GIScience & Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgrs20

A review of building detection from very high resolution optical remote sensing images

Jiayi Li, Xin Huang, Lilin Tu, Tao Zhang & Leiguang Wang

To cite this article: Jiayi Li, Xin Huang, Lilin Tu, Tao Zhang & Leiguang Wang (2022) A review of building detection from very high resolution optical remote sensing images, GIScience & Remote Sensing, 59:1, 1199-1225, DOI: 10.1080/15481603.2022.2101727

To link to this article: https://doi.org/10.1080/15481603.2022.2101727

GISCIENCE & REMOTE SENSING
2022, VOL. 59, NO. 1, 1199–1225
https://doi.org/10.1080/15481603.2022.2101727

Taylor & Francis
Taylor & Francis Group

A review of building detection from very high resolution optical remote sensing images

Jiayi Li[a], Xin Huang[a,b], Lilin Tu[a], Tao Zhang[c] and Leiguang Wang[d]

[a]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, PR China; [b]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China; [c]Department of survey and design, Tianjin Survey and Design Institute Group Co., Ltd, Tianjin, China; [d]Forestry College of Southwest Forestry University, Institutes of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming, China

ABSTRACT

Building detection from very high resolution (VHR) optical remote sensing images, which is an essential but challenging task in remote sensing, has attracted increased attention in recent years. However, despite the many methods that have been developed, an in-depth review of the recent literature on building extraction from VHR optical images is still lacking. In this article, we present a comprehensive review of the recent advances (since 2000) in this field. In total, we survey and summarize 417 articles in terms of the building detection method, post-processing, and accuracy assessment. The building detection methods are categorized into physical rule based methods, image segmentation based methods, and traditional and advanced machine learning (i.e. deep learning) methods. Furthermore, four promising related research directions of building polygon delineation, building change detection, building type classification, and height retrieval from monocular optical images are also discussed. Overall, building detection from VHR optical images is a popular research topic that has received extensive attention, due to its great significance. It is hoped that this review will help researchers to have a better understanding of this topic, and thus assist them to conduct related work.

ARTICLE HISTORY
Received 21 March 2022
Accepted 10 July 2022

KEYWORDS
Building detection; building extraction; machine learning; data fusion; remote sensing

1. Introduction

Buildings are the most prominent man-made structure and geographical feature in urban areas (Huang and Zhang 2012). Accurate and up-to-date building information plays a vital role in many applications, e.g. urban planning, environmental monitoring, real-estate management, population estimation, and disaster risk evaluation (Krayenhoff et al. 2018; Huang and Wang, 2019). According to Sritarapipat and Takeuchi (2017), in the remote sensing interpretation field, building detection refers to the extraction of individual building parcels from remote sensing imagery. Aerial and satellite very high resolution (VHR) images such as IKONOS, QuickBird, GeoEye, WorldView, Pleiades, Ziyuan-3, and Gaofen-2 can provide us with abundant detail information in the spatial domain (Paci, Chini, and Emery 2009). The increased spatial resolution helps to improve the ability to separate the different objects in urban areas, and allows individual building information extraction. Accordingly, building extraction from VHR images has become a popular topic (Uzar 2017).

Accurate building extraction from VHR images is not an easy task and still remains a challenge, due to the complexity of buildings and their surroundings (Huang and Zhang 2018; Swan et al. 2022). Firstly, buildings have significant differences in size, shape, height, and function, and they also present large variations in high-resolution images caused by the illumination, viewing angle, occlusions, and shadows. Moreover, complicated urban scenes consisting of spectrally similar objects such as roads, bare ground, and parking lots bring difficulties to accurate building extraction (Huang 2011). To deal with these problems, numerous studies have investigated this topic and proposed many methods from different perspectives. Some review studies have also been conducted to describe the work carried out during the processing stages. For example, Mayer (1999) reported and summarized the studies that have used aerial images and were published before the year 2000. Brenner (2005) also reviewed building reconstruction from optical images and light detection and ranging (LiDAR)

CONTACT Xin Huang xhuang@whu.edu.cn
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15481603.2022.2101727

data. Haala and Kada (2010) summarized the building reconstruction techniques based on airborne LiDAR data. However, a comprehensive and in-depth review of the advances in building detection from VHR optical images is still lacking. Thus, in this article, we aim to present a comprehensive review of the recent progress (since 2000) in building detection from VHR optical images. Some other sensor types, such as LiDAR and synthetic aperture radar (SAR), have also been used for building extraction (Zhao, Zhou, and Kuang 2013; Zhou and Gong 2018). However, optical images are still the leading data source, considering the higher cost of LiDAR data collection (Cao et al. 2020) and the great difficulties involved with building interpretation in SAR imagery (Deng et al. 2019), due to its unique imaging principle.

Accordingly, this review mainly focuses on VHR optical imaging systems, where "VHR optical imagery" refers to remote sensing data from the visible to infrared spectrum, with a spatial resolution ranging from 0.05 m to 4 m. The literature search was performed using an online database of peer-reviewed literature. The query was performed with the predefined keywords of "building detection" or "building extraction" in the article title, abstract, and keywords, under the category of "remote sensing" and using a time range from 1 January 2000, to 31 December 2021. A set of strategies was then further used to manually select the papers of interest from the query results. Finally, a literature database was formed, covering 417 articles published in 87 international journals. The BibTeX file containing all 417 indexed articles is included in Supplementary Material I. The details of the online database as well as the strategies used are provided in Supplementary Material II. These articles have mainly been published in the mainstream top journals of remote sensing (Figure 1), indicating that building extraction is a popular research direction for VHR image interpretation. Figure 2 shows the date of publication of these 417 articles, where it can be seen that the number of publications per year has increased exponentially with time. Before 2010, the number of publications per year was less than 10, mainly due to the poor availability of VHR (i.e. meter-level and even higher) remote sensing images. Benefiting from the more and more easily accessible data and the booming development of deep learning, the article amount has increased significantly since 2018, and reached more than 100 in 2021.

According to the retrieved articles, as shown in Figure 2, over the last two decades, a large number of methods have been developed for building detection from aerial and satellite images, which can be categorized into physical rule based methods, image segmentation based methods, and traditional and advanced machine learning (i.e. deep learning) methods. Some works can be regarded as a combination of the above methods, and we counted them in each relevant category in Figure 2. These three categories of physical rule based methods, image segmentation based methods, and traditional and advanced machine learning

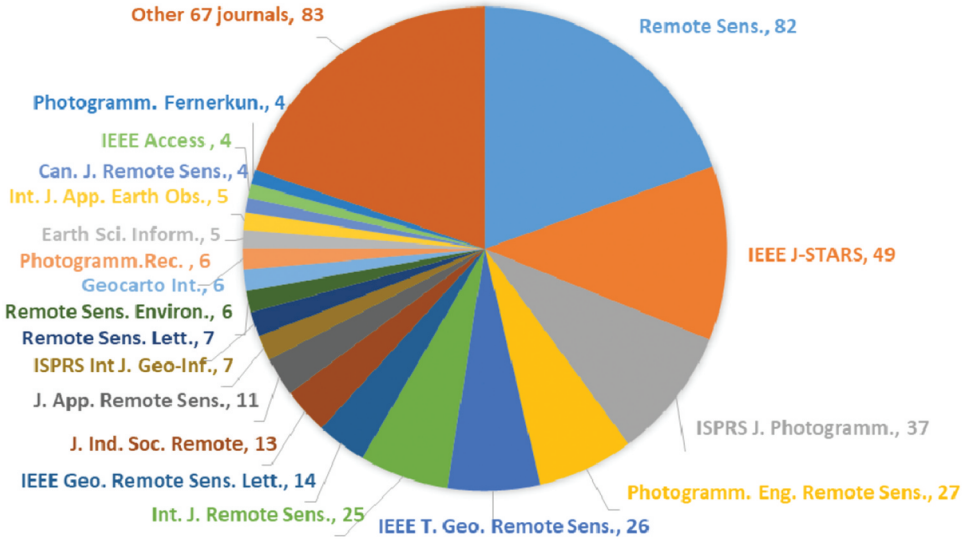

Figure 1. Relevant journals in the 417 articles. The number after the abbreviation of the journal name represents the total number of publications.

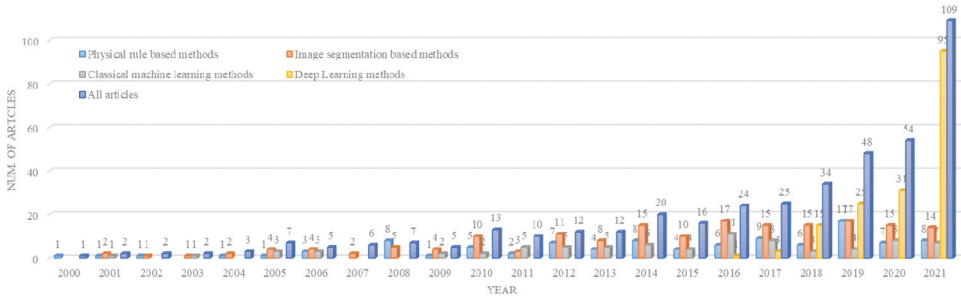

Figure 2. Dates of publication of the building detection related articles (and the major categories) from 2000 to 2021, as indexed by Scopus.

methods account for 24%, 42%, and 49.5%, respectively. Before the era of deep learning, extracting buildings was conducted by rules according to the physical characteristics of buildings in VHR imagery, and making a virtue out of the high spatial resolution of VHR imagery to interpret buildings as image segments. The deep learning methods that were first introduced in 2016 account for 40.7%, demonstrating a strong competitiveness. At the same time, accuracy assessment as well as post-processing are also important parts of building detection. Pixel-wise (277 of 417 articles) and object-wise (126 articles) assessment are the most widely used approaches, while geometric-based assessment (26 papers) is less common. Moreover, post-processing methods including correctness and completeness improvement (53 out of the 417 articles) are also important techniques for building detection.

This comprehensive review aims to address the status, challenges, and prospects for building detection from VHR optical images, to provide a better understanding of this research field for researchers. According to the survey, the three mainstream building detection techniques, i.e. physical rule based methods, image segmentation based methods, and machine learning based methods, are reviewed in Section 2. Section 3 presents the post-processing techniques and accuracy assessment approaches for building detection. Furthermore, on the basis of building detection, some building-related remote sensing image interpretation tasks are also discussed, including building polygon delineation, building change detection, and building height retrieval. Finally, Section 5 presents our conclusions with regard to building detection using VHR optical remote sensing imagery.

2. Building detection methods for optical images

Over the last two decades, a large number of methods have been developed for building detection from aerial and satellite images, which can be categorized into physical rule based methods, image segmentation based methods, and traditional and advanced machine learning (i.e. deep learning) methods.

2.1. Physical rule based detection methods

The physical rule based methods extract the objects according to the knowledge of buildings in high-resolution optical images, and they do not rely on the collection of building samples (Attarzadeh and Momeni 2018), so these methods are of great importance. Moreover, the physical rule based methods are able to reduce the amount of manual work and save on the cost of the building extraction process (Attarzadeh and Momeni 2018). As each building feature descriptor can record the candidates that have a high building probability, these methods identify the objects that meet the multiple building characteristics as a building by conducting probability binarization with a set of thresholds (Huang 2011). Accordingly, in the following, the building characteristics are analyzed item by item, and we present several representative methods that synthesize multiple features to detect buildings.

2.1.1 Building characteristics

A) The geometric characteristics (i.e. shape and size) are some of the most important properties of buildings. Although buildings appear in a variety of sizes and shapes, the most common building shape is a rectangle or a combination of several rectangles (e.g.

L-, T-, and U-shaped) (Ngo et al. 2017). Compared to elongated roads, most buildings within a certain size range are more spatially isotropic (Huang 2011).

B) The spectral characteristics refer to the reflected energy of buildings. Most modern residential and commercial buildings, which are typically made of bright materials (e.g. glass and marble), usually have a higher spectral reflectance than their surroundings (Guo and Du 2017). However, some other buildings with dark materials (e.g. old concrete and bitumen) in the traditional residential and industrial areas can have a relatively low spectral reflectance (Zhang and Huang 2018). As a result, such buildings are usually more difficult to identity from high-resolution optical images.

C) The textural characteristics refer to the visual patterns produced by tonal variation over spatial areas (i.e. individual buildings or building clusters). In satellite images with a resolution of around 2 m (e.g. QuickBird, WorldView-2), buildings often feature homogeneous reflectance with little variance. However, aerial images can provide a finer spatial resolution (e.g. 0.05 m), in which the various installations on buildings, such as chimneys, antennas, domes, and water tanks, can be clearly observed (Lee, Lee, and Lee 2008). Such detailed objects usually result in relatively high heterogeneity for buildings in VHR images. In addition, the buildings and their surroundings, such as the cast shadows and vegetation, often lead to high local contrast (Huang and Zhang 2012).

D) The contextual characteristics denote the spatial constraints or relationships between the target objects and their neighborhoods (Cheng and Han 2016). Considering that buildings cast shadows on the ground (Ngo et al. 2017), shadow can be regarded as a strong clue for building detection. However, the existence, shape, and size of shadows can be influenced by the solar azimuth, satellite viewing angle, and building density (Zhang et al. 2017).

E) The vertical characteristic (i.e. above-terrain) is also an important property for buildings (Gilani, Awrangjeb, and Lu 2018). In an urban environment, height can be utilized to distinguish buildings from terrain objects (e.g. roads and rivers). Due to the influence of the satellite viewing angle, the occlusion of dense buildings at different heights indicates the existence of buildings, but destroys the co-occurrence relationship of buildings and shadows, and interrupts building footprint extraction (Chen et al. 2007).

In summary, as an above-terrain object, buildings have high local contrast, spatial isotropy, an artificial shape, small size, and are usually surrounded by shadows (Huang and Zhang 2012). Most of the physical rule based methods are based on a combination of multiple clues for buildings (Liu et al. 2019). In the following, several representative methods for use with stereo optical images are introduced.

2.1.2 Physical rule based methods using stereo optical images

Stereo imaging is a photogrammetric technique that was initially developed for creating the illusion of depth in an image, according to the basic principle of the human visual system. As suggested in Huang, Cao, and Li (2020), for stereo optical images, two or more pictures of an object are taken from different viewing angles, in order to make the vertical information perceivable when observing the images. Currently, many VHR remote sensing satellites, such as the WorldView series and Ziyuan-3, are capable of producing stereo images that can provide additional useful clues for building detection (Huang et al. 2017a).

The generation of a digital surface model (DSM) is an essential step for detecting buildings using stereo optical images. Currently, the stereo matching algorithms (Gruen 2012), such as hierarchical semi-global matching (SGM) (Qin 2014), make it possible to produce a DSM for a larger area with a lower cost, compared to using LiDAR point cloud data. The DSM indicates the vertical information of the Earth's surface, and a normalized DSM (nDSM) can then be generated by a top-hat morphological operation, to describe the height of the objects above the Earth's surface. Figure 3 presents some typical urban scenes for building detection from stereo optical images, using various methods. The nDSM, which is the most commonly used feature derived from multi-view images, indicates the presence of buildings according to their height information. However, its performance can be affected by two factors: (1) the error of the digital terrain model (DTM), as an nDSM is often computed as DSM −DTM; and (2) inaccurate stereo matching, which leads to incompleteness in some complicated urban areas or blurred boundaries, especially for high-rise buildings (Qin, Tian, and Reinartz 2016a).

Moreover, the angular properties of buildings, based on the geometric and spectral variations of buildings from different viewing angles (Huang, Chen, and Gong 2018a), are also a useful indicator. For instance, Liu et al.

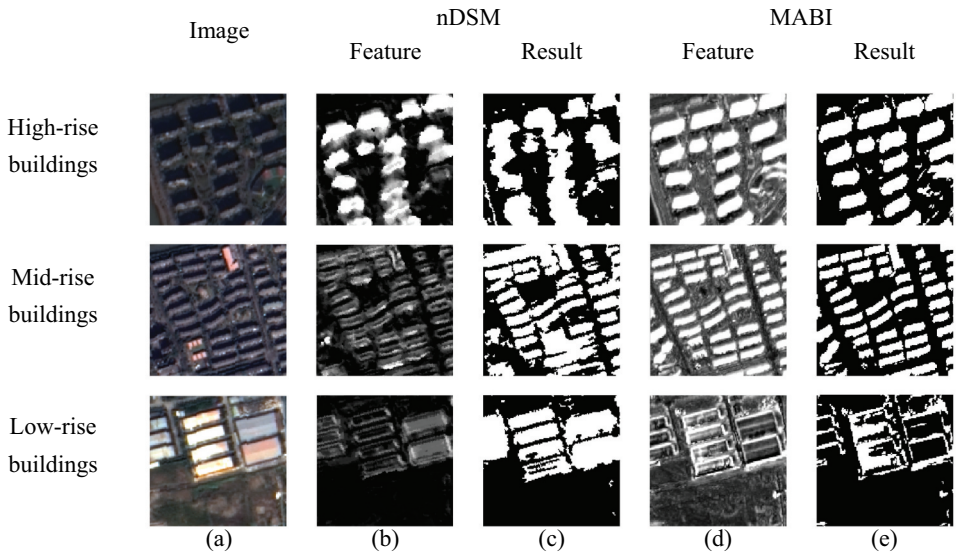

Figure 3. Typical examples of building detection using stereo optical images (Ziyuan-3): (a) image scenes for high-, mid-, and low-rise buildings; (b) and (d) represent feature images for an nDSM and the multi-angular built-up index (MABI), respectively; and (c) and (e) represent the building detection results obtained using the nDSM and MABI, respectively, i.e. pixels with an nDSM (MABI) value larger than a manual threshold were extracted as buildings. The detailed technical steps involved in generating this figure are described in Supplementary Material III.

(2019) designed the multi-angular built-up index (MABI) for use with multi-view images, which is calculated as the maximum value of the difference normalized by the reflectance values of the stereo image pairs. The MABI can highlight the elevated objects (e.g. buildings) with viewing angle differences. From Figure 3, it can be seen that the MABI achieves a good performance for high-rise and mid-rise buildings as it can reflect the angular information inherent in stereo images. However, with respect to low-rise buildings, the MABI does not perform as well, due to the insignificant angular difference in these areas.

2.1.3 Physical rule based methods using monocular optical images

By delineating the spectral (i.e. local contrast and high intensity) and planar spatial (i.e. shape, size, texture, and context) characteristics of buildings in optical images, these methods can be categorized into building component based methods and building-based methods (Mishra, Pandey, and Baghel 2016).

The line segments, which serve as potential components of building candidates, can be first extracted by various methods, such as the Hough transform (Turker and Koc-San 2015), the Canny edge detector (Canny 1986), and the EDLines detector (Akinlar and Topal 2011). On the basis of the geometric shape of buildings, the extracted features, i.e. lines and edges, are then grouped and merged into complete building boundaries (Yan et al. 2017). Although the geometric-based methods are intuitive, they still have some limitations. Firstly, it is difficult to distinguish building features (e.g. building edges) from non-buildings (e.g. road edges) without any prior knowledge. Furthermore, only buildings with specific shapes (i.e. rectangular and the combination of multiple rectangles) can be extracted by such a method (Guo et al. 2016).

In terms of the building-based methods, for the buildings that occupy a small size in VHR optical images, the Harris detector with a high response to building corners is preferred (Liu et al. 2019). For individual bright buildings with a larger size and detailed structure, the morphological building index (MBI) (Huang 2011), which utilizes morphological top-hat by reconstruction to delineate the features of buildings (i.e. the reflected energy, size, and contrast), works well. For building area extraction, a building index named PanTex (Pesaresi, Gerhardinger, and Kayitakire 2008), which is based on the contrast metric of the gray-level co-occurrence matrix (GLCM) for different directions and displacements, can characterize built-up areas and their neighborhoods. Figure 4 presents some typical scenes (i.e. buildings with different heights, sizes, and reflected energy; buildings with sparse/dense distributions) for building detection from optical images using various methods. Finally, the MBI, which is a representative

automatic building index based on morphological operators, as mentioned previously, served as a comparison method, and it achieved reasonable results in all the scenes (see Figure 4).

2.2. Image segmentation based building detection

The pure physical rule based building extraction methods often have difficulties in dealing with practical scenes (Mayunga, Zhang, and Coleman 2005) as the building features can be affected by the sensor type, spatial resolution, weather, illumination, and the complicated urban environment. In some practical applications, as buildings can be represented as image segments with specific characteristics, object-based image processing (OBIA) techniques can achieve satisfactory performances, with only moderate manual input (Tan et al. 2016; Shen, Ai, and Li 2019; Bialas, Oommen, and Havens 2019). OBIA refers to the process of dividing a VHR scene into non-overlapping segments and identifying the land-cover objects of interest. For the OBIA-based methods, most of the mainstream segmentation methods (e.g. region-based methods such as seeded region growing (SRG) (Liu, Cui, and Yan 2008)), graph-based methods (such as graph cut methods (Ok 2013)), and gradient-based methods (such as mean shift (Sirmacek 2011)) are applicable, and the manual processing included in the segmentation is specifically for building extraction. According to the type of manual processing, these methods can be categorized into initialization-based methods and methods based on optimization of the segmentation process.

2.2.1 Building extraction oriented segmentation initialization

Some segmentation methods start with an initialization that requires a certain amount of human interaction (Li, Zhang, and Zhang 2014). A typical example is the SRG method, which is a kind of region-based image segmentation method that has been used for building extraction. This approach examines the neighborhood of the initial seed points (i.e. building pixels) and then judges whether the pixels should be merged to the segment. However, the seed point selection, which is the first step in region growing, usually involves user interaction. For example, Liu, Cui, and Yan (2008) developed a general framework using SRG segmentation to extract simple and small rectilinear buildings from their background.

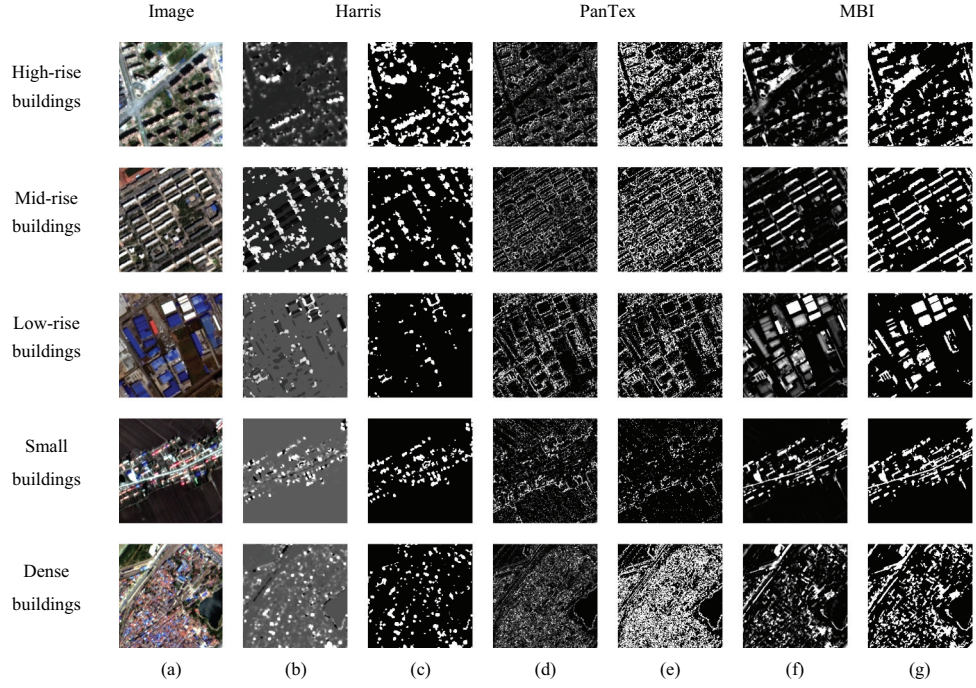

Figure 4. Typical examples of building detection using a single optical image: (a) image scenes for high-rise, mid-rise, low-rise, and small and dense buildings; (b), (d), and (f) represent the feature images for the Harris detector, PanTex index, and the MBI, respectively; (c), (e), and (g) represent the building detection results of the Harris detector, PanTex index, and the MBI, respectively, i.e. pixels with a feature value larger than a manual threshold were extracted as buildings. The detailed technical steps involved in generating this figure are provided in Supplementary Material III.

Similarly, Müller and Zaum (2005) presented an approach that starts with the use of the SRG method to segment the imagery, where the spectral and geometric features are then extracted to differentiate buildings and non-buildings.

In addition to the SRG method, the snake models (which are also known as active contour models), are often used as an alternative approach to solve the task of building detection (Liasis and Stavrou 2016; Chandra 2022). Snake models are defined by energy function minimization, which directs the curves to move to the boundaries of objects. However, the performance of the snake models is strongly influenced by two key points, i.e. the initialization of the snake model and the formulation of the energy function (Mayunga, Coleman, and Zhang 2010). For example, when the initialization cannot cover the building objects effectively, the models have difficulty in selectively detecting the target object (i.e. building structures) in the imagery. To tackle this problem, manual interaction has been introduced into the process of snake curve initialization. For instance, Mayunga, Coleman, and Zhang (2007) developed a semi-automatic detection method based on a snake model. In this method, the seed point (i.e. center of the building) is manually specified and a radial casting algorithm is employed for the initialization of the snake contours, to extract the buildings. The same method has also been employed to extract buildings in dense settlement areas (Mayunga, Coleman, and Zhang 2010)

2.2.2 Manual interaction in optimization of the segmentation process

Besides the initialization for the image segmentation process, manual interaction can also be conducted after the images are completely segmented. With the assistance of manual interpretation, the segmentation results can be further optimized, and some foreground objects (i.e. buildings) and background objects (i.e. non-buildings) can be marked for the subsequent building detection. For instance, Jiang et al. (2008) designed a semi-automatic method to detect buildings by combining segmentation and region selection. In this method, the image is first divided into several object segments using the mean shift algorithm, and then the over-segmented building objects are merged through a manual interaction step. In addition, Tan et al. (2016) developed a semi-automatic right-angle building extraction method. In this method, a seed line at the center of the buildings in the image is manually drawn, and the over-segmented building objects are merged as the foreground area (Tan et al. 2016). With the marked foreground areas and background objects that are generated near the image boundary, the graph cut model is then finally used to detect the buildings.

It should be noted that some of the common methods can be either automatic or semi-automatic (Mishra, Pandey, and Baghel 2016), depending on the degree of manual interaction involved in a specific framework. For instance, the snake models can serve as automatic systems by realizing automatic contour initialization and reformulating the energy function in terms of the properties of buildings (Kabolizade et al. 2014). An example of multi-scale mean shift based segmentation is shown in Figure 5, where every building is manually identified at its optimal scale, and all the building objects are finally combined to generate the result.

2.3 Machine learning based building detection

Machine learning is aimed at making a computer system learn the ability to resolve a specific task from the provided training data (Huang et al. 2018b). Building extraction can be regarded as a binary classification task, and many approaches have employed machine learning based methods to distinguish buildings and non-buildings (Cohen et al. 2016). In the case of sufficient training data which stores prior knowledge of the buildings, the machine learning methods can deal with complicated problems more effectively (Qin, Tian, and Reinartz 2016b).

The machine learning methods can be categorized into classical "shallow learning" and "deep learning" methods, according to the depth of the model structure used for the building extraction (Li et al. 2017b). The traditional machine learning methods typically use shallow models and handcrafted features, while the deep learning based methods are characterized by deep model structures and feature learning. In the following, these two categories of methods for building detection are described in detail.

2.3.1 Classical machine learning methods with a shallow model structure

The classical methods consist of several steps, i.e. feature extraction, feature selection/fusion, and classifier training (Du, Zhang, and Zhang 2015), which we focus on in this subsection.

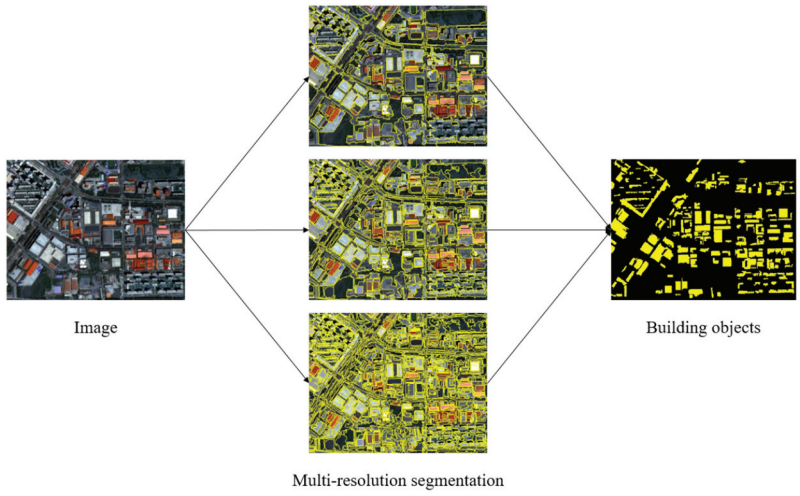

Figure 5. An example of mean shift based building detection (Huang and Zhang 2008). Left: image scene. Middle: three segmentation maps at large, middle, and small scales. Right: the final building detection results.

Feature extraction is the first key component. An expert considers the characteristics of the targets and then designs effective features that can distinguish the targets from the background (Chen and Han, 2016). For the task of building detection, the commonly extracted features are the spectral, textural, morphological, and contextual features (Turker et al., 2015). Such features can describe the building characteristics from various perspectives. For instance, differential morphological profiles (DMPs) can be used to describe the local spectral-structural features (e.g. bright and dark components) of the objects, which can characterize the co-occurrence relationship between buildings and shadows (Zhang et al. 2017). In addition, several other features, such as local binary patterns (LBPs) (Dornaika et al. 2016), the histogram of oriented gradients (HOG) features (Li, Cheng, and Yu 2016), and GLCM measures (Guo and Du 2017), have been widely applied in building extraction. Moreover, some non-building features, such as the normalized difference vegetation index (NDVI) and normalized difference water index (NDWI), which can improve the building-background separability, have also been used (Li et al. 2010).

After the process of feature extraction, the complementary features are fused (Taha and Ibrahim 2022). One widely used feature fusion method is feature vector concatenation. However, a simple stacked feature set can result in a high-dimensional feature space with information redundancy, so that feature selection is necessary in such cases (Zhang et al. 2017). Feature selection is aimed at generating a feature subset that is composed of the most informative features, which account for only a small proportion of the feature dimensionality, but have a classification performance that is comparable to that of the original feature set. For instance, Guo and Du (2017) first generated a high-dimensional feature set for building description, including spectral, geometric, and textural features. The correlation-based feature selection (CFS) strategy was then used to select an optimal feature subset and reduce the dimensionality. Zhang et al. (2017) first extracted multiple candidate features for building density estimation, and then evaluated the importance of each feature using recursive feature elimination (RFE) embedded in support vector regression (SVR). Finally, only one-third of the initial features were selected, according to the feature importance.

With informative features and collected labeled samples, a classifier can be trained to achieve building detection (Chen and Han, 2016). The widely used feature learning approaches include support vector machine (SVM) (Dornaika et al. 2016), AdaBoost (Cohen et al. 2016), random forest (RF) (Huang, Chen, and Gong 2018a), *k*-nearest neighbor (kNN) (Chandra and Ghosh 2018), and artificial neural networks (ANNs) (Teimouri, Mokhtarzade, and Zoej 2016). Besides training a single classifier for building classification, the fusion of multiple classifiers at the decision level is also an effective approach to further improve the detection performance. For instance, Senaras, Ozay, and Yarman Vural (2013) trained several classifiers using different features (color, texture, and shape), and then fused the results produced by each classifier using a hierarchical architecture.

2.3.2 Deep learning: the advanced end-to-end building detection methods

Handcrafted features are usually selected and generated by experts, in order to obtain optimal results. The performance of the classical methods is therefore heavily reliant on human ingenuity in the feature design (Xu et al. 2018). Moreover, the so-called "semantic gap" exists between the image features (e.g. color and shape) and the semantics of buildings (e.g. the co-occurrence between building and shadow). Considering the complicated urban scenes with increasingly difficult building extraction, the descriptive ability of these shallow features can become limited, or even completely exhausted (Cheng and Han, 2016).

Recently, deep learning has shown its impressive feature representation ability and has obtained notable successes in various applications (Zhang, Zhang, and Du 2016). Compared to the classical works based on handcrafted features, the recently designed deep networks are classical data-driven models, which automatically learn high-level and hierarchical features using the massive training data (Ps and Aithal 2022). In the following, the benchmark building detection datasets and the deep learning based models for building detection are introduced.

A)Building detection datasets

Since the Zeebrugge dataset was published as part of the 2015 IEEE GRSS Data Fusion Contest, dozens of building detection datasets have been released. Note that the datasets used to evaluate the traditional methods are typically small in size, and the training and test sets are collected from the same local region (or image), resulting in a poor generalization ability. In the era of deep learning, the more advanced datasets (see Table 1) ensure that the training and test sets are spatially independent, the spatial coverage is wider, and the data volume is larger, which is in accordance with reality.

Considering the size of the buildings (>10 m^2), Table 1 lists some benchmark satellite/aerial datasets, for most of which the spatial resolution ranges from the centimeter level to 2 m, except for the relatively coarse resolution of SpaceNet 7 (i.e. 4 m). In addition to the widely used RGB channels (for instance, Figure 6(a)), some datasets also provide extra useful information for further depicting buildings. With regard to the spectral information, the Potsdam and WHU-Satellite datasets have RGB/near-infrared (NIR) bands, and the SpaceNet and SpaceNet 4 datasets are made up of eight spectral bands of the WorldView 2/3 sensors. The specific network modules with the inclusion of multispectral information can help to distinguish natural surfaces from buildings (Huang et al. 2021b). With regard to the vertical information, the Potsdam, Vaihingen, Zeebrugge, and DFC19-JAX datasets provide airborne LiDAR derived nDSMs, and the SpaceNet 4 dataset is made up of 27 unique views, for which the viewing angles range from −32.5° to 54.0° (Weir et al. 2019). Several datasets (e.g. DFC19-JAX) have further attempted to boost the deep learning networks by combining planar and stereo remote sensing observations (Cao and Huang 2021). With regard to the temporal characteristics, the WHU Building Change Detection, SECOND, Hi-UCD, and ZKXT_2021 datasets contain multi-temporal remote sensing observations, building footprints for each date, and building change records. The SpaceNet 7 dataset of Planet satellite imagery mosaics includes 24 temporal images (one per month). With regard to the differences of the imaging angles of multi-temporal observations and the static property of buildings, a multi-temporal dataset can also be seen as a form of multi-view dataset. Thus, by the use of the multi-temporal datasets, it is possible to delineate the vertical information, to improve the detection performance (Papadomanolaki, Vakalopoulou, and Karantzalos 2021). In addition, given the dense time-series datasets, such as SpaceNet 7, it is also worthwhile designing phenological information related modules to distinguish buildings (i.e. the static man-made structures) from natural surfaces.

In terms of the labels of the datasets, despite the fact that building extraction is generally regarded as being equivalent to detecting roof contours, most datasets (e.g. Inria and SpaceNet) directly treat the building footprints as the evaluation targets (Chen et al. 2020). It should be noted that the complicated misalignment between the footprints and roof outlines worsens the difficulty of spatial positioning. Moreover, the label quality of these benchmark datasets can be uneven, mainly due to the burdensome cost of labeling large-scale data (Zhang et al. 2020). There are three major label sources: crowdsourced geospatial data (Jung et al., 2021), government open-source data (Chen et al. 2020), and interactive annotation with artificial intelligence tools (Acuna et al. 2018). For instance, OpenStreetMap (OSM),[1] which is a widely used source of crowdsourced geospatial data, was used to construct the labels of the SpaceNet, SpaceNet 4,

Table 1. Comparison of the representative optical image building detection datasets.

Dataset name	Spatial				Spectral bands	Vertical feature	Platform	Website
	Resolution	Num. of patches	Patch size	Coverage				
Zeebrugge	5 cm	5	10,000 × 10,000	12 km^2	RGB	LiDAR	Aerial	http://www.classic.grss-ieee.org/community/technical-committees/data-fusion/2015-ieee-grss-data-fusion-contest/
ISPRS -Potsdam	5 cm	38	6000 × 6000	2 km^2	RGB-NIR	DSM	Aerial	https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam/
-Vaihingen	9 cm	33	>1200 × 1800	11 km^2	RGB-NIR	DSM	Aerial	https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-vaihingen/
AIRS	7.5 cm	1047	10,000 × 10,000	475 km^2	RGB	No	Aerial	https://www.airs-dataset.com/
DLRSD	0.3 m	2100	256 × 256	-	RGB	No	Aerial	https://sites.google.com/view/zhouwx/dataset#h.p_hQS2jYeaFpV0
Massachusetts	1 m	151	1500 × 1500	340 km^2	RGB	No	Aerial	http://www.cs.toronto.edu/~vmnih/data/
WHU -Aerial	0.3 m	8188	512 × 512	450 km^2	RGB	No	Aerial	http://gpcv.whu.edu.cn/data/building_dataset.html
-Satellite	0.3–2.5 m	17,592	512 × 512	950 km^2	RGB-NIR	No	Satellite	
Inria Aerial Image	0.3 m	180	5000 × 5000	810 km^2	RGB	No	Satellite	https://project.inria.fr/aerialimagelabeling/
CrowdAI	0.3 m	341,058	300 × 300	-	RGB	No	Satellite	https://www.crowdai.org/challenges/mapping-challenge
SpaceNet -Rio de Janeiro	0.5 m	1400	650 × 650	2,544 km^2	8-band multispectral of WV2	No	Satellite	https://oldpan.me/archives/download-aws-spacenet-dataset
-Vegas-Paris-Shanghai-Khartoum	0.3 m	10,863	650 × 650	7110 km^2	8-band multispectral of WV3			
SpaceNet 4	0.3 m	62,000	900 × 900	665 km^2	8-band multispectral of WV3	Multi-view	Satellite	https://spacenet.ai under a CC-BY SA 4.0 License
DFC19-JAX	0.3 m	3083	2048 × 2048	>1160 km^2	RGB	nDSM	Satellite	http://www.classic.grss-ieee.org/community/technical-committees/data-fusion/2019-ieee-grss-data-fusion-contest/
EvLab-SS	0.1–0.25 m	60	4500 × 4500	-	RGB	No	Satellite and aerial	http://earthvisionlab.whu.edu.cn/zm/SemanticSegmentation/index.html
BDCI	2 m	140,000	256 × 256	-	RGB	No	Satellite	https://www.datafountain.cn/competitions/475/datasets
DSTL	-	-	-	25 km^2	RGB	No	Satellite	https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection
WHU Building Change Detection	0.2 m	1	15,354 × 32,507	20 km^2	RGB	No	Aerial	https://study.rsgis.whu.edu.cn/pages/download/building_dataset.html
SpaceNet 7	4 m	2389	1024 × 1024	41,250 km^2	RGB	No	Satellite	https://spacenet.ai/sn7-challenge/
SECOND	0.3 m	4662	512 × 512	-	RGB	No	Aerial	http://www.captain-whu.com/project/SCD/
Hi-UCD	0.1 m	1293	1024 × 1024	30 km^2	RGB	No	Aerial	https://arxiv.org/abs/2011.03247
ZKXT_2021	<1 m	2500	512 × 512	-	RGB	No	Satellite	http://gaofen-challenge.com

(a) SpaceNet 7 (b) Zeebrugge (c) Vaihingen (d) Potsdam

Figure 6. Building patches in the open-access datasets: (a) two true-color parcels from Paris in the SpaceNet dataset, (b) two true-color parcels in the Zeebrugge dataset, (c) two false-color parcels in the Vaihingen dataset, and (d) a true-color parcel in the Potsdam dataset. These datasets can be freely downloaded from the websites listed in .Table 1

SpaceNet 7, and DFC19-JAX datasets (i.e. the most large-scale datasets in Table 1). Land Information New Zealand (LINZ)[2] released a set of aerial imagery and the corresponding building map, which were used in the WHU-Aerial and AIRS datasets to manually refine the building map and construct the benchmark datasets. More recently, Hao et al. (2021) released an interactive tool for building semantic annotation tasks.[3] Generally speaking, the quality of the building labels is affected by the outdated labels, spatial misalignment, wrong labels from the sources (e.g. OSM), and the degree of manual refinement. Among the benchmark datasets listed in Table 1, the AIRS, Zeebrugge, Potsdam, and Vaihingen datasets have been manually edited in a strict manner, and their labels have the highest accuracy.

In terms of generality and diversity, buildings under different geographical, economic, religious, and cultural conditions can be very different, and the global spatial distribution of the building detection datasets is very uneven (i.e. mainly located in China, Europe, New Zealand, and the UK). Although the SpaceNet 7 dataset is spread out across the globe and covers six continents, its generality is limited by its relatively rough spatial resolution (i.e. 4 m). For a deep learning network, since its detection ability is heavily dependent on the training samples, it is of great significance to further construct a large-scale and representative dataset.

B) Deep models for building detection

In the early period, based on the image recognition paradigm (e.g. VGG/ResNet) in computer vision, researchers used sliding windows to crop whole remote sensing images into regular patches (e.g. sized 64×64), learned the features hierarchically, and labeled each patch (i.e. "building" or not) (Alshehhi et al. 2017). Although some progress has been made in these methods, the "mosaic-like" phenomenon and high computational redundancy inevitably occur in such patch-wise processing (Alshehhi et al. 2017).

To overcome this problem, fine-grained inference, i.e. semantic segmentation, which assigns a label (i.e. "building" or "background") to every pixel in a given image, has become more popular for building detection (Xu et al. 2018). There are two mainstream network architectures for semantic segmentation: (1) the encoder-decoder U-shaped architecture (Navab et al. 2015); and (2) multi-scale subnetworks in parallel (e.g. HRNet (Sun et al. 2019)). As the name implies, the U-shaped architecture consists of two parts: an encoder path (i.e. the left-hand side of the U) to downsample and aggregate the semantic feature representations at multiple different levels, and a decoder path (i.e. the right-hand part of the U) to gradually upsample and allocate the semantic information to each level. Generally speaking, the encoder can be modeled by the afore-mentioned patch-wise network, the decoder is carried out as the reverse of the encoder, and skip connections connect the features with the same spatial scale in the two afore-mentioned paths,

to keep the fine details. Although the U-shaped networks can achieve a desirable overall accuracy and building area coverage estimation, blob-like segments (Ji, Wei, and Lu 2019) indicate a poor performance in the target edge parts, and the spatial details of the results can be inferior to the input optical imagery. Thus, the second type of architecture, e.g. HRNet, aims to keep the high-resolution feature representations by repeatedly fusing the parallel high-to-low resolution subnetworks (Sun et al. 2019). Benefiting from the repeated multi-scale fusion, HRNet has become a popular sematic segmentation architecture, and has achieved high spatial delineation for building extraction (Zhu et al. 2021a).

As the studies have moved forward, researchers have found that the omission of small buildings and the holes in large buildings cannot be solved well with only a semantic segmentation network, as the inference of the pixel labels is independent of each other (Ji et al. 2019). Thus, in contrast to semantic segmentation, instance segmentation, which is aimed at allocating a unique label for each instance (i.e. individual buildings), has been considered. Mask RCNN is the most widely used instance segmentation architecture in building extraction (Liu et al. 2021b). Mask RCNN consists of object detection and mask segmentation. By treating each building as an individual object, the first module uses a ResNet-like subnetwork to extract high-level semantic information and identify the location of building objects. Mask segmentation is then used to determine the class label of each pixel (i.e. building or not) within the identified proposal. Considering the spatial details of buildings, a recent natural approach is to use fine-grained features in semantic segmentation to replace ResNet in instance segmentation (Zhao, Persello, and Stein 2021). In fact, this approach that combines instance segmentation and semantic segmentation could be the future direction of building detection.

Moreover, the characteristics of buildings can be formulated as plug-and-play modules and inserted into the afore-mentioned architectures, to achieve task-specific improvements. Given the limited space available, several representative studies from four aspects are described here. Firstly, considering that buildings appear in a variety of sizes, the Siamese U-Net method (Ji, Wei, and Lu 2019) combines segmentation maps of different resolutions to produce scale-invariant predictions, and the pyramid scene parsing network (PSPNet) method (Zhao et al. 2017) utilizes a pyramid pooling module to combine the multi-level features extracted by the encoder. Secondly, in view of the edge information of buildings, Jiang et al. (2020) used structural similarity to evaluate the predicted and real boundary pixels and alleviate the blob-like segments. Furthermore, Wen et al. (2021) designed a multi-scale erosion network coupled with a semantic decoding module for building edge detection. Thirdly, in view of the geometry of building rooftops, adversarial learning for shape regularization has been used to model the shape patterns of buildings (Ding et al., 2021a), and a modified PointNet method (Qi et al. 2017) was proposed to learn the vertex deformation, to further refine the shape of buildings. In addition, Chen et al. (2020) simulated the process of manual delineation of rooftop outlines by a convolutional recurrent neural network (CRNN), to enable the boundaries of the rooftops to be generated with straight lines and sharp corners. Lastly, for the vertical characteristic of buildings, Ferrari et al. (2021) designed RGB and co-located DSM streams to extract specific building features, and then fused them together in a cross-modal stream. In addition, a lot of works have simultaneously taken several sources of prior information about buildings into consideration (Ahmed et al. 2021; Ding et al. 2021bb; Feng et al. 2021; Wang et al. 2021; ZZhao, Persello, and Stein 2021; Zhu et al. 2021b; Chattopadhyay and Kak 2022; Chen et al. 2022; Li et al. 2022), resulting in complementary improvements in the delineation of buildings.

3. Post-processing and accuracy assessment for building detection

On the basis of the initially extracted building results, post-processing can significantly improve the correctness and completeness of the detected buildings. The main post-processing methods are summarized in the following. The standard metrics used for building detection evaluation are also introduced. Finally, an experimental comparison as well as a discussion are provided to analyze the detection methods and the post-processing step.

3.1. Post-processing for building detection

Post-processing is aimed at refining the original detection result in order to enhance its accuracy (Huang et al. 2017b). The initial building detection

results produced by various methods often suffer from false alarms and omissions. The false alarms are usually related to roads, open ground, and bright soil, as these terrain objects often present a similar spectral reflectance to buildings. The omissions mainly refer to dark and heterogeneous roofs (Zhang, Zhang, and Du 2016). To deal with these problems, a post-processing framework is essential.

3.1.1 A) Correctness improvement

Soil and roads are the main sources of false alarms during building detection (Huang et al. 2017b). Such false alarms can be eliminated by effective prior knowledge (e.g. spectral, geometric, and contextual constraints). Specifically, the spectral constraints, e.g. the NDVI or the hue component of the hue-saturation-value (HSV) color space, can alleviate the false alarms caused by soil (Huang et al. 2017b). Geometric constraints are based on connected component analysis of the initial binary building map. The commonly used geometric metrics, such as area or the length-width ratio, can remove small noisy items or elongated objects (e.g. roads). Finally, a contextual constraint usually refers to shadow verification for building candidates (Manno-Kovacs and Sziranyi 2015). By imposing the shadow constraint on the initially detected building results, ground objects such as parking lots and open areas can be removed. Although shadows may have different shapes and sizes in high-resolution optical images, it can be easier to extract shadow information than to directly detect buildings (Li et al. 2017a). To date, most researchers have used thresholding methods based on the panchromatic, visible, and NIR channels, or color space transformation, to achieve shadow extraction (Ghanea, Moallem, and Momeni 2016; Liasis and Stavrou 2016). Several examples of false alarm mitigation are presented in Figure 7.

3.1.2 B) Completeness improvement

In the initial building results, some buildings may be identified only partly (Huang et al. 2017b), due to the heterogeneity of building roofs. In addition, some small holes may appear in the building objects, as shown in Figure 4(c,e, and g). To deal with such problems, morphological operations and the region growing method can be used to fill the holes and supplement the incomplete buildings (Gao et al. 2018), as shown in the examples in Figure 8. However, it should be considered that it is difficult to recover buildings that are completely undetected (Huang et al. 2017b). Such omission errors often appear in very challenging scenes, e.g. dark roofs with low contrast with the surroundings. Accordingly, in practice, it is reasonable to preserve more building candidates that can be further verified, according to the knowledge and rule constraints (Huang et al. 2017b). Such a strategy can achieve a balance between the correctness and completeness of building detection results (Li et al. 2017a).

For a given detection region, although the post-processing approaches can result in an accuracy improvement, the post-processing rules and thresholds are currently set by trial by error, and the generalization and automation level are poor. For instance, the geometric constraints, such as the area or length-width ratio, are sensitive to the spatial resolution of VHR imagery and the landscape of the target region. Considering the maturity level of the afore-mentioned building detection techniques and their applicability for downstream users, post-processing is still necessary. Meanwhile, the better the initial building detection results, the easier the post-processing will be.

3.2. Accuracy assessment for building detection

3.2.1 A) Pixel-wise assessment

Three standard and commonly used quantitative metrics are used in this article for evaluating the building detection results at the pixel level, namely, the intersection over union (IoU), precision (P), recall (R), and F-measure (F, also called the F-score) (Ok, Senaras, and Yuksel 2013). The evaluation measures are based on the overlapping areas between the reference data and detection result. For the pixel-based classification, every pixel is labeled into one of four classes: true positive (TP), false positive (FP), false negative (FN), or true negative (TN). TP refers to a building pixel that is correctly identified. FP indicates a non-building pixel that is identified as building. FN represents a building pixel that is identified as non-building. TN corresponds to a non-building pixel that is correctly identified.

$$P = \frac{\mathrm{card}(TP)}{\mathrm{card}(TP) + \mathrm{card}(FP)} \tag{1}$$

$$R = \frac{\mathrm{card}(TP)}{\mathrm{card}(TP) + \mathrm{card}(FN)} \tag{2}$$

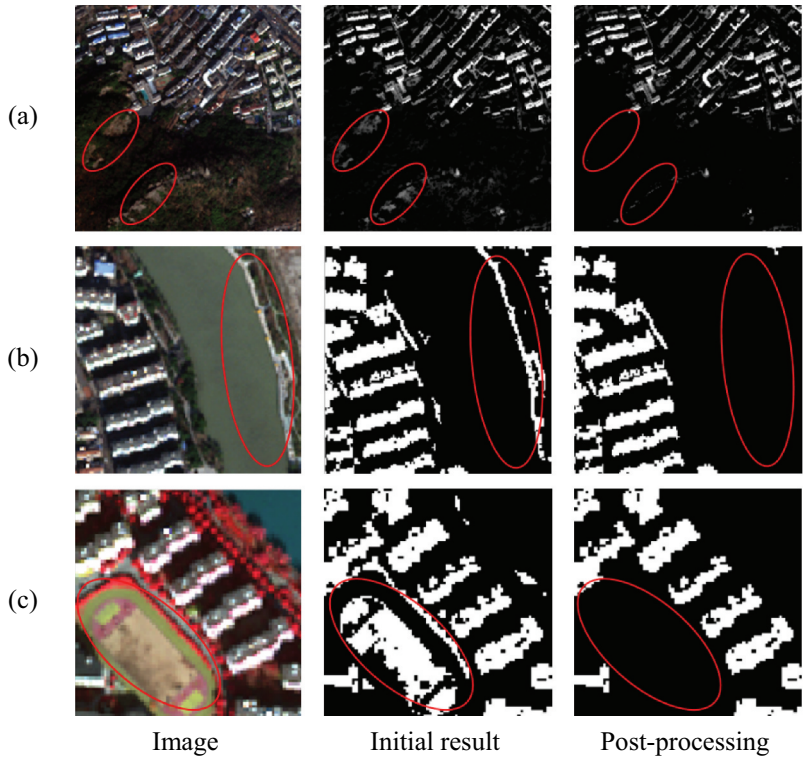

Figure 7. Examples of post-processing for building detection: (a) spectral constraint, (b) shape verification, and (c) shadow verification. Left column: false-color high spatial resolution images; middle column: initial building footprint generated by the method proposed by Huang and Zhang (2011).

$$F = \left(1 + \alpha^2\right) \frac{P \times R}{P \times \alpha^2 + R} \tag{3}$$

$$\text{pixel wise } IoU = \frac{A_r \cap A_p}{A_r \cup A_p} \tag{4}$$

where card(.) means the number of pixels classified to each type; α is a non-negative hyper-parameter to balance the two metrics; and A_r and A_p represent the footprints delineated by the reference and the prediction, respectively. Thus, the pixel-wise IoU measures whether each building pixel in the prediction map is labeled in the reference map.

The precision/recall indicates the correctness/completeness of the predicted buildings. The F-measure combines these two metrics into a single metric to comprehensively evaluate the quality of the detection result. Assigning $\alpha^2 < 1$ in Equation (3) can give more importance to the precision metric. As suggested in many studies, precision and recall are usually considered to be equally important, so setting α as 1 is appropriate (Jozdani and Chen 2020). The IoU is defined as the overlap ratio of the prediction and the reference, and is another widely used overall metric.

3.2.2 B) Object-wise assessment

For the object-based accuracy assessment, the mean average precision (AP) and mean average recall (AR) over multiple IoU scores are used (Jozdani and Chen 2020). The object-wise IoU is defined as the ratio between the area of the intersection and the area of the union of each extracted individual building footprint and its closest reference building mask. Specifically, the AP and AR are averaged over the object-wise IoU values, with thresholds from 0.50 to 0.95 and an interval of 0.05:

$$AP = \frac{AP_{0.5} + AP_{0.55} + \ldots + AP_{0.95}}{10} \tag{5}$$

where AP_{thr} means the correctness of the predicted buildings at the object-wise IoU>*thr*, $thr \in \{0.5, \ldots 0.95\}$. The AR can be formulated in a similar fashion. In addition, considering the variety of building sizes, $AP_{(S,M,L)}$ and $AR_{(S,M,L)}$ evaluate the detection results for small, medium, and large buildings, respectively (Li, Di Wegner, and Lucchi 2019). As defined in ZZhao, Persello, and Stein (2021), small, medium, and large represent an area of < 32^2 pixels, an area between 32^2 pixels and 96^2 pixels, and an area > 96^2 pixels, respectively,

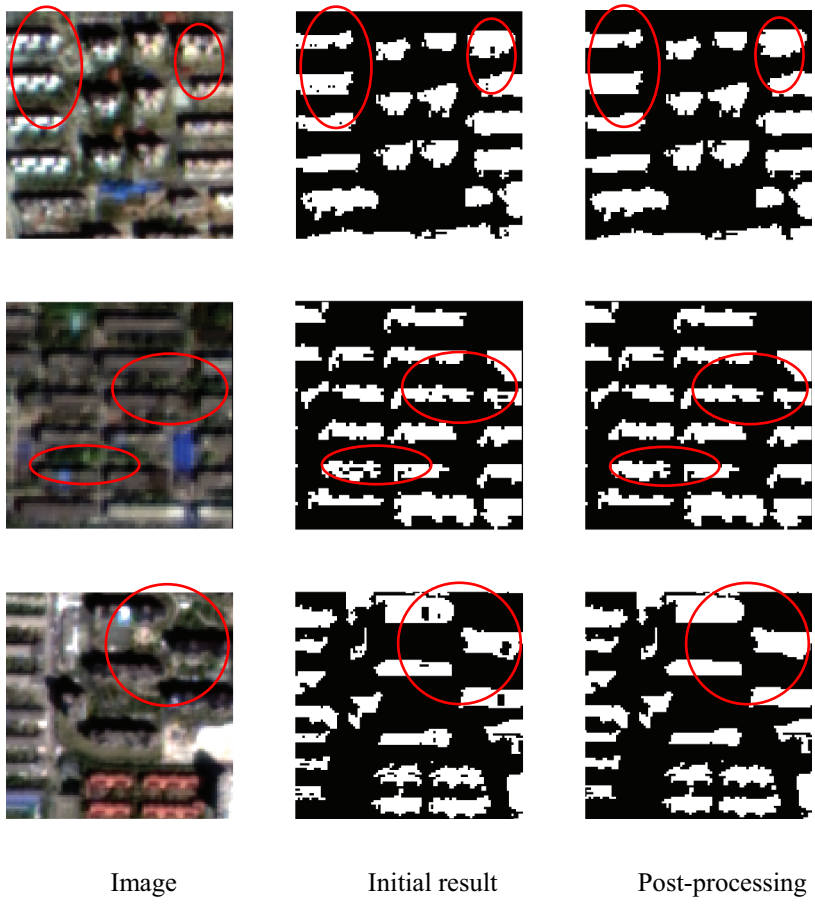

Figure 8. Examples of filling the holes and supplementing the incomplete buildings. The detailed technical steps involved in generating this figure are provided in Supplementary Material III.

3 Geometric assessment. As an artificial structure, individual buildings have specific geometric characteristics. The geometric assessment metrics can be grouped into vertex-, boundary-, and shape-based metrics.

Since each individual extracted building and its closest reference building mask can be assigned as two sets of vertices, the vertex-based metrics (Chen et al. 2020) can be computed as:

$$\begin{aligned} VerF &= \frac{1}{k}\sum_{\mathrm{thr}=1}^{k}\frac{2P_{\mathrm{thr}}\times R_{\mathrm{thr}}}{P_{\mathrm{thr}}+R_{\mathrm{thr}}},\\ VerP &= \frac{1}{k}\sum_{\mathrm{thr}=1}^{k}P_{\mathrm{thr}},\\ VerR &= \frac{1}{k}\sum_{\mathrm{thr}=1}^{k}R_{\mathrm{thr}} \end{aligned} \tag{6}$$

where *thr* is the threshold value around the reference building mask; thr $\in \{1, \ldots, k\}$ (k = 5); and P_T and R_T are, respectively, the vertex precision and recall under threshold thr. At the same time, boundary metrics such as the boundary based F-score (*BoundF*), precision (*BoundP*), and recall (*BoundR*) can also be computed in a similar fashion (Cheng et al. 2019; Huang, Tang, and Xu 2022).

Another widely used vertex-based metric is the Hausdorff distance (Hd) (Xie et al. 2020; Huang, Tang, and Xu 2022), which measures the maximal-minimal distance between two sets of vertices:

$$\begin{aligned} \mathrm{Hd}(A_{\mathrm{v}},B_{\mathrm{v}}) &= \max[h(A_{\mathrm{v}},B_{\mathrm{v}}),h(B_{\mathrm{v}},A_{\mathrm{v}})],\ \text{where}\ h(A_{\mathrm{v}},B_{\mathrm{v}})\\ &= \max_{a\in A_{\mathrm{v}}}\left[\min_{b\in B_{\mathrm{v}}}\|a-b\|\right],\ h(B_{\mathrm{v}},A_{\mathrm{v}})\\ &= \max_{b\in B_{\mathrm{v}}}\left[\min_{a\in A_{\mathrm{v}}}\|b-a\|\right], \end{aligned} \tag{7}$$

where A_{v} and B_{v} are two sets of vertices from the reference mask and estimated parcel, respectively. $\|\|$ denotes the spatial distance between the two vertices.

The E_{curve} metric, which was developed by Ding et al. (2021a), can measure the misalignment in building boundaries:

$$E_{curve}(A_v, B_v) = \|g_c(A_v) - g_c(B_v)\| \quad (8)$$

where g_c refers to the contour curvature function (Gonzalez, Woods, and Masters 2009), and a large E_{curve} indicates that the boundary of the predicted building boundary is uneven.

To estimate the average shape dissimilarity for an extracted building footprint compared to the corresponding reference building, a metric named E_{shape} can be used (Ding et al. 2021a):

$$\begin{aligned} E_{shape}(A_v, B_v) &= \|f_s(A_v) - f_s(B_v)\| \text{ where } f_s(\cdot) \\ &= 4\pi\left(\frac{\text{card}(\cdot)}{p(\cdot)^2}\right) \end{aligned} \quad (9)$$

where $f_s(.)$ is the widely used perimeter-area ratio metric to estimate the compactness of the object, and $p(.)$ and card(.) are the perimeter and the area of the object, respectively. Thus, Equation (9) can be used to estimate the shape dissimilarity (in terms of compactness) for an extracted building footprint (A_v) compared to the corresponding reference building (B_v).

According to the 417 articles considered in this review, pixel-wise assessment (277 of the 417 articles) is the most widely used approach, and many object oriented building detection methods can be evaluated by the pixel-wise metrics (Zeng, Wang, and Lehrbass 2013). Pixel-wise assessment is more convenient for area-based evaluation and statistics (Olofsson et al. 2014), while it can be biased for buildings with a large size. In contrast, object-wise assessment and geometric-based assessment are more robust to the building size, while the cost of the reference labels is higher. It is noted that precision, recall, F-measure, and IoU can be conducted at both the pixel level and object level. Precision is preferred when evaluating the possibility of correctly detecting buildings by a proposed method; recall is preferred when the task focuses on reducing building omissions, in applications such as land resource surveying; the F-measure combines these two metrics into a single metric to comprehensively evaluate the quality of the detection result; and the IoU is a standardized criteria for deep learning based tasks. Moreover, geometric-based assessment is more useful for the tasks focused on building layout/morphology (Zeng, Wang, and Lehrbass 2013).

3.3. Comparison and discussion

By surveying the 417 articles, it was found that the results of the ISPRS Semantic 2D Labeling Contest (https://www.isprs.org/education/benchmarks/UrbanSemLab/results/) can be used to evaluate the mainstream approaches in a fair manner. The details of the results of this contest are provided in Supplementary Material IV. Based on the contest results, the current situation, characteristics, and potential of the mainstream detection methods can be compared and analyzed, as follows:

Firstly, there is only one result ("UT_MEV" in the list) that is based on a physical rule (i.e. the fusion of the spectral, spatial, and vertical characteristics of buildings), which was rated no. 137 overall in the Vaihingen 2D Labeling Contest (140 results in total). This suggests that, from the perspective of accuracy, the physical rule based detection methods are inferior to the other approaches. The diversity of buildings in complicated urban areas is the main bottleneck for this kind of method. Although the features can indicate the existence probability for buildings, the omission and commission errors are still severe. Thus, the use of physical rule based methods is falling. Meanwhile, post-processing by the use of physical rules for buildings does have potential, as the initial detection results will have had many of the complicated background signals filtered out.

Except for the UT_MEV method, the other results in both the Vaihingen and Potsdam 2D labeling contests are all based on supervised classification, and most of the top-ranked results are based on deep learning. As indicated by the average F-score of all the related results, the differences between the deep learning based methods and the classical machine learning based methods are 1.94% (Vaihingen) and 4.75% (Potsdam), respectively. On the basis of the spectral-spatial-vertical man-made building features, most of the recent classical machine learning based works that are equipped with image segmentation techniques train the models with dozens of parameters. The man-made features are low level and not specially designed for the target building region, and the models are too simple to capture the complicated patterns of buildings. When dealing with a new building task, the existing non-deep learning models are difficult to generalize or transfer. In contrast, under the guidance of the training data in a given region, the deep

learning approaches can simultaneously learn the features of buildings and the detection model in an end-to-end manner. Deep learning networks with millions of parameters can capture building patterns well. Moreover, many plug-and-play modules that focus on the buildings characteristic in a target region can be easily added to (or removed from) a pre-trained network to construct a new model, indicating the high flexibility of the deep learning approach. Although deep learning techniques have achieved promising performances in building detection, several challenges remain. Firstly, deep networks need a huge volume of labeled data (Zhao et al. 2017), but it is impractical to acquire abundant and precise data for deep architecture training. Secondly, the feature representation in deep learning is difficult to explain and, accordingly, hyper-parameter tuning can be subjective and inefficient. In the future, much effort should be made to construct an effective and robust deep learning network for building detection. Another potential research direction would be to develop a deep learning network in the manner of fusing data- and model-driven formalization, i.e. investigating the network architectures and understanding the characteristics of buildings in VHR optical images.

Lastly, according to the results, image segmentation, vertical information, and post-processing result in improvements of 0.34%, 0.64%, and 0.38% in F-measure, indicating the positive function of these techniques (see Supplementary Material IV for more details). In the 417 articles, there are 92, 106, and 53 studies focusing on these techniques, respectively. It is shown that the vertical information is the most beneficial, while the post-processing makes a considerable positive contribution. As the initial building detection will have filtered out many of the complicated background signals, pixel-wise correctness and completeness improvement can be achieved in the post-processing step.

4. Research perspectives: related optical remote sensing interpretation tasks

4.1. Building polygon delineation

For many geographic information system (GIS) applications, building detection is an intermediate step in a more comprehensive workflow (Zebedin et al. 2006), which is aimed at achieving an abstract and vectorized representation of the building contours. Before the comeback of deep learning, building polygon delineation was usually formulated as a multi-step and bottom-up workflow, as follows: (1) building detection (see Sections 2.1–2.3.1); (2) building parcel refinement (see Section 3.1); (3) vectorization; and (4) simplification. Vectorization involves producing vertices of the polygons outlining the building instances, which converts the raster result into a building polygon shapefile format. Simplification involves reducing the redundant vertices while maintaining the building contours. A well-known polygon simplification method is the Douglas-Peucker algorithm (Douglas and Peucker 1973), which regularizes the building curve by approximately representing the curve as a series of points and then reducing the points on the curve. The basic steps of the Douglas-Peucker algorithm are illustrated in Figure 9 and introduced as follows:

1) Connect two points between the head and tail of the curve (e.g. 1 and 9 in Figure 8(a)) to measure the distance from the middle points to the straight line.

2) If the maximum distance (e.g. the distance of 4 to the straight line in Figure 8(a)) is greater than a predefined threshold, then the corresponding points are retained; otherwise, all the points between the head and tail points are discarded (e.g. 3, 5, and 7 in Figure 8(c)).

3) Partition the reserved points, i.e. the head and tail points, into two parts (e.g. in Figure 8(b), 1–4 as the first part and 4–9 as the second part), and repeat the above steps until there are no discardable points (Figure 8(c-d)).

Although the multi-step and bottom-up workflow is popular (Dawen et al., 2021; Daranagama and Witayangkurn 2021), the necessary human intervention (such as setting a threshold by trial and error) comes at a high cost (Partovi et al. 2017), and omission errors may still exist along the boundaries (ZZhao, Persello, and Stein 2021). In contrast, deep learning networks that directly predict the polygons from the imagery (Chen et al. 2020; Wei and Ji 2021; ZZhao, Persello, and Stein 2021) are a relatively new and promising research direction.

A deep building polygon generation network is formulated as an end-to-end framework, where the first method to be developed, which was named PolygonRNN (Castrejón et al. 2017), was made up of a basic architecture with two subnetworks: (1)

a building instance segmentation subnetwork to encode the image features; and (2) a building outline polygon prediction subnetwork to infer the positions of the vertices and sequentially connect them. In PolygonRNN, the first subnetwork is a modified VGG architecture, and the second subnetwork is a sequential recurrent network: i.e. a two-layer convolutional long short-term memory (LSTM) network using the image features and the information of the first and last vertices to predict the current vertex. In this manner, the geometric relationship between the vertices and boundary lines can be embedded. On the basis of PolygonRNN, recent studies have introduced several improvements, such as adding building refinement modules between the two subnetworks (ZZhao, Persello, and Stein 2021), using more advanced recurrent neural network (RNN) modules (Castrejón et al. 2017; Huang, Tang, and Xu 2022), and replacing the RNN with other layers, such as graph convolutional blocks (Wei and Ji 2021) or convolutional neural networks (CNNs) (Chen et al. 2020). However, to date, although the related studies have vigorously expanded, a lot of effort is still needed to achieve a practical level of building detection.

4.2. Building change detection

As an interpretation task based on building extraction, building change detection has many applications, such as disaster assessment (Chen et al. 2021) and urbanization monitoring (Huang et al. 2017a). Generally speaking, the building change detection task includes determination of both the location and type of the change (Papadomanolaki, Vakalopoulou, and Karantzalos 2021). According to Huang, Cao, and Li (2020), for the change types of buildings, these can also be summarized into three categories: new construction, demolition, and reconstruction. However, as the identification of the building change type is relatively difficult, the current research has mainly focused on determining the change location (Liu et al. 2021a).

Recently, building change detection methods have gradually developed from traditional to deep learning based methods (Sun et al. 2020; Daranagama and Witayangkurn 2021; Song and Jiang 2021). The traditional approaches can be summarized into: (1) post-classification methods; and (2) direct classification methods. The first type of method independently produces multi-temporal building maps from independent classifications, and then the changed buildings are identified by comparing these maps (Wen et al. 2019). The second type of method accomplishes this task by directly measuring the change magnitude (Huang et al. 2014). Both approaches have their respective advantages and disadvantages. For instance, illumination variations and spatial misalignment are major obstacles for the first category of approaches, and the accumulation of errors from each independent building detection is also a challenge for the second category (Li, Huang, and Chang 2020). Supervised deep learning methods with

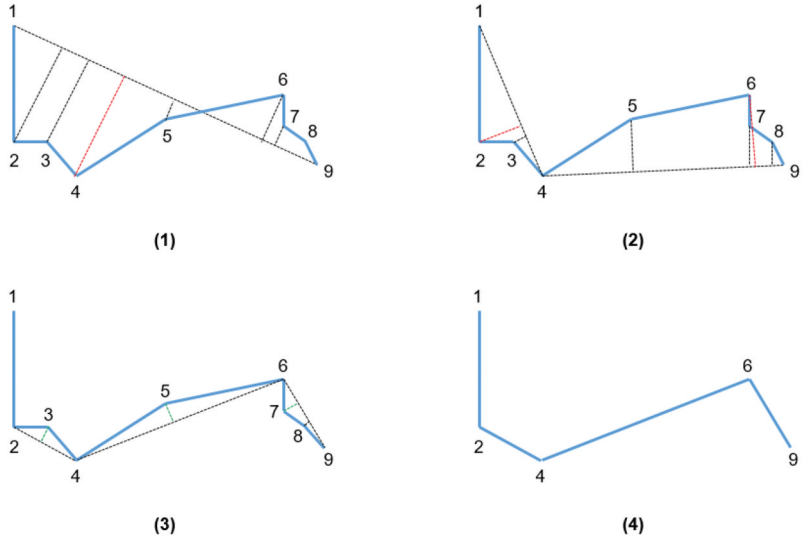

Figure 9. An illustration of the Douglas-Peucker algorithm (Douglas and Peucker 1973).

massive labeled data can achieve more promising performances. According to the method used to manage the bi-temporal images, these studies can be divided into early fusion methods and late fusion methods (Daudt, Le Saux, and Boulch 2018). In the early fusion methods, VHR optical images collected at different times are stacked and then fed into a semantic segmentation based network to determine the change magnitude. In the late fusion methods, the building features of bi-temporal images are separately extracted by two independent encoders, and then combined and finally fed into a decoder to generate a change map. Based on these two architectures, several attempts have been made to formulate a specific network to deal with the problems such as the complicated backgrounds in dense urban scenes (Liu et al. 2021a), pseudo-changes such as roofs with different colors, vehicles, and containers (Song and Jiang 2021), geometric misalignment due to relief displacement (Zhang et al. 2021), the rarity and sparsity of changed building samples (Daranagama and Witayangkurn 2021), and the spectral differences and scale variations of bi-temporal images (Liu et al. 2021a). It is worth noting that the uncertainty of multi-temporal VHR georeferencing can introduce significant errors. Even if, on the whole, the VHR images achieve a subpixel georeferencing accuracy, the spatial misalignment of high-rise buildings will still be serious. To deal with this issue, georeferencing error tolerated methods that make full use of the spatial contextual information have been used for VHR building change detection (Huang et al. 2014). Moreover, when the training samples can inform us whether the object of interest is a building object at each time point, a multi-task network that simultaneously performs the building change detection and building detection can further improve the change detection performance (Sun et al. 2020).

Compared to the studies that have explored building change detection in the two-dimensional (2D) domain (e.g. Yu et al. 2016; Huang et al. 2017a), three-dimensional (3D) building change extraction remains a relatively new topic. 3D building information refers to the height and volume of the buildings. In addition, some other building parameters, such as the floor area ratio and sky view factor, can also be calculated. The typical 3D information can be obtained from various data sources, including LiDAR point cloud data, digital elevation models, multi-view images, and even monocular images. The building change detection task can be expanded to 3D space, in applications such as building construction process tracking, urban vertical growth monitoring, and 3D city model updating (Wen et al. 2019). Although 3D building change detection has great potential in various applications, some challenges still remain. The first issue is the uncertainty of the 3D data. For instance, with respect to multi-view images, the image matching may fail to detect high-rise buildings, and thus result in inaccurate height information. Accordingly, more advanced algorithms are still needed to improve the accuracy of 3D data. Furthermore, the access to multi-temporal 3D data is often costly. It is thus necessary to develop methods for time-series 3D building data generation and updating in a relatively economic manner. An alternative solution may be the fusion of multi-temporal stereo images (e.g. ZY-3) and an accurate 3D building model derived at one specific date.

4.3. Building type classification

Most of the existing research has focused on building detection (Liu et al. 2021a), i.e. distinguishing building objects from non-building objects. In recent applications, building types rather than the location and geometric information have been the focus, in various aspects, such as energy consumption modeling, disaster risk assessment, and seismic building vulnerability evaluation (Wurm, Schmitt, and Taubenböck 2017). The aim of building type classification is to classify the buildings into different classes according to their characteristics. For instance, according to their height, buildings can be classified into low-, mid-, and high-rise buildings; they can also be categorized into block developments, terraced buildings, and detached buildings, based on their geometry and morphology; and they can be divided into residential, commercial, and industrial buildings, considering their urban functions (Du, Zhang, and Zhang 2015; Bo, Bei, and Song 2018). Generally speaking, compared to building type classification in terms of building height and morphology, identifying the semantic type (e.g. residential and commercial) can be more challenging, owing to the semantic gap between the image features and the building function in the real world (Taubenböck et al. 2013).

A few studies have investigated building type classification from remote sensing data, for which the framework usually includes two main steps: (1) building extraction; and (2) type classification. Since the precise extraction of building objects is an essential prerequisite for building type classification, the existing studies have tended to use LiDAR and GIS data that can provide accurate building footprints and height (Mariana et al. 2014). 2D and 3D metrics, such as area, perimeter, and volume, have then been applied to describe the geometric characteristics of the different building types (Wurm, Schmitt, and Taubenböck 2017l; Huang et al., 2021a). With respect to building function classification, multi-source data fusion is an appropriate approach. For instance, Du, Zhang, and Zhang (2015) undertook building semantic classification by integrating high-resolution optical images and GIS data. Various features, i.e. the spectra, texture, and geometry, were then combined with an improved RF classifier, which was used to classify the buildings according to their function, such as apartments, and industrial buildings. Huang, Chen, and Gong (2018a) employed multi-view images and proposed a series of angular difference features at multiple levels that can not only reflect the height information of buildings, but can also provide the possibility to discriminate buildings with similar heights. In their study, the results demonstrated that the local angular variations are beneficial for semantic building classification, e.g. distinguishing residential apartments, factories, and cottages.

4.4. Building height retrieval from monocular images

Height, which is one of the most significant characteristics of buildings, is of great importance in understanding building morphology and function within urban scenes (Amirkolaee et al. 2019). Currently, height information can be obtained from LiDAR point cloud data or multi-view images using dense matching (Saeidi et al. 2014). However, such datasets are not always available, and monocular optical images are still the most dominant data source for building information extraction (Karatsiolis, Kamilaris, and Cole 2021). Thus, a promising research direction would be to estimate building height information from monocular images, in the absence of other auxiliary data sources (e.g. LiDAR and stereo images) (Liasis and Stavrou 2016).

One useful clue that can be considered for height estimation is the contextual information, e.g. the shadows cast by buildings (Bosch et al. 2019). Generally speaking, a geometric prior between a building and its corresponding shadow can be quantitatively modeled (e.g. by a linear function) by the elevation angles of the sun and satellite and the length of the shadow (Liasis and Stavrou 2016). By establishing such a relationship, the building height can be estimated. In this case, accurate shadow detection is a critical basis for building height retrieval. Some widely used methods for shadow detection are the histogram threshold technique, invariant color models, morphological transformation, and the active contour model (Huang and Zhang 2012; Ok 2013; Liasis and Stavrou 2016). It should be noted that building height estimation based on shadow information is more suitable for scenes with sparsely distributed buildings, so that the shadow structures can be completely observed. In fact, high-density building regions can result in fragmented and incomplete shadow components, due to the occlusions, which can seriously affect the accuracy of building height retrieval.

In addition to shadow-based methods, the recently developed deep learning models have great potential for building height retrieval from monocular images. Generally speaking, height estimation from monocular images can be technically difficult as there is an inherent ambiguity in transforming the intensity or color properties into height information (Mou and Zhu 2018). However, it is easier for humans to infer the depth information from monocular images in terms of the visual cues, including object size, texture, context, occlusion, and orientation (Amirkolaee and Arefi 2019). The deep learning models have great potential to extract mid- and high-level abstract features that can serve as cues for depth perception, and thus have potential for height estimation from monocular images. To date, limited efforts have been made to investigate this issue (Amirkolaee and Arefi 2019), where the deep network architectures were made up of encoding and decoding parts, which were used for abstract feature extraction and height value transformation, respectively. These frameworks achieved reasonable accuracies with respect to low- and mid-rise buildings, but some challenges still remain. Firstly, the accuracy for high-rise buildings was not fully satisfactory. An alternative solution may be the fusion of traditional features and deep features. For instance, the contextual

information, e.g. shadow properties, combined with multispectral features, could be jointly used as inputs for deep learning models. The other problem is the reliance on prepared training data, e.g. reference nDSM data. Under this circumstance, the transferability of a deep learning model should be further investigated, in order to improve its efficiency for large-area applications.

5. Conclusion

Building extraction from VHR optical images is an essential but challenging research topic, and much effort has been made in this field. However, a deep and comprehensive review of this topic is still lacking in the current research community. Accordingly, in this article, we have presented a review of the recent advances (since 2000) in building extraction from VHR optical images. We surveyed a large number of studies and summarized them in terms of the workflow of building detection from high-resolution images, including the detection method, post-processing, and accuracy assessment. Specifically, the various methods for building detection were categorized into physical rule based methods, image segmentation based methods, and traditional and advanced machine learning (i.e. deep learning) methods. Furthermore, we further discussed multi-source data fusion for building detection (i.e. the fusion of optical images and LiDAR/SAR data). Finally, we suggested four promising research directions, i.e. building polygon delineation, detailed building type classification according to the building morphology and function, building height retrieval from monocular images, and building change detection. It is our hope that that this review will provide researchers with a better understanding of the issues in building extraction from VHR optical images.

Notes

1. https://www.openstreetmap.org/
2. https://www.linz.govt.nz/data/linz-data
3. https://github.com/PaddlePaddle/PaddleSeg/blob/develop/EISeg/docs/remote_sensing.md

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Special Fund of Hubei Luojia Laboratory under Grant 220100031, the National Natural Science Foundation of China under Grants 42071311 and 41971295, the Foundation for Innovative Research Groups of the Natural Science Foundation of Hubei Province under Grant 2020CFA003, the Wuhan 2022 Shuguang Project under Grants 2022010801020123.

Data availability statement

The data used to generate Figures 1 and 2 are provided in Supplementary Material I. The details of the online database as well as the strategies used to query the 417 articles are detailed in Supplementary Material II. The detailed technical steps involved in generating the figures are described in Supplementary Material III. The data used to assess the performance of the mainstream detection techniques are provided in Supplementary Material IV. The data used to generate Figures 3–5 and Figures 7–8 are available from the corresponding author by request. The data used to generate Figure 6 were provided by a third party, and have been listed in Table 1 of the submitted article.

References

Acuna, D., H. Ling, A. Kar, and S. Fidler 2018. "Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++." In 2018 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, Salt Lake City, USA. 859–868. doi:10.1109/CVPR.2018.00096.

Ahmed, N., R. M. Rahman, M. S. G. Adnan, and B. Ahmed. 2021. "Dense Prediction of Label Noise for Learning Building Extraction from Aerial Drone Imagery." *International Journal of Remote Sensing* 42: 8906–8929. doi:10.1080/01431161.2021.1973685.

Akinlar, C., and C. Topal. 2011. "EDLines: A real-time Line Segment Detector with A False Detection Control." *Pattern Recognition Letters* 32: 1633–1642. doi:10.1016/j.patrec.2011.06.001.

Alshehhi, R., P. R. Marpu, W. L. Woon, and M. D. Mura. 2017. "Simultaneous Extraction of Roads and Buildings in Remote Sensing Imagery with Convolutional Neural Networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 139–149. doi:10.1016/j.isprsjprs.2017.05.002.

Amirkolaee, H. A., and H. Arefi. 2019. "Height Estimation from Single Aerial Images Using a Deep Convolutional encoder-decoder Network." *ISPRS Journal of Photogrammetry and Remote Sensing* 149: 50–66. doi:10.1016/j.isprsjprs.2019.01.013.

Attarzadeh, R., and M. Momeni. 2018. "Object-Based Rule Sets and Its Transferability for Building Extraction from High Resolution Satellite Imagery." *Journal of the Indian Society of Remote Sensing* 46: 169–178. doi:10.1007/s12524-017-0694-6.

Bialas, J., T. Oommen, and T. C. Havens. 2019. "Optimal Segmentation of High Spatial Resolution Images for the Classification of Buildings Using Random Forests, Int." *International Journal of Applied Earth Observation and Geoinformation* 82: 101895. doi:10.1016/j.jag.2019.06.005.

Bo, H., Z. Bei, and Y. Song. 2018. "Urban land-use Mapping Using a Deep Convolutional Neural Network with High Spatial Resolution Multispectral Remote Sensing Imagery." *Remote Sensing of Environment* 214: 73–86. doi:10.1016/j.rse.2018.04.050.

Bosch, M., K. Foster, G. Christie, S. Wang, G. D. Hager, and M. Brown, 2019. Semantic Stereo for Incidental Satellite Images. Proc. - IEEE Winter Conf. Appl. Comput. Vision, WACV, Beijing China, 1524–1532. doi:10.1109/WACV.2019.00167.

Brenner, C. 2005. "Building Reconstruction from Images and Laser Scanning." *International Journal of Applied Earth Observation and Geoinformation* 6: 187–198. doi:10.1016/j.jag.2004.10.006.

Canny, J. 1986. "A Computational Approach to Edge Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8: 679–698. doi:10.1109/TPAMI.1986.4767851.

Cao, S., Q. Weng, M. Du, B. Li, R. Zhong, and Y. Mo. 2020. "Multi-scale three-dimensional Detection of Urban Buildings Using Aerial LiDAR Data." *GIScience & Remote Sensing* 57 (8): 1125–1143. doi:10.1080/15481603.2020.1847453.

Cao, Y., and X. Huang. 2021. "A Deep Learning Method for Building Height Estimation Using high-resolution multi-view Imagery over Urban Areas: A Case Study of 42 Chinese Cities." *Remote Sensing of Environment* 264: 112590. doi:10.1016/j.rse.2021.112590.

Castrejón, L., K. Kundu, R. Urtasun, and S. Fidler 2017. "Annotating Object Instances with a polygon-RNN." In Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, Honolulu, HI, USA. 4485–4493. doi:10.1109/CVPR.2017.477.

Chandra, N., and J. K. Ghosh. 2018. "A Cognitive Viewpoint on Building Detection from Remotely Sensed Multispectral Images." *IETE Journal of Research* 64: 165–175. doi:10.1080/03772063.2017.1351320.

Chandra, N. 2022. A Review of Building Detection Methods from Remotely Sensed Images." https://www.currentscience.ac.in/data/forthcoming/414.pdf

Chattopadhyay, S., and A. C. Kak. 2022. "Uncertainty, Edge, and Reverse-Attention Guided Generative Adversarial Network for Automatic Building Detection in Remotely Sensed Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15: 3146–3167. doi:10.1109/JSTARS.2022.3166929.

Chen, L. C., T. A. Teo, J. Y. Wen, and J. Y. Rau. 2007. "Occlusion Compensated True Orthorectification for high-resolution Satellite Images." *The Photogrammetric Record* 22: 39–52. doi:10.1111/j.1477-9730.2007.00416.x.

Chen, Q., L. Wang, S. L. Waslander, and X. Liu. 2020. "An end-to-end Shape Modeling Framework for Vectorized Building Outline Generation from Aerial Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 170: 114–126. doi:10.1016/j.isprsjprs.2020.10.008.

Chen, M., J. Wu, L. Liu, W. Zhao, F. Tian, Q. Shen, B. Zhao, and R. Du. 2021. "Dr-net: An Improved Network for Building Extraction from High Resolution Remote Sensing Image." *Remote Sens* 13: 1–19. doi:10.3390/rs13020294.

Chen, J., Y. Jiang, L. Luo, and W. Gong. 2022. "ASF-Net: Adaptive Screening Feature Network for Building Footprint Extraction from Remote-Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–13. doi:10.1109/TGRS.2022.3165204.

Cheng, G., and J. Han. 2016. "A Survey on Object Detection in Optical Remote Sensing Images." *ISPRS J. Photogramm. Remote Sens* 117: 11–28. doi:10.1016/j.isprsjprs.2016.03.014.

Cheng, G., and J. Han. 2016. "A Survey on Object Detection in Optical Remote Sensing Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 117: 11–28. doi:10.1016/j.isprsjprs.2016.03.014.

Cheng, D., R. Liao, S. Fidler, and R. Urtasun 2019. "Darnet: Deep Active Ray Network for Building Segmentation." In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Long Beach, CA, USA. 7423–7431. doi:10.1109/CVPR.2019.00761

Cohen, J. P., W. Ding, C. Kuhlman, A. Chen, and L. Di. 2016. "Rapid Building Detection Using Machine Learning." *Applied Intelligence* 45: 443–457. doi:10.1007/s10489-016-0762-6.

Daranagama, S., and A. Witayangkurn. 2021. "Automatic Building Detection with Polygonizing and Attribute Extraction from High-Resolution Images." *ISPRS Int. J. Geo-Information* 10. doi:10.3390/ijgi10090606.

Daudt, R., B. Le Saux, and A. Boulch 2018. "Fully Convolutional Siamese Networks for Change Detection." In 2018 25th IEEE Conf. Image Process. ICIP, Quebec, Canada, 4063–4067. doi:10.1109/ICIP.2018.8451652.

Dawen, Y., S. Ji, J. Liu, and S. Wei. 2021. "Automatic 3D Building Reconstruction from multi-view Aerial Images with Deep Learning." *ISPRS Journal of Photogrammetry and Remote Sensing* 171: 155–170. doi:10.1016/j.isprsjprs.2020.11.011.

Deng, L., Y. N. Yan, Y. He, Z. H. Mao, and J. Yu. 2019. "An Improved Building Detection Approach Using L-band POLSAR two-dimensional time-frequency Decom position over Oriented built-up Areas." *GIScience & Remote Sensing* 56 (1): 1–21. doi:10.1080/15481603.2018.1484409.

Ding, L., H. Tang, Y. Liu, Y. Shi, X. X. Zhu, and L. Bruzzone 2021a. "Adversarial Shape Learning for Building Extraction in VHR Remote Sensing Images." In IEEE Trans. Image Process, 7149, 1–13. doi:10.1109/TIP.2021.3134455.

Ding, Q., Z. Shao, X. Huang, and O. Altan. 2021b. "DSA-Net: A Novel Deeply Supervised attention-guided Network for Building Change Detection in high-resolution Remote Sensing Images." *International Journal of Applied Earth Observation and Geoinformation* 105. doi:10.1016/j.jag.2021.102591.

Dornaika, F., A. Moujahid, Y. El Merabet, and Y. Ruichek. 2016. "Building Detection from Orthophotos Using a Machine Learning Approach: An Empirical Study on Image Segmentation and Descriptors." *Expert Systems with Applications* 58: 130–142. doi:10.1016/j.eswa.2016.03.024.

Douglas, H., and K. Peucker. 1973. "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature." *Cartographica: the International Journal for Geographic Information and Geovisualization* 10 (2): 112–122. doi:10.1002/9780470669488.ch2.

Du, S., F. Zhang, and X. Zhang. 2015. "Semantic Classification of Urban Buildings Combining VHR Image and GIS Data: An Improved Random Forest Approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 105: 107–119. doi:10.1016/j.isprsjprs.2015.03.011.

Feng, D., Y. Xie, S. Xiong, J. Hu, M. Hu, Q. Li, and J. Zhu. 2021. "Regularized Building Boundary Extraction from Remote Sensing Imagery Based on Augment Feature Pyramid Network and Morphological Constraint." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 12212–12223. doi:10.1109/JSTARS.2021.3130038.

Ferrari, L., F. Dell'acqua, P. Zhang, and P. Du. 2021. "Integrating EfficientNet into an HAFNet Structure for Building Mapping in high-resolution Optical Earth Observation Data." *Remote Sensing* 13: 1–15. doi:10.3390/rs13214361.

Gao, X., M. Wang, Y. Yang, and G. Li. 2018. "Building Extraction From RGB VHR Images Using Shifted Shadow Algorithm." *IEEE Access* 6: 22034–22045. doi:10.1109/ACCESS.2018.2819705.

Ghanea, M., P. Moallem, and M. Momeni. 2016. "Building Extraction from high-resolution Satellite Images in Urban Areas: Recent Methods and Strategies against Significant Challenges." *International Journal of Remote Sensing* 37: 5234–5248. doi:10.1080/01431161.2016.1230287.

Gilani, S. A. N., M. Awrangjeb, and G. Lu. 2018. "Segmentation of Airborne Point Cloud Data for Automatic Building Roof Extraction." *GIScience & Remote Sensing* 55 (1): 63–89. doi:10.1080/15481603.2017.1361509.

Gonzalez, R. C., R. E. Woods, and B. R. Masters 2009. "Digital Image Processing."

Gruen, A. 2012. "Development and Status of Image Matching in Photogrammetry." *The Photogrammetric Record* 27: 36–57. doi:10.1111/j.1477-9730.2011.00671.x.

Guo, Z., S. Du, M. Li, and W. Zhao. 2016. "Exploring GIS Knowledge to Improve Building Extraction and Change Detection from VHR Imagery in Urban Areas." *International Journal of Image and Data Fusion* 7: 42–62. doi:10.1080/19479832.2015.1051138.

Guo, Z., and S. Du. 2017. "Mining Parameter Information for Building Extraction and Change Detection with Very high-resolution Imagery and GIS Data." *GIScience & Remote Sensing* 54: 38–63. doi:10.1080/15481603.2016.1250328.

Haala, N., and M. Kada. 2010. "An Update on Automatic 3D Building Reconstruction." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 570–580. doi:10.1016/j.isprsjprs.2010.09.006.

Hao, Y., Y. Liu, Z. Wu, L. Han, Y. Chen, G. Chen, L. Chu, et al. 2021. "EdgeFlow: Achieving Practical Interactive Segmentation with Edge-Guided Flow." In Proceedings of the IEEE International Conference on Computer Vision. 1551–1560. doi:10.1109/ICCVW54120.2021.00180

Huang, X., and L. Zhang. 2008. "An Adaptive mean-shift Analysis Approach for Extraction and Classification from Urban Hyperspectral Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 46: 4173–4185. doi:10.1109/TGRS.2008.2002577.

Huang, X, and Zhang, L. 2011. "A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery." *Photogrammetric Engineering & Remote Sensing* 77: 721–732. doi:10.14358/PERS.77.7.721.

Huang, X., and L. Zhang. 2011. "A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery." *Photogrammetric Engineering and Remote Sensing* 77: 721–732. doi:10.14358/PERS.77.7.721.

Huang, X., and L. Zhang. 2012. "Morphological Building/Shadow Index for Building Extraction from High-Resolution Imagery over Urban Areas." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5: 161–172. doi:10.1109/JSTARS.2011.2168195.

Huang, X., T. Zhu, L. Zhang, and Y. Tang. 2014. "A Novel Building Change Index for Automatic Building Change Detection from high-resolution Remote Sensing Imagery." *Remote Sensing Letters* 5: 713–722. doi:10.1080/2150704X.2014.963732.

Huang, X., D. Wen, J. Li, and R. Qin. 2017a. "Multi-level Monitoring of Subtle Urban Changes for the Megacities of China Using high-resolution multi-view Satellite Imagery." *Remote Sensing of Environment* 196: 56–75. doi:10.1016/j.rse.2017.05.001.

Huang, X., W. Yuan, J. Li, and L. Zhang. 2017b. "A New Building Extraction Postprocessing Framework for High-Spatial-Resolution Remote-Sensing Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 654–668. doi:10.1109/JSTARS.2016.2587324.

Huang, X., and T. Zhang 2018. "Morphological Building Index (MBI) and Its Applications to Urban Areas." 33–49.

Huang, X., H. Chen, and J. Gong. 2018a. "Angular Difference Feature Extraction for Urban Scene Classification Using ZY-3 multi-angle high-resolution Satellite Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 135: 127–141. doi:10.1016/j.isprsjprs.2017.11.017.

Huang, X., T. Hu, J. Li, Q. Wang, and J. A. Benediktsson. 2018b. "Mapping Urban Areas in China Using Multisource Data with a Novel Ensemble SVM Method." *IEEE Transactions on Geoscience and Remote Sensing* 56: 4258–4273. doi:10.1109/TGRS.2018.2805829.

Huang, X., and Y. Wang. 2019. "Investigating the Effects of 3D Urban Morphology on the Surface Urban Heat Island Effect in Urban Functional Zones by Using high-resolution Remote Sensing Data: A Case Study of Wuhan, Central China." *ISPRS Journal of Photogrammetry and Remote Sensing* 152: 119–131. doi:10.1016/j.isprsjprs.2019.04.010.

Huang, X., Y. Cao, and J. Li. 2020. "An Automatic Change Detection Method for Monitoring Newly Constructed Building Areas Using time-series multi-view high-resolution Optical Satellite Images." *Remote Sensing of Environment* 244: 111802. doi:10.1016/j.rse.2020.111802.

Huang, X., S. Li, J. Li, X. Jia, J. Li, X. X. Zhu, and J. A. Benediktsson. 2021a. "A Multispectral and Multiangle 3-D Convolutional Neural Network for the Classification of ZY-3 Satellite Images over Urban Areas." *IEEE Transactions on Geoscience and Remote Sensing* 59: 10266–10285. doi:10.1109/TGRS.2020.3037211.

Huang, X., J. Yang, J. Li, and D. Wen. 2021b. "Urban Functional Zone Mapping by Integrating High Spatial Resolution Nighttime Light and Daytime multi-view Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 175: 403–415. doi:10.1016/j.isprsjprs.2021.03.019.

Huang, W., H. Tang, and P. Xu 2022. "OEC-RNN: Object-Oriented Delineation of Rooftops with Edges and Corners Using the Recurrent Neural Network from the Aerial Images." IEEE Trans. Geosci. Remote Sens, 60, 1–12. doi:10.1109/TGRS.2021.3076098.

Ji, S., S. Wei, and M. Lu. 2019. "Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set." *IEEE Transactions on Geoscience and Remote Sensing* 57: 574–586. doi:10.1109/TGRS.2018.2858817.

Ji, S., S. Wei, and M. Lu. 2019. "Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set." *IEEE Transactions on Geoscience and Remote Sensing : A Publication of the IEEE Geoscience and Remote Sensing Society* 57: 574–586. doi:10.1109/TGRS.2018.2858817.

Jiang, N., J. X. Zhang, H. T. Li, and X. G. Lin 2008. "Semi-automatic Building Extraction from High Resolution Imagery Based on Segmentation." In 2008 International Workshop on Earth Observation and Remote Sensing Applications, Beijing, China, 1–5. doi:10.1109/EORSA.2008.4620311.

Jiang, X., X. Zhang, Q. Xin, X. Xi, and P. Zhang. 2020. "Arbitrary-Shaped Building Boundary-Aware Detection with Pixel Aggregation Network." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 1–13. doi:10.1109/JSTARS.2020.3017934.

Jozdani, S., and D. Chen. 2020. "On the Versatility of Popular and Recently Proposed Supervised Evaluation Metrics for Segmentation Quality of Remotely Sensed Images: An Experimental Case Study of Building Extraction." *ISPRS Journal of Photogrammetry and Remote Sensing* 160: 275–290. doi:10.1016/j.isprsjprs.2020.01.002.

Jung, H., H. Choi, and M. Kang, 2021. Boundary Enhancement Semantic Segmentation for Building Extraction from Remote Sensed Image. IEEE Trans. Geosci. Remote Sens. doi:10.1109/TGRS.2021.3108781.

Karatsiolis, S., A. Kamilaris, and I. Cole. 2021. "IMG2nDSM: Height Estimation from Single Airborne RGB Images with Deep Learning." *Remote Sensing* 13: 2417. doi:10.3390/rs13122417.

Krayenhoff, E. S., M. Moustaoui, A. M. Broadbent, V. Gupta, and M. Georgescu. 2018. "Diurnal Interaction between Urban Expansion, Climate Change and Adaptation in US Cities." *Nature Climate Change* 8: 1097–1103. doi:10.1038/s41558-018-0320-9.

Lee, D. H., K. M. Lee, and S. U. Lee. 2008. "Fusion of Lidar and Imagery for Reliable Building Extraction." *Photogrammetric Engineering & Remote Sensing* 74: 215–225. doi:10.14358/PERS.74.2.215.

Li, Y., L. Zhu, H. Shimamura, and K. Tachibana 2010. "An Integrated System on Large Scale Building Extraction from DSM." Image (Rochester, N.Y.) XXXVIII, 35–39.

Li, J., H. Zhang, and L. Zhang. 2014. "Supervised Segmentation of Very High Resolution Images by the Use of Extended Morphological Attribute Profiles and a Sparse Transform." *IEEE Geoscience and Remote Sensing Letters* 11: 1409–1413. doi:10.1109/LGRS.2013.2294241.

Li, B., K. Cheng, and Z. Yu. 2016. "Histogram of Oriented Gradient Based GIST Feature for Building Recognition." *Computational Intelligence and Neuroscience* 2016: 1–9. doi:10.1155/2016/6749325.

Li, E., S. Xu, W. Meng, and X. Zhang. 2017a. "Building Extraction from Remotely Sensed Images by Integrating Saliency Cue." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 906–919. doi:10.1109/JSTARS.2016.2603184.

Li, Y., B. He, L. Teng, and X. Bai 2017b. "Evaluation the Performance of Fully Convolutional Networks for Building Extraction Compared with Shallow Models." In IGARSS 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, Texas, USA.

Li, Z., J. Di Wegner, and A. Lucchi 2019. "Topological Map Extraction from Overhead Images." In Proceedings of the IEEE International Conference on Computer Vision. Institute of Electrical and Electronics Engineers Inc, Seoul, South Kerean, 1715–1724. 10.1109/ICCV.2019.00180

Li, J., X. Huang, and X. Chang. 2020. "A label-noise Robust Active Learning Sample Collection Method for multi-temporal Urban land-cover Classification and Change Analysis." *ISPRS Journal of Photogrammetry and Remote Sensing* 163: 1–17. doi:10.1016/j.isprsjprs.2020.02.022.

Li, Q., S. Zorzi, Y. Shi, F. Fraundorfer, and X. X. Zhu. 2022. "RegGAN: An End-to-End Network for Building Footprint Generation with Boundary Regularization." *Remote Sensing* 14 (8): 1835. doi:10.3390/rs14081835.

Liasis, G., and S. Stavrou. 2016. "Building Extraction in Satellite Images Using Active Contours and Colour Features." *International Journal of Remote Sensing* 37: 1127–1153. doi:10.1080/01431161.2016.1148283.

Liu, Z., S. Cui, and Q. Yan 2008. "Building Extraction from High Resolution Satellite Imagery Based on multi-scale Image Segmentation and Model Matching." In 2008 International Workshop on Earth Observation and Remote Sensing Applications, Beijing, China, 1–7. doi:10.1109/EORSA.2008.4620321.

Liu, C., X. Huang, Z. Zhu, H. Chen, X. Tang, and J. Gong. 2019. "Automatic Extraction of built-up Area from ZY3 multi-view Satellite Imagery: Analysis of 45 Global Cities." *Remote Sensing of Environment* 226: 51–73. doi:10.1016/j.rse.2019.03.033.

Liu, T., M. Gong, D. Lu, Q. Zhang, H. Zheng, F. Jiang, and M. Zhang. 2021a. "Building Change Detection for VHR Remote Sensing Images via Local-Global Pyramid Network and Cross-Task Transfer Learning Strategy." *IEEE Transactions on Geoscience and Remote Sensing* 2892: 1–1. doi:10.1109/tgrs.2021.3130940.

Liu, Y., D. Chen, A. Ma, Y. Zhong, F. Fang, and K. Xu. 2021b. "Multiscale U-Shaped CNN Building Instance Extraction Framework with Edge Constraint for High-Spatial-Resolution Remote Sensing Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 59: 6106–6120. doi:10.1109/TGRS.2020.3022410.

Manno-Kovacs, A., and T. Sziranyi. 2015. "Orientation-selective Building Detection in Aerial Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 108: 94–112. doi:10.1016/j.isprsjprs.2015.06.007.

Mariana, B., T. Ivan, L. Thomas, B. Thomas, and H. Bernhard. 2014. "Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data." *Remote Sensing* 6: 1347–1366. doi:10.3390/rs6021347.

Mayer, H. 1999. "Automatic Object Extraction from Aerial imagery —a Survey Focusing on Buildings." *Computer Vision and Image Understanding* 74: 138–149. doi:10.1006/cviu.1999.0750.

Mayunga, S., Y. Zhang, and D. Coleman. 2005. "Semi-automatic Building Extraction Utilizing QuickBird Imagery." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36. doi:10.17265/2162-5263/2016.04.004.

Mayunga, S. D., D. J. Coleman, and Y. Zhang. 2007. "A semi-automated Approach for Extracting Buildings from QuickBird Imagery Applied to Informal Settlement Mapping." *International Journal of Remote Sensing* 28: 2343–2357. doi:10.1080/01431160600868474.

Mayunga, S. D., D. J. Coleman, and Y. Zhang. 2010. "Semi-automatic Building Extraction in Dense Urban Settlement Areas from high-resolution Satellite Images." *Survey Review* 42: 50–61. doi:10.1179/003962609X451690.

Mishra, A., A. Pandey, and A. S. Baghel. 2016. "Building Detection and Extraction Techniques: A Review." In 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 3816–3821.

Mou, L., and X. Zhu 2018. "IM2HEIGHT: Height Estimation from Single Monocular Imagery via Fully Residual Convolutional-Deconvolutional Network."

Müller, S., and D. Zaum 2005. "Robust Building Detection in Aerial Images." 36.

Navab, N., J. Hornegger, W. M. Wells, and A. F. Frangi 2015. "Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015." In 18th International Conference Munich, Germany, October 5- 9,2015 proceedings, part III. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Munich, Germany, 9351, 12–20. doi:10.1007/978-3-319-24574-4.

Ngo, -T.-T., V. Mazet, C. Collet, and P. de Fraipont. 2017. "Shape-Based Building Detection in Visible Band Images Using Shadow Information." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 920–932. doi:10.1109/JSTARS.2016.2598856.

Ok, A. O., C. Senaras, and B. Yuksel. 2013. "Automated Detection of Arbitrarily Shaped Buildings in Complex Environments from Monocular VHR Optical Satellite Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 51: 1701–1717. doi:10.1109/TGRS.2012.2207123.

Ok, A. O. 2013. "Automated Detection of Buildings from Single VHR Multispectral Images Using Shadow Information and Graph Cuts." *ISPRS Journal of Photogrammetry and Remote Sensing* 86: 21–40. doi:10.1016/j.isprsjprs.2013.09.004.

Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder. 2014. "Good Practices for Estimating Area and Assessing Accuracy of Land Change." *Remote Sensing of Environment* 148: 42–57. doi:10.1016/j.rse.2014.02.015.

Paci, F., M. Chini, and W. J. Emery. 2009. "A Neural Network Approach Using multi-scale Textural Metrics from Very high-resolution Panchromatic Imagery for Urban land-use Classification." *Remote Sensing of Environment* 113: 1276–1292. doi:10.1016/j.rse.2009.02.014.

Papadomanolaki, M., M. Vakalopoulou, and K. Karantzalos 2021. "A Deep Multitask Learning Framework Coupling Semantic Segmentation and Fully Convolutional LSTM Networks for Urban Change Detection." In IEEE Trans. Geosci. Remote Sens, 59, 7651–7668. doi:10.1109/TGRS.2021.3055584.

Partovi, T., R. Bahmanyar, T. Kraus, and P. Reinartz. 2017. "Building Outline Extraction Using a Heuristic Approach Based on Generalization of Line Segments." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 933–947. doi:10.1109/JSTARS.2016.2611861.

Pesaresi, M., A. Gerhardinger, and F. Kayitakire. 2008. "A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1: 180–192. doi:10.1109/JSTARS.2008.2002869.

Ps, P., and B. H. Aithal. 2022. "Building Footprint Extraction from Very high-resolution Satellite Images Using Deep Learning." *Journal of Spatial Science* 1-17. doi:10.1080/14498596.2022.2037473.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas 2017. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation." In Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognition, Honolulu, HI, USA, 77–85. doi:10.1109/CVPR.2017.16.

Qin, R. 2014. "Change Detection on LOD 2 Building Models with Very High Resolution Spaceborne Stereo Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 96: 179–192. doi:10.1016/j.isprsjprs.2014.07.007.

Qin, R., J. Tian, and P. Reinartz. 2016a. "3D Change Detection – Approaches and Applications." *ISPRS Journal of Photogrammetry and Remote Sensing* 122: 41–56. doi:10.1016/j.isprsjprs.2016.09.013.

Qin, R., J. Tian, and P. Reinartz. 2016b. "Spatiotemporal Inferences for Use in Building Detection Using Series of very-high-resolution space-borne Stereo Images." *International Journal of Remote Sensing* 37: 3455–3476. doi:10.1080/01431161.2015.1066527.

Saeidi, V., B. Pradhan, M. O. Idrees, and Z. A. Latif. 2014. "Fusion of Airborne LiDAR with Multispectral SPOT 5 Image for Enhancement of Feature Extraction Using Dempster-Shafer Theory." *IEEE Transactions on Geoscience and Remote Sensing* 52: 6017–6025. doi:10.1109/TGRS.2013.2294398.

Senaras, C., M. Ozay, and F. T. Yarman Vural. 2013. "Building Detection With Decision Fusion." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6: 1295–1304. doi:10.1109/JSTARS.2013.2249498.

Shen, Y., T. Ai, and C. Li. 2019. "A Simplification of Urban Buildings to Preserve Geometric Properties Using Superpixel Segmentation." *International Journal of Applied Earth Observation and Geoinformation* 79: 162–174. doi:10.1016/j.jag.2019.02.008.

Sirmacek, B. 2011. "Graph Theory and Mean Shift Segmentation Based Classification of Building Facades." In 2011 Joint Urban Remote Sensing Event, Munich, Germany, 409–412. doi:10.1109/JURSE.2011.5764806.

Song, K., and J. Jiang. 2021. "AGCDetNet:AnAttention-Guided Network for Building Change Detection in High-Resolution Remote Sensing Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 4816–4831. doi:10.1109/JSTARS.2021.3077545.

Sritarapipat, T., and W. Takeuchi. 2017. "Building Classification in Yangon City, Myanmar Using Stereo GeoEye Images, Landsat Image and night-time Light Data." *Remote Sensing Applications: Society and Environment* 6: 46–51. doi:10.1016/j.rsase.2017.04.001.

Sun, K., B. Xiao, D. Liu, and J. Wang 2019. "Deep high-resolution Representation Learning for Human Pose Estimation." In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, 2019-June, Long Beach, CA, USA, 5686–5696. doi:10.1109/CVPR.2019.00584.

Sun, Y., X. Zhang, J. Huang, H. Wang, and Q. Xin. 2020. "Fine-Grained Building Change Detection from Very High-Spatial-Resolution Remote Sensing Images Based on Deep Multitask Learning." *IEEE Geoscience and Remote Sensing Letters* 19: 1–5. doi:10.1109/lgrs.2020.3018858.

Swan, B., M. Laverdiere, H. Yang, and A. Rose. 2022. "Iterative self-organizing SCEne-LEvel Sampling (ISOSCELES) for large-scale Building Extraction." *GIScience & Remote Sensing* 59: 1–16. doi:10.1080/15481603.2021.2006433.

Taha, L. G. E. D., and R. E. Ibrahim. 2022. "A Machine Learning Model for Improving Building Detection in Informal Areas: A Case Study of Greater Cairo." *Geomatics and Environmental Engineering* 16 (2): 39–58. doi:10.7494/geom.2022.16.2.39.

Tan, Y., Y. Yu, S. Xiong, and J. Tian 2016. "Semi-automatic Building Extraction from Very High Resolution Remote Sensing Imagery via Energy Minimization Model." In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 657–660. doi:10.1109/IGARSS.2016.7729165.

Taubenböck, H., M. Klotz, M. Wurm, J. Schmieder, B. Wagner, M. Wooster, T. Esch, and S. Dech. 2013. "Delineation of Central Business Districts in Mega City Regions Using Remotely Sensed Data." *Remote Sensing of Environment* 136: 386–401. doi:10.1016/j.rse.2013.05.019.

Teimouri, M., M. Mokhtarzade, and M. J. V. Zoej. 2016. "Optimal Fusion of Optical and SAR high-resolution Images for Semiautomatic Building Detection." *GIScience & Remote Sensing* 53: 45–62. doi:10.1080/15481603.2015.1116140.

Turker, M., and D. Koc-San. 2015. "Building Extraction from high-resolution Optical Spaceborne Images Using the Integration of Support Vector Machine (SVM) Classification, Hough Transformation and Perceptual Grouping." *International Journal of Applied Earth Observation and Geoinformation* 34: 58–69. doi:10.1016/j.jag.2014.06.016.

Uzar, M. 2017. "Automatic Building Extraction with Multi-sensor Data Using Rule-based Classification Automatic Building Extraction with Multi-sensor Data Using Rule-based Classification." *European Journal of Remote Sensing* 7254. doi:10.5721/EuJRS20144701.

Wang, H., J. Qi, Y. Lei, J. Wu, B. Li, and Y. Jia. 2021. "A Refined Method of high-resolution Remote Sensing Change Detection Based on Machine Learning for Newly Constructed Building Areas." *Remote Sensing* 13. doi:10.3390/rs13081507.

Wei, S., and S. Ji. 2021. "Graph Convolutional Networks for the Automated Production of Building Vector Maps from Aerial Images." *IEEE Transactions on Geoscience and Remote Sensing*. doi:10.1109/TGRS.2021.3060770.

Weir, N., D. Lindenbaum, A. Bastidas, A. Etten, V. Kumar, S. McPherson, J. Shermeyer, and H. Tang 2019. "SpaceNet MVOI: A multi-view Overhead Imagery Dataset." In Proc. IEEE Int. Conf. Comput. Vis, Seoul, South Kerean, 2019-October, 992–1001. doi:10.1109/ICCV.2019.00108.

Wen, D., X. Huang, A. Zhang, and X. Ke. 2019. "Monitoring 3D Building Change and Urban Redevelopment Patterns in Inner City Areas of Chinese Megacities Using multi-view Satellite Imagery." *Remote Sensing* 11: 763. doi:10.3390/rs11070763.

Wen, X., X. Li, C. Zhang, W. Han, E. Li, W. Liu, and L. Zhang. 2021. "Me-net: A multi-scale Erosion Network for Crisp Building Edge Detection from Very High Resolution Remote Sensing Imagery." *Remote Sensing* 13: 3826. doi:10.3390/rs13193826.

Wurm, M., A. Schmitt, and H. Taubenböck. 2017. "Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9: 1901–1912. doi:10.1109/JSTARS.2015.2465131.

Xie, Y., J. Zhu, Y. Cao, D. Feng, M. Hu, W. Li, Y. Zhang, and L. Fu. 2020. "Refined Extraction of Building Outlines from High-Resolution Remote Sensing Imagery Based on a Multifeature Convolutional Neural Network and Morphological Filtering." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 1842–1855. doi:10.1109/JSTARS.2020.2991391.

Xu, Y., L. Wu, Z. Xie, and Z. Chen. 2018. "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters." *Remote Sensing* 10: 144. doi:10.3390/rs10010144.

Yan, Y., Z. Tan, N. Su, and C. Zhao. 2017. "Building Extraction Based on an Optimized Stacked Sparse Autoencoder of Structure and Training Samples Using LIDAR DSM and Optical Images." *Sensors (Switzerland)* 17: 1957. doi:10.3390/s17091957.

Yu, W., W. Zhou, Y. Qian, and J. Yan. 2016. "A New Approach for Land Cover Classification and Change Analysis: Integrating Backdating and an Object-Based Method." *Remote Sensing of Environment* 177: 37–47. doi:10.1016/j.rse.2016.02.030.

Zebedin, L., A. Klaus, B. Gruber-Geymayer, and K. Karner. 2006. "Towards 3D Map Generation from Digital Aerial Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 60: 413–427. doi:10.1016/j.isprsjprs.2006.06.005.

Zeng, C., J. Wang, and B. Lehrbass. 2013. "An Evaluation System for Building Footprint Extraction from Remotely Sensed Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6: 1640–1652. doi:10.1109/JSTARS.2013.2256882.

Zhang, L., L. Zhang, and B. Du. 2016. "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art." *IEEE Geoscience and Remote Sensing Magazine* 4: 22–40. doi:10.1109/MGRS.2016.2540798.

Zhang, T., X. Huang, D. Wen, and J. Li. 2017. "Urban Building Density Estimation from High-Resolution Imagery Using Multiple Features and Support Vector Regression." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 3265–3280. doi:10.1109/JSTARS.2017.2669217.

Zhang, T., and X. Huang. 2018. "Monitoring of Urban Impervious Surfaces Using Time Series of High-Resolution Remote Sensing Images in Rapidly Urbanized Areas: A Case Study of Shenzhen." *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens* 11: 2692–2708. doi:10.1109/JSTARS.2018.2804440.

Zhang, Z., W. Guo, M. Li, and W. Yu. 2020. "GIS-Supervised Building Extraction With Label Noise-Adaptive Fully Convolutional Neural Network." *IEEE Geoscience and Remote Sensing Letters* 17 (12): 2135–2139. doi:10.1109/LGRS.2019.2963065.

Zhang, Y., M. Deng, F. He, Y. Guo, G. Sun, and J. Chen. 2021. "FODA: Building Change Detection in High-Resolution Remote Sensing Images Based on Feature-Output Space Dual-Alignment." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 8125–8134. doi:10.1109/JSTARS.2021.3103429.

Zhao, L., X. Zhou, and G. Kuang. 2013. "Building Detection from Urban SAR Image Using Building Characteristics and Contextual Information." *EURASIP Journal on Advances in Signal Processing* 2013: 1–16. doi:10.1186/1687-6180-2013-56.

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia 2017. "Pyramid Scene Parsing Network." In Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, Honolulu, HI, USA, 2017-Janua, 6230–6239. doi:10.1109/CVPR.2017.660.

Zhao, W., C. Persello, and A. Stein. 2021. "Building Outline Delineation: From Aerial Images to Polygons with an Improved end-to-end Learning Framework." *ISPRS Journal of Photogrammetry and Remote Sensing* 175: 119–131. doi:10.1016/j.isprsjprs.2021.02.014.

Zhou, Z., and J. Gong. 2018. "Automated Residential Building Detection from Airborne LiDAR Data with Deep Neural Networks." *Advanced Engineering Informatics* 36: 229–241. doi:10.1016/j.aei.2018.04.002.

Zhu, Q., C. Liao, H. Hu, X. Mei, and H. Li. 2021a. "MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction from Remote Sensed Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 59: 6169–6181. doi:10.1109/TGRS.2020.3026051.

Zhu, Y., Z. Liang, J. Yan, G. Chen, and X. Wang. 2021b. "E-D-Net: Automatic Building Extraction from High-Resolution Aerial Images with Boundary Information." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 4595–4606. doi:10.1109/JSTARS.2021.3073994.