Village-Level Homestead and Building Floor Area Estimates Based on UAV Imagery and U-Net Algorithm

Xueyan Zhang [1,2]

[1] Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; xyzhang@igsnrr.ac.cn; Tel.: +86-101-6488-8673
[2] Centre for Chinese Agricultural Policy, Chinese Academy of Sciences, Beijing 100101, China

check for
updates

Abstract: China's rural population has declined markedly with the acceleration of urbanization and industrialization, but the area under rural homesteads has continued to expand. Proper rural land use and management require large-scale, efficient, and low-cost rural residential surveys; however, such surveys are time-consuming and difficult to accomplish. Unmanned aerial vehicle (UAV) technology coupled with a deep learning architecture and 3D modelling can provide a potential alternative to traditional surveys for gathering rural homestead information. In this study, a method to estimate the village-level homestead area, a 3D-based building height model (BHM), and the number of building floors based on UAV imagery and the U-net algorithm was developed, and the respective estimation accuracies were found to be 0.92, 0.99, and 0.89. This method is rapid and inexpensive compared to the traditional time-consuming and costly household surveys, and, thus, it is of great significance to the ongoing use and management of rural homestead information, especially with regards to the confirmation of homestead property rights in China. Further, the proposed combination of UAV imagery and U-net technology may have a broader application in rural household surveys, as it can provide more information for decision-makers to grasp the current state of the rural socio-economic environment.

Keywords: rural homestead; household survey; U-net; UAV; China

1. Introduction

Massive rural–urban migration has accelerated the process of urbanization and industrialization in China in the last few decades. From 2000 to 2016, China's rural resident population decreased from 808 million to 589 million, showing a decline of 27.1% [1]. However, the area under rural homesteads has increased rather than decreased because newly evicted farmers prefer to keep rural homes [2–4]; it has expanded from 14.5 to 19.9 million hectares, translating into an increase of 37.2% [1]. A vast number of farmers treat their homesteads as inherited wealth and not just as land for construction. At the same time, when farmers settle in cities, the transfer of homesteads to others is restricted [5]. Many challenges persist regarding the use and management of rural homesteads. On the one hand, a rural homestead serves as housing security for the farmer [6]. On the other hand, this sense of security has given rise to irrational phenomena, such as the over-occupation of land, leaving land idle, and the under-utilization of land [7]. To promote rural development, the Chinese government's proposal for the construction of beautiful villages focuses on the preparation of village plans according to local conditions, in-depth surveys of the farmers, and the rational layout and conservation of land. The use and management of homesteads is a key part of this exercise, and, thus, additional data and field surveys to those currently available are required. Household surveying is a common method of collecting relevant socio-economic and thematic information, with homestead and floor areas forming the core of this information [8]. However, a small proportion of farmers often have an incentive to

misrepresent data to receive higher government subsidies or to avoid exposing over-occupied land [9]. Moreover, most of the villages in China are densely populated with homesteads, requiring extensive and time-consuming surveys. Therefore, additional approaches are needed to collect more accurate spatial data to properly monitor the condition of the rural homesteads.

The use of unmanned aerial vehicles (UAVs) offers new opportunities for monitoring rural homesteads, as they facilitate real-time and high-resolution data collection [10]. Due to the centimeter-scale resolution of the ground texture, UAV images are beneficial for the visual interpretation of rural homesteads [11]. Yang et al. measured the building density and floor area ratio of rural settlements using a Dajiang UAV with visual interpretation [12]. However, visual interpretation is inadequate to support rural surveys in China, which usually cover thousands of villages. The height of a building may be detected in different ways using UAV images. Li et al. proposed a method for estimating building heights using sentinel-1 data, which focused on the urban scale [13]. Wang et al. reconstructed a 3D building based on UAV tilt photography [14], which is not suitable for dense rural homestead communities because their calculations were based on a single building.

In recent years, deep learning methods have also been used to identify rural buildings [15]. Li et al. employed AlexNet and support vector machine algorithms to detect hollow village buildings based on high-resolution remote sensing images [16]. These approaches are based on object detection methods [17], whose primary task is to find all the objects of interest in the image and determine their locations [18]. Object detection techniques use rectangular frames to locate objects, but both the roof distribution and the roof shape of rural buildings are irregular; thus, the identification accuracy of these methods is limited [19]. Furthermore, in the homestead identification task, the desired output should include homestead building boundaries, and each pixel should be assigned a class label [20]. Pixel-based technology implies that the network learns to provide predictions for each pixel [21]. U-net is a convolutional autoencoder widely used in the medical field and other industries; it performs high precision pixel-based segmentation on images [20]. However, the use of U-net to recognize rural homesteads is still uncommon.

In the estimation of the homestead and building floor areas at the village-level, it is still a challenge to explore a method applicable to rural China to achieve real-time image acquisition, pixel-based identification, and 3D modeling for rural buildings, one that provides a potential alternative to time-consuming and laborious household surveys. In this study, the objectives were: (1) to extract the spatial distribution of homesteads from UAV images, mainly relying on a pixel-based image classification using the U-net algorithm; (2) to develop and validate a building height model (BHM) to determine the number of floors and the floor area of rural buildings based on 3D modelling; and (3) to develop and test a village-level method to estimate homestead and floor areas in a rapid and low-cost manner, which is useful for rural surveys in China and other developing countries.

2. Data and Methods

2.1. Study Area

The study area is in the Jianfeng village, east of Qishan town, Qimen county, in the Anhui province of China (Figure 1a). Anhui is among the first batch of provinces in China to pilot the reform of the rural collective property rights system. Rural homesteads comprise the core of the next step of rural reform and development. Qimen county is mountainous. The survey area measures 52,578.12 m^2 and is a narrow strip of land on the whole, with mountainous terrain located to its north and a river channel to its south. A total of 19 household survey sites exist in the village. Records exist for the sizes of the homesteads and the number of floors and floor areas of the rural buildings (Figure 1b).
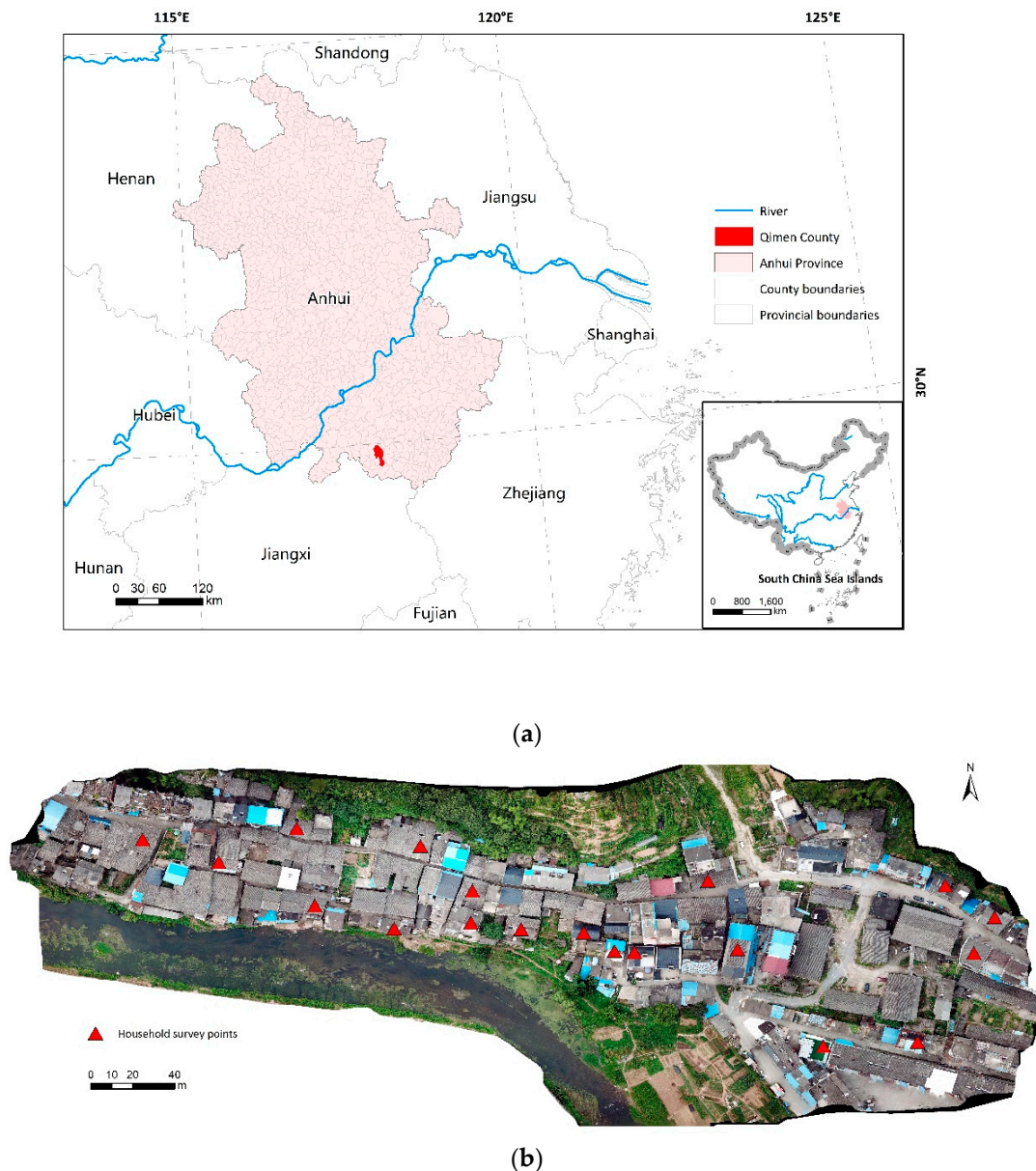
(**a**)

(**b**)

Figure 1. Study area: (**a**) Qimen county, Anhui province, (**b**) Jianfeng village, Qimen county. The red triangles in Figure 1b indicate household survey locations.

2.2. Image Acquisition Using UAV Data

To acquire UAV images, we employed a DJI Mavic Pro UAV, which is a quadcopter with a four-wheel drive motor and a complementary metal-oxide semiconductor (CMOS) camera with a focal length of 28 mm and an effective pixel count of 12.35 million for the 1″ CMOS. The maximum speed and flight time were 18 m/s and 27 min, respectively. The UAV data were obtained on 16 August 2019. A dry, windless day was chosen to avoid any distortions caused by the undulations of the UAV camera. Autonomous flight planning was conducted for the study area. The flight path was in the north–south direction and split into two flights of approximately 12 min each at a speed of 11 m/s. The camera has an F-shift of 2.2, and a shutter speed of 1/2000 s. The ISO value, which can be adjusted automatically according to the light conditions, was set between 100 and 1600. During the UAV image acquisition, the camera angle was −90°, and the safe mode was turned on. The sensor produced images of 20 MP in the red, green, and blue (RGB) wavelengths. During the flight, the camera was automatically released every 7 m, while the position of the device was simultaneously recorded by the

internal GPS/GLONASS dual-mode satellite positioning system. In the study area, the flight longitude ranged from 117°42′57.60″ E to 117°43′12.00″ E and the flight latitude ranged from 29°51′07.20″ N to 29°51′10.80″ N. The images were obtained from an altitude of 100.4 m with a 70% lateral overlap, 90% forward overlap, and an optical ground sample distance of 3.1 cm. The pixel size of the images was 2.7 cm. The covered flight area amounted to 52,578.12 m^2. A total of 130 RGB images were obtained during the survey.

2.3. Estimation of Rural Homestead Area Using the U-Net Algorithm

2.3.1. U-Net Architecture and Parameter Settings

The U-net architecture is illustrated in Figure 2, following the equivalent diagram developed by Ronneberger et al. [20]. The U-net architecture consists of two parts: the contraction path and the expansion path. The contraction path follows a typical convolutional network architecture, with many feature channels that allow the network to propagate context information to higher resolution layers.
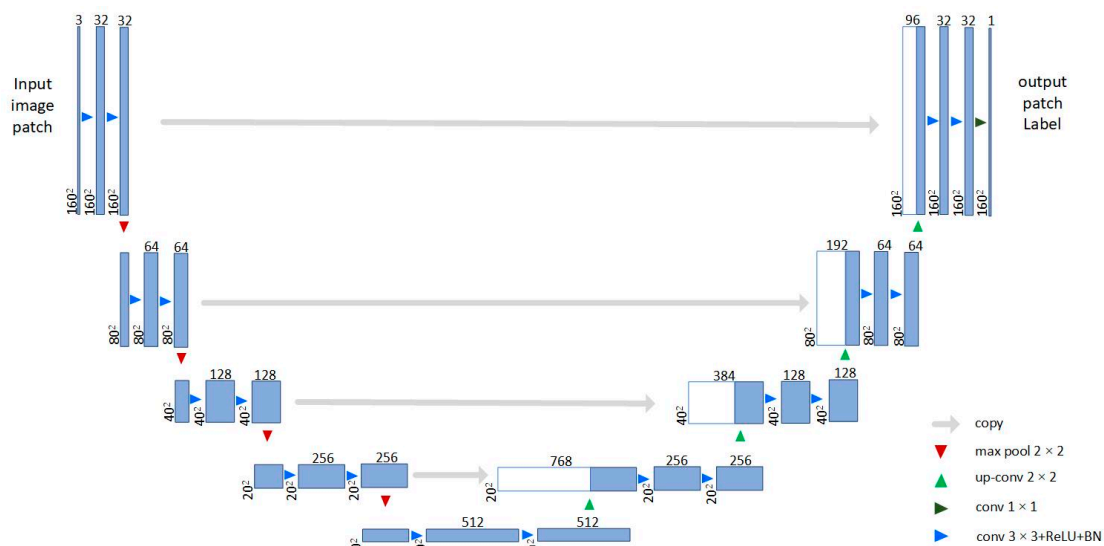

Figure 2. U-net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

The U-net in this study consists of convolutional layers with a convolution kernel size of 3 × 3, followed by a rectifier linear unit (ReLU). To achieve a numerically stable training procedure, a batch normalization (BN) layer was incorporated after every convolution layer. Then, 2 × 2 steps of 2 maximum pooling layers were followed to complete the down-sampling, while the size of the feature map decreased. The feature channels were increased by an order of two at every downsampling step and the feature channels were halved at each upsampling step. The same-padding hyperparameter was used to control the spatial size of the output volumes. In the final layer, a convolution layer with a convolution kernel size of 1 × 1 mapped the 32–channel feature map to the required number of categories, using a sigmoid function as the neuronal activation function. The network had a total of 23 layers.

2.3.2. Training

Here, a total of 188 RGB image samples of 650 × 650 pixels were prepared based on UAV data. The homesteads and other features in the study area were visually interpreted as the label data. In this case, only two classes were required, "homestead" and "non-homestead", representing the presence

or absence of homesteads. Deep neural networks typically perform better with more training data. Models trained on small datasets do not generalize well and suffer from overfitting. It is imperative to exploit data augmentation to increase the total number of training images. This issue was addressed by segmenting all 650×650 images into 160×160 patches. The 188 image pairs (image and respective label raster) were augmented to a total of 4324 image pairs. Data processing and spatialization were conducted using Python 3.6 and ArcGIS 10.5, respectively. The data split cross-validation parameter was equal to 0.2, and the shuffle was True. For the training dataset, 3450 images were used and the remaining 874 images were used for validation during the training. The Adam optimizer had a momentum of 0.9 and a learning rate 0.0001. The network was trained for 100 epochs using binary cross-entropy as a loss function. Similarity was measured using the Jaccard coefficient [22]. All the experiments were run using Keras 2.2.2 with TensorFlow 1.10.0 using python 3.6.

2.3.3. Validation

The following quantitative indicators were used to evaluate performance in statistical analysis: overall accuracy, precision, recall, and *F*1 score. These indicators are presented as calculated true positives (*TP*s), false positives (*FP*s), true negatives (*TN*s), and false negatives (*FN*s). For a class *l*, TP is the number of pixels that are correctly classified as *l*. FP is the number of pixels that are misclassified as *l*. Finally, FN represents pixels that belong to *l* but are associated by the model with some other classes.

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

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

$$F1 = \frac{2 \cdot \text{Precisoin} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Precision and recall are common indicators used to evaluate classification performance [23]. However, these two indicators are sometimes contradictory. Therefore, we employed *F*1 for the synthesis [24]. Moreover, to further evaluate the performance of the developed approach, we used intersection-over-union (*IoU*), which represents the proximity of the predicted object to the ground truth. In Equation (5), *A* and *B* are two different data samples [21].

$$IoU(A, B) = \frac{A \cap B}{A \cup B} \tag{5}$$

2.4. Generation of Building Height Model and Estimation of Homestead Floor Area

The floor area estimations should be based on point clouds and the 3D structures of the rural buildings. Two types of remote sensing techniques are suitable for application on UAV platforms: airborne laser scanning (ALS) and structure from motion (SfM). SfM photogrammetry techniques underperform in terms of accuracy, whereas ALS can provide more accurate estimates of the vertical structures of buildings [25]. However, SfM is more readily available than ALS because SfM is inexpensive for users in developing countries [26]. Therefore, UAVs using SfM technology can detect data at an acceptable spatial and temporal resolution, making them a more cost-effective solution [10]. The workflow of the SfM method consists of two main processes: aligning the images and constructing the geometry. The 3D point cloud was generated using the Agrisoft Photoscan Professional Edition software (Agisoft LLC, St. Petersburg, Russia) [27]. First, the camera position of each image was located and matched to the common points in the image; this allowed the identification of calibration parameters for image comparison. Based on the estimated camera position and the image itself [28], a point cloud was then built and a digital terrain model (DTM) was generated [26].

The height of a rural building can be approximated as the height of the BHM. Theoretically, the BHM can be obtained by subtracting the digital surface model (DSM) from the DTM. The DSM was obtained by Krieger spatial interpolation, based on the points selected from non-homestead areas. After interpolation, the point pairs of the interpolated DSM data and the observed DTM data were obtained and used to test the fitting accuracy between the kriging interpolation surface DSM and the UAV DTM. The floor area of the homestead is the product of the area of the homestead and the number of floors of the building. The area of the homestead was identified by the U-net algorithm. The building height was calculated using the difference between the elevation data obtained by the UAV and the interpolated ground surface data. To obtain the number of floors, thresholds were formed to measure the height of the rural buildings according to the local reality. The stratification thresholds were set as follows: a surface layer less than 1.0 m, an elevation difference of 1.0–4.0 m for buildings with only a ground floor, an elevation difference of 4.0–8.0 m for a building with two floors, an elevation difference of 8.0–12.0 m for three floors, and an elevation difference above 12.0 m for buildings with more than three floors. Then, the number of building floors was obtained based on the usual floor heights of local buildings, such as the ground (only) floor, two floors, three floors, and more.

$$Area_{Homestead} = Area_{Base} \times N_{Floors} \qquad (6)$$

where $Areas_{Homestead}$ is the total floor area of the homestead, m^2; $Areas_{Base}$ is the area of the homestead base, m^2; and N_{Floors} is the number of floors in the rural building.

3. Results

3.1. Homestead Recognition Based on U-Net Algorithm

Figure 3 shows the detailed results of the U-net identification of the homestead floor area with UAV images for the indicative area. The minimum loss rate of the model was 0.01 and the Jaccard efficiency equaled 0.98. The validation results of the U-net algorithm-based homestead identification are presented in Table 1. The overall accuracy, namely 0.92, was higher than that for the others.

Table 1. Validation of the U-net algorithm identification results.

Indicators	Precision	Recall	F1	Overall Accuracy	IoU	TP	FP	TN	FN
U-net	0.91	0.86	0.88	0.92	0.80	21,638,559	2,046,555	40,908,779	3,530,350

Figure 4 shows a comparison of the area and spatial distribution of the homesteads between the ground truth and the value estimated by U-net at the village level. A clear separation between the homestead and non-homestead categories is obvious. Compared to the ground truth, the village roads, vegetable gardens, and trees can be effectively classified as belonging to the non-homestead category. There is an obvious high degree of consistency between the two, with only the edges of the U-net identification results being somewhat irregular.
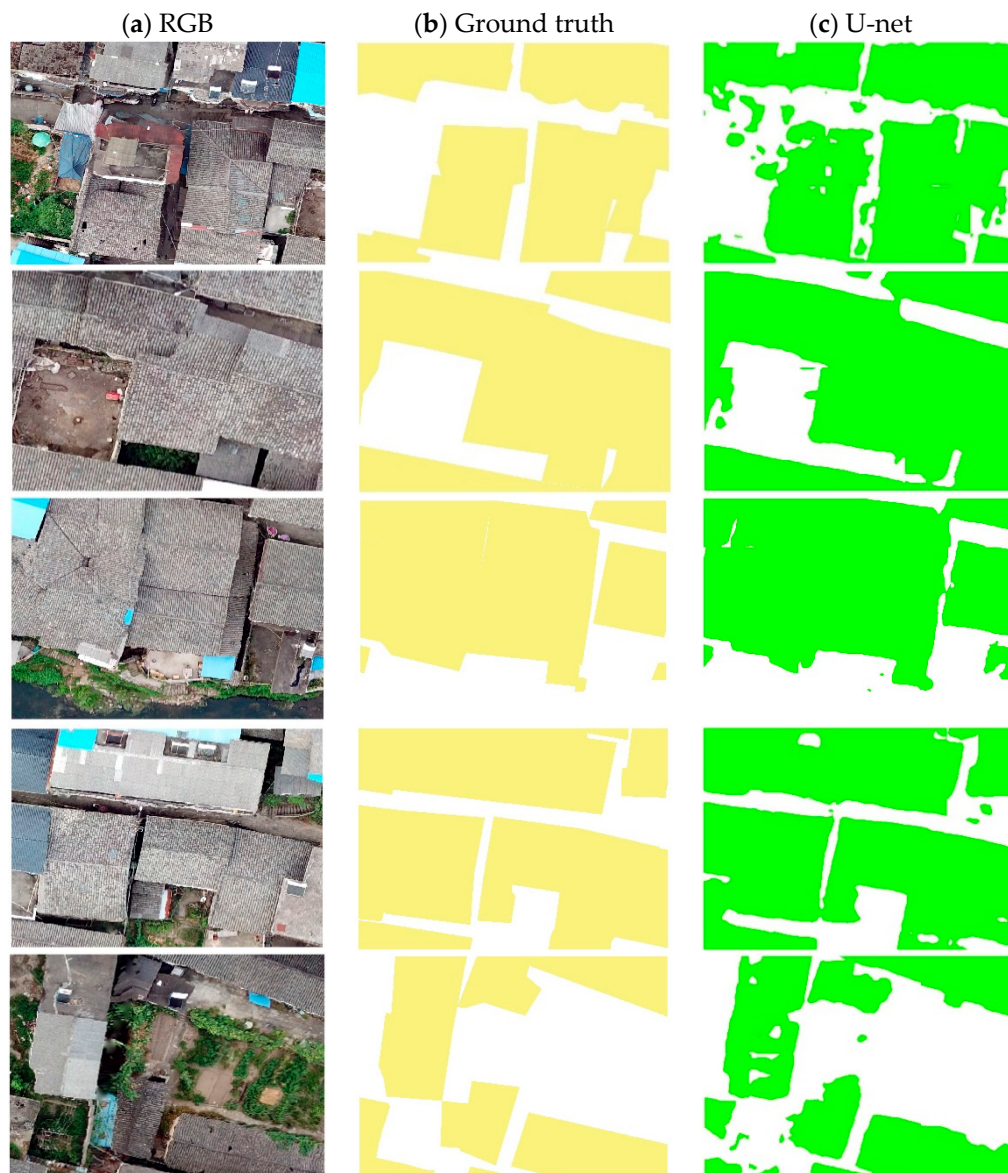
(**a**) RGB (**b**) Ground truth (**c**) U-net

Figure 3. Comparison of the unmanned aerial vehicle (UAV) image, ground truth, and U-net algorithm recognition results in indicative regions. (**a**) UAV red, green, and blue wavelength (RGB) images, (**b**) corresponding ground truth images (yellow for homesteads and white for other areas), and (**c**) results identified by the U-net algorithm (green for homestead and white for other areas).

Figure 4. Village-level estimates of the areas and spatial distributions of rural homesteads. (**a**) Ground truth of homesteads, and (**b**) identification results based on the U-net algorithm.

3.2. Floor Estimates for Rural Buildings Based on UAV DTM

Figure 5a shows the DTM established by the SfM method with an elevation difference of 50.11 m in the study area (i.e., 50.11 m at the highest point on the northern side and 0 m at the river channel on the southern side). The DSM was obtained by Krieger spatial interpolation, based on 633 points selected from non-homestead areas (Figure 5b). After interpolation, 633-point pairs of the interpolated DSM and the observed DTM data were obtained and used to test the fitting accuracy between the kriging interpolation surface DSM and the UAV DTM. Figure 6 shows a scatter plot of the 633-point pairs of the DSM and DTM, which has an R^2 value of 0.9875. This indicates that the DSM and DTM had good fitting accuracies. Then, the digital height model was obtained using the height difference between DTM and DSM in the homestead area identified by the U-net algorithm (Figure 5c). The BHM was divided into different floors (Figure 5c). Nineteen household survey datasets were available to test the consistency of the number of floors obtained from the BHM. The consistency compared to the survey data was 0.89 (Figure 7). This is an acceptable overall accuracy, and only the data for the 6th and 16th household survey sites along the *x*-axis were underestimated.
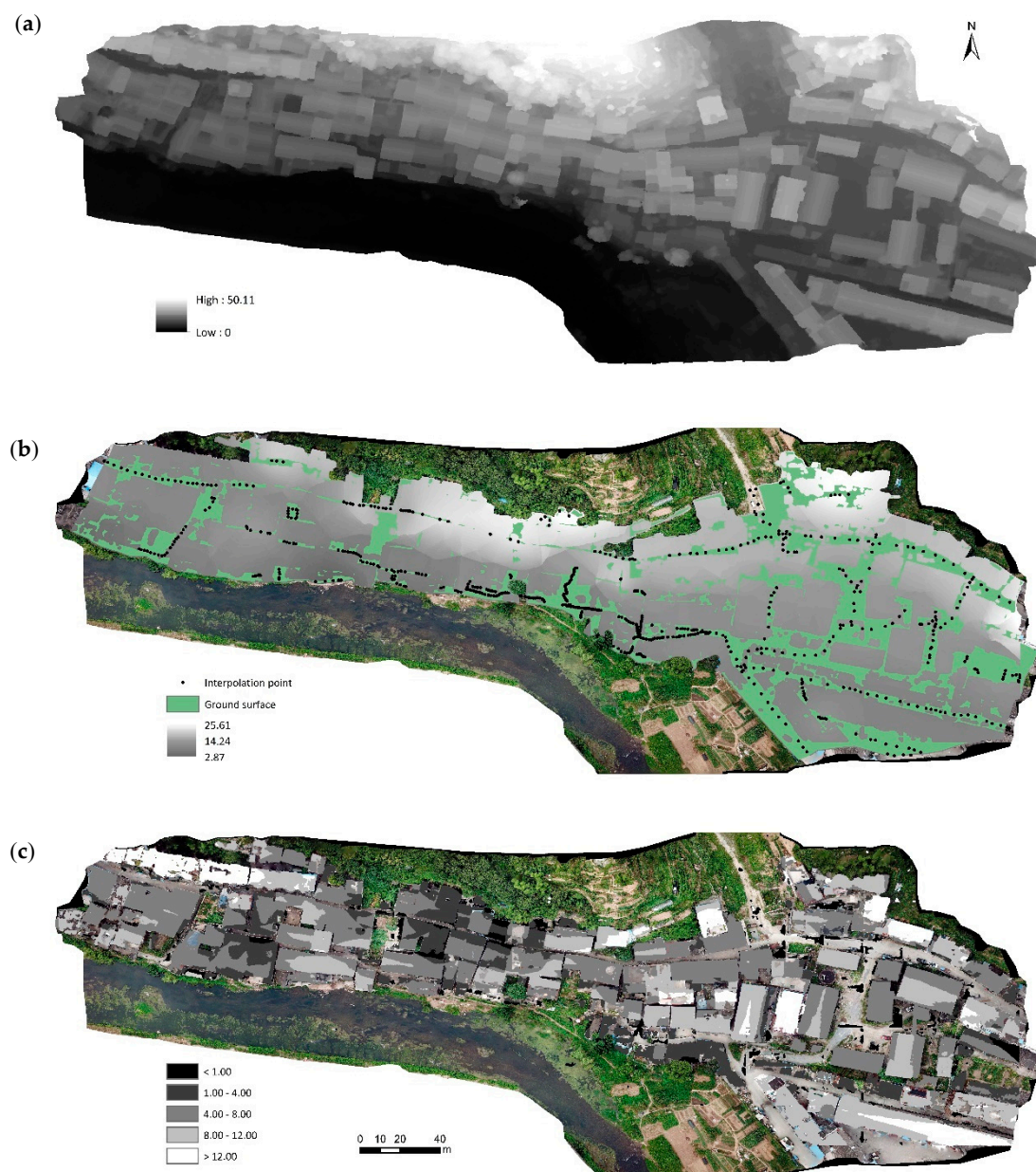
Figure 5. UAV-based estimates of the number of floors in rural buildings. (**a**) Digital terrain model (DTM) built based on the SfM method from the UAV images, (**b**) digital surface model (DSM) based on kriging interpolation with 633 control points, and (**c**) building height model (BHM) divided into different floors (unit: m).
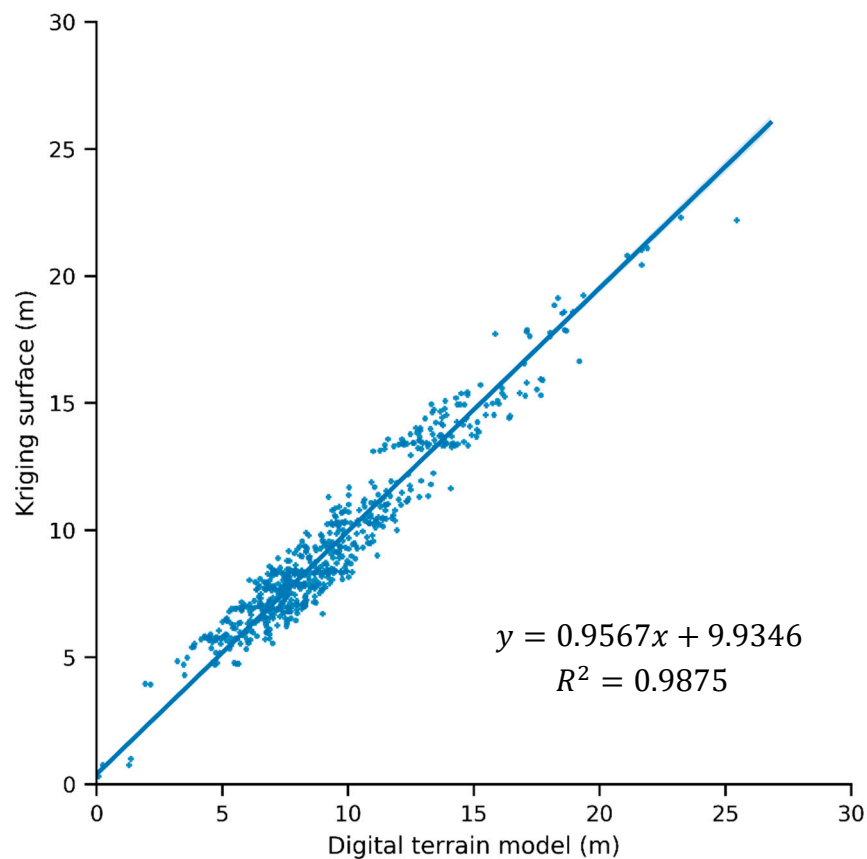
Figure 6. Verification of the DSM with a scatter plot of the testing points of the DSM and DTM. The *x*- and *y*-coordinates represent the testing points of the DTM and DSM, respectively.
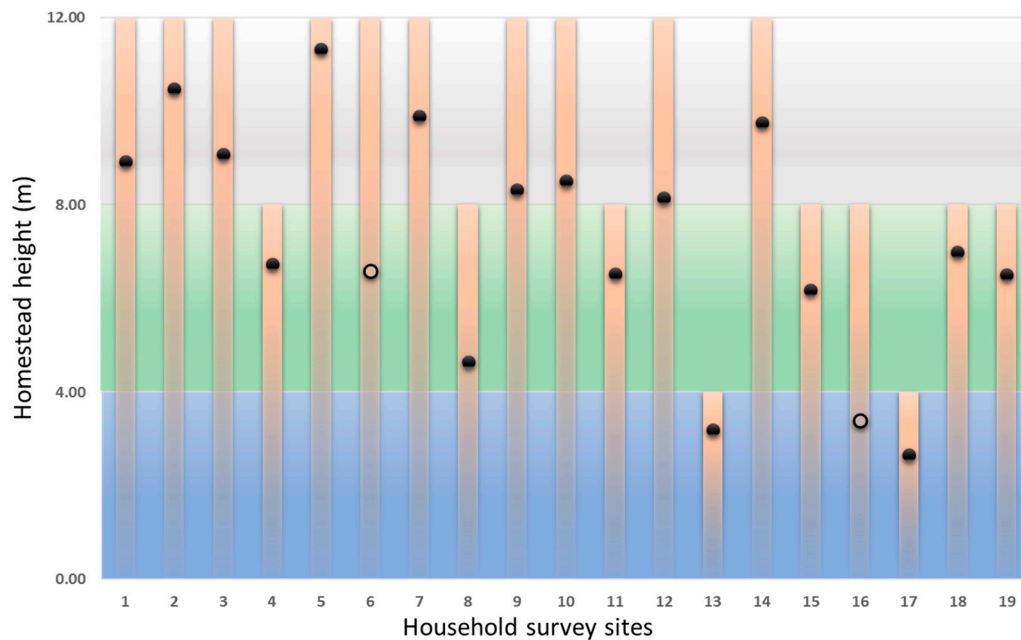

Figure 7. Consistency between the number of floors recorded in the household survey and the number of floors estimated with the proposed technique. The bars show the number of floors surveyed, while the scatter plot represents the calculated homestead floor heights. The black dots indicate that the estimated number of floors is consistent with the number of floors recorded in the survey, and the hollow dots indicate inconsistency. The blue, green, and gray regions in the background represent the number of floors (ground, two, and three) of the homestead, respectively.

3.3. Estimated Floor Area at the Village Level

The total area of the homesteads identified by the U-net algorithm was 17,477.52 m^2. The number of building floors is shown in Figure 5c. According to Equation (6), the constructed area in the village equals 37,965.25 m^2, of which the area covered by the ground floors alone accounts for 12.02% of the total homestead area, while the second floors, third floors, and floors beyond the third floor account for 34.11%, 33.79%, and 20.08% of the total homestead area, respectively. Thus, these results show that most of the buildings in the surveyed area contain two and three floors, which is consistent with the architectural conventions in rural southern China.

4. Discussion

A method based on UAV imagery and the U-net algorithm was developed for the estimation of village-level homestead and floor areas, with the advantage of real-time image acquisition, pixel-based identification, and 3D modeling recognition. The overall resulting accuracies were 0.92 and 0.89 for the homestead area and the number of building floors, respectively. Thus, our experience of using a combination of UAV and U-net technologies to identify village-level objects provides a potential alternative to time-consuming and laborious household surveys, which has important implications for the ongoing homestead use and management reform in China, especially for homestead ownership confirmation.

In Table 1, U-net showed high accuracy in identifying the buildings in this study. Many attempts have been made to use convolution neural networks (CNNs) to improve the performance of building detection based on object detection technology [15,16]. However, object detection techniques use rectangular frames to locate objects, and the distribution of homestead buildings and the irregular shapes of the roof planes limit the identification accuracy of these methods [1]. Konstantinidis et al. proposed a modular CNN architecture to identify buildings with pixel-based detection technology [29], wherein the network learns to provide some dense predictions for each pixel [21]. The pixel-based architecture is fully convolutional; therefore, in this work, we employed the commonly used pixel-based architecture. Papadomanolaki et al. compared multiple methods based on CNN architecture and enforced pixels that belonged to the same object to be classified under the same semantic category [21]. Therefore, the results of this study prove the advantage of U-net and a pixel-based architecture for estimating the area of rural homesteads.

However, some error sources remain. The BHM estimates are the key to determining the floor areas of the rural buildings. In theory, the height of an object can be calculated from its UAV image using photogrammetry, by subtracting the DSM from the DTM [30]. However, it is difficult to extract a BHM from a UAV-derived DTM because the terrain surface is obscured by the roof [31]. Furthermore, the DSM was estimated to use the elevation control points located within the range of the country trails around the homesteads. Since rural trails in the southern part of China are generally narrower, the DSM interpolation surface errors were slightly higher for the narrow trails than the other open areas. Figure 4b shows a comparison of the interpolated ground control points and UAV-derived DSM; both showed excellent agreement, as R^2 equaled 0.99. However, as explained previously, the DSM elevation values generated with the UAV images were usually overestimated for the narrow roads surrounded by the buildings. Therefore, we set the DSM elevation surface threshold to less than 1 m for the ground surface area.

Moreover, the average slope of the area is approximately 30°, increasing the difficulty of an accurate interpolation. If the study area is located in a plain, the uncertainty caused by the slope will be relatively small. However, in complex terrains, such as the one in this study, improvement in accuracy will require an increase in the surveyed and measured sampling points.

U-net provides an advantage in terms of the number of training samples, as the algorithm requires a small amount of data to train the model [20]. Due to the limited range of the UAV flights, the 188 image pairs were augmented to a total of 4324 image pairs. The overall accuracy of the U-net deep learning network recognition was generally good (0.92), and rapid image segmentation was possible with the

established model. The data augmentation played a vital role, allowing a few annotated images and a very reasonable training time to complete image recognition. However, the question remains whether there is a lower limit of annotated images for U-net to work accurately. In subsequent studies, we plan to decrease the training image pairs to test the robustness of U-net.

In addition, this study referred to 19 ground survey sites, and the proposed technique provided a consistency of 89.47%. This number may be an overestimate or an underestimate; however, during household surveys, most farmers reflect the true situation, and evidence of homestead ownership was confirmed, but it is possible that some of the descriptions may have been biased.

5. Conclusions

In this study, a method based on UAV imagery and the U-net algorithm was developed for village-level homestead and building floor area estimation, with the advantage of real-time image acquisition, pixel-based identification, and 3D modeling recognition. The resulting overall accuracy for the estimation of the homestead area and the number of building floors was 0.92 and 0.89, respectively. This method is a potential alternative to time-consuming and costly household surveys and is, thus, of great significance not only for the use and management of homesteads, but also for the ongoing homestead ownership confirmation in China. The combination of UAV imagery and the U-net algorithm may also have broader applications in the area of homestead use and management. For instance, the number of greenhouses, irrigation facilities, and even agricultural machinery are important components of rural household surveys. The proposed method can assist decision-makers to grasp the current state of the rural socio-economic environment and make policy recommendations accordingly. In the future, the accuracy of the model for use in areas with complex topography and dense housing will be further improved.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Liu, S.Y.; Xiong, X.F. Property rights and regulation: Evolution and reform of China's homestead system. *China Econ. Stud.* **2019**, *6*, 17–27.
2. Long, H.L.; Li, Y.R.; Liu, Y.S.; Woods, M.; Zou, J. Accelerated restructuring in rural China fueled by increasing vs. decreasing balance land-use policy for dealing with hollowed villages. *Land Use Policy* **2012**, *29*, 11–22. [CrossRef]
3. Liu, Y.S.; Fang, F.; Li, Y.H. Key issues of land use in China and implications for policy making. *Land Use Policy* **2014**, *40*, 6–12. [CrossRef]
4. Chen, H.X.; Zhao, L.M.; Zhao, Z.Y. Influencing factors of farmers' willingness to withdraw from rural homesteads: A survey in Zhejiang, China. *Land Use Policy* **2017**, *68*, 524–530. [CrossRef]
5. Tian, Y.; Kong, X.; Liu, Y.; Wang, H. Restructuring rural settlements based on an analysis of inter-village social connections: A case in Hubei Province, Central China. *Habitat Int.* **2016**, *57*, 121–131. [CrossRef]
6. Cao, Q.; Sarker, M.N.I.; Sun, J.Y. Model of the influencing factors of the withdrawal from rural homesteads in China: Application of Grounded theory method. *Land Use Policy* **2019**, *85*, 285–289. [CrossRef]
7. Xu, H.; Liu, Y. Policy implications and impact of household registration system on peasants' willingness to return rural residential lands: Evidence from household survey in rural China. *Panoeconomicus* **2016**, *63*, 135–146.
8. Watmough, G.R.; Marcinko, C.L.J.; Sullivan, C.; Tschirhart, K.; Mutuo, P.K. Socioecological informed use of remote sensing data to predict rural household poverty. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1213–1218. [CrossRef]
9. Sun, L.; Perter, H. Formalizing informal homes, a bad idea: The credibility thesis applied to China's "extra-legal" housing. *Land Use Policy* **2018**, *79*, 891–901. [CrossRef]

10. Puliti, S.; Ene, L.T.; Gobakken, T.; Næsset, E. Use of partial-coverage UAV data in sampling for large scale forest inventories. *Remote Sens. Environ.* **2017**, *194*, 115–126. [CrossRef]

11. Deng, F.; Dou, A.X.; Wu, W.Y.; Chen, Z.H.; Yuan, X.X. Rapid Investigation of disaster situation in extreme disaster area of Jiuzhaigou earthquake in Sichuan based on UAV remote sensing. *J. Catastrophology* **2018**, *33*, 210–215.

12. Yang, C.; Li, H.; Xu, G.; Xiang, X.; Yang, D. A measure to the building density and floor area ratio of rural settlements based on Da Jiang unmanned aerial vehicle remote sensing. *Mt. Res.* **2019**, *37*, 144–150.

13. Li, X.C.; Zhou, Y.Y.; Gong, P.; Seto, K.C.; Clinton, N. Developing a method to estimate building height from Sentinel-1 data. *Remote Sens. Environ.* **2020**, *240*, 111705. [CrossRef]

14. Wang, J.Z.; Lin, Z.J.; Li, C.M.; Hong, Z.G. 3D reconstruction of buildings with single UAV image. *Remote Sens. Inf.* **2004**, *4*, 11–15.

15. Ren, Y.Y.; Zhang, X.F.; Ma, Y.J.; Yang, Q.Y.; Wang, C.J.; Dai, J.G.; Zhao, Q.Z. Target detection of Rural Buildings in UAV remote sensing images based on convolutional neural network. *J. Nanjing Norm. Univ. (Eng. Technol. Ed.)* **2019**, *19*, 29–36.

16. Li, Z.; Li, Y.S.; Wu, X.; Liu, G.; Lu, H.; Tang, M. Hollow village building detection method using high resolution remote sensing image based on CNN. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 160–165.

17. Protopapadakis, E.; Voulodimos, A.; Doulamis, A.; Doulamis, N.; Stathaki, T. Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing. *Appl. Intell.* **2019**, *49*, 2793–2806. [CrossRef]

18. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; p. I.

19. Liu, Z.Q.; Cao, Y.W.; Wang, Y.Z.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional network. *Autom. Constr.* **2019**, *104*, 129–139. [CrossRef]

20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

21. Papadomanolaki, M.; Vakalopoulou, M.; Karantzalos, K. A novel object-based deep learning framework for semantic segmentation of very high-resolution remote sensing data comparison with convolutional and fully convolutional networks. *Remote Sens.* **2019**, *11*, 684. [CrossRef]

22. Rahman, M.; Hassan, M.R.; Buyya, R. Jaccard index based availability prediction in enterprise grids. *Procedia Comput. Sci.* **2012**, *1*, 2707–2716. [CrossRef]

23. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

24. Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [CrossRef] [PubMed]

25. Wallance, L.; Lucieer, A.; Malenovsky, Z.; Truner, D.; Vopenka, P. Assessment of forest structure using two UAV techniques: A comparison of airborne laser scanning and structure from motion (SfM) point clouds. *Forests* **2016**, *7*, 62. [CrossRef]

26. González-Jaramillo, V.; Fries, A.; Bendix, J. AGB estimation in a tropical mountain forest (TMF) by means of RGB and multispectral images using an unmanned aerial vehicle (UAV). *Remote Sens.* **2019**, *11*, 1413. [CrossRef]

27. Agisoft. *Agisoft Photoscan UserManual*; Agisoft LLC: St. Petersburg, Russia, 2014; Available online: http://www.agisoft.com/downloads/user-manuals (accessed on 10 April 2020).

28. Torres-sánchez, J.; Castro, A.I.; Peña, J.M.; Jiménez-Brenes, F.M.; Arquero, O.; Lovera, M.; López-Granados, F. Mapping the 3D structure of almond trees using UAV acquired photogrammetric point clouds and object-based image analysis. *Biosyst. Eng.* **2018**, *176*, 172–184.

29. Konstantinidis, D.; Argyriou, V.; Stathaki, T.; Grammalidis, N. A modular CNN-based building detector for remote sensing images. *Comput. Netw.* **2019**, *168*, 107034. [CrossRef]

30. Weidner, U.; Förstner, W. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS J. Photogramm. Remote Sens.* **1995**, *50*, 38–49. [CrossRef]
31. Wimala, V.I.; Menno, S.; Elisabeth, A.; Hans, M. Monitoring height and greenness of non-woody floodplain vegetation with UAV time series. *ISPRS J. Photogram. Remote Sens.* **2018**, *141*, 112–123.