Article

An Improved Boosting Learning Saliency Method for Built-Up Areas Extraction in Sentinel-2 Images

Zhenhui Sun [1,2] , Qingyan Meng [3,*] and Weifeng Zhai [4]

[1]   Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; sunzhcas@gmail.com
[2]   University of Chinese Academy of Sciences, Beijing 100049, China
[3]   Sanya Institute of Remote Sensing, Hainan 572029, China
[4]   School of Sciences, Qiqihar University, Qiqihar 161006, China; weifengzhai@sohu.com
[*]   Correspondence: mengqy@radi.ac.cn; Tel.: +86-010-644852195

check for updates

Abstract: Built-up areas extraction from satellite images is an important aspect of urban planning and land use; however, this remains a challenging task when using optical satellite images. Existing methods may be limited because of the complex background. In this paper, an improved boosting learning saliency method for built-up area extraction from Sentinel-2 images is proposed. First, the optimal band combination for extracting such areas from Sentinel-2 data is determined; then, a coarse saliency map is generated, based on multiple cues and the geodesic weighted Bayesian (GWB) model, that provides training samples for a strong model; a refined saliency map is subsequently obtained using the strong model. Furthermore, cuboid cellular automata (CCA) is used to integrate multiscale saliency maps for improving the refined saliency map. Then, coarse and refined saliency maps are synthesized to create a final saliency map. Finally, the fractional-order Darwinian particle swarm optimization algorithm (FODPSO) is employed to extract the built-up areas from the final saliency result. Cities in five different types of ecosystems in China (desert, coastal, riverside, valley, and plain) are used to evaluate the proposed method. Analyses of results and comparative analyses with other methods suggest that the proposed method is robust, with good accuracy.

Keywords: built-up areas; saliency detection; improved boosting learning; Sentinel-2

1. Introduction

Population density and resource utilization intensity tend to be very high in built-up areas. Rapid urbanization has resulted in several problems, including the urban heat island effect, air pollution, and unreasonable land use. Therefore, extracting built-up areas is a major topic of interest across numerous fields, including sustainability, remote sensing, and the social sciences. To efficiently distribute information regarding built-up areas to various research disciplines, remote sensing technology is widely used to extract and monitor these areas. The term "built-up areas" is widely used in the literature, and refers to the spatial extent of urbanized areas on a regional scale, but this is a nebulous and inconsistent definition [1]. For the purpose of this study, built-up areas are defined as areas dominated by buildings, streets, and impervious surfaces; golf courses, green urban parks, sparse buildings in suburbs, and rural settlements are not included within this working definition.

Over the last few decades, many methods for extracting built-up areas have been proposed [1–15]. These methods can be broadly categorized into four groups: classification-based, index-based, texture-based, and multisensor-based methods. Classification-based methods [1–3] primarily consider how suitable classifiers may be used to extract built-up areas. However, these methods are fraught with challenges and limitations when applied at regional and global scales; for example, scene-to-scene

data analyses are subjective, while the overall process is time-consuming, and entails complicated computing [4]. Index-based methods [4–9] are designed based on spectral bands within which built-up areas exhibit their highest and lowest reflectance values amidst a multispectral dataset. Usually, these methods are unsuccessful in addressing the difficulties in distinguishing between built-up areas and other land cover types. Texture-based methods [10,11] can extract built-up areas based on high texture granularity. However, such methods may fail when ground objects with similar texture features to the built-up areas are encountered. Multisensor-based methods [1,12–14] combine the various characteristics of multiple sensors to extract built-up areas. However, owing to difficulties associated with synthesizing different data types, such methods have not been widely used [15].

Saliency detection provides a unique perspective for ground object extraction, because it selects only interesting information related to the current behavior or task to be processed, ignoring irrelevant information [16]. Saliency detection methods have been gradually introduced into the field of remote sensing, in recent years, to detect specific objects [17–19] in very high spatial resolution images, or high spatial resolution images, and they have proven to be effective. The Sentinel-2 satellites have a multispectral instrument with 13 spectral bands, and its richer spectrum ensures superior detection and extraction of built-up areas. Due to the unique spectral characteristics of the built-up areas, they can be highlighted in some band combinations, so the built-up areas can easily be identified by saliency detection methods. The saliency detection methods can be mainly categorized as bottom-up and top-down methods. Numerous bottom-up saliency detection methods have been proposed, and these can be broadly divided into four groups: those based on contrast, graph theory, and information theory, prior knowledge, and low-rank matrix recovery theory.

Contrast-based methods consist of local contrast and global contrast methods. The former is used to investigate the rarity of image regions with respect to nearby neighborhoods; the classic saliency method proposed by Itti et al. [20] is a typical local contrast method. Since its proposal, several approaches have also adopted the center-surround contrast strategy to calculate saliency, including the graph-based visual saliency detection method [21], the fuzzy growing method [22], and the discriminant center-surround hypothesis model [23,24], and so on. The global contrast methods calculate saliency using the contrast of a pixel or region with respect to the entire image. Cheng et al. [25] proposed a regional contrast-based saliency extraction algorithm that simultaneously evaluates global contrast differences and spatial coherence. Perazzi et al. [26] developed a contrast-based filtering method, and Shi et al. [27] proposed a generic and fast computational framework called pixelwise image saliency aggregating (PISA), that uses prior spatial information to weight color contrast, and direction contrast to generate a final saliency map.

Graph theory-based methods firstly use a graph model to represent an image, and then apply an established undirected or directed graph to predict the saliency value of each region. Gopalakrishnan et al. [28] performed random walks on graphs to detect salient objects. Wei et al. [29] proposed a model based on the geodesic method. Jiang et al. [30] proposed a saliency detection method using absorbing Markov chains. Yan et al. [30,31] applied a hierarchical model to analyze saliency cues from multiple levels of structure that were then integrated to infer a final saliency map. Qin et al. [32] introduced cellular automata to intuitively detect salient objects based on a dynamic evolution model. Based on information theory-based model, Bruce and Tsotsos [33] used Shannon's self-information to measure saliency, and proposed a visual attention model based on information maximization. Zhang et al. [34] used the self-information of local features in an image to represent object rarity, and then measured the saliency value.

Prior knowledge-based methods improve the accuracy of saliency detection methods via the general image properties that researchers have summarized through observation or experiment. Common prior knowledge types include center, background, objectness, semantic, color, spatial distribution, and sparse priors, with the center and background priors particularly widely used. Several studies [26,35,36] employed the center prior, and demonstrated that it can enhance model performance. The background prior has also been used in several studies [37,38].

Matrix recovery-based methods, aiming at decomposing a matrix into a low-rank matrix and a sparse one, have shown the potential to address the problem of saliency detection, where the decomposed low-rank matrix naturally corresponds to the background, and the sparse one captures salient objects. Peng et al. [39] introduced sparse structures into the rotated principal component analysis (RPCA) model, and proposed a saliency detection method based on low-rank representation and structural sparse matrix decomposition. Lang et al. [40] proposed a multitask sparse pursuit algorithm based on low-rank representation.

Compared to bottom-up methods, little research has hitherto been conducted regarding the top-down saliency model. Jiang et al. [41] proposed a learning-based method by regarding saliency detection as a regression problem, where the saliency detection model was constructed based on the integration of numerous descriptors extracted from training samples with ground truth labels. Zhang et al. [34] integrated the top-down and bottom-up information to construct a Bayesian-based top-down model, where saliency is computed locally. Yang et al. [42] proposed a method combining conditional random field and sparse coding theory. Cholakkal et al. [43] regarded top-down saliency detection as an image-classification problem, and proposed a saliency detection method based on an image-classification framework.

As each group has different advantages, Tong et al. [44] proposed a bootstrap learning (BL) method to enhance performance; it exploits the strength of both bottom-up contrast-based saliency methods and top-down learning methods. However, more research is needed if the BL method is used to extract the built-up area. First, BL introduces a dark channel prior into the coarse saliency detection model to generate a coarse saliency map, but this prior is not suitable for all images. In images of darker backgrounds or brighter foregrounds, it may produce the opposite effect. Although the authors used adaptive weights to attenuate the inverse effects of the dark channel prior, remote sensing images are highly complex, especially where water bodies occur as dark background; in such cases, the BL algorithm may fail. Second, BL does not take into account the spatial information of ground objects, which may result in the detection of large amounts of background information. In addition, it simply superimposes multiscale saliency maps without fully integrating the information they provide.

In this study, an improved boosting learning saliency method for extracting built-up areas from remote sensing images is proposed. First, we determine the optimal band combination for extracting built-up areas. To overcome the shortcomings associated with a dark channel prior, we introduce a multi-cue fusion and water removal strategy into the coarse saliency model, to improve the accuracy of the coarse saliency map. Then, the GWB model [45] is used to consider spatial information, thereby eliminating the impact of land cover surrounding built-up areas, to further improve accuracy and provide reliable training samples for the strong saliency model. After that, the CCA [46] is employed to effectively integrate multiscale saliency maps to optimize the accuracy of the final saliency map. Finally, FODPSO algorithm [47] is used to segment the final saliency map to accurately capture information on built-up areas.

The contribution of this paper is threefold: (1) We improve the BL saliency method based on the characteristics of the remote sensing image for extracting the built-up areas. (2) GWB and CCA are introduced in the proposed method to suppress background regions and attach more importance to regions which are more likely to be parts of built-up areas. (3) We determine the optimal band combination of Sentinel-2 for built-up areas detection.

The rest of this paper is organized as follows: the proposed method is illustrated in Section 2, Section 3 focuses on the experimental results, and Sections 4 and 5 provide the discussion and conclusions, respectively.

2. Proposed Method

A flowchart of the proposed method is shown in Figure 1. The method consists of four stages. First, the image is sharpened to 10 m, and the optimal band combination for the built-up area extraction is determined. Then, the false color image generated by optimal band combination is segmented into a

group of segmented objects. Subsequently, a coarse saliency map is constructed based on multiple clues fusion, and GWB, to generate training samples for a strong model. Based on the representation of three features, a strong classifier is trained to measure saliency. Next, the coarse and refined saliency maps are weighted, in combination, to generate the final saliency map. Finally, the built-up areas are extracted using the FODPSO method.
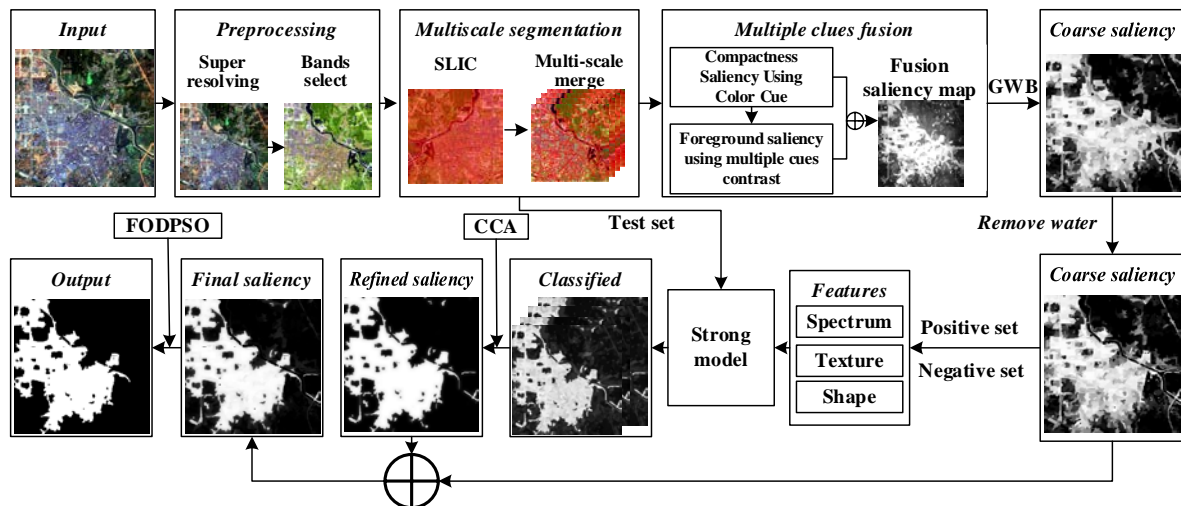

Figure 1. Flowchart of built-up area extraction.

2.1. Image Preprocessing

2.1.1. Sentinel-2 Constellation

The Sentinel-2 constellation consists of two polar-orbiting satellites (Sentinel-2A and Sentinel-2B) placed in the same orbit. Sentinel-2A and Sentinel-2B equip with multispectral instruments capable of acquiring 13 bands information at different spatial resolutions (10, 20, and 60 m). Sentinel-2 provides more details in the near-infrared (NIR) band range and short wavelength infrared (SWIR) band range, which is helpful for land cover, land monitoring, and emergency response [48]. A high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions, which results in 2–3 days at mid-latitudes) provides more cloudless images, and is a good support for the built-up areas extraction.

2.1.2. Atmospheric Correction and Image Sharpening

The bottom-of-atmosphere (surface) reflectance is a basic input to many earth observation applications, ranging from land surface phenology to land cover classification and change detection [49]. To process top-of-atmosphere Level-1C data into atmospherically corrected bottom-of-atmosphere data, the Sen2Cor processor (version 2.4), developed by ESA to perform atmospheric correction, was employed [50].

In higher spatial resolution images, built-up areas tend to be more easily detected because higher spatial resolution images can clearly define the boundaries of the built-up areas, uniformly highlight built-up areas, and eliminate redundant backgrounds in the extracted built-up areas [19]. To sharpen the bands of a Sentinel-2A image with spatial resolutions of 20 m and 60 m to a spatial resolution of 10 m, the modified selected and synthesized band scheme [51] was employed.

2.1.3. Optimal Band Selection

Built-up areas yield a higher reflectance response in the SWIR than in other bands [52], and it may help in alleviating the problem of confusion between built-up areas and other types of land cover, such as artificial open spaces, river gravel, and sand dunes [53]. As Sentinel-2 has two SWIR bands, it is,

therefore, inherently advantageous when applied to built-up areas extraction. In this study, both SWIR bands were selected to form the optimal band combination for built-up areas extraction. To select the third band of the optimal band combination, the optimum index factor (OIF) [54] was employed. OIF is a statistic value that can be used to select the optimum combination of three bands in a satellite image with which you want to create a color composite. The optimal band combination of bands, out of all possible 3-band combinations, is the one with the highest amount of "information" (highest sum of standard deviations), with the least amount of duplication (lowest correlation among band pairs). Band 9 and Band 10 are the water vapor band and the cirrus band. Band 8 and band 8A have very high correlations, and the information overlap is large. The standard deviation of Band 1, Band 2, and Band 3 is relatively low and contains less information. Therefore, these bands are not considered. Five candidate band combinations are Bands 12, 11, 8; Bands 12, 11, 7; Bands 12, 11, 6; Bands 12, 11, 5; and Bands 12, 11, 4. The OIF values of candidate band combinations were calculated based on the TIFF images exported by SNAP in ENVI software. The OIF method can reflect the amount of information in the band combination, but it still has some limitations. We needed to further analyze the separability of candidate band combinations for built-up areas and non-built-up areas. Here, the Jeffries–Matusita (J-M) distance [55,56] was used as a separability criterion for optimal band combination selection, whereby the J-M value ranges from 0 to 2. First, we used ENVI to select samples of built-up areas and non-built-up areas from the TIFF images exported by SNAP. Then, we calculated the J-M value and determined the band combination with the largest J-M value as the optimal band combination. The optimal results are shown in Section 3.

2.2. Multiscale Segmentation

The saliency map accuracy level is sensitive to the segmentation scale, so a multiscale strategy was employed. The false color image generated by optimal band combination was first segmented into homogenous and compact regions using a simple linear iterative clustering (SLIC) superpixel segmentation method [57]. In SLIC, the number of superpixels, N', can affect the effect of segmentation. If N' is too small, it is often impossible to accurately separate the built-up areas from the background; if N' is too large, it needs much more computing time [58]. As can be seen from Figure 2a,b, in some superpixels, both the built-up area pixels and the non-built-up area pixels are included, and the contour of the built-up areas cannot be accurately captured. In Figure 2c, the contour of the built-up areas can be accurately captured. Due to the scenes of remote sensing images being highly complex, determining the optimal N' for each image is very time-consuming. To simplify the problem, we given a large N' to avoid under segmentation. In this study, we empirically set N' to 20,000. Then, Hu's method [59] was adopted to merge similar superpixels into a set of objects O_i, $i = 1, ..., N$, where N is the number of segmented objects. Figure 2d,e show the results of merging 20,000 superpixels into 2000 and 4000 objects, respectively. It not only reduces the number of superpixels, but also ensures that the outline of the built-up areas is well captured.

    (**a**)   (**b**)   (**c**)   (**d**)   (**e**)
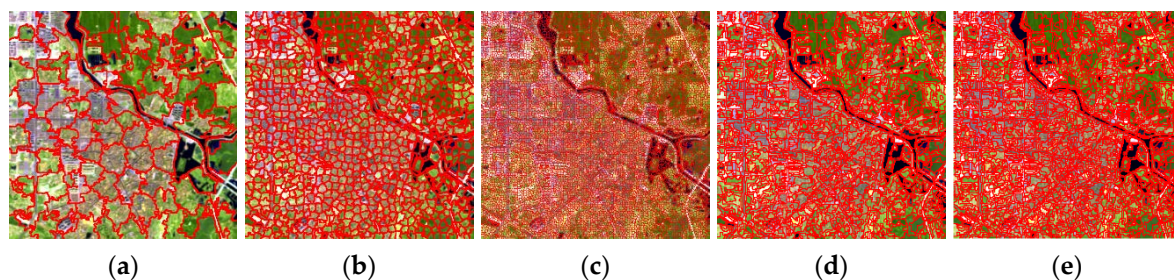
Figure 2. (**a**) Segmentation result of SLIC (N = 100) (**b**) Segmentation result of SLIC (N = 1000) (**c**) Segmentation result of SLIC (N = 10,000) (**d**) Result of merging 20,000 superpixels into 2000 objects (**e**) Result of merging 20,000 superpixels into 4000 objects.

2.3. Feature Selection

In this paper, three descriptors, including the color, texture, and spatial features, are used to describe each segmented object. Color feature is an important feature of the saliency detection methods; almost all saliency methods utilize the color feature. In particular, CIELab [60] aspires to perceptual uniformity, and its L component closely matches human perception of lightness, while the a and b channels approximate the human chromatic opponent system, and RGB is often the default choice for scene representation and storage [36]. Both of them are complementary and widely used for saliency detection [41,44,61]. Hence, we calculated the average pixel value of each segmented object O_i in RGB space and CIELab space, and the color feature of the segmented object O_i can be described as

$$F_{color} = [c_{r,g,b}, c_{L,a,b}], \tag{1}$$

where $c_{r,g,b}$ and $c_{L,a,b}$ represent the average value of each color channel of the pixels in the segmented object O_i in the RGB and CIELab color spaces.

Built-up areas usually have unique texture features. Local binary patterns (LBPs) [62] was utilized to calculate the texture feature of segmented objects. First, the LBP encoding for each pixel in the image was calculated using a 3×3 window and, in the uniform pattern [62], each pixel was assigned a value between 0 and 56. It is worth pointing out that although larger window size (such as 5×5 or 7×7) can utilize more information in the neighborhood, the noise corruption from the pixel away from the center can be more severe, which inevitably deteriorates the discriminative ability of the LBPs' feature [63]. Then, an LBP histogram for each segmented object O_i was constructed, and the texture feature of the segmented object O_i can be described as

$$F_{texture} = \mathbf{H}_{LBP} = \{h_0, , h_i, , h_{56}\}, \tag{2}$$

where h_i is the value of the i-th bin in an LBP histogram.

For spatial features, the eccentricity and area properties were used to eliminate segmented objects with a large eccentricity and a large area; these segmented objects are often strips of bare rock and river banks (such as yellow river bank). This can be described as

$$F_{spatial} = [O_{Area} > th_1 \cap O_{Ecce} < th_2], \tag{3}$$

where O_{Area} is the area and O_{Ecce} is the eccentricity of segmented object O_i, and O_{Ecce} is between 0 and 1. To avoid erroneously eliminating the road inside the built-up areas, we only consider the long strip-shaped segmented objects with a large area. We experimentally set the th_1 to 500 pixels, and th_2 was set to 0.95. The feature vector of segmented object O_i can be obtained by

$$F = [F_{color}, F_{texture}, F_{spatial}] \tag{4}$$

2.4. Coarse Saliency Map

In this part, we mainly explain how to obtain coarse saliency maps and training samples for a strong model. In Section 2.4.1, we mainly used the clues, such as the color and texture, to obtain the initial saliency map of the built-up areas. However, the map often has some background and water information. In Section 2.4.2, we considered the spatial information of ground objects, and introduced the GWB model to eliminate the background information similar to the built-up areas, and obtained the coarse saliency map. In Section 2.4.3, we eliminated the water information from the coarse saliency map. In Section 2.4.4, we selected the training samples from the coarse saliency map for the strong model.

2.4.1. Multiple Cues Fusion

Compactness Saliency Using Color Cues

Following the image segmentation, a graph $G = (V, E)$ with N nodes $\{v_1, v_2, \cdots, v_N\}$ was constructed, and edges E weighted by an affinity matrix $W = [w_{ij}]_{N \times N}$. Node v_i corresponds to the ith segmented object, and edge e_{ij} link nodes v_i and v_j to each other, and the CIELab color space distance l_{ij} between nodes v_i and v_j is defined as

$$l_{ij} = ||c_i - c_j||, \tag{5}$$

where c_i and c_j are the mean of segmented objects corresponding to nodes v_i and v_j in the CIELab color space. Note that the distance matrix $\mathbf{L} = [l_{ij}]_{N \times N}$ is normalized to the interval [0, 1]. The affinity matrix w_{ij} is defined as

$$w_{ij} = \begin{cases} e^{-\frac{l_{ij}}{\sigma^2}}, & if\ j = \Omega_i \\ 0, & otherwise \end{cases}, \tag{6}$$

where σ is a constant, and Ω_i denotes the set of neighbors of node v_i. If v_i and v_j are adjacent, v_j is treated as a neighbor of v_i, and the set of neighbors is equal to the number of v_j.

Salient objects typically have compact spatial distributions, whereas background regions are widely distributed across the entire image. Therefore, compactness may be determined by calculating the spatial variances of the segmented objects to calculate the compactness saliency map [64]. First, the similarity a_{ij} between a pair of segmented objects, v_i and v_j, is defined as

$$a_{ij} = e^{-\frac{l_{ij}}{\sigma^2}} \tag{7}$$

The similarity based on the manifold ranking through the constructed graph is as follows:

$$\mathbf{H}^T = (\mathbf{D} - \alpha \cdot \mathbf{W})^{-1}\mathbf{A}, \tag{8}$$

where $\mathbf{A} = [a_{ij}]_{N \times N}$, $\mathbf{D} = diag\{d_{11}, d_{22}, \ldots, d_{NN}\}$, d_{ii} is the degree of nodes v_i, and $\mathbf{H} = [h_{ij}]_{N \times N}$ is the similarity matrix after the diffusion process, α balances the smooth and fitting constraints of the manifold ranking algorithm and, empirically, α was set to 0.99 as in [65]. The spatial variance of segmented objects can be calculated as

$$sv(i) = \frac{\sum_{j=1}^{N} h_{ij} \cdot n_j \cdot ||b_j - \mu_i||}{\sum_{j=1}^{N} h_{ij} \cdot n_j}, \tag{9}$$

where n_j represents the number of pixels that belong to segmented object v_j, $b_j = [b_j^x, b_j^y]$ represents the centroid of the segmented object v_j, and the $\mu_i = [\mu_i^x, \mu_i^y]$ represents the spatial mean.

Considering that segmented objects at the center of an image are more noticeable, the spatial distances between segmented objects and the image's center can be calculated as follows

$$sd(i) = \frac{\sum_{j=1}^{N} h_{ij} \cdot n_j \cdot ||b_j - p||}{\sum_{j=1}^{N} h_{ij} \cdot n_j}, \tag{10}$$

where $p = [p_x, p_y]$ is the spatial coordinate of the image center.

The saliency map based on compactness is defined as

$$S_{com}(i) = 1 - \text{Norm}(sv(i) + sd(i)), \tag{11}$$

where $\text{Norm}(x)$ is a function that normalizes x to [0, 1].

Foreground Saliency Using Multiple Cues Contrast

Although the compactness saliency method tends to perform well, as attested to by a previous study [64], it only used the spatial variance of the color in the image space. As it primarily depends on color information, when the foreground and background objects are similar in color, the saliency deteriorates. To address this limitation, further aspects, such as texture and position, should be incorporated to refine the results.

First, the foreground seed set was determined by segmenting the compactness saliency map. Then, the contrast of each segmented object with the seeds was calculated using multiple cues, including information on texture and position. The foreground saliency is computed as follows:

$$S_{fore}(i) = \sum_{j \in \Omega_s} [D_t(i,j) \cdot \exp(-||\boldsymbol{b}_i - \boldsymbol{b}_j||/\sigma^2) \cdot n_j], \tag{12}$$

where Ω_s is the foreground seed set, D_t is the texture similarity between segmented objects based on LBP, and $||\boldsymbol{b}_i - \boldsymbol{b}_j||$ is the Euclidean distance between position of segmented objects.

Next, the S_{FG} map was propagated using manifold ranking and, then, the propagated map was normalized to [0, 1], and denoted $S_{fore}(i)$. The $S_{com}(i)$ and $S_{fore}(i)$ maps are complementary to one another, and both saliency maps were integrated to define the initial saliency map,

$$S_{ic}(i) = \eta \cdot S_{com}(i) + (1 - \eta) \cdot S_{fore}(i), \tag{13}$$

where η balances the compactness saliency map $S_{com}(i)$ and foreground saliency map $S_{fore}(i)$. In the optimal band combination, the built-up areas can be better identified using color features, while the built-up areas are also sensitive to texture features. Both of them have important contributions, so η was set to 0.5.

2.4.2. Geodesic Weighted Bayesian

Spatial information is a key aspect of geographic information system (GIS) and remote sensing fields, and while spatial relationships have increasingly been incorporated into satellite image processes, less attention has been given to the use of higher-level spatial relationships [66]. To rectify this, a GWB model [45] was introduced to optimize the initial saliency map. The Bayesian inference for estimating the saliency map [67] is calculated as

$$p(sal|v) = \frac{p(sal) \cdot p(v|sal)}{p(sal) \cdot p(v|sal) + p(bk) \cdot p(v|bk)} , \tag{14}$$

$$p(bk) = 1 - p(sal) , \tag{15}$$

where $p(sal)$ is the prior probability of being salient at pixel v, $p(bk)$ is the prior probability of a pixel belonging to the background, $p(v|sal)$ and $p(v|bk)$ are the likelihood of observations, v is the feature vector of a given pixel. When the spatial relationships were considered, $p(v|sal)$ and $p(v|bk)$ can be rewritten as

$$p(v|sal) = \sum_{s_i \in sal} p_{geo}(s_i|sal)p(v|s_i) , \tag{16}$$

$$p(v|bk) = \sum_{s_i \in bk} p_{geo}(s_i|bk)p(v|s_i), \tag{17}$$

where s_i is segmented object, *sal* is initial set of salient regions, and *bk* is the initial set of background regions. $p_{geo}(s_i)$ denotes the probability of s_i, namely, the weight of segmented object s_i.

Given pixel x, the feature vector was represented by its CIELab color and LBP texture features, and the observation likelihood of the given pixel x in segmented object O_i can be calculated as

$$p(v|s_i) = \prod_{f \in \{L,a,b,LBP\}} \frac{n_j(f(x))}{n_j},$$ (18)

where n_j denotes the number of pixels within segmented object O_i, $n_j(f(x))$ denotes the number of $f(x)$ values contained in segmented object O_i, and $f \in \{L, a, b, LBP\}$ denotes the component of feature vector v, substituting observation likelihood (16) and (17) into (14), and utilizing the initial saliency map as a prior distribution to generate a more precise saliency map. Then, the initial saliency map was further refined to obtain the coarse saliency map S_{course} based on graph cut method [68].

2.4.3. Removing the Water Bodies

Water bodies in remote sensing images usually belong to dark targets, and are more easily identified as salient objects, which can render several saliency detection methods unsuccessful. To avoid the interference caused by water bodies, they must be removed from the coarse saliency map. In [69], Xu noticed that water bodies have a stronger absorbability, and the built-up class has greater radiation in the SWIR band. Based on this characteristic, we set the segmented objects whose average pixel values are smaller than the given threshold, T_w, to 0, thereby achieving the purpose of removing water bodies. To determine T_w, the histogram of SWIR band was first generated. For cities with more water bodies, water bodies occupy a larger area, so there is a peak on the left side of the histogram. The gray values of other ground objects are usually greater than water bodies, and their peaks on the histogram are to the right of the water peak. We determined the value corresponding to the first trough to the right of the water peak as T_w. Based on statistical results of multiple images, we determined that the threshold T_w is 0.15. The cities with less water are almost unaffected by water bodies, and T_w was set to 0.01. The gray value of the building shadow is also low, but its area is small. To avoid removing building shadow, we only removed segmented objects with a large area and the gray value less than T_w. Considering that buildings in some areas are dense, the area of the shadow is relatively large. We empirically set the area threshold for removing water to 100 pixels.

2.4.4. Training Sample Selection

To select accurate training samples from the coarse saliency map, a set of selection rules was established: first, the average saliency value of each segmented object was computed, and two thresholds, T_h and T_l (T_h is greater than T_l), were set to generate initial built-up area and non-built-up area training samples. Both thresholds can be adaptively determined by the mean value ϖ of the coarse saliency map, T_h was set to ϑ times ϖ, and T_l was set to ϖ, where ϑ is a parameter and ϑ was set to 1.8; more discussions about the values of ϑ can be found in Section 3.1.3. The segmented objects with saliency values above the T_h were selected as initial built-up area samples, while those with saliency values below the T_l were selected as initial non-built-up area samples. Next, we constrained the initial training sample set using the spatial feature $F_{spatial}$ to obtain the training samples $\{s_i, l_i\}_{i=1}^P$, where s_i is the i-th training sample from the coarse saliency map S_{course}, l_i is the binary label of the training sample, P is the number of the samples, built-up areas samples are labeled +1, and non-built-up areas samples are labeled −1.

2.5. Refined Saliency Map

One of the main difficulties using a support vector machine (SVM) is to determine the appropriate kernel for the given image. To select the appropriate kernel function for any input image, a multiple kernel boosting method [70] was employed. In this method, SVMs with different kernels are selected as weak classifiers, then, a strong classifier is learned, based on the boosting method. In this paper, we used N_f ($N_{feature}$ by N_{kernel}) different standard SVM classifiers, where $N_{feature}$ is the number of

features and N_{kernel} is the number of kernel functions. The four different kernel functions are linear, polynomial, radial basis function, and sigmoid. For different feature sets, the decision function can be defined as

$$Y(r) = \sum_{n=1}^{N_f} \beta_n \sum_{i=1}^{P} w_i l_i k_n(s, s_i) + \overline{b}, \ \sum_{n=1}^{N_f} \beta_n = 1, \quad \beta_n \in \mathbb{R}_+ \tag{19}$$

where β_n is the kernel weight, w_i is the Lagrange multiplier, and $\overline{b}$ is the bias in the standard SVM algorithm. Equation (19) is a conventional function for the multiple kernel learning method; when the boosting algorithm was used to replace the simple combination of single-kernel SVMs in the multiple kernel learning, Equation (19) can be rewritten as

$$Y(r) = \sum_{n=1}^{N_f} \beta_n (\boldsymbol{w}^T \boldsymbol{k_n}(\boldsymbol{s}) + \overline{\boldsymbol{b}}), \tag{20}$$

where $\boldsymbol{k_n}(\boldsymbol{s}) = [k_n(s,s_1), k_n(s,s_2), \dots, k_n(s,s_P)]^T$, $w = [w_1l_1, w_2l_2, \dots, w_Pl_P]^T$, and $b = \sum\limits_{n=1}^{Nf} \overline{b}$. By setting the decision function as $Z_n(S) = w^T k_n(S) + \overline{b}_n$, the AdaBoost method may be employed to train a strong classifier, and Formula (20) can be rewritten as

$$Y(s) = \sum_{j=1}^{J} \beta_j z_j(s) \tag{21}$$

The AdaBoost method was used to calculate β_j, and J represents the number of iterations of the boosting process. The process is as follows:

Step 1: Begin with uniform weights, $\omega_1(i) = 1/P$, $i = 1, 2, \dots, P$, and assign a set of decision functions $\{Z_n(S), n = 1, 2, \dots, N_f\}$ to each weak classifier.

Step 2: Compute the classification error $\{\varepsilon_n\}$ for each of the weak classifiers, and ascertain the decision function $z_j(s)$ with the minimum error ε_j; then, the combination coefficient β_j is computed by

$$\beta_j = \frac{1}{2} \cdot \ln \cdot \frac{1 - \varepsilon_j}{\varepsilon_j} \cdot \frac{1}{2} \cdot (\text{sgn}(\ln \frac{1 - \varepsilon_j}{\varepsilon_j}) + 1), \tag{22}$$

where sgn(x) is the sign function, which equals 1 when x > 0, and is −1 otherwise, and β_j must exceed 0.

Step 3: Update the weight according to Equation (23), and repeat step 2 for the next iteration until J iterations are completed.

$$w_{j+1}(s_i) = \frac{w_j(s_i)e^{-\beta_j l_i z_j(s_i)}}{2\sqrt{\varepsilon_j(\varepsilon_j - 1)}} \tag{23}$$

Following J iterations, all of the β_j and $z_j(s)$ can be obtained, and then the strong classifier was learned. Subsequently, a pixel-wise saliency map was generated using the strong classifier. Finally, the refined saliency map $S_{refined}$ was improved based on the graph cut method [68] and guided filter [71].

2.6. Multiscale Saliency

Since the size of the ground objects in the image is different, the saliency objects can appear on a variety of scales. In other words, the accuracy of the saliency map is sensitive to the number of segmented objects [70], a multiscale strategy was employed. In this study, seven layers (*M* is set to 7) of segmented objects with different granularities were generated, where *N* was set to 1000, 1500, 2000, 2500, 3000, 3500, and 4000, in each of the layers. More discussions about the values of *M* can be found in Section 3.1.3. To effectively integrate the results of the multiple scales *M*, the CCA method [46] was employed, whereby each cell corresponds to a pixel, and the saliency values of all pixels constitute the set of cells' states. For any cell in a saliency map, there should be $5M - 1$ neighbors, including pixels

with the same coordinates from different saliency maps, in addition to their 4 connected pixels [46]. The saliency value of pixel i in the m-th saliency map at time t stands for its probability to be the foreground F, represented as $S_{m,i}^{(t)}$, while its possibility to be the background B is denoted as $1 - S_{m,i}^{(t)}$. Otsu's method was used to binarize each map using an adaptive threshold. The threshold did not change, and was only related to the initial image. The threshold of the m-th saliency map is represented as γ_m. Following segmentation, a pixel i can be classified as foreground or background. If a pixel, i, is foreground, the probability that one of its neighboring pixels, j, is measured as foreground is λ, while μ is the probability that j is measured as background when i belongs to the background. We assumed that λ is equal to μ if it was considered equally probable that the pixel belongs to the foreground or to the background. The posterior probability can be denoted as $S_{m,i}^{(t)} \cdot \lambda$, which represents the probability of pixel i belonging to the foreground F, on the condition that its neighboring pixel j in the m-th saliency map was binarized as foreground at time t, and posterior probability $S_{m,i}^{(t+1)}$ can also be used to represent the probability of pixel i belonging to the foreground F at time $t + 1$. Based on the prior ratio in [46], we have

$$\frac{S_{m,i}^{(t+1)}}{1 - S_{m,i}^{(t+1)}} = \frac{S_{m,i}^{(t)}}{1 - S_{m,i}^{(t)}} \cdot \frac{\lambda}{1 - \lambda} \tag{24}$$

To deal with the logarithm of Equation (24), we have

$$l(s_{m,i}^{(t+1)}) = l(s_{m,i}^{(t)}) + \Lambda, \tag{25}$$

where $l(s_{m,i}^{(t+1)})$ is $\ln(s_{m,}^{(t+1)}/1 - s_{m,}^{(t+1)})$, and Λ is $\ln(\lambda/1 - \lambda)$. Assuming that each neighbor's contribution was conditionally independent, the synchronous updating rule can be defined as

$$l(s_{m,i}^{(t+1)}) = l(s_{m,i}^{(t)}) + \sum_m^{(t)} \cdot \Lambda, \tag{26}$$

$$\sum_m^{(t)} = \sum_{j=1}^{5} \sum_{k=1}^{M} \delta(k = m, j > 1) \cdot sign(s_{j,k}^{(t)} - \gamma_k \cdot \mathbf{1}), \tag{27}$$

where $s_m^{(t)}$ is the m-th saliency map at time t, M is the number of multiscale saliency maps, $s_{j,k}^{(t)}$ is the vector containing the saliency values of the j-th neighbor for all pixels in the m-th saliency map at time t, and $\mathbf{1} = [1, 1, \ldots, 1]$. After T_C iterations, the integrated saliency map $s^{(TC)}$ can be integrated as

$$s^{(T_C)} = \frac{1}{M} \sum_{m=1}^{M} s_m^{(T_C)} \tag{28}$$

2.7. Integration

Coarse saliency maps boast several advantages for detecting details and capturing local structural information, while refined saliency maps are more adept at describing global shapes. To maximize the complementarity of both salient maps, we integrated them using a weighted combination

$$S_{final} = \kappa \cdot S_{course} + (1 - \kappa) \cdot S_{refine}, \tag{29}$$

where κ is a balance factor for the combination. In the extraction of built-up areas, greater attention was paid to the outer contours of a city, and the fine saliency map is more applicable, so κ was set as 0.2, κ is between 0 and 1.

2.8. Bulit-Up Area Extraction

In the final saliency map, S_{final}, built-up areas usually have the highest values, ground objects similar to built-up areas have the next highest values, and other ground objects have very low values. As such, the final saliency map can also be broadly segmented into three parts based on the gray value. To extract accurate built-up areas, an appropriate threshold segmentation image needs to be set. In [72], the genetic algorithm was used to determine the optimal segmentation threshold and achieved good results. In our paper, a multi-threshold segmentation algorithm, FODPSO [47], was employed. Following segmentation, the highest value part is the binary map of the built-up areas. The pseudo-code for FODPSO is presented in Table 1.

Table 1. Pseudo-code for the fractional-order Darwinian particle swarm optimization (FODPSO) algorithm.

Start
Set Initial parameters v_n [0], x_n [0], χ_{1n} [0], χ_{2n} [0]
// v_n is position parameter, x_n is velocity parameter, χ_{1n} is local best, χ_{2n} is global best
for i = 1:1:Max. Number of the iteration
Generated swarm matrix
for n = 1:1:Number of swarm matrix row
Calculate fitness function of each row
end
Obtain min. fitness function's parameter configuration
If min. fitness function(i)<min. fitness function ($i-1$)
Update $\chi_{1n}^i[t]$, $\chi_{2n}^i[t]$
Update $v_n^i[t+1]$, $x_n^i[t+1]$
else
Kill all swarm matrix member
Go to "generated swarm matrix"
end
end
end

3. Experimental Results

To date, there has been little investigation into saliency detection in remote sensing images; thus, there are no classic testing datasets with existing ground truth (GT) that may be consulted when categories of Sentinel-2 images are introduced to evaluate the effectiveness and novelty of the proposed method. The GT map is obtained by manual segmentation based on the definition of the built-up areas in the first section. Since the built-up areas extraction is greatly affected by the surrounding land cover, based on the ecosystems, the experimental cities are divided into five types: desert, coastal, riverside, valley, and plain cities (Figure 3). Desert cities are distributed in the northwest of China, surrounded by desert and loess, with little vegetation. These conditions have a significant impact on the built-up area extraction. Coastal cities are located in the eastern coastal areas of China, and are the most economically active areas, requiring more timely monitoring of built-up areas. The riverside cities are distributed along the banks of the Yangtze and Yellow Rivers. As dark ground object, water bodies are likely to be detected as saliency targets by the saliency model, which may affect the built-up area extraction by the saliency method. Valley cities are located in western China, and are often affected by bare rock, rendering built-up areas extraction extremely challenging. Plain cities are distributed in eastern China, where the landform of the built-up areas is plain, and surrounded by significant amounts of farmland (including bare land), which may have an impact on the extraction of built-up areas. Details of the study areas are presented in Table 2.
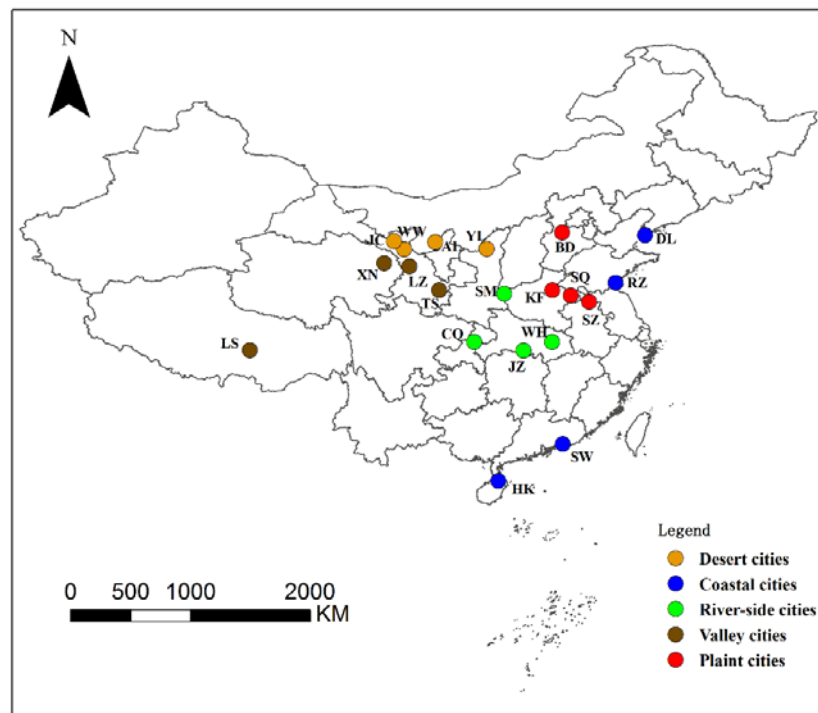
Figure 3. Study areas: Alxa (AL), Jinchang (JC), Wuwei (WW), Yulin (YL), Dalian (DL), Haikou (HK), Rizhao (RZ), Shanwei (SW), Chongqing (CQ), Jingzhou (JZ), Sanmenxia (SM), Wuhan (WH), Lanzhou (LZ), Lhasa (LS), Tianshui (TS), Xining (XN), Baoding (BD), Kaifeng (KF), Shangqiu (SQ), and Suzhou (SZ).

Table 2. Study areas.

City Type	City	Size (pixels)	Date
Desert cities	Alxa (Inner Mongolia)	1000 × 900	11/01/2016
	Jinchang (Gansu)	858 × 858	11/24/2016
	Wuwei (Gansu)	1000 × 1000	11/04/2016
	Yulin (Shanxi)	1500 × 1500	04/01/2016
Coastal cities	Dalian (Liaoning)	1300 × 1300	11/21/2016
	Haikou (Hainan)	1250 × 1250	12/20/2016
	Rizhao (Shandong)	1000 × 1000	11/03/2016
	Shanwei (Guangdong)	500 × 500	12/13/2016
Riverside cities	Chongqing	1650 × 1650	04/14/2017
	Jinzhou (Hubei)	1250 × 1250	08/01/2016
	Sanmenxia (Henan)	1000 × 1000	09/23/2016
	Wuhan (Hubei)	1800 × 1200	08/28/2016
Valley cities	Lanzhou (Gansu)	1750 × 900	12/01/2016
	Lhasa (Tibet)	1500 × 1500	10/24/2016
	Tianshui (Gansu)	1020 × 1020	06/06/2017
	Xining (Qinghai)	1750 × 1750	11/04/2016
Plain cities	Baoding (Hebei)	1400 × 1400	09/01/2016
	Kaifeng (Henan)	1200 × 1200	08/23/2016
	Shangqiu (Henan)	1250 × 1250	08/28/2016
	Suzhou (Anhui)	1200 × 1200	08/28/2016

Based on the OIF and J-M methods, the selection of the optimal band combination for different images is shown in Table 3. From Table 3, the optimal band combination for most cities consists of bands 12, 11, and 7, while the optimal band combination for most valley cities is composed of bands 12, 11, and 5.

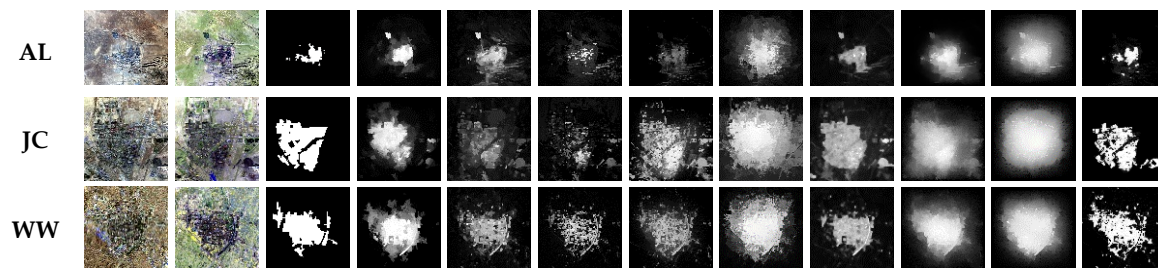**Table 3.** Optimal band combination.

| Desert City	Cities	Alxa	Jinchang	Wuwei	Yulin
	Optimal	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 5	Bands 12, 11, 7
Coastal City	Cities	Dalian	Haikou	Shanwei	Rizhao
	Optimal	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 5
Riverside Cities	Cities	Chongqing	Jingzhou	Sanmenxia	Wuhan
	Optimal	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 7
Valley Cities	Cities	Lanzhou	Lhasa	Tianshui	Xining
	Optimal	Bands 12, 11, 5	Bands 12, 11, 5	Bands 12, 11, 5	Bands 12, 11, 7
Plain Cities	Cities	Baoding	Kaifeng	Shangqing	Suzhou
	Optimal	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 7	Bands 12, 11, 7

To evaluate the performance of the proposed method with respect to saliency detection, we compared it to the eight most recent saliency detection methods: dense and sparse reconstruction (DSR) [73], discriminative regional feature integration (DRFI) [41], regional principal color (RPC) [74], diffusion-based compactness and local contrast (DCLC) [64], inner and inter label propagation (LPS) [75], bootstrap learning (BL) [44], diffusion process on a two-layer sparse graph (DPTLSG) [76], and reversion correction and regularized random walks ranking (RCRR) [77]. To ensure fairness, all methods used the optimal band combination as the original image for saliency detection. To further evaluate the extraction accuracy of the proposed method, the saliency map was segmented to obtain the binary map of the built-up areas based on FODPSO; the results were compared to some built-up area extraction methods. Due to index-based methods being sensitive to the built-up areas [72], they are widely used in built-up areas extraction. In this study, two index-based methods, NDBI [7] and NBI [5], were selected. PanTex [11] is a method for extracting built-up areas based on texture feature, and has been evaluated by many experiments [78], so the method was also selected.

3.1. Comparison to the State-of-the-Art Saliency Methods

3.1.1. Qualitative Experiment

The results of saliency maps generated by nine different methods are presented for qualitative comparison in Figure 4. It is clear that our method efficiently detects built-up areas and identifies the contours of built-up areas most accurately, while the results of the other eight methods are inferior. For riverside cities, several of the methods, including DSR, DRFI, RPC, BL, and RCRR, identify water bodies as salient objects rather than built-up areas. Of these methods, DCLC and DPTLSG have superior performance, but still fall short in terms of accuracy. The BL method also produces satisfactory saliency maps, but its performance is generally poor in coastal and riverside cities. DSR, DRFI, and RPC fail to highlight the built-up areas in their entireties. LPS focuses too much on an image's central information, overlooking the information on built-up areas throughout the image. The RCRR method detects and highlights unnecessary and irrelevant background information.

Figure 4. *Cont.*
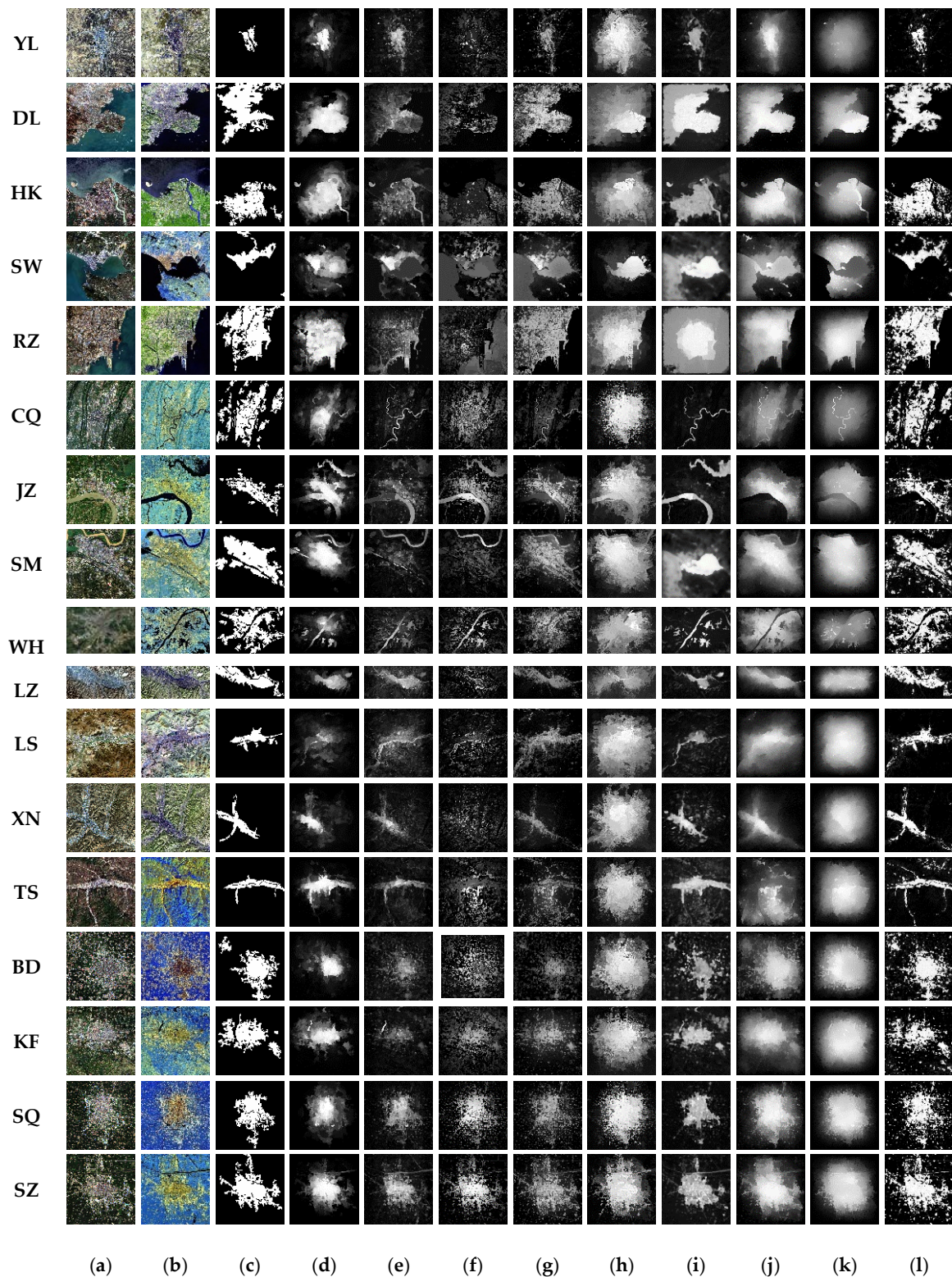
Figure 4. Saliency maps produced by our proposed model and eight competing models. (**a**) Original images; (**b**) optimal band combinations; (**c**) ground truth; (**d**) dense and sparse reconstruction (DSR); (**e**) discriminative regional feature integration (DRFI); (**f**) regional principal color (RPC); (**g**) diffusion-based compactness and local contrast (DCLC); (**h**) inner and inter label propagation (LPS); (**i**) bootstrap learning (BL); (**j**) diffusion process on a two-layer sparse graph (DPTLSG); (**k**) reversion correction and regularized random walks ranking (RCRR); (**l**) Ours.

3.1.2. Quantitative Experiment

To quantitatively evaluate the performance of each saliency method, the receiver operating characteristic (ROC)-area under the curve (AUC) metric, precision, recall, and F-measure, and time comparison were used.

ROC-AUC Metric

The receiver operating characteristic (ROC) curve is derived by thresholding a saliency map at the threshold within the range [0, 255], and further classifying the saliency map into the saliency objects and the background [79]. The ROC graph is generated by plotting the true positive rate (on the *y*-axis) against the false positive rate (on the *x*-axis). The true positive and false positive rates are expressed as

$$TPR = \frac{TP}{TP + FN}, \tag{30}$$

$$FPR = \frac{FP}{FP + TN}, \tag{31}$$

where *TPR* is the true positive rate, *FPR* is the false positive rate, *TP* is the true positives and it is the number of correctly identified built-up areas, *FN* is the false negatives and it is the number of incorrectly rejected, FP is false positives and it is the number of the incorrectly identified, and *TN* is the true negatives and it is the number of the correctly rejected. When the *FPR* value is the same, the higher the *TPR*, the better a method's performance; the larger the area under the curve (AUC), the better the performance. The AUC values for the different methods are presented in Table 4, from which it may be seen that the proposed method has the highest AUC value. The ROC curves for the different methods are shown in Figure 5a. We can conclude that the ROC curve generated by our model demonstrates superior performance.

Table 4. The area under the curve (AUC) for different methods.

Method	DSR	DRFI	RPC	DCLC	LPS	BL	DPTLSG	RCRR	Ours
AUC	0.8890	0.8788	0.7521	0.9146	0.8485	0.8366	0.9167	0.8405	0.9687

Precision, Recall, and F-Measure

To further evaluate the quality of the saliency maps, precision, recall, and F-measure were employed. They can be computed by

$$precision = \frac{\sum_x \sum_y t(x, y) s(x, y)}{\sum_x \sum_y s(x, y)}, \tag{32}$$

$$recall = \frac{\sum_x \sum_y t(x, y) s(x, y)}{\sum_x \sum_y t(x, y)}, \tag{33}$$

$$F_n = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}, \tag{34}$$

where the term (x, y) denotes the coordinates of the images, $t(x, y)$ is the ground truth, and $s(x, y)$ is the binary image after the thresholding saliency maps. The threshold was set to twice the average gray value; the segmented objects whose average gray value are greater than the threshold was designated foreground, with all others designated background. High recall means that a model returned most of the built-up areas, whereas high precision means that a model returned substantially more built-up areas than background regions. The F-measure is the harmonic mean of precision and recall, and β^2 was set as 1 to balance the importance of precision and recall.

Figure 5b shows the precision, recall, and F-measure values of the evaluated methods. According to Figure 5b, our method achieves the highest precision, which means that the redundant background in the residential areas acquired by our method is the smallest among the nine methods. The recall of our method is the highest among the nine methods, which means that our methods can complete built-up areas. The built-up areas extracted by the RPC and BL methods are greatly affected by water bodies and have a low recall and precision. LPS focuses too much on an image's central information, so it also has a low recall and precision. DCLC, DPTLSG have good accuracy and recall rate, and can also identify the built-up areas well. Although RCRR has a good recall rate, it extracts more background information, so the accuracy is lower. Overall, the proposed method performs well against the state-of-the-art methods.
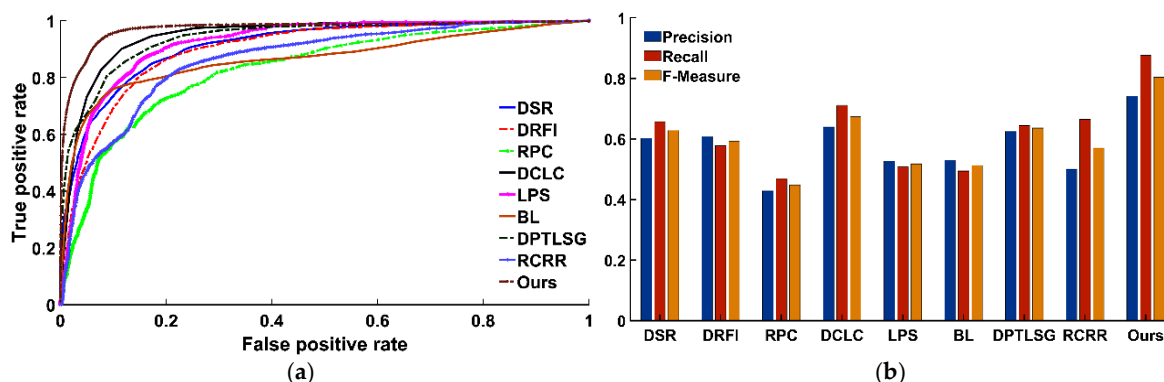
Figure 5. Quantitative evaluation results of different methods: (**a**) ROC curves of different methods on Sentinel-2 images (**b**) Precision, recall, and F-measure of different methods on Sentinel-2 images.

Time Comparison

We compared the computational time for each method using MATLAB on a PC with 8 G RAM, Intel Core i5-4590 CPU @ 3.30 GHz. The average time comparison of our method and the other competing methods is given in Table 5. As can be seen from Table 5, our method is very time-consuming, mainly due to multiscale segmentation and an ensemble learning strategy.

Table 5. Running time comparisons for nine methods.

Method	DSR	DRFI	RPC	DCLC	LPS	BL	DPTLSG	RCRR	Ours
Time(s)	178.33	1547.65	514.23	325.42	137.92	135.24	1583.00	1363.15	1226.22

3.1.3. Important Parameter Settings

There are two important parameters in the proposed method: M, layers of segmented objects; the parameter ϑ used to calculate the thresholds T_h and T_l. To determine the optimal parameter M, we compared the accuracy and time of the different M. The results are shown in Table 6. From Table 6, as M increases, the accuracy increases, but the calculation time also rises sharply. If M continues to increase, the computational time cannot be accepted.

Table 6. Comparison results of different M.

M	$M = 3$ ($N = 1000, 1500, 2000$)	$M = 5$ ($N = 1000, 1500, 2000, 2500, 3000$)	$M = 7$ ($N = 1000, 1500, 2000, 2500, 3000, 3500, 4000$)
AUC	0.9645	0.9664	0.9687
F-Measure	0.7853	0.7868	0.8038
Time(s)	259.60	626.42	1226.22

To determine the optimal parameter ϑ, we compared the accuracy of different ϑ. The results are shown in Table 7. From Table 7, we can see that when ϑ is 1.8, the accuracy is the highest. Although ϑ = 1.8 was determined to be the optimal ϑ, it was not suitable for images of Xining and Yulin. For the images of Xining and Yulin, ϑ was set to 2.8.

Table 7. Comparison results of different ϑ.

ϑ	1.2	1.4	1.6	1.8	2.0
AUC	0.9534	0.9581	0.9592	0.9597	0.9487
F-Measure	0.7147	0.7212	0.7299	0.7378	0.7148

3.2. Comparison to the State-of-the-Art Built-Up Areas Extraction Methods

To evaluate the overall performance of the proposed method in extracting built-up areas, we further compared our method to some advanced built-up areas extraction methods, including two index-based methods [5,7] and one texture-based method [11]. They were calculated according to Equations (35)–(37). To make these equations available, we used ESA's Sen2Cor atmospheric correction module to process Sentinel-2 Level-1C images into Level-2A bottom of atmosphere (BOA) reflectance images [50].

$$NDBI = \frac{b_{swir} - b_{nir}}{b_{swir} - b_{nir}},\tag{35}$$

$$NBI = \frac{b_{red} \cdot b_{swir}}{b_{bir}},\tag{36}$$

$$f(built\_up) = \wedge\{tx_1, tx_i, tx_n\},\ i \in [1,2,,n],\tag{37}$$

where b_{swir} is the reflectance of the SWIR band, b_{nir} is the reflectance of the NIR band, b_{red} is the reflectance of the red band, $tx_i = f(w = 9, v_i, m = CON)$, $i \in [\alpha_1, d_1; a_2, d_2; \ldots; a_n, d_n]$. w = window size, α and d are the distance and angle defining the displacement vector v required to select the pairs producing the co-occurrence matrix, and m is the textural measure applied to the given co-occurrence matrix distribution. $CON = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i-j)^2 \cdot P_{ij}$, with N_g is the number of gray levels present in the image, and $P_{i,j}$ is the (i,j)th entry of the co-occurrence matrix.

Figure 6 presents the results achieved by the different methods. The binary maps shown in Figure 6f,g,h were obtained by automatically determining the segmentation threshold using Otsu's algorithm to segment maps in Figure 6b,c,d. The binary maps shown in Figure 6j were obtained by segmenting the maps in Figure 6f. based on the optimal thresholds determined by FODPSO algorithm. Since the built-up areas are usually large, we keep the area larger than the T_{area} and remove the area smaller than the T_{area}, where the T_{area} is empirically set to 3000 pixels. From Figure 6, it is clear that the two index-based methods perform poorly on the images of the desert cities and the valley cities, they are almost impossible to use to identify and extract the built-up areas, while the desert and bare rock are clearly extracted. However, they perform very well on the images of coastal cities, riverside cities, and plain cities, because of the high vegetation coverage or large water area in these cities. The PanTex method performs better than both index-based methods in detecting built-up areas, with the locations clearly identifiable. However, the PanTex method only utilizes texture features; some areas with texture features similar to built-up areas may also be extracted. For example, in desert cities, they are surrounded by a large number of loess and desert areas, which have similar texture features to the built-up areas and are, therefore, incorrectly extracted. Although the land cover around cities varies, our proposed method can still efficiently identify the locations and boundaries of built-up areas, and can accurately extract them. To quantitatively evaluate the various methods, three statistical measures were used: overall accuracy, commission error, and omission error. The commission error represents the percentage of pixels that belong to non-built-up areas but have been classified as built-up areas. Omission error represents the percentage of pixels that belong to built-up areas, but have been classified as non-built-up areas. Table 8 shows the average statistical measurement results for five

different types of cities. The overall accuracies of our proposed method in five different types of cities are higher than the other three methods, and the commission and omission errors of the proposed method are the lowest among the four methods. NDBI and NBI have low overall accuracy and high commission errors and omission error on images in desert and valley cities. This suggests that these two index-based methods are not suitable for extracting built-up areas surrounded by bare rock and desert. However, they perform well on images in the other three types of cities and achieve high overall accuracy. PanTex performs very well, second only to the proposed method. Its omission error is also relatively low, while the commission error is high in desert cities. This indicates that when PanTex extracted the built-up areas of the desert city, a large amount of non-built-up areas are incorrectly extracted. In summary, the proposed method takes into account the different features of the built-up areas, based on visual salience, and can achieve good results in different types of cities.
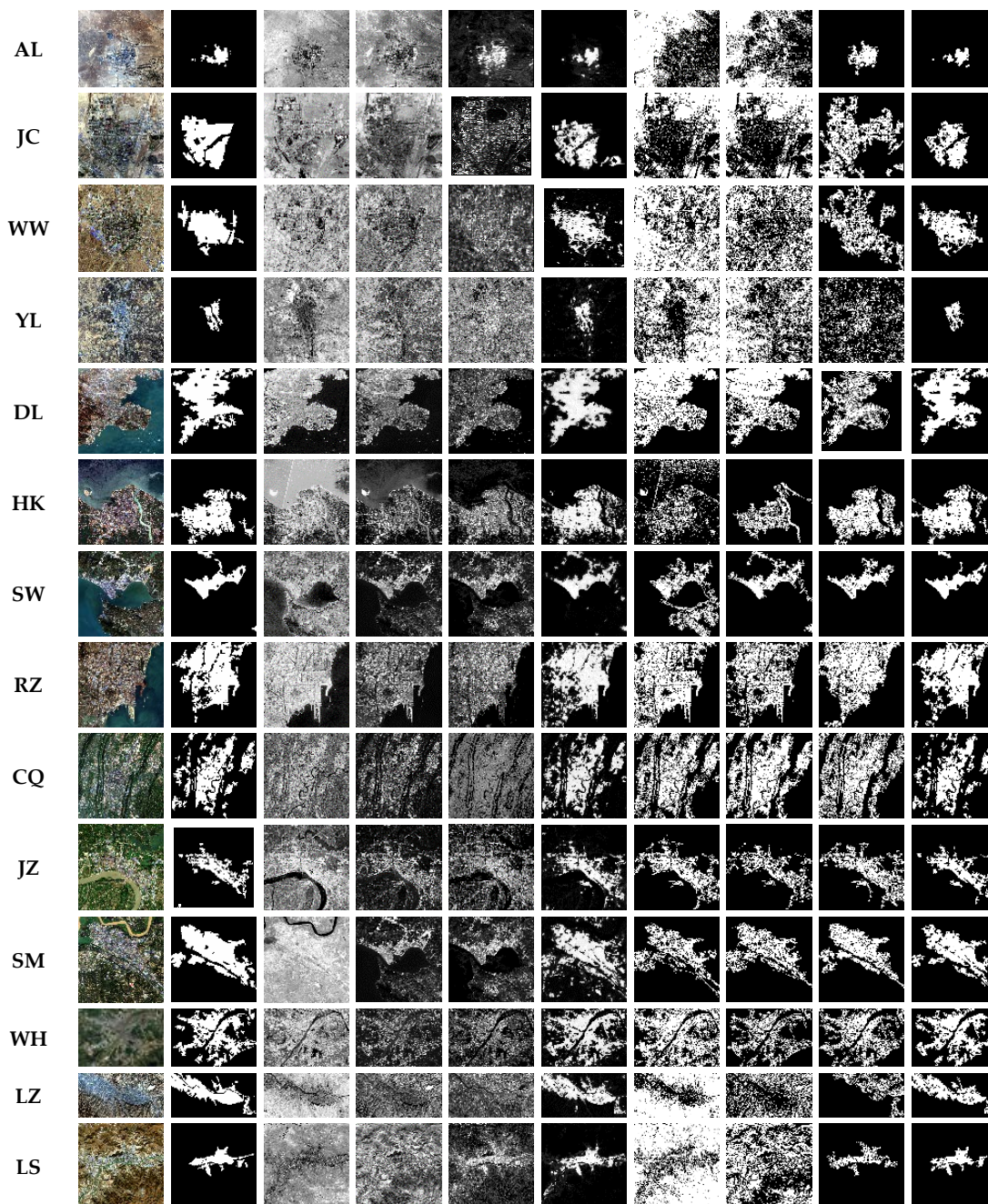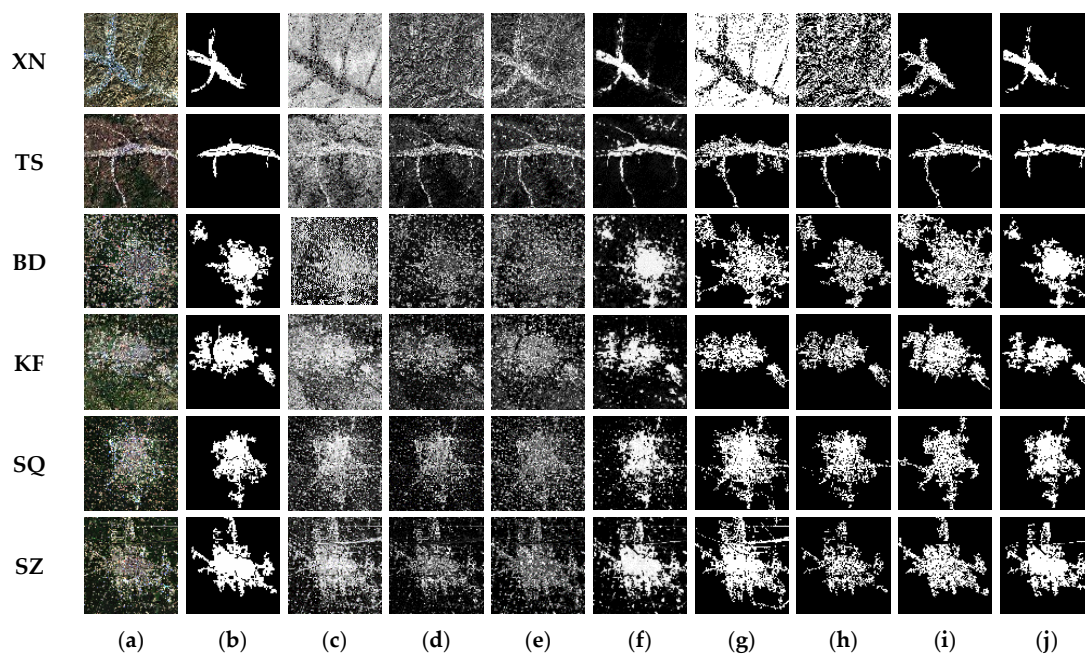

Figure 6. *Cont.*

Figure 6. Comparison of the results of the four methods. (**a**) RGB images, (**b**) ground truth, (**c**) NDBI maps, (**d**) NBI maps, (**e**) PanTex maps, (**f**) our saliency maps, (**g**) built-up areas maps (NDBI), (**h**) built-up areas maps (NBI), (**i**) built-up areas maps (PanTex), (**j**) built-up areas maps (proposed method).

Table 8. Accuracy assessment of the resultant images.

	NDBI			NBI			PanTex			Ours		
	OA	**Co**	**Om**	**OA**	**Co**	**Om**	**OA**	**Co**	**Om**	**OA**	**Co**	**Om**
Desert	36.40	90.85	65.66	42.52	90.48	64.32	81.27	65.18	33.87	95.56	11.25	26.26
Costal	78.22	48.27	30.59	86.06	27.66	27.18	87.88	36.66	36.39	95.13	8.04	11.17
Riverside	80.96	38.64	24.07	84.51	29.34	28.92	81.57	37.21	29.03	93.35	10.49	17.84
Valley	48.62	83.01	48.86	64.62	71.72	43.35	87.81	29.54	25.91	96.37	11.68	18.79
Plain	88.43	30.87	16.87	88.56	21.61	40.45	87.80	30.27	22.26	94.77	8.72	17.97

OA = overall accuracy, Co. = commission error, Om = omission error.

4. Discussion

In this paper, a new method for extracting built-up areas from images based on the principles of salient object detection is proposed. Compared to other saliency detection methods, the unique band information of remote sensing is fully exploited. For example, the optimal band combination composed of unique bands of satellite images can highlight the built-up areas very effectively. Water bodies in the SWIR band can be removed to prevent them from being detected as salient objects. Compared to existing built-up area extraction methods using optical images, our method devotes greater attention to the most conspicuous built-up areas in the image, and not only the spectral and textural information of the built-up areas. Therefore, it is more robust and not easily distorted by surrounding ground objects, such as bare rock or desert. Compared to the BL saliency method, we consider a greater number of cues and introduce the GWB model to improve the detection accuracy of the coarse saliency map, thus providing a more reliable training sample for training strong models. CCA is employed to integrate multiscale detection results, as distinct from the simple superposition effect using the BL saliency method, and it can improve the ultimate detection accuracy and reduce the background information. In addition, shape information of ground objects is also utilized. For the multiscale segmentation strategy, we do not set different SLIC parameters for segmenting the image multiple times but, rather, adopt a fixed segmentation parameter, then merge the superpixels based on the various merge parameters to obtain multiscale segmentation images, improving segmentation and algorithm efficiency.

Although our method is precise and robust, it does have some shortcomings and limitations that cannot be ignored. First, it employs multiscale segmentation and an ensemble learning strategy that affects its processing efficiency, resulting in a computational time that is several times as long as some other methods. However, we can select fewer scales if the requirements can be met and the calculation time is shortened. Second, our method also incorrectly detects some non-built-up areas. For example, in the Lhasa image, the river bank is detected. The third limitation concerns the detection accuracy of the coarse saliency map, which affects the reliability of the training sample and the final result. The detection accuracy of the coarse saliency map depends on the detection method and the input image. Overall, the proposed method has the potential to extract built-up areas in different types of cities with adequate accuracy.

5. Conclusions

This paper proposes a new built-up area extraction method based on an improved BL saliency model. First, the band combinations that highlight the built-up areas are explored. Then, we produce coarse saliency maps based on multiple cues, and the GWB model to generate training samples for a strong classification model that is subsequently used to produce a refined saliency map. To further improve detection performance, multiscale saliency maps are integrated by CCA. The final saliency result combines the coarse and refined saliency maps. Finally, the information pertaining to built-up areas is extracted using the FODPSO algorithm. Comparative experimentation using other advanced saliency detection methods indicates that our method outperforms the other eight models in extracting built-up areas from various complex background environments. Comparative analyses with three advanced built-up area extraction methods confirm the superior performance of our proposed method. Therefore, the proposed mothed not only has good precision and robustness, but also has practical value in the extraction of built-up areas.

Future research will focus primarily on three aspects: First, we intend to optimize our method further, shortening the computational duration of the process. Second, we will take more features into consideration to avoid the extraction of other ground objects. Third, we will determine the optimal band combinations of other satellites for extracting built-up areas, ultimately extending our method for use with more satellites.

Author Contributions: Z.S. had the original idea for the study. Q.M. supervised the research and contributed to the article's organization. W.Z. analyzed the data. Z.S. performed the experiments and drafted the manuscript, which was revised by all of the authors. All of the authors read and approved the submitted manuscript.

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

References

1. Taubenböck, H.; Esch, T.; Felbier, A.; Wiesner, M.; Roth, A.; Dech, S. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.* **2012**, *117*, 162–176. [CrossRef]
2. Yu, S.; Sun, Z.; Guo, H.; Zhao, X.; Sun, L.; Wu, M. Monitoring and analyzing the spatial dynamics and patterns of megacities along the maritime silk road. *J. Remote Sens.* **2017**, *21*, 169–181.
3. Sun, Z.; Guo, H.; Li, X.; Lu, L.; Du, X. Estimating urban impervious surfaces from landsat-5 tm imagery using multilayer perceptron neural network and support vector machine. *J. Appl. Remote Sens.* **2011**, *5*, 053501. [CrossRef]
4. Deng, C.; Wu, C. Bci: A biophysical composition index for remote sensing of urban environments. *Remote Sens. Environ.* **2012**, *127*, 247–259. [CrossRef]

5. Jieli, C.; Manchun, L.; Yongxue, L.; Chenglei, S.; Wei, H.U. Extract residential areas automatically by new built-up index. In Proceedings of the 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010.

6. Xu, H. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 4269–4276. [CrossRef]

7. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]

8. Sun, G.; Chen, X.; Jia, X.; Yao, Y.; Wang, Z. Combinational build-up index (cbi) for effective impervious surface mapping in urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2081–2092. [CrossRef]

9. Zhang, P.; Sun, Q.; Liu, M.; Li, J.; Sun, D. A strategy of rapid extraction of built-up area using multi-seasonal landsat-8 thermal infrared band 10 images. *Remote Sens.* **2017**, *9*, 1126. [CrossRef]

10. Shao, Z.; Tian, Y.; Shen, X. Basi: A new index to extract built-up areas from high-resolution remote sensing images by visual attention model. *Remote Sens. Lett.* **2014**, *5*, 305–314. [CrossRef]

11. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [CrossRef]

12. Wentz, E.A.; Stefanov, W.L.; Gries, C.; Hope, D. Land use and land cover mapping from diverse data sources for an arid urban environments. *Comput. Environ. Urban Syst.* **2006**, *30*, 320–346. [CrossRef]

13. Leinenkugel, P.; Esch, T.; Kuenzer, C. Settlement detection and impervious surface estimation in the mekong delta using optical and sar remote sensing data. *Remote Sens. Environ.* **2011**, *115*, 3007–3019. [CrossRef]

14. Zhu, Z.; Woodcock, C.E.; Rogan, J.; Kellndorfer, J. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using landsat and sar data. *Remote Sens. Environ.* **2012**, *117*, 72–82. [CrossRef]

15. Zhang, J.; Li, P.; Wang, J. Urban built-up area extraction from landsat tm/etm+ images using spectral information and multivariate texture. *Remote Sens.* **2014**, *6*, 7339–7359. [CrossRef]

16. Borji, A.; Cheng, M.; Jiang, H.; Li, J. Salient object detection: A benchmark. *IEEE Trans. Imag. Process.* **2015**, *24*, 5706–5722. [CrossRef] [PubMed]

17. Dong, C.; Liu, J.; Xu, F. Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor. *Remote Sens.* **2018**, *10*, 400. [CrossRef]

18. Zhang, Y.; Wang, X.; Xie, X.; Li, Y. Salient object detection via recursive sparse representation. *Remote Sens.* **2018**, *10*, 652. [CrossRef]

19. Zhang, L.; Li, A.; Zhang, Z.; Yang, K. Global and local saliency analysis for the extraction of residential areas in high-spatial-resolution remote sensing image. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3750–3763. [CrossRef]

20. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]

21. Harel, J.; Koch, C.; Perona, P. Graph-based visual saliency. In *Advances in Neural Information Processing Systems*; Touretzky, D.S., Mozer, M.C., Hasselmo, M.E., Eds.; MIT Press: Cambridge, MA, USA, 2007.

22. Ma, Y.; Zhang, H. Contrast-based image attention analysis by using fuzzy growing. In Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, CA, USA, 2–8 November 2003.

23. Gao, D.; Mahadevan, V.; Vasconcelos, N. The discriminant center-surround hypothesis for bottom-up saliency. In *Advances in Neural Information Processing Systems*; Platt, J.C., Koller, D., Singer, Y., Roweis, S.T., Eds.; Curran Associates Icn: Vancouver, BC, Canada, 2008.

24. Gao, D.; Vasconcelos, N. Bottom-up saliency is a discriminant process. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.

25. Cheng, M.; Mitra, N.J.; Huang, X.; Torr, P.H.; Hu, S.-M.; Intelligence, M. Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 569–582. [CrossRef] [PubMed]

26. Perazzi, F.; Krähenbühl, P.; Pritch, Y.; Hornung, A. Saliency filters: Contrast based filtering for salient region detection. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.

27. Shi, K.; Wang, K.; Lu, J.; Lin, L. Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 25–27 June 2013.

28. Gopalakrishnan, V.; Hu, Y.; Rajan, D. Random walks on graphs for salient object detection in images. *IEEE Trans. Imag. Process.* **2010**, *19*, 3232–3242. [CrossRef] [PubMed]

29. Wei, Y.; Wen, F.; Zhu, W.; Sun, J. Geodesic saliency using background priors. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012.

30. Jiang, B.; Zhang, L.; Lu, H.; Yang, C.; Yang, M.-H. Saliency detection via absorbing markov chain. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 3–6 December 2013.

31. Yan, Q.; Xu, L.; Shi, J.; Jia, J. Hierarchical saliency detection. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Oregon, Portland, 25–27 June 2013.

32. Qin, Y.; Lu, H.; Xu, Y.; Wang, H. Saliency detection via cellular automata. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015.

33. Bruce, N.; Tsotsos, J. Saliency based on information maximization. In *Advances in Neural Information Processing Systems*; Jordan, M.I., LeCun, Y., Solla, S.A., Eds.; MIT Press: Cambridge, MA, USA, 2006; pp. 155–162.

34. Zhang, L.; Tong, M.H.; Marks, T.K.; Shan, H.; Cottrell, G.W. Sun: A bayesian framework for saliency using natural statistics. *J. Vis.* **2008**, *8*, 32. [CrossRef] [PubMed]

35. Shen, X.; Wu, Y. A unified approach to salient object detection via low rank matrix recovery. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 18–20 June 2012.

36. Borji, A.; Itti, L. Exploiting local and global patch rarities for saliency detection. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 18–20 June 2012.

37. Wang, Q.; Zheng, W.; Piramuthu, R. Grab: Visual saliency via novel graph model and background priors. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.

38. Zhu, W.; Liang, S.; Wei, Y.; Sun, J. Saliency optimization from robust background detection. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014.

39. Peng, H.; Li, B.; Ji, R.; Hu, W.; Xiong, W.; Lang, C. Salient object detection via low-rank and structured sparse matrix decomposition. *IEEE Trans. Patt. Anal. Mach. Intell.* **2013**, *39*, 796–802.

40. Lang, C.; Liu, G.; Yu, J.; Yan, S. Saliency detection by multitask sparsity pursuit. *IEEE Trans. Imag. Process.* **2012**, *21*, 1327–1338. [CrossRef] [PubMed]

41. Jiang, H.; Wang, J.; Yuan, Z.; Wu, Y.; Zheng, N.; Li, S. Salient object detection: A discriminative regional feature integration approach. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Oregon, Portland, 25–27 June 2013.

42. Yang, J.; Yang, M. Top-down visual saliency via joint crf and dictionary learning. *IEEE Trans. Patt. Anal. Mach. Intell.* **2017**, *39*, 576–588. [CrossRef] [PubMed]

43. Cholakkal, H.; Rajan, D.; Johnson, J. Top-Down Saliency with Locality-Constrained Contextual Sparse Coding. Available online: http://www.bmva.org/bmvc/2015/papers/paper159/paper159.pdf (accessed on 4 August 2018).

44. Tong, N.; Lu, H.; Ruan, X.; Yang, M.-H. Salient object detection via bootstrap learning. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015.

45. Wang, X.; Ma, H.; Chen, X. Geodesic weighted bayesian model for salient object detection. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 397–401.

46. Qin, Y.; Feng, M.; Lu, H.; Cottrell, G.W. Hierarchical cellular automata for visual saliency. *Int. J. Comput. Vis.* **2018**, 1–20. [CrossRef]

47. Couceiro, M.; Ghamisi, P. *Fractional Order Darwinian Particle Swarm Optimization: Applications and Evaluation of an Evolutionary Algorithm*; Springer: Berlin, Germany, 2015.

48. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

49. Vuolo, F.; Żółtak, M.; Pipitone, C.; Zappa, L.; Wenng, H.; Immitzer, M.; Weiss, M.; Baret, F.; Atzberger, C. Data service platform for sentinel-2 surface reflectance and value-added products: System use and examples. *Remote Sens.* **2016**, *8*, 938. [CrossRef]

50. Mueller-Wilm, U. Sentinel-2 msi—Level-2a Prototype Processor Installation and User Manual. Available online: http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf (accessed on 6 July 2018).

51. Park, H.; Choi, J.; Park, N.; Choi, S. Sharpening the vnir and swir bands of sentinel-2a imagery through modified selected and synthesized band schemes. *Remote Sens.* **2017**, *9*, 1080. [CrossRef]

52. Valdiviezo-N, J.C.; Téllez-Quiñones, A.; Salazar-Garibay, A.; López-Caloca, A.A. Built-up index methods and their applications for urban extraction from sentinel 2a satellite data: Discussion. *J. Opt. Soc. Am. A* **2018**, *35*, 35–44. [CrossRef] [PubMed]

53. Pesaresi, M.; Corbane, C.; Julea, A.; Florczyk, A.J.; Syrris, V.; Soille, P.; Sensing, R. Assessment of the added-value of sentinel-2 for detecting built-up areas. *Remote Sens.* **2016**, *8*, 299. [CrossRef]

54. Chavez, P.; Berlin, G.L.; Sowers, L.B. Statistical method for selecting landsat mss ratios. *J. Appl. Photogr. Eng.* **1982**, *8*, 23–30.

55. Richards, J.A.; Richards, J. *Remote Sensing Digital Image Analysis*; Springer: Berlin, Germany, 1999.

56. Swain, P.H.; Davis, S.M. Remote sensing: The quantitative approach. *IEEE Trans. Patt. Anal. Mach. Intell.* **1981**, 713–714. [CrossRef]

57. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Patt. Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

58. Moya, M.M.; Koch, M.W.; Perkins, D.N.; West, R.D.D. Superpixel segmentation using multiple sar image products. In *Radar Sensor Technology XVIII*; International Society for Optics and Photonics: Bellingham, WA, USA, 2014; p. 90770R.

59. Hu, Z.; Wu, Z.; Zhang, Q.; Fan, Q.; Xu, J. A spatially-constrained color–texture model for hierarchical vhr image segmentation. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 120–124. [CrossRef]

60. Connolly, C.; Fleiss, T. A study of efficiency and accuracy in the transformation from rgb to cielab color space. *IEEE Trans. Imag. Process.* **1997**, *6*, 1046–1048. [CrossRef] [PubMed]

61. Hu, P.; Wang, W.; Zhang, C.; Lu, K. Detecting salient objects via color and texture compactness hypotheses. *IEEE Trans. Imag. Process.* **2016**, *25*, 4653–4664. [CrossRef] [PubMed]

62. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Patt. Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]

63. Jia, S.; Deng, B.; Zhu, J.; Jia, X.; Li, Q. Local binary pattern-based hyperspectral image classification with superpixel guidance. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 749–759. [CrossRef]

64. Zhou, L.; Yang, Z.; Yuan, Q.; Zhou, Z.; Hu, D. Salient region detection via integrating diffusion-based compactness and local contrast. *IEEE Trans. Imag. Process.* **2015**, *24*, 3308–3320. [CrossRef] [PubMed]

65. Zhou, D.; Weston, J.; Gretton, A.; Bousquet, O.; Schölkopf, B. Ranking on data manifolds. In *Advances in Neural Information Processing Systems*; Mit Press: Cambridge, MA, USA, 2004; pp. 169–176.

66. Qiao, C.; Wang, J.; Shang, J.; Daneshfar, B. Spatial relationship-assisted classification from high-resolution remote sensing imagery. *Int. J. Dig. Earth* **2015**, *8*, 710–726. [CrossRef]

67. Xie, Y.; Lu, H.; Yang, M. Bayesian saliency via low and mid level cues. *IEEE Trans. Imag. Process.* **2013**, *22*, 1689–1698.

68. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Patt. Anal. Mach. Intell.* **2004**, *26*, 1124–1137. [CrossRef] [PubMed]

69. Xu, H. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]

70. Lu, H.; Zhang, X.; Qi, J.; Tong, N.; Ruan, X.; Yang, M.-H. Co-bootstrapping saliency. *IEEE Trans. Imag. Process.* **2017**, *26*, 414–425. [CrossRef] [PubMed]

71. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Patt. Anal. Mach. Intell.* **2013**, 1397–1409. [CrossRef] [PubMed]

72. Li, K.; Chen, Y. A genetic algorithm-based urban cluster automatic threshold method by combining viirs dnb, ndvi, and ndbi to monitor urbanization. *Remote Sens.* **2018**, *10*, 277. [CrossRef]

73. Li, X.; Lu, H.; Zhang, L.; Ruan, X.; Yang, M.-H. Saliency detection via dense and sparse reconstruction. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013.

74. Lou, J.; Ren, M.; Wang, H. Regional principal color based saliency detection. *PLoS ONE* **2014**, *9*, e112475. [CrossRef] [PubMed]

75. Li, H.; Lu, H.; Lin, Z.; Shen, X.; Price, B. Inner and inter label propagation: Salient object detection in the wild. *IEEE Trans. Imag. Process.* **2015**, *24*, 3176–3186. [CrossRef] [PubMed]

76. Zhou, L.; Yang, Z.; Zhou, Z.; Hu, D. Salient region detection using diffusion process on a two-layer sparse graph. *IEEE Trans. Imag. Process.* **2017**, *26*, 5882–5894. [CrossRef] [PubMed]

77. Yuan, Y.; Li, C.; Kim, J.; Cai, W.; Feng, D.D. Reversion correction and regularized random walk ranking for saliency detection. *IEEE Trans. Imag. Process.* **2018**, *27*, 1311–1322. [CrossRef] [PubMed]

78. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L. A global human settlement layer from optical hr/vhr rs data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [CrossRef]

79. Zhang, L.; Lv, X.; Liang, X. Saliency analysis via hyperparameter sparse representation and energy distribution optimization for remote sensing images. *Remote Sens.* **2017**, *9*, 636. [CrossRef]