


## Article

# RERB: A Dataset for Residential Area Extraction with Regularized Boundary in Remote Sensing Imagery for Mapping Application

Songlin Liu <sup>1,2,\*</sup> , Li Zhang <sup>2</sup>, Wei Liu <sup>2</sup>, Jun Hu <sup>3</sup>, Hui Gong <sup>2</sup>, Xin Zhou <sup>4</sup> and Danchao Gong <sup>2</sup><sup>1</sup> School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, China<sup>2</sup> State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China<sup>3</sup> School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510006, China<sup>4</sup> Tian-Hui Satellite Center of China, Beijing 100094, China

\* Correspondence: songlinliu@whu.edu.cn; Tel.: +86-18992830817

**Abstract:** Due to the high automaticity and efficiency of image-based residential area extraction, it has become one of the research hotspots in surveying, mapping, and computer vision, etc. For the application of mapping residential area, the extracted contour is required to be regular. However, the contour results of existing deep-learning-based residential area extraction methods are assigned accurately according to the actual range of residential areas in imagery, which are difficult to directly apply to mapping due to the extractions being messy and irregular. Most of the existing ground object extraction datasets based on optical satellite images mainly promote the research of semantic segmentation, thereby ignoring the requirements of mapping applications. In this paper, we introduce an optical satellite images dataset named RERB (Residential area Extraction with Regularized Boundary) to support and advance end-to-end learning of residential area mapping. The characteristic of RERB is that it embeds the prior knowledge of regularized contour in the dataset. In detail, the RERB dataset contains 13,892 high-quality satellite images with a spatial resolution of 2 m acquired from different cities in China, and the size of each image is approximately  $256 \times 256$  pixels, which covers an area of more than 3640 square kilometers. The novel published RERB dataset encompasses four superiorities: (1) Large-scale and high-resolution; (2) well annotated and regular label contour; (3) rich background; and (4) class imbalance. Therefore, the RERB dataset is suitable for both semantic segmentation and mapping application tasks. Furthermore, to validate the effectiveness of the RERB, a novel end-to-end regularization extraction algorithm of residential areas based on contour cross-entropy constraints is designed and implemented, which can significantly improve the regularization degree of extraction for the mapping of residential areas. The comparative experimental results demonstrate the preponderance and practicability of our public dataset and can further facilitate future research.

**Keywords:** residential area extraction; mapping requirement; contour regularization; end-to-end deep model; Tian-Hui 1 satellite images



**Citation:** Liu, S.; Zhang, L.; Liu, W.; Hu, J.; Gong, H.; Zhou, X.; Gong, D. RERB: A Dataset for Residential Area Extraction with Regularized Boundary in Remote Sensing Imagery for Mapping Application. *Electronics* **2022**, *11*, 2790. <https://doi.org/10.3390/electronics11172790>

Academic Editor: Stefano Ferilli

Received: 1 August 2022

Accepted: 2 September 2022

Published: 5 September 2022

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



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

## 1. Introduction

Topographic map data with 1:50,000 scale are one of the most basic geographic information data, which play a significant and strategic role in national economy and national defense construction [1,2]. With the rapid development of society, users have increasingly higher requirements for the current situation of topographic maps, and the updating of topographic maps [3,4] has become the primary and urgent work. The content of topographic maps mainly includes ground objects and the undulating form of terrain. Because the changes in terrain data are generally relatively small, the updated objects of topographic map mainly consist of ground objects.

Optical remote sensing imagery is one of the main data sources for updating ground objects in topographic maps. Ground object mapping mainly refers to the acquisition of

object information in remote sensing imagery according to the corresponding scale graphic specification [1]. At present, ground object mapping using remote sensing imagery is mainly completed manually, with high precision but low efficiency, making this kind of fashion tedious, expensive, and labor intensive, so it is difficult to meet the needs of rapid applications, such as land planning and automatic driving. According to official statistics [5], it takes at least 40 days and 10 thousand RMB to produce 1:50,000 topographical map data. The number of global 1:50,000 topographical maps is about 400 thousand, and each update requires an investment of about 4 billion RMB.

With the launch of Zi-Yuan 3 (ZY-3) and Tian-Hui 1 (TH-1) stereo mapping satellites [6,7], China has the capability of measuring and updating 1:50,000 topographic maps with satellite remote sensing imagery. Among the ground feature elements, residential areas are one of the most important elements in topographic map content. A survey demonstrated that in most areas, the workload of extracting residential areas accounts for more than 60% of all the work of extracting ground features [2]. Therefore, studying the automatic extraction method of residential areas for mapping applications is of considerable significance to improving the efficiency of mapping work.

With the development of this automatic extraction technology, many institutions around the world have developed digital mapping systems integrated with automatic technology for recognizing ground feature elements [2]. For examples, both the eCognition [8] of Definiens and the EasyFeature [9] of Handleray have integrated the ground feature recognition technology. Specifically, this kind of method mainly includes two steps: extraction and post-processing.

Semantic segmentation [10,11] is a typical and an efficient technology to accomplish the extraction step, which indicates, dividing the image into pixel groups with specific semantics and recognizing each region's category. In recent years, the development of deep learning techniques, such as convolution neural network (CNN), has injected new vitality into the study of semantic segmentation. However, due to the complexity of ground features and background in remote sensing imagery, the classification results of residential areas extracted by semantic segmentation method are usually not perfect, especially at residential area boundaries [12], which are irregular contours. Consequently, these classification results cannot be directly employed in mapping applications. In addition, the post-processing technology is exploited to obtain the regularized object boundary contour. The popular operations used to identify the boundary of a raster dot group include smoothing, line segment fitting, denudation under complex constraints, and conditional random field (CRF) method, etc. In addition, there are also some methods using an end-to-end network to process the boundary of objects. The abovementioned innovative works focus on improving extraction accuracy but without consideration of the matching degree between the extraction results and the mapping requirement.

Obviously, the mapping method with two steps is cumbersome, and the post-processing step also greatly reduces the overall intelligence of the mapping method. End-to-end fashion can realize intelligent mapping without manual intervention. To promote the end-to-end mapping method, we present and introduce an optical satellite images dataset named RERB (Residential area Extraction with Regularized Boundary). To the best of the authors' knowledge, there is no dataset released for the application of mapping residential area, which limits the research of end-to-end residential area regularization extraction. Compared to existing datasets, the contour of label image in RERB dataset consists of regular line segments. Given this point, it can facilitate the research for end-to-end training of residential area regularized extraction. Specifically, the public RERB dataset consists of 13,892 satellite images in  $256 \times 256$  size, covering an area of more than 3640 square kilometers.

To summarize, our contributions are as follows:

- (1) According to the specifications for cartographic symbols of 1:50,000 topographic map, our work summarizes the requirements of regular extraction in the residential area mapping application.

- (2) We construct a residential area mapping dataset called RERB with regular contour labels based on TH-1 [7] satellite images, which is the first dataset released for the residential feature mapping application. Furthermore, in order to measuring the compliance of the extraction results with the mapping requirements when using RERB dataset, we design a special evaluation index named CMI (contour matching index) based on contour matching. Extensive experiments demonstrate the superiority of RERB dataset.
- (3) We sufficiently explore the contour constraint with regular contours in label images by integrating the contour cross-entropy constraint and the original loss function into an end-to-end network, which can significantly improve the regularization degree of extraction results for the mapping of residential areas.

The remainder of this paper is organized as follows: Section 2 introduces the related works. Section 3 presents the constructed RERB dataset in detail. Section 4 details the experimental results along with in-depth analysis. Section 5 finishes the paper with conclusions and our future perspective.

## 2. Related Works

In this section, we first describe the development of datasets for ground object extraction based on optical image and then introduce semantic segmentation methods. Finally, we introduce post-processing technology.

### 2.1. Datasets for Ground Object Extraction

Recently, with the advancement of deep learning technology, datasets have played an important part in ground object extraction. Any effective deep learning model is obtained by training with many original images and their corresponding labels. As shown in Table 1, the widely used open-source datasets with optical image pixel level annotation include WHU [13], LandCoverNet [14], GID [15], LoveDA [16], SSD [17], etc.

**Table 1.** Overall comparison of some satellite image datasets.

	Year	Resolution (m)	Image Size	Samples	Categories	Task-Semantic Segmentation	Task-Mapping
WHU	2019	0.45	512 × 512	17,388	2	✓	✓ (building)
	2019	0.3–2.3	512 × 512	204	2	✓	✓ (building)
LandCoverNet	2020	10	256 × 256	1980	7	✓	
GID	2020	4	7200 × 6800	150	6	✓	
LoveDA	2021	0.3	1024 × 1024	5987	7	✓	
SSD	2021	2	7400 × 4950	23	5	✓	
RERB (ours)	2022	2	256 × 256	13,892	2	✓	✓

WHU dataset is released by Wuhan University, and it includes one land-cover category, namely, buildings. WHU dataset can be used to construct a building extraction model in a topographic map with a scale of 1:10,000 or larger and cannot be directly applied to residential area mapping in 1:50,000 scale topographic maps.

The Gaofen image dataset (GID) contains 150 high-quality GF-2 images from more than 60 cities in China, with a spatial resolution of 4 m. The size of each image is approximately 7200 × 6800 pixels, and it includes six land cover categories, namely, built-up, farmland, forest, meadow, water, and others, which represents all categories other than the former five categories. Similarly, LandCoverNet, LoveDA, and SSD are also constructed for land use and land cover (LULC) classification. If they are used for topographic mapping, post-processing steps still need to be added after model inference.

To study end-to-end regularized extraction technology of residential area, we propose the RERB dataset in this paper.

## 2.2. Semantic Segmentation

Semantic segmentation is a long-standing research topic that assigns a label to each pixel, known as pixel-level classification. In 2015, Long et al. [18] proposed full connected networks (FCNs), whose excellent performance led researchers to change their understanding of semantic segmentation from regional clustering to pixel classification. At present, CNN-based methods have completely exceeded the segmentation accuracy of traditional methods. However, the training steps of FCNs are complex, and it is easy to lose pixel position information during up-sampling. After that, U-Net [19], SegNet [20], PSPNet [21], the DeepLab family [22–24], and FastFCN [25] were developed. U-Net can effectively fuse multilevel feature maps, and small objects and large objects are processed by using shallow and deep information, respectively. U-Net is essentially a structure based on multiscale context and multilevel feature fusion. SegNet improves the segmentation accuracy by recording the position of pooled values in the original feature map and accurately mapping the relevant values to the corresponding positions in the up-sampling step. However, SegNet still fails to recover the object boundary very well. PSPNet integrates the multiscale background information with a pyramid pooling module. To obtain a larger receptive field, PSPNet improves the backbone network by using dilated convolutions [26]. Furthermore, additional losses can provide the intermediate supervision information in PSPNet. The DeepLab series leads research on semantic segmentation. DeepLab v3+ [24], which integrates more local information in low-level features and replaces the feature extractor with a more complex Xception network [27], performs well on several public datasets. In addition, the atrous spatial pyramid pooling (ASPP) structure proposed by the DeepLab network has been widely employed in semantic segmentation research literature. FastFCN uses the joint pyramid up-sampling (JPU) module to improve the dilated convolution and obtains faster speed and higher accuracy.

Especially, semantic segmentation technology has been applied to remote sensing imagery and medical image [28,29] in recent years, which has greatly improved the research level of methods used to automatically extract ground feature elements. For example, Ying Sun et al. [30] used optical images and light detection and ranging (LiDAR) data to construct multichannel input data and designed a convolution neural network (CNN) model with multiscale encoder–decoder architecture to achieve enhanced segmentation results. Cui et al. [31] also improved the accuracy of building extraction by using the multiscale information of images. Y. Liu et al. [32] jointly used LiDAR data and introduced a higher-order CRF to increase the accuracy of ground object segmentation. In addition, several researchers designed two-stage training approaches [33], modified loss function [34], self-attention modules [35,36], edge information [37], or both self-attention and edge enhancement modules [17] to fully exploit the context information of remote sensing imagery from a larger perspective.

## 2.3. Post-Processing Technology

The processing object of post-processing technology is the raster dot group, which is obtained by semantic segmentation. Traditional operations used to identify the boundary of a raster dot group include smoothing, line segment fitting, and denudation under complex constraints [38,39]. Most of these methods belong to the field of traditional image processing, and the degree of automation and intelligence is low.

Moreover, CRF [40–42] methods are also widely used in post-processing of semantic segmentation results. Using CRF, the segmentation results can be corrected, especially at ground object borders. However, these CRF methods require the introduction of samples to the CRF control process, and this operation cause the CRF methods to lose their end-to-end characteristics. For mapping tasks based on automatic extraction technology, when the end-to-end characteristics are lost, the ground object mapping work must add

an additional manual post-processing operation, which greatly reduces the overall intelligence of mapping tasks. Hence, new end-to-end methods must be introduced to solve this problem.

Ying Sun et al. [43] first constructed multichannel input data using optical images and LiDAR data and then achieved a better result than SegNet by designing an end-to-end encoding–decoding structure. Meanwhile, the object boundary is strengthened. There are also some methods using an end-to-end network to process the boundary of objects, such as ACE2P [44], Gated-SCNN [45], and EaNet [46]. The ACE2P model realizes end-to-end high-precision training by fully integrating the bottom characteristics, global contextual information, and edge details in the human body parsing task. Gated-SCNN is a double branch structure, in which the target shape information is embedded into the semantic segmentation network by a shape branch. Except for traditional semantic segmentation labels, image boundary labels are also needed in Gated-SCNN. To effectively separate confusing objects with sharp contours, EaNet is constructed based on a large kernel pyramid pooling (LKPP) module and a dice-based edge-aware loss function.

### 3. The RERB Dataset and Model Construction

This section first describes the contour requirements for mapping applications and then introduces the RERB dataset. Finally, we analyze the statistics for RERB dataset and describe the construction of a residential area regularized extraction model.

#### 3.1. Contour Requirements for Mapping Applications

Different topographic maps are distinguished by scale and commonly used scales generally include 1:2000, 1:5000, 1:10,000, and 1:50,000. The 1:50,000 topographic map data are one of the most basic geographic information data. At present, ground object mapping using remote sensing imagery is mainly completed manually, with high precision but low efficiency, making this kind of fashion tedious, expensive, and labor intensive [5]. As a result, it is very important to analyze the mapping requirements and build a mapping dataset to improve the intelligence of mapping work.

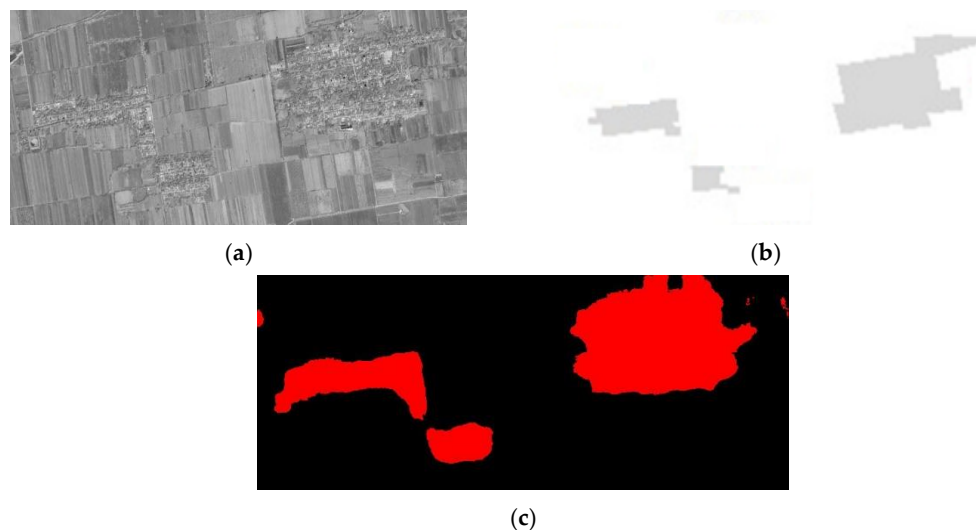
Topographic maps of different scales are constrained by corresponding graphic specifications, which mainly stipulates the symbols, annotations, and contour decoration of various ground objects and geomorphic elements represented on topographic maps, as well as the methods and basic requirements of using these symbols. This paper mainly focuses on 1:50,000 scale, and its corresponding current national standard [1] was issued on 14 October 2017 and implemented on 1 May 2018.

Ground object mapping in surveying and mapping field mainly refers to the collection of the ground object information from remote sensing imagery according to the corresponding specification for cartographic symbols [1]. Figure 1 is an example of residential area extraction and mapping based on optical images. Figure 1b is an illustration of a residential area in the 1:50,000 topographic map corresponding to the original image in Figure 1a. Mapping is to obtain the contour of ground objects that meet the requirements of graphic specifications from remote sensing imagery.

Residential areas [2] refer to houses that are contiguous to each other in cities, towns, and villages. There are obvious outer contours and primary and secondary streets in residential areas. The graphic specification [1] stipulates that the convex and concave parts should be comprehensively represented when their length is less than 0.5–1 mm on the maps. In the 1:50,000 topographic map, 1 mm on the map represents the actual 50 m, and the length of 50 m is 25 pixels in the image with a resolution of 2 m. Therefore, the graphic specification requires that the convex and concave parts should be smoothed when their length is less than 12.5–25 pixels.

Figure 1c is a direct extraction result of residential areas based on semantic segmentation algorithms. The contour line is messy and has a high degree of border redundancy. Figure 1b shows an illustration of the residential area layer in a topographic map, and it is a standard representation corresponding to the cartographic symbols used in topographic

mapping. Its outer contour is multiple straight-line segments. The comparison indicated that the results of traditional semantic segmentation algorithms are different from the requirements of the cartographic symbols, and the contour of the extracted results must be regularized as much as possible.



**Figure 1.** An example of residential area mapping based on images: (a) Original image; (b) the residential area layer in topographic map; and (c) direct extraction results of residential areas.

To sum up, the extracted contour is required to be regular when images are used for residential area mapping. Each segment of extracted contour is generally a straight-line segment, which is relatively regular. Therefore, when building a dataset that supports the end-to-end regularization extraction of residential areas, it is necessary to ensure that the label image contour meets the regularization requirements.

### 3.2. Overview and Data Properties

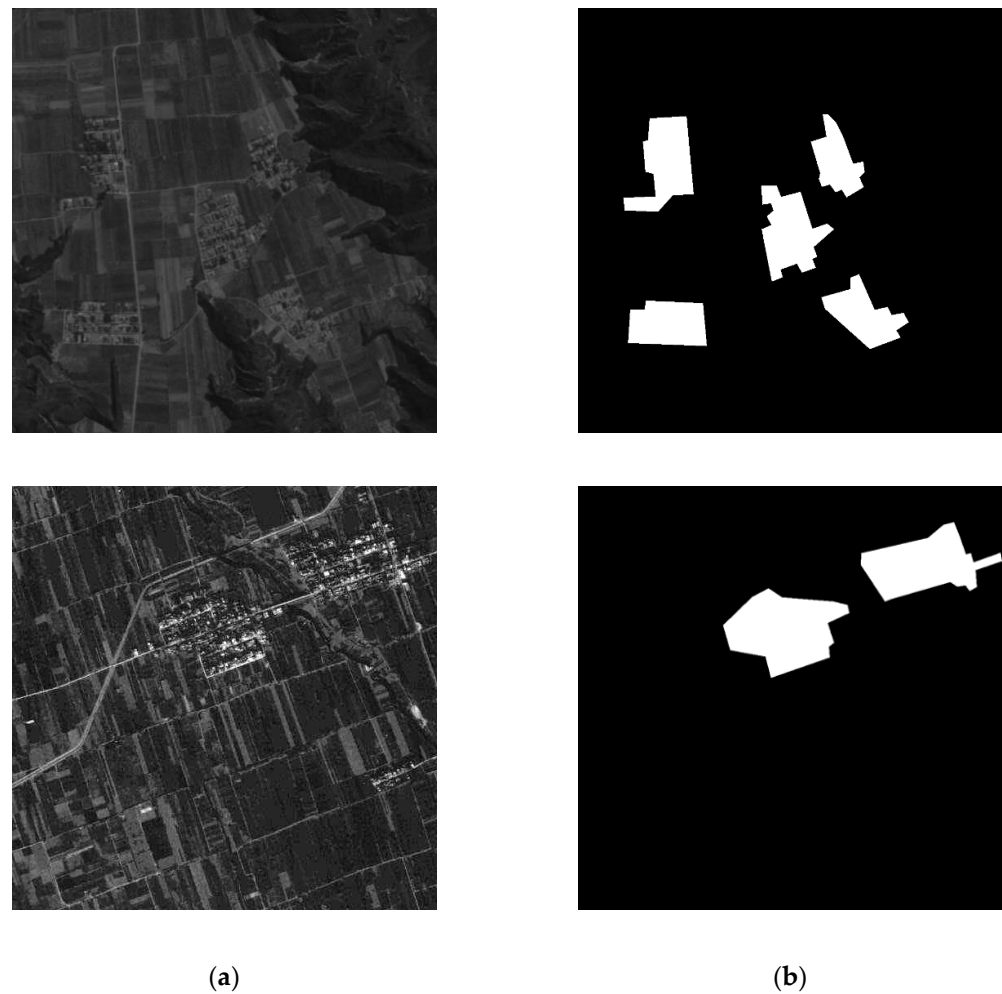
In order to create RERB dataset, we collected 13,892 high-resolution TH-1 images [7], and the size of each image is approximately  $256 \times 256$  pixels. Figure 2 shows the label visualization result in this dataset. The TH-1 satellite is the first stereo mapping transmission satellite in China, and its goal is to achieve topographic mapping at a 1:50,000 scale without using ground control points. It consists of a high-resolution camera with ground pixel size of 2 m and a multispectral camera with a ground pixel size of 10 m. Images with a spatial resolution of 2 m are applied in this dataset, and these images cover a geographical area of more than 3640 square kilometers.

The proportions of residential area and other land cover categories in RERB dataset are shown in Table 2. It is obvious that the proportion of the residential area is lower than that of the other categories, which is consistent with the distribution of large-scale remote sensing imagery scenes.

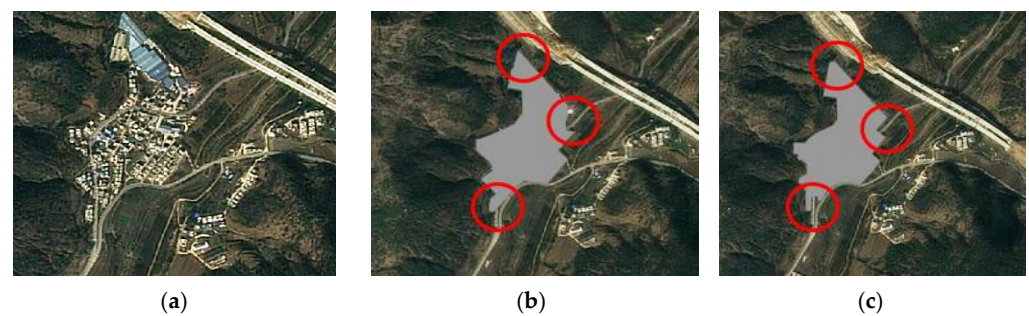
**Table 2.** The proportion of residential area in our dataset.

	Proportion (%)	Label Number	Color
Residential area	15.89	1	(255,255,255)
Background	84.11	0	(0,0,0)

The labels used for traditional semantic segmentation usually do not have regularization characteristics, as shown in Figure 3b. This kind of label is assigned accurately according to the actual range of residential areas in the image [42]. Different from semantic segmentation labels, according to the contour regularization requirements in mapping application, we need to ensure the labels of residential areas into a regular format.

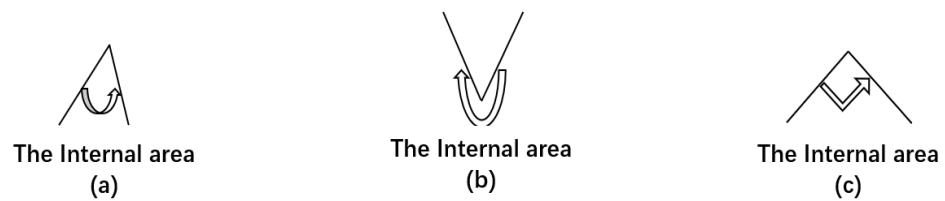


**Figure 2.** Label visualization for a remote sensing image: (a) Original image and (b) label image.



**Figure 3.** Comparison of traditional semantic segmentation labels and mapping application labels: (a) Original image; (b) semantic segmentation labels; and (c) mapping application labels.

In addition to the regularization requirements of contour line segments, special attention should also be paid to the treatment of the included angle between line segments when labeling. The main principles include small contour protrusion removal and small contour concave part filling. As shown in Figure 4, using the interior of the patch as the reference direction, the contour protrusion and the concave part of the contour are defined when the angle between contour segments is too small ( $<45^\circ$ ) and excessively large ( $>90^\circ$ ), respectively. These situations will be corrected with blunt or right angles. For example, in Figure 3b, there are small acute angles as shown in the red circles at the corner of residential areas. Therefore, as shown in Figure 3c, we edit these angles by using a right angle or an obtuse angle in mapping application labels.



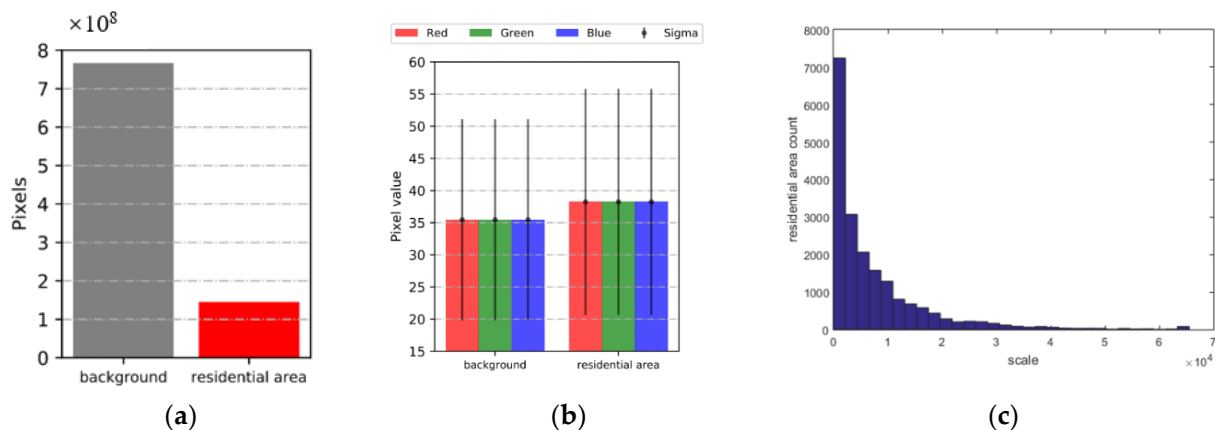
**Figure 4.** Diagrams of raised and sunken areas: (a) The angle is too small; (b) the angle is too large; and (c) the angle is  $90^\circ$ .

We split 85% of these images into the train set and leave the remaining 15% as the test set. As for annotation, RERB dataset provides pixel-level labels for two important categories, including background and residential area. They are labeled with black (0) and white (1).

We also analyze RERB dataset and find it has four properties: (1) Large-scale and high-resolution. As shown in Table 1, RERB contains 13,892 high-quality satellite images acquired from different cities in China. It covers an area of more than 3640 square kilometers. (2) Well annotated and regular label contour. For each satellite image, we provide accurate pixel-wise mapping application labels for two categories ('background' and 'residential' area), which are annotated by a group of experts. (3) Rich background. The remote sensing mapping task is always faced with the diverse background samples (i.e., ground objects that are not of interest). The high-resolution and different scenes bring more rich details for the background samples. (4) Class imbalance. As shown in Table 2, two categories have very different proportions, which lead to a class imbalance problem. This problem poses a special challenge for the regularization extraction of the residential areas task.

### 3.3. Statistics for RERB Dataset

Some statistics of the RERB dataset are analyzed in this section. The number of labeled pixels has been counted. As is shown in Table 2 and Figure 5a, the background class contains the most pixels with rich and diverse background samples, which cause special challenge for residential areas extraction.



**Figure 5.** Statistics for the RERB dataset: (a) Histogram of the number of pixels for each class; (b) spectral statistics. The mean and standard deviation (sigma) for background and residential area are reported; and (c) distribution of the residential area sizes.

For the spectral statistics (Figure 5b), the background category has a lower mean value (color column) and standard deviation (vertical line). Because of the high-resolution images of TH-1 satellite are single channel, the values of red, green, and blue are same. As is shown in Figure 5c, most of the residential areas have relatively small scales. Through calculation, the average size of the minimum 30% residential areas is about 479.71 pixels, and the average size of the maximum 30% residential areas is about 18,851 pixels. The multiscale residential areas require the models to have multiscale capture capabilities.



### 3.4. Construction of Residential Area Regularized Extraction Model

The common semantic segmentation network is generally a symmetric network with encoding–decoding structure [19,20]. The encoding operations mainly include convolution and pooling. Convolution is used to extract high-dimensional features of the input image, and pooling is used to make the image smaller. The decoding operations mainly include deconvolution and up-sampling. Deconvolution makes the features of the image reappear after classification, and up-sampling can restore the original size of the image. Finally, the classification results of each pixel are output. In terms of loss function, cross-entropy [46] has been the most widely used loss function in semantic segmentations of images.

In order to test the effectiveness of the RERB dataset, we designed an end-to-end regularized extraction network by analyzing the regularization characteristics of label contour and the constraints of loss function.

As shown in Figure 6, compared with the traditional semantic segmentation network in Figure 7, our method extracts the contour of the label image first, and realizes regularization extraction by adding the cross-entropy constraint of the label contour image and model prediction image to the original loss function. The baseline method chosen in this article can be any semantic segmentation network, such as U-Net [19] or DeepLab v3+ [24].

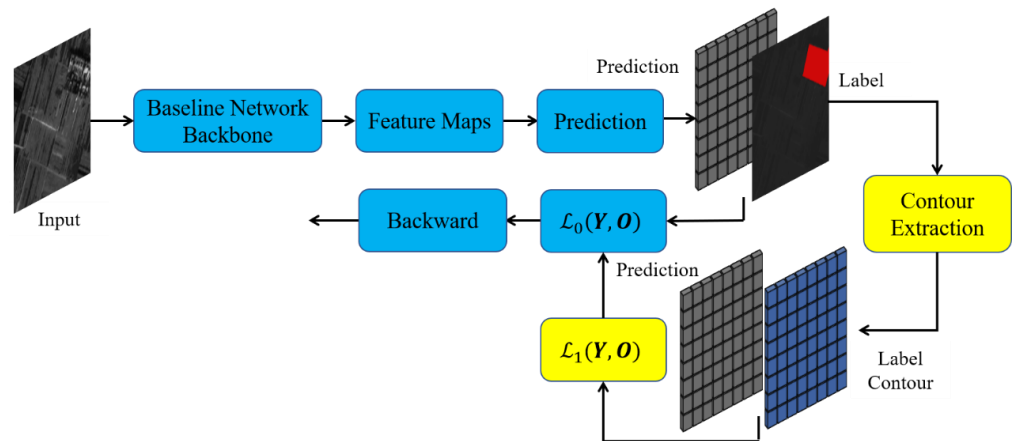


Figure 6. Regularized extraction model structure of residential area (ours).

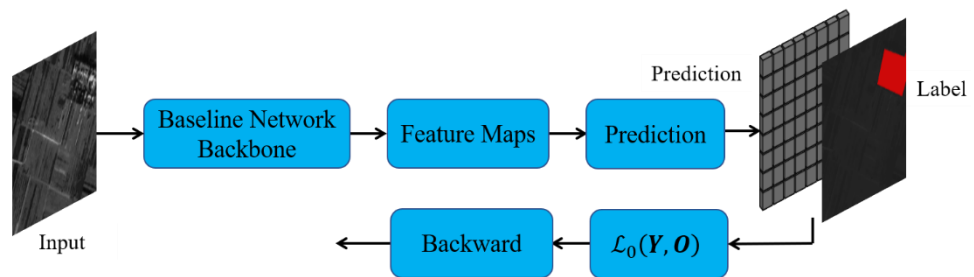


Figure 7. Traditional semantic segmentation network (previous).

The cross-entropy loss function can make the predicted image learned from the training data similar to the real label image. Considering that the label image contour of RERB dataset already has good, regularized contour characteristics, we first extract the contour of label image and then constrain the contour regularization degree of the network prediction image by calculating the cross-entropy loss between the label contour image and the network prediction image, as shown in Equation (1):

$$\mathcal{L}_1(\mathbf{Y}, \mathbf{O}) = F_{ce}(\mathbf{Y}, G_{con}(\mathbf{O})) \tag{1}$$

where  $\mathbf{O} \in \{0, 1\}^{W \times H}$  denotes the label image,  $W \times H$  represents the image size,  $\mathbf{Y}$  is the network inference result image, which has the same size as image  $\mathbf{O}$ .  $F_{ce}$  and  $G_{con}(\mathbf{O})$  represent the cross-entropy loss function and the contour extraction function, respectively.

$\mathcal{L}_1(\mathbf{Y}, \mathbf{O})$  represents the degree of inconsistency between the contours of the two images. Through the calculation and back propagation of  $\mathcal{L}_1(\mathbf{Y}, \mathbf{O})$  in the training process, the contour of the prediction image can be made more and more regular. The cross-entropy loss function is expressed as follows:

$$F_{ce}(\mathbf{Y}, \mathbf{G}) = - \sum_x g_x \log(y_x) + (1 - g_x) \log(1 - y_x) \tag{2}$$

where  $y_x$  and  $g_x$  denotes the value at position  $x$  in the image  $\mathbf{Y}$  and  $\mathbf{G}$ , respectively.

The contour of the label image can be extracted by the corrosion of a  $3 \times 3$  structuring element. Corrosion is a commonly used morphological operation in an image processing file, and it can be expressed as follows:

$$G_{con}(\mathbf{O}) = |\mathbf{O} - Erosion_{3 \times 3}(\mathbf{O})| \tag{3}$$

In the above formula, the corrosion operation  $Erosion_{3 \times 3}(\mathbf{O})$  can remove the area contour in the image  $\mathbf{O}$ , and then the contour image can be obtained by subtracting the corroded image from the original image.

In the training stage, the Adam [47] optimizer is adopted, and it is a first-order optimization algorithm. The best model can be obtained by minimizing joint loss function  $\mathcal{L}(\mathbf{Y}, \mathbf{O})$ , which is shown in the following formula:

$$\mathcal{L}(\mathbf{Y}, \mathbf{O}) = \alpha \mathcal{L}_0(\mathbf{Y}, \mathbf{O}) + \beta \mathcal{L}_1(\mathbf{Y}, \mathbf{O}) \tag{4}$$

where  $\mathcal{L}_0(\mathbf{Y}, \mathbf{O})$  is the original loss function of the baseline network, and the functions used in this paper include cross-entropy and Lovász [48]. The existence of  $\mathcal{L}_0(\mathbf{Y}, \mathbf{O})$  can ensure the segmentation accuracy of the original semantic segmentation network.  $\alpha$  and  $\beta$  are the two weights of loss functions, which are experimentally determined.

#### 4. Experiment and Analysis

In this section, we carried out experimental verification and tested the effectiveness of the RERB dataset by using the model constructed in Sections 3 and 4. We first introduced the evaluation metrics. Then, we performed an ablation study to determine some parameters. In the contrast experiment, the baseline networks were U-Net and DeepLab v3+. All experiments were carried out on a platform with an Intel Core (TM) i9 3.60 GHz CPU, 32 GB RAM, GeForce GTX 2080 GPU, and 11 GB video memory. These algorithms were implemented using PyTorch 1.0 and Python 3.7.

##### 4.1. Design of Evaluation Metrics

The traditional semantic segmentation evaluation indexes, such as mean intersection over union (*mIoU*) [18], mainly evaluated the extraction accuracy in pixel units, which cannot reflect the regularization degree of contours as a whole. In detail, the calculation of *mIoU* was based on the confusion matrix, as shown in Table 3. There were  $n_{cl}$  different classes in total, including backgrounds, where  $n_{ij}$  was the number of pixels of class  $i$  predicted to belong to class  $j$  and  $t_i = \sum_j n_{ij}$  was the total number of pixels of class  $i$ .

**Table 3.** Confusion matrix.

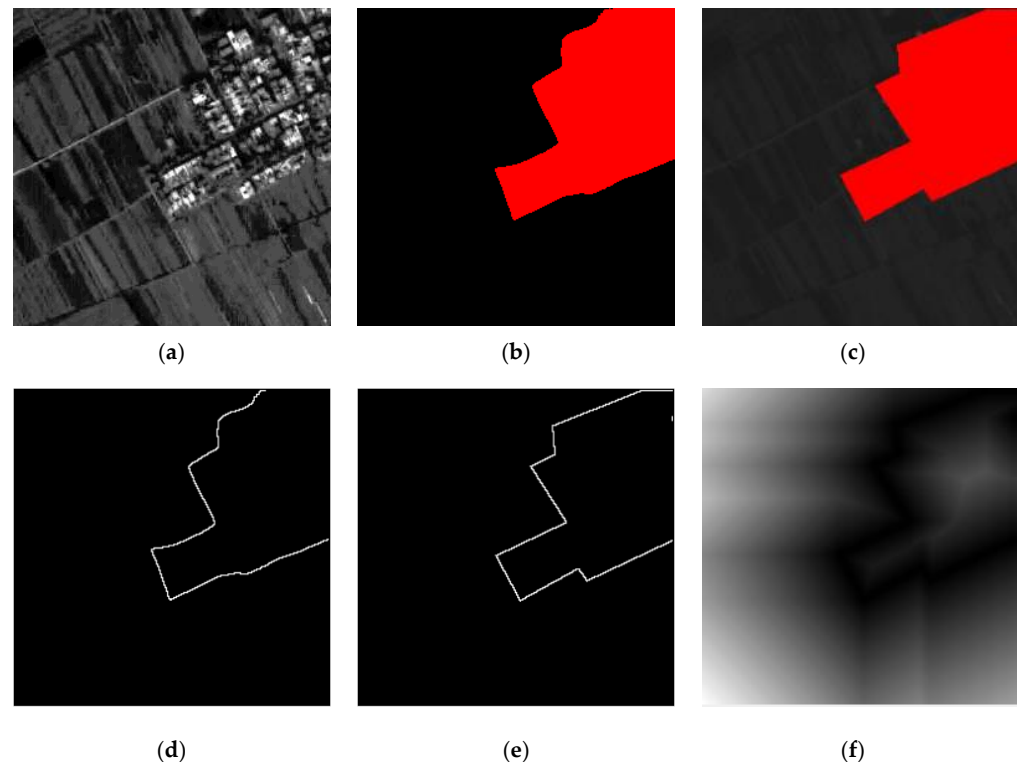
Confusion Matrix		Ground Truth Labels				
		class 1	...	class $i$	...	class $n_{cl}$
prediction	class 1	$n_{11}$	...	$n_{1i}$	...	$n_{1n_{cl}}$
	...	...	...	...	...	...
	class $i$	$n_{i1}$	...	$n_{ii}$	...	$n_{in_{cl}}$
	...	...	...	...	...	...
	class $n_{cl}$	$n_{n_{cl}1}$	...	$n_{n_{cl}i}$	...	$n_{n_{cl}n_{cl}}$

Therefore,  $mIoU$  is calculated as follows.

$$mIoU = (1/n_{cl}) \text{mathlarger} \sum_i n_{ii} / (t_i + \text{mathlarger} \sum_j n_{ji} - n_{ii}) \quad (5)$$

To quantitatively evaluate the regularization extraction results, a contour matching index (CMI) was designed to measure the performance of the algorithm in this paper. The specific steps of the CMI calculation are as follows.

- (1) The contours of the model prediction image (Figure 8b) and the label image (Figure 8c) were extracted, and the results are shown in Figure 8d,e;
- (2) The distance transform of the contour of label image was computed, as shown in Figure 8f;
- (3) A contour matching value was obtained by matching the contour of the model prediction image (Figure 8d) with the distance transformed image (Figure 8f);
- (4) The CMI value of this image was obtained by dividing the matched value by the number of pixels in the contour of the label image.



**Figure 8.** Schematic illustration of the procedure used to calculate the CMI: (a) Image; (b) model prediction image; (c) label image; (d) contour image of (b); (e) contour image of (c); and (f) transformed image of (e).

The critical factor of the distance transform [49] was the definition of distance. In this paper, a block distance transform was adopted. The pixel value of the true contour point was 0 in the image after the distance transform was computed. The farther away from the true contour point, the larger the pixel value of the transformed image. Thus, the matching value can be obtained by calculating the sum of pixel values in the transformed image corresponding to the position of the contour points in the model prediction image. Since the contour image was a binary image that contained only the residential area point (pixel value 1) and the background point (pixel value 0), the matching values between Figure 8d,f can be calculated as follows:

$$S = \sum_{(i,j)} C_{i,j} \cdot e_{i,j} / \text{sum}(G_t) \quad (6)$$

where  $(i, j)$  is the pixel coordinate,  $C_{i,j}$  and  $e_{i,j}$  are pixel values at  $(i, j)$  in Figure 8d,f, respectively, and  $sum(G_t)$  represents the total number of contour points in label contour image  $G_t$  (Figure 8e).

The value  $S$  reflected the matching degree between the prediction result and the value image. The smaller the value was, the higher the matching degree. Furthermore, the average CMI of all images was used as the evaluation result when a whole test set was evaluated.

Considering that the background class occupied most of the image, we removed the background class in the calculation of mIoU to prevent it from affecting the evaluation of other ground features. Therefore, the evaluation indexes included the CMI and IoU of residential areas.

#### 4.2. Parameters Settings and Ablation Study

##### 4.2.1. Parameters Settings

In the experiment, we divided the training set and test set according to the ratio of 17:3. Finally, the training set and the test set contained approximately 13,611 image slices and 281 image slices, respectively. To verify the adaptability of the proposed method to different loss functions, Lovász was used for  $\mathcal{L}_0(Y, O)$  when the baseline network was U-Net, and cross-entropy was used for  $\mathcal{L}_0(Y, O)$  when the baseline network was DeepLab v3+. The number of ground feature elements  $c$  was set as 2.

The polynomial learning rate policy was employed where the initial learning rate was multiplied by  $(1 - iter / total\_iter)^{1.5}$  after each iteration. The maximum number of training cycles was 100 epochs, and thus,  $total\_iter = 100$ . The optimal model was determined by testing the model epoch by epoch during training. The weight decay coefficient was set to 0.0005. In terms of the optimization method, the Adam [47] optimizer was used for training.

Batch size value had a great impact on model training and quality of results. Usually, we selected the maximum value according to the network parameters and the hardware configuration (mainly the video memory of GPU). In this paper, we carried out experiments with a batch size of 8, which was determined by model size and video memory. The selection principle was to make the video memory not overflow.

##### 4.2.2. Ablation Study

In this section, we first studied the influence of the initial learning rate on the test set of the RERB dataset. To perform this ablation study, we adopted the semantic segmentation network and the metric mIoU. We evaluated the performance pertaining to the abovementioned parameters, as described in Table 4.

**Table 4.** Ablation study for the initial learning rate.

The Initial Learning Rate	U-Net (Lovász)	DeepLab v3+ (Cross-Entropy)
	mIoU	mIoU
$1 \times 10^{-3}$	0.76099	0.76919
$1 \times 10^{-4}$	0.77895	0.79395
$5 \times 10^{-5}$	0.77924	0.79749
$2 \times 10^{-5}$	0.78533	0.80450
$1 \times 10^{-5}$	0.77406	0.79527
$1 \times 10^{-6}$	0.77384	0.77298

The experiments specified in Table 4 were conducted with training batch size = 8. As shown in Table 4, the *mIoU* peaked when the initial learning rate was  $2 \times 10^{-5}$ .

Next, we studied the influence of the weight  $\alpha$  and  $\beta$  on the test set of the RERB dataset. The cross-entropy loss between the label contour image and the network prediction image was inserted to the above semantic segmentation models. The metric CMI was adopted in these experiments.

The experiments specified in Table 5 were conducted with training batch size = 8 and the initial learning rate  $2 \times 10^{-5}$ . As shown in Table 5, the CMI index of the proposed method reached the minimum value when  $\alpha = 0.3$  and  $\beta = 0.7$ .

**Table 5.** Ablation study for  $\alpha$  and  $\beta$ .

$(\alpha, \beta)$	Ours (U-Net) Lovász+ $\mathcal{L}_1(Y, O)$	Ours (DeepLab v3+) Cross-Entropy+ $\mathcal{L}_1(Y, O)$
	CMI	CMI
(0.1, 0.9)	50.647	69.378
(0.2, 0.8)	51.083	72.654
(0.3, 0.7)	39.687	32.074
(0.4, 0.6)	46.174	58.461
(0.5, 0.5)	54.378	36.794
(0.6, 0.4)	56.376	49.586
(0.7, 0.3)	72.545	50.277
(0.8, 0.2)	53.475	40.433
(0.9, 0.1)	43.602	44.169

#### 4.3. Results and Analysis

We parameter tuned some parameters of U-Net, DeepLab v3+ and our proposed method, and the quantitative evaluation results on the test set of RERB dataset are shown in Table 6.

**Table 6.** Training parameters and quantitative evaluation results.

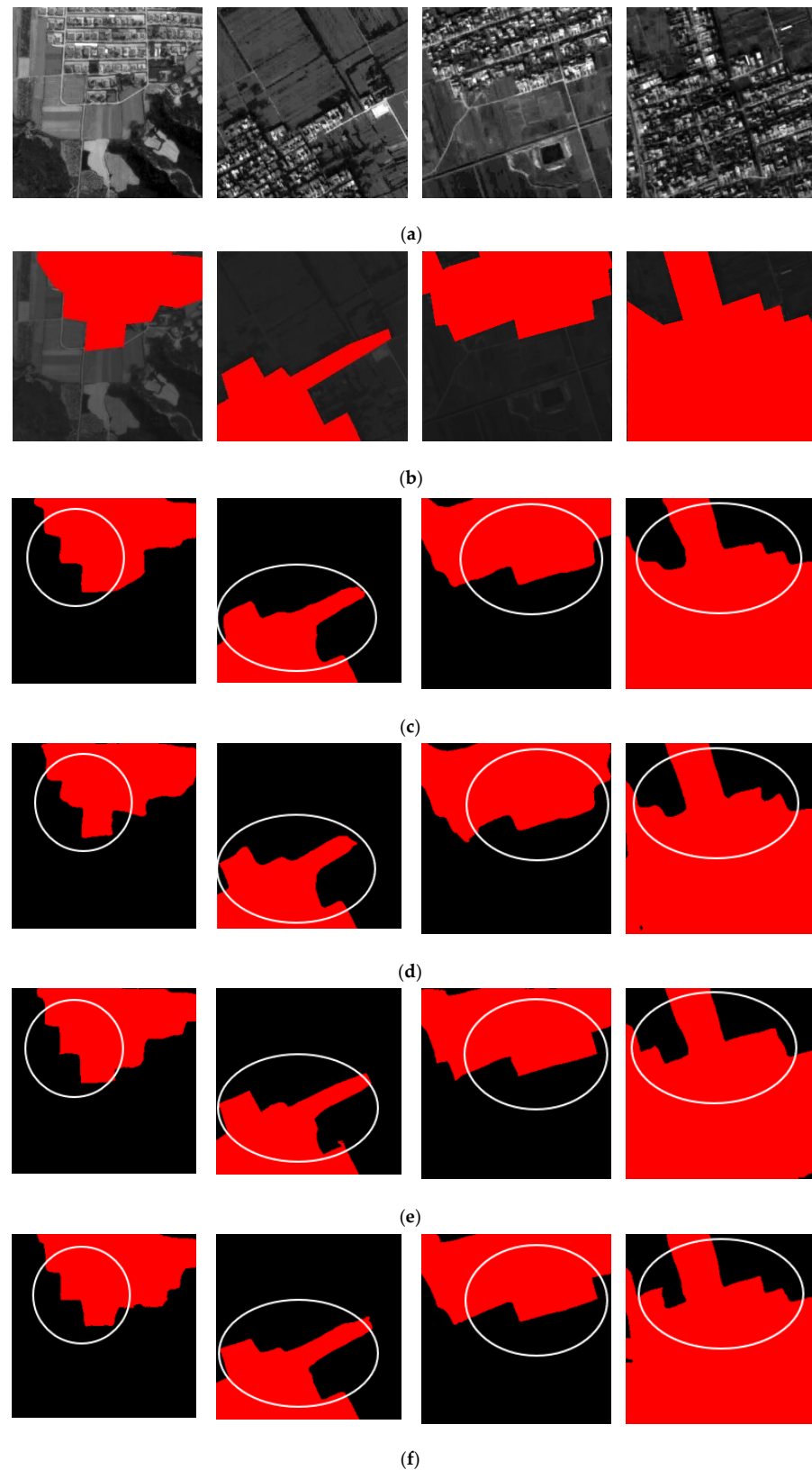
	U-Net	Ours (U-Net)	DeepLab v3+	Ours (DeepLab v3+)
loss	Lovász	Lovász+ $\mathcal{L}_1(Y, O)$	Cross-Entropy	Cross-Entropy + $\mathcal{L}_1(Y, O)$
initial lr	$2 \times 10^{-5}$	$2 \times 10^{-5}$	$2 \times 10^{-5}$	$2 \times 10^{-5}$
batch size	8	8	8	8
$(\alpha, \beta)$	--	(0.3, 0.7)	--	(0.3, 0.7)
IoU	0.7853	0.7813 (−0.51%)	0.7953	0.8003 (+0.63%)
CMI	65.638	39.686 (+39.54%)	43.051	32.074 (+25.50%)
train epoch	18	41	41	52
test time	15.37 s	15.55 s	42.79 s	42.78 s

The contrasting experimental results are shown in Figure 9. As seen from Table 6 and Figure 9, the regularization level of residential area contours extracted by our proposed method had increased greatly, especially those areas marked by white circles. When the baseline network was U-Net, the IoU of residential areas decreased by 0.51%, but the CMI increased by 39.54%. Moreover, both the IoU of residential areas and the CMI increased by 0.63% and 25.5%, respectively, when the baseline network was DeepLab v3+.

Compared with the semantic segmentation dataset, the label image contour in RERB dataset had the regularization characteristic and provided additional information, so it could support the construction and training of end-to-end regularization extraction model of residential areas. The experimental results demonstrated the preponderance and practicability of the RERB dataset.

In terms of computational complexity, according to the model construction method in Sections 3 and 4, the increased calculation amount of this method compared with the basic network mainly included label image edge extraction and contour cross-entropy loss calculation during training. The operations of contour extraction included a corrosion operation with  $3 \times 3$  structuring element and a subtraction. Contour cross-entropy loss calculation included logarithmic calculation and accumulation, which was the same as the original cross-entropy function. Therefore, during the training phase, the computational complexity of the proposed method was slightly larger than that of the traditional semantic segmentation network. Consequently, the runtime of training and the optimal epoch

number of our method were higher than those before model modification. In the case of test time, our proposed method was at the same level with traditional models.



**Figure 9.** Extraction results of some residential areas: (a) Original images; (b) label images; (c) U-Net; (d) DeepLab v3+; (e) our model (the baseline network is U-Net); and (f) our model (the baseline network is DeepLab v3+).

## 5. Conclusions

For residential areas, the difference between semantic segmentation labels and mapping application labels limits the possibility of end-to-end regularization extraction training. In order to address this problem, we built a dataset named RERB (Residential area Extraction with Regularized Boundary) for the end-to-end regularization extraction of residential areas. To ensure the rationality of RERB dataset, we analyzed the contour representation requirements for residential area mapping according to the graphic specification of 1:50,000 topographic map, and then transformed it into the following annotation requirements: the contour of label image should be regular, and the included angle of contour line segments should be as right angle as possible. Based on these principles, we have completed the annotation of residential areas in 13,892 image patches based on TH-1 images. The size of each image is approximately  $256 \times 256$  pixels. The RERB dataset encompasses four properties: (1) Large-scale and high-resolution; (2) well annotated and regular label contour; (3) rich background; and (4) class imbalance. In reality, high resolution, complex background, and category imbalance represent three challenges in residential area mapping. Finally, a residential area regularization extraction model is constructed with a contour cross-entropy constraint by using the regular contour label of a residential area. Experimental results showed that the proposed algorithm can improve the regularization degree of the extracted contour of residential areas while maintaining nearly the same extraction accuracy. This fully proves the effectiveness of RERB dataset. In the future, we will expand and improve the dataset of mapping residential area and conduct in-depth research on the end-to-end model for mapping.

**Author Contributions:** Conceptualization, L.Z., S.L. and D.G.; methodology, S.L., W.L. and D.G.; resources, D.G.; data curation, S.L., H.G., X.Z. and J.H.; writing—original draft preparation, S.L. and J.H.; writing—review and editing, S.L., W.L. and J.H.; supervision, L.Z. and D.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (NSFC) under grant 62101395 and the independent research project of State Key Laboratory of Geo-Information Engineering (Grants No. SKLGIE2020-ZZ-1).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available, as the research group's mapping model construction related research is still being carried on.

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

## References

1. GB/T 20257.3-2017; Cartographic Symbols for National Fundamental Scale Maps-Part 3: Specifications for Cartographic Symbols 1:25000 1:50000 & 1:100000 Topographic Maps. Standardization Administration of the P.R.C.: Beijing, China, 2017; pp. 19–21. (In Chinese)
2. Feng, W. *Remote Sensing Image Interpretation*, 1st ed.; Science Press: Beijing, China, 1998; pp. 162–165. (In Chinese)
3. Peterle, J. A concept for topographic map updating using digital orthophotos. *Photogrammetria* **1985**, *40*, 87–94. [[CrossRef](#)]
4. Holland, D.A.; Boyd, D.S.; Marshall, P. Updating topographic mapping in Great Britain using imagery from high-resolution satellite sensors. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 212–223. [[CrossRef](#)]
5. Ministry of Finance of the PRC; State Bureau of Surveying and Mapping of the PRC. *Detailed Rules for Quota Calculation of Surveying and Mapping Production Costs*; Ministry of Finance of the PRC: Beijing, China, 2009; p. 24. (In Chinese)
6. Li, D. China's first civilian three-line-array stereo mapping satellite: ZY-3. *Acta Geod. Cartogr. Sin.* **2012**, *41*, 317–322. (In Chinese)
7. Wang, J.; Wang, R.; Hu, X.; Su, Z. The on-orbit calibration of geometric parameters of the Tian-Hui 1 (TH-1) satellite. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 144–151. [[CrossRef](#)]
8. eCognition. Available online: <https://geospatial.trimble.com/products-and-solutions/ecognition> (accessed on 26 August 2021).
9. Han, F.; Su, Y.; Zheng, J. Research on method of extracting discovery based on EasyFeature elements. *Geomat. Spat. Inf. Technol.* **2020**, *43*, 234–236. (In Chinese)

10. Alokasi, H.; Ahmad, M.B. Deep learning-based frameworks for semantic segmentation of road scenes. *Electronics* **2022**, *11*, 1884. [[CrossRef](#)]
11. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* **2022**, *493*, 626–646. [[CrossRef](#)]
12. Fu, G.; Liu, C.J.; Zhou, R.; Sun, T.; Zhang, Q.J. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [[CrossRef](#)]
13. Ji, S.; Wei, S. Building extraction via convolutional neural networks from an open remote sensing building dataset. *Acta Geod. Cartogr. Sin.* **2019**, *48*, 448–459. (In Chinese) [[CrossRef](#)]
14. Alemohammad, H.; Booth, K. LandCoverNet: A global benchmark land cover classification training dataset. *arXiv* **2020**, arXiv:2012.03111.
15. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [[CrossRef](#)]
16. Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; Zhong, Y. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. In Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, Virtual, 6–14 December 2021. [[CrossRef](#)]
17. Liu, S.; Gao, K.; Qin, J.; Gong, H.; Wang, H.; Zhang, L.; Gong, D. SE<sup>2</sup>Net: Semantic segmentation of remote sensing images based on self-attention and edge enhancement modules. *J. Appl. Remote Sens.* **2021**, *15*, 026512. [[CrossRef](#)]
18. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [[CrossRef](#)]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241. [[CrossRef](#)]
20. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
21. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. [[CrossRef](#)]
22. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
23. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
24. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conf. Computer Vision (ECCV), Part VII, Munich, Germany, 8–14 September 2018; pp. 833–851. [[CrossRef](#)]
25. Wu, H.; Zhang, J.; Huang, K. FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation. *arXiv* **2019**, arXiv:1903.11816.
26. Yu, F.; Koltun, V. Multi-Scale context aggregation by dilated convolutions. In Proceedings of the 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
27. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [[CrossRef](#)]
28. Hasan, A.H.; Al-Kremy, N.A.R.; Alsaffar, M.F.; Jawad, M.A.; Al-Terehi, M.N. DNA Repair Genes (APE1 and XRCC1) Polymorphisms-Cadmium Interaction in Fuel Station Workers. *J. Pharm. Negat. Results* **2022**, *13*, 32–37.
29. Alsaffar, M.F. Elevation of Some Biochemical and Immunological Parameters in Hemodialysis Patients Suffering from Hepatitis C Virus Infection in Babylon Province. *Indian J. Forensic Med. Toxicol.* **2021**, *15*, 2354–2362.
30. Sun, Y.; Zhang, X.; Xin, Q.; Huang, J. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 3–14. [[CrossRef](#)]
31. Cui, W.; Xiong, B.; Zhang, L. Multi-scale fully convolutional neural network for building extraction. *Acta Geod. Cartogr. Sin.* **2019**, *48*, 597–608. (In Chinese) [[CrossRef](#)]
32. Liu, Y.; Piramanayagam, S.; Monteiro, S.T.; Saber, E. Semantic segmentation of multisensory remote sensing imagery with deep ConvNets and high-order conditional random fields. *J. Appl. Remote Sens.* **2019**, *13*, 016501. [[CrossRef](#)]
33. Ding, L.; Bruzzone, L. A deep architecture based on a two-stage learning for semantic segmentation of large-size remote sensing images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5228–5231. [[CrossRef](#)]
34. Zheng, X.; Huan, L.; Xia, G.; Gong, J. Parsing very high resolution urban scene images by learning deep ConvNets with edge-aware loss. *ISPRS J. Photogramm. Remote Sens.* **2020**, *170*, 15–28. [[CrossRef](#)]
35. Liu, J.; Xiong, X.; Li, J.; Wu, C.; Song, R. Dilated residual network based on dual expectation maximization attention for semantic segmentation of remote sensing images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1825–1828. [[CrossRef](#)]



36. Zhang, X.; Du, L.; Tan, S.; Wu, F.; Zhu, L.; Zeng, Y.; Wu, B. Land use and land cover mapping using RapidEye imagery based on a novel band attention deep learning method in the Three Gorges reservoir area. *Remote Sens.* **2021**, *13*, 1225. [[CrossRef](#)]
37. Abdollahi, A.; Pradhan, B. Integrating semantic edges and segmentation information for building extraction from aerial images using UNet. *Mach. Learn. Appl.* **2021**, *6*, 100194. [[CrossRef](#)]
38. Liu, J.; Zhang, J.; Li, Z.; Zhang, G.; Du, W.; Zhao, W.; Liu, J. Technical framework of 1:10000 cartographic element extraction based on GF-7 satellite. *Geomat. World* **2018**, *25*, 58–61. (In Chinese) [[CrossRef](#)]
39. Zhao, M.; Liu, S.; Xu, G.; Yang, M. A method of residential area contours regularization in remote sensing image based on straight line segment fitting. *Geomat. Sci. Eng.* **2019**, *39*, 29–33. (In Chinese)
40. Pan, X.; Zhao, J.; Xu, J. An end-to-end and localized post-processing method for correcting high-resolution remote sensing classification result images. *Remote Sens.* **2020**, *12*, 852. [[CrossRef](#)]
41. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Fully convolutional neural networks for remote sensing image classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5071–5074. [[CrossRef](#)]
42. He, C.; Fang, P.Z.; Zhang, Z.; Xiong, D.H.; Liao, M.S. An end-to-end conditional random fields and skip-connected generative adversarial segmentation network for remote sensing images. *Remote Sens.* **2019**, *11*, 1604. [[CrossRef](#)]
43. Sun, Y.; Zhang, X.; Zhao, X.; Xin, Q. Extracting building boundaries from high resolution optical images and LiDAR data by integrating the convolutional neural network and the active contour model. *Remote Sens.* **2018**, *10*, 1459. [[CrossRef](#)]
44. Ruan, T.; Liu, T.; Huang, Z.; Wei, Y.; Wei, S.; Zhao, Y. Devil in the details: Towards accurate single and multiple human parsing. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 28–30 January 2019; pp. 4814–4821.
45. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. Gated-SCNN: Gated shape CNNs for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 5228–5237. [[CrossRef](#)]
46. YA, D.M.; Liu, Q.; Qian, Z.B. Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; pp. 743–746. [[CrossRef](#)]
47. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3th International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
48. Berman, M.; Triki, A.R.; Blaschko, M.B. The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4413–4421. [[CrossRef](#)]
49. Embrechts, H.; Roesse, D. Parallel Algorithms for the Distance Transformation. In *Parallel Processing: CONPAR 92—VAPP V*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 387–391. [[CrossRef](#)]