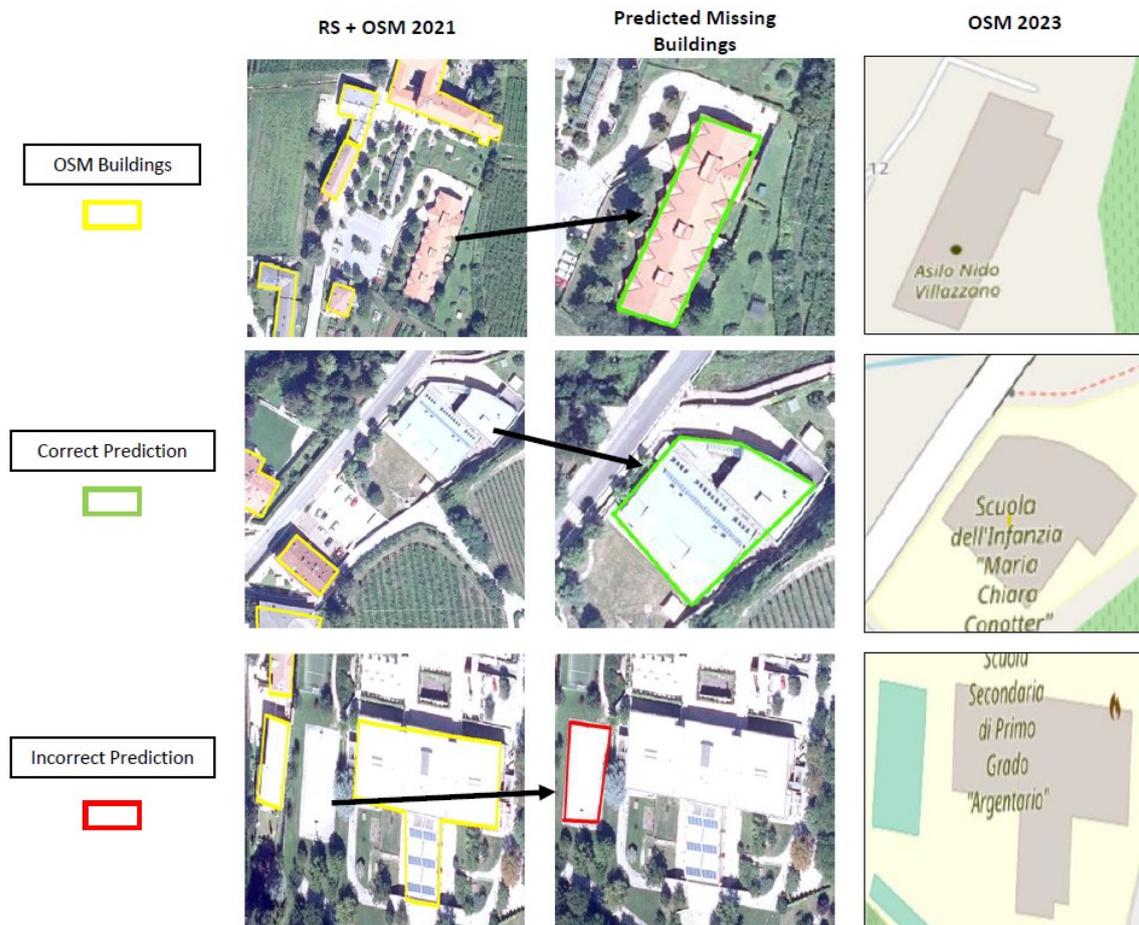
Figure 8. Missing building Trento case (validation): OSM 2021, predicted missing buildings, and OSM 2023.

5.2. Change Detection Case

In addition to binary class semantic segmentation, we consider the change detection problem. In an emergency, like a natural disaster, when buildings are destroyed, updated maps are very important to respond to the critical situation. Despite significant improvements in global human well-being, humanitarian crises, and natural calamities continue to strike the planet. Despite the long-term conflict and the need to influence people in many regions of the world, reliable maps of impacted areas are sometimes unavailable or obsolete due to disaster or war consequences. Satellite imagery might be useful in such applications, but converting images to maps is time-consuming. Today, maps exist that are created by individual companies/organizations or by volunteer initiatives such as Maphathon (coordinated mapping event), Google Maps, or OSM, and contain information on roads, buildings, agriculture, rivers, and other features [6].

To complement the proposed approach and to better understand the robustness of this methodology, the model outputs can be partially confirmed by taking into account the Beirut accident (change detection analysis). The largest non-nuclear explosion ever recorded happened on 4 August 2020, at 6:07 p.m. near Beirut, Lebanon, when an estimated 2750 tons of ammonium nitrate that had been incorrectly stored exploded. After the collision, there was a huge explosion and shock wave felt throughout Beirut, which was captured by many onlookers. The explosion was observed nearly 200 km away on the island of Cyprus, and shock waves with a magnitude of 3.3 were recorded [39]. The proposed approach works under the hypothesis that trained neural networks can perform well in areas having similar appearances to the training one, and it is assumed that the model performance decreases significantly when applied in areas different from it. However, we test how the proposed technique can assist in paving the way for autonomous satellite imagery processing to build useful, real-time map updating when the urban skyline gets upset by a catastrophic event. Here, pre- and post-change images are given to the baseline model, and change detection analysis is performed. This area is different and spatially disconnected from the training one, and pre and post-change analyses are shown in Figure 9. We applied this method to detect buildings with different resolutions (30 cm), as the network was trained with 100 cm resolution. There were almost 370 buildings in the ground-truth data (OSM buildings), and 243 buildings were detected as true positive and 4 buildings as true negative, with about 71% overall accuracy. We fine-tuned the model with OSM labels, and the results are shown in Figure 10 with 87% overall accuracy. It also justifies the reproducibility of the method.

Figure 9. Beirut change detection case: red color polygons are true predictions, while green and blue are false predictions.

Input

(a)

Output

(b)
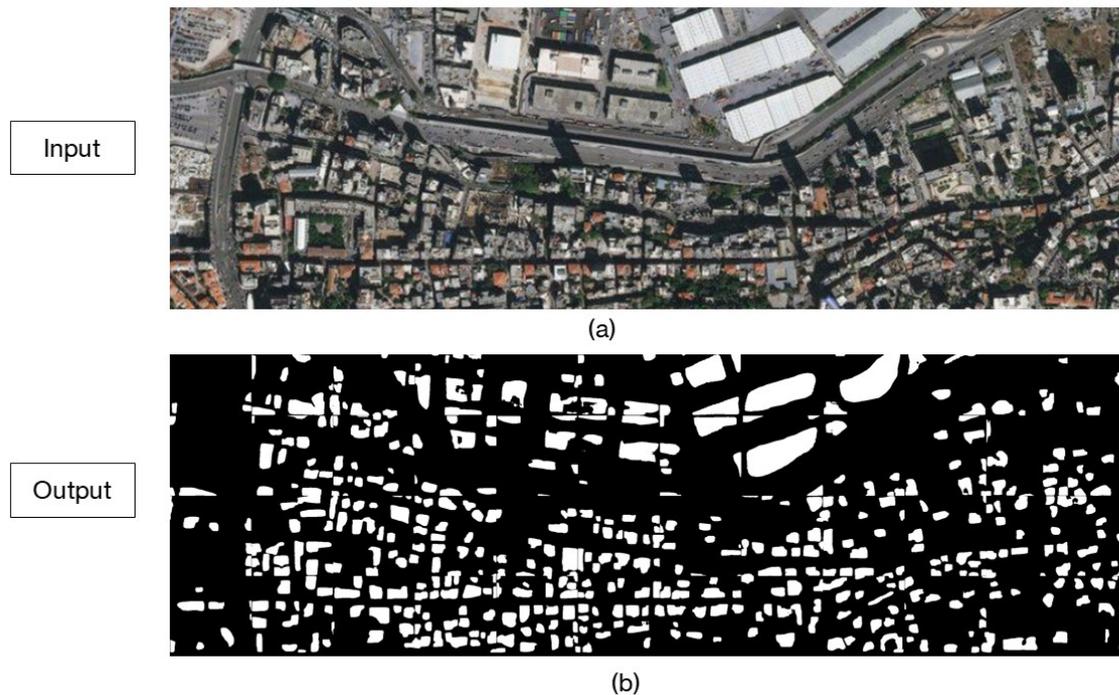
Figure 10. Beirut segmentation: (**a**) remote sensing imagery, (**b**) building segmentation.

We can summarize that the results are handling three noises of OSM data:

1. The registration noise is handled by the CCR approach in Section 3, and Mask R-CNN achieves better performance with a quality improved dataset in building extraction than using noisy OSM with VHR images.
2. The omission noise is mitigated and described in Section 5.1, and the results promote the use of VHR-updated imagery to highlight the missing objects in OSM.
3. The updated mapping issue is described in Section 5.2, which leads to the change detection analysis of an area. We perform a case study on the Beirut area for change detection mapping, where an explosion occurred on 4 August 2020. The results show that this approach mitigated the problem of updated mapping and proposed a new aspect of change detection and OSM quality.

Supervised classification on the considered dataset (RS and OSM), with a deep learning approach, performs well. The model can detect individual buildings, and the results are validated through OQT and updated OSM data, which show the significance of the applied methodology.

6. Discussion

This study shows how advanced deep learning handles remote sensing and OSM for accurate building segmentation. OSM open data possess many intrinsic and extrinsic quality issues, which may lead to wrong information in real and ground truth scenarios. The study focuses on solutions to three specific OSM quality issues: misalignment with satellite imagery (registration error), missing buildings (omission error), and updated mapping (destroyed buildings).

The image registration method employs an area-based approach, to achieve similarity between source and target images using the CCR method. The proposed approach highlights missing buildings in OSM after performing instance segmentation, which not only detects building boxes but also their footprints (solving a limitation mentioned in [40]). The OSM missing building and change detection maps generated by the proposed approach show high mapping accuracy in two target areas. The proposed method of the buildings segmentation model trained on remote sensing and open data offers a viable alternative

to difficulties with humanitarian mapping. The goal is to help volunteers by estimating and showing missing built-up areas in OSM, which will be useful for future detailed mapping by both individual mappers and humanitarian groups and possibly consider adding human settlement footprints directly or updating if buildings are destroyed. Also, recent discussions presented crucial insights from either the remote sensing or OSM community viewpoints by thinking about how ML methods best serve OSM [41], and our approach is a valuable contribution to this point. Then the proposed approach can be used to estimate the effort needed to complete the missing buildings in a given area as a function of the amount of missing buildings. This can be taken into account when planning volunteer mapping efforts and humanitarian mapping campaigns. The earlier research by [18] revealed an encouraging discovery that cutting-edge ML algorithms may be able to increase the accuracy and quicken the pace of the present humanitarian mapping technique. Where the present volunteer-based humanitarian mapping (like HOTOSM) initiatives are only responsible for disasters, our suggested method can be used to regularly investigate OSM missing areas, which aids in local community disaster preparedness.

We tested the study area with the OQT tool, and indicators are correlated with the predicted results. However, further testing is required to encourage the acceptance and (appropriate) use of new data products built on OSM history. Early efforts by [42,43] revealed encouraging insights to reconsider the usefulness and resilience of the method when scaling up to a worldwide humanitarian mapping scenario as well as a comprehensive analysis of how such an approach might work in various geographically isolated rural locations (such as rural contexts, climate, vegetation, and settlement types). It could be possible to tackle this challenge with the proposed approach due to its generalization capability to handle different areas (rural or urban).

But there are a few restrictions in our work that need to be taken into account. First, we chose locations in our case study where there are almost updated OSM data. To further assess the performance of the proposed strategy in this context, a wider validation in locations where OSM shows poor building information will be applied. With regard to OSM buildings, the suggested technique offers new insight into "Error of Omission" mitigation. However, more investigation is required to improve OSM building geometry directly. Because OSM data lack historical information, it can only be used in time-series mapping to a certain extent, and the majority of earlier research was only performed for a single year [44]. The suggested method in this work may be utilized to quickly process training samples from historical years, offering a fresh approach to long time-series land cover mapping. The use of OSM data in long time-series large-scale land cover mapping is anticipated to considerably increase the adaptability and efficiency of long-term land cover mapping by avoiding repetitive manual labor in the training data processing. Another step should take into consideration merging automatic mapping techniques with the current deep learning techniques with human assistance to ensure accuracy and validation to better capture distributed OSM missing built-up regions, like an early attempt with the RapidID tool. Our method could help to provide an additional data layer for RapidID that the OpenStreetMap community can validate and import as described in a case with Microsoft buildings in [45]. Future works are therefore urged to adopt or expand the approach suggested in this article to address the aforementioned constraints.

Using the knowledge gained from writing this study, we intended to underline the importance of machine-assisted mapping in this context. By integrating humans and machines, our upcoming work will focus on what thematic information we can derive from remote sensing and OSM to improve the data quality, how tagged information can be added to increase its reliability, and how machine learning and remote sensing can be fused for advanced analysis of noisy OSM.

7. Conclusions

Urban planning and city sensing are being transformed by the increased availability of geospatial data. Most of the geographic data were previously proprietary, rare, and in many cases unavailable during major catastrophes. Volunteered geographic information (VGI) has been playing a growing role in the support of humanitarian relief since the early 2000s [36]. OSM, the world's first openly licensed geospatial dataset created by volunteers, has been repeatedly shown to be highly suitable for GIS mapping and other environmental applications. There is a growing demand for a completely automated program that can detect locations where OSM features are not fully mapped. While earlier research used OSM data as a reference for classifying built-up land cover using satellite images [45], we show how remotely sensed data may be used as predictors of OSM missing buildings footprint for spatial coverage and how deep learning plays a significant role to process noisy OSM. Apart from dealing with OSM registration noise, the supervised approach detects missing buildings in OSM (omission noise) by utilizing updated high-resolution images that are both efficient and accurate. A building extraction approach using Mask R-CNN and building boundary regularization is presented. However, unlike Noisy OSM, which creates irregularly shaped polygons, our approach yields regularized polygons that may be used in a variety of city-sensing applications. Our results and comparisons show that remote sensing data are updated; meanwhile, they promote the fact that OSM data are being updated and provide new insights to be leveraged successfully for semantic labeling with deep learning methods. Our model can detect individual buildings through the generation of bounding boxes. Its main goal is to enhance and support the OSM mapping community by estimating the number of missing buildings and prioritizing unmapped regions. This work contributes to the efforts aimed at using machines to assist in humanitarian mapping methods.

Author Contributions: M.U.: conceptualization; software; methodology; validation; data creation; writing—original draft; results analysis. F.B.: supervision; writing—review and editing; formal analysis. M.N.: formal analysis; review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The OSM data presented in this study are openly available and the website (GEOFEBRIK) is mentioned in the manuscript. The remote sensing data presented in this study can be available on request from the corresponding author.

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

Appendix A. Ohsome Quality Analyst

The OQT gives an estimate of the quality of OpenStreetMap data for specific regions using a set of quality indicators that can be combined for quality reports. Each indicator produces a normalized value between 0 and 1 as their output. This value is divided into a green–yellow–red labeling schema to simplify understanding. The outcomes provide a quality score for the topic and area of interest, as well as for each of the individual indicators. They are presented with the help of a straightforward traffic light system (green–yellow–red), a verbal explanation, and individual graphs. Considering the Trento area, we calculate OQT indicators, and the graphs and their descriptions (Figure A1). These analyses are carried out to show the complexity of finding missing buildings in the Trento data. Although the area of interest is updated, there are yet missing buildings to map. Therefore, this area is a challenging one in the validation of the proposed method.

Graph	Description
Mapping Saturation (plot: OSM data; Nls Four-Parameter Logistic Model, 2008–2022)	"The saturation of the last 3 years is 100%. High saturation has been reached (97%<Saturation<100%)." "value": 0.9999997534058088 "label": "green" "class": 5
Total Contributions: 8925 (bar chart: Percentage of contributions vs Year, 2008–2022; Median Year: 2016)	In the last 3 years, 99.16% of the elements were edited. The remaining 0.63% were last edited more than 8 years ago. It is likely that most of the features are up-to-date." "value": 2 "label": "green" "class": 4

Figure A1. OQT analysis for Trento Study Area: mapping saturation and total contributions.

References

1. Horita, F.E.A.; Degrossi, L.C.; de Assis, L.F.G.; Zipf, A.; de Albuquerque, J.P. The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: A systematic literature review. In Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, IL, USA, 15–17 August 2013.
2. Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.
3. Poorazizi, M.E.; Hunter, A.J.; Steiniger, S. A volunteered geographic information framework to enable bottom-up disaster management platforms. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1389–1422. [CrossRef]
4. Chen, H.; Zhang, W.; Deng, C.; Nie, N.; Yi, L. Volunteered geographic information for disaster management with application to earthquake disaster databank & sharing platform. *Iop Conf. Ser. Earth Environ. Sci.* **2017**, *57*, 012015.
5. Mirbabaie, M.; Bunker, D.; Stieglitz, S.; Marx, J.; Ehnis, C. Social media in times of crisis: Learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *J. Inf. Technol.* **2020**, *35*, 195–213. [CrossRef]
6. Goldblatt, R.; Jones, N.; Mannix, J. Assessing OpenStreetMap completeness for management of natural disaster by means of remote sensing: A case study of three small island states (Haiti, Dominica and St. Lucia). *Remote Sens.* **2020**, *12*, 118. [CrossRef]
7. Barrington-Leigh, C.; Millard-Ball, A. Correction: The world's user-generated road map is more than 80% complete. *PLoS ONE* **2019**, *14*, e0224742. [CrossRef]
8. Zhou, Q.; Zhang, Y.; Chang, K.; Brovelli, M.A. Assessing OSM building completeness for almost 13,000 cities globally. *Int. J. Digit. Earth* **2022**, *15*, 2400–2421. [CrossRef]
9. Barron, C.; Neis, P.; Zipf, A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* **2014**, *18*, 877–895. [CrossRef]
10. Basiri, A.; Jackson, M.; Amirian, P.; Pourabdollah, A.; Sester, M.; Winstanley, A.; Moore, T.; Zhang, L. Quality assessment of OpenStreetMap data using trajectory mining. *Geo-Spat. Inf. Sci.* **2016**, *19*, 56–68. [CrossRef]
11. Hecht, R.; Kunze, C.; Hahmann, S. Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 1066–1091. [CrossRef]
12. Törnros, T.; Dorn, H.; Hahmann, S.; Zipf, A. Uncertainties of completeness measures in OpenStreetMap–A case study for buildings in a medium-sized German city. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 353. [CrossRef]
13. Senaratne, H.; Mobasheri, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [CrossRef]
14. Zhang, L.; Huang, X.; Huang, B.; Li, P. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2950–2961. [CrossRef]
15. Iglovikov, V.; Seferbekov, S.; Buslaev, A.; Shvets, A. Ternausnetv2: Fully convolutional network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 233–237. [CrossRef]

16. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote. Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

17. See, L.; Mooney, P.; Foody, G.; Bastin, L.; Comber, A.; Estima, J.; Fritz, S.; Kerle, N.; Jiang, B.; Laakso, M.; et al. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 55. [CrossRef]

18. Herfort, B.; Li, H.; Fendrich, S.; Lautenbach, S.; Zipf, A. Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sens.* **2019**, *11*, 1799. [CrossRef]

19. Anderson, J.; Sarkar, D.; Palen, L. Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 232. [CrossRef]

20. Li, Q.; Shi, Y.; Huang, X.; Zhu, X.X. Building footprint generation by integrating convolution neural network with feature pairwise conditional random field (FPCRF). *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7502–7519. [CrossRef]

21. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 645–657. [CrossRef]

22. Marcos, D.; Tuia, D.; Kellenberger, B.; Zhang, L.; Bai, M.; Liao, R.; Urtasun, R. Learning deep structured active contours end-to-end. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8877–8885.

23. Jilani, M.; Corcoran, P.; Bertolotto, M. Probabilistic graphical modelling for semantic labelling of crowdsourced map data. In *Intelligent Systems Technologies and Applications: Volume 2*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 213–224.

24. Fleischmann, P.; Pfister, T.; Oswald, M.; Berns, K. Using openstreetmap for autonomous mobile robot navigation. In *Intelligent Autonomous Systems 14: Proceedings of the 14th International Conference IAS-14 14*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 883–895.

25. Wang, Z.; Zipf, A. Using openstreetmap data to generate building models with their inner structures for 3d maps. *ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2017**, *4*, 411. [CrossRef]

26. Bittner, K.; Cui, S.; Reinartz, P. Building extraction from remote sensing data using fully convolutional networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2017**, *42*, 481–486. [CrossRef]

27. Mnih, V.; Hinton, G.E. Learning to label aerial images from noisy data. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), Edinburgh, UK, 26 June–1 July 2012; pp. 567–574.

28. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3226–3229.

29. Kaiser, P.; Wegner, J.D.; Lucchi, A.; Jaggi, M.; Hofmann, T.; Schindler, K. Learning aerial image segmentation from online maps. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6054–6068. [CrossRef]

30. Li, H.; Herfort, B.; Huang, W.; Zia, M.; Zipf, A. Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 41–51. [CrossRef]

31. Cui, K.; Fu, P.; Li, Y.; Lin, Y. Bayesian fully convolutional networks for brain image registration. *J. Healthc. Eng.* **2021**, *2021*, 5528160. [CrossRef] [PubMed]

32. Glocker, B.; Sotiras, A.; Komodakis, N.; Paragios, N. Deformable medical image registration: Setting the state of the art with discrete methods. *Annu. Rev. Biomed. Eng.* **2011**, *13*, 219–244. [CrossRef] [PubMed]

33. Marcos, D.; Hamid, R.; Tuia, D. Geospatial correspondences for multimodal registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5091–5100.

34. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1137–1149. [CrossRef]

35. Iglovikov, V.; Mushinskiy, S.; Osin, V. Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv* **2017**, arXiv:1706.06169.

36. Zhao, K.; Kang, J.; Jung, J.; Sohn, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 247–251.

37. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

38. Herfort, B.; Troilo, R. Analyzing changes in OSM over time - full history access to OSM data through the ohsome framework. In Proceedings of the Talk at the State of the Map Conference 2022, Florence, Italy, 19–21 August 2022.

39. El Sayed, M.J. Beirut ammonium nitrate explosion: A man-made disaster in times of the COVID-19 pandemic. *Disaster Med. Public Health Prep.* **2022**, *16*, 1203–1207. [CrossRef]

40. Li, H.; Herfort, B.; Lautenbach, S.; Chen, J.; Zipf, A. Improving OpenStreetMap missing building detection using few-shot transfer learning in sub-Saharan Africa. *Trans. GIS* **2022**, *26*, 3125–3146. [CrossRef]

41. Mooney, P.; Galvan, E. What has machine learning ever done for us? In Proceedings of the Academic Track at the State of the Map 2021, Online, 9–11 July 2011. [CrossRef]

42. Huck, J.J.; Perkins, C.; Haworth, B.T.; Moro, E.B.; Nirmalan, M. Centaur VGI: A hybrid human–machine approach to address global inequalities in map coverage. *Ann. Am. Assoc. Geogr.* **2021**, *111*, 231–251. [CrossRef]

43. Herfort, B.; Lautenbach, S.; Porto de Albuquerque, J.; Anderson, J.; Zipf, A. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nat. Commun.* **2023**, *14*, 3985. [CrossRef] [PubMed]
44. Viana, C.M.; Encalada, L.; Rocha, J. The value of OpenStreetMap historical contributions as a source of sampling data for multi-temporal land use/cover maps. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 116. [CrossRef]
45. Brinkhoff, T. Open street map data as source for built-up and urban areas on global scale. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 557. [CrossRef]