A Review of Building Extraction from Remote Sensing Imagery: Geometrical Structures and Semantic Attributes

Qingyu Li, Lichao Mou, Yao Sun, Yuansheng Hua, Yilei Shi, *Member, IEEE*, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*—In the remote sensing community, extracting buildings from remote sensing imagery has triggered great interest. While many studies have been conducted, a comprehensive review of these approaches that are applied to optical and synthetic aperture radar (SAR) imagery is still lacking. Therefore, we provide an in-depth review of both early efforts and recent advances, which are aimed at extracting geometrical structures or semantic attributes of buildings, including building footprint generation, building facade segmentation, roof segment and superstructure segmentation, building height retrieval, building type classification, building change detection, and annotation data correction. Furthermore, a list of corresponding benchmark datasets is given. Finally, challenges and outlooks of existing approaches as well as promising applications are discussed to enhance comprehension within this realm of research.**

***Index Terms*—building extraction, deep learning, optical imagery, review, synthetic aperture radar (SAR),**

I. Introduction

Although cities occupy 3% of the Earth's land surface, they are responsible for 60-80% of energy usage and 70% of greenhouse gas emissions [1]. The frequent city renewal and rapid urban growth lead to substantial changes within cities [2]. These alterations can have adverse repercussions on the environment and ecology, e.g., urban heat island, the greenhouse effect, and resource depletion [3] [4]. Urban structures are characterized by buildings in both planar and vertical dimensions, offering insights into urban development. For example, the area and volume of buildings correlate with population distribution [5] [6], greenhouse gas emission [7] [8], and energy consumption [9] [10]. Consequently, up-to-date information about buildings is the key element to environmentally sustainable urbanization. Moreover, geometrical structures and semantic attributes of buildings can be exploited in various domains, including 1) undocumented building detection, 2) emergency responses and rescue operations, 3) autonomous vehicle navigation, and 4) facility management.

The most reliable geometrical structures and semantic attributes of buildings can be achieved by field surveying and mapping [11]; however, these methods are labor-intensive owing to substantial workloads. In contrast, remote sensing techniques capable of extracting buildings in a cost-effective manner have become a mainstream strategy. Remote sensing imagery usually consists of two types: 1) optical imagery, and 2) synthetic aperture radar (SAR) imagery. A wide variety of optical sensors with different spatial resolutions are available for building extraction. The benefit of SAR imagery lies in its ability to penetrate through clouds, thus alleviating the limitation of sun illumination and weather.

Nevertheless, several challenges are associated with building extraction from optical imagery and SAR imagery. An essential issue is the intra-class variance and inter-class similarity of buildings on remote sensing imagery [12] [13] [14]. Intra-class variance denotes buildings are diverse in scale, appearance, and structure on remote sensing imagery, which is due to differences in architectural designs (e.g., size, height, and color), materials (e.g., metal, clay, concrete, and stone), and land use functions (e.g., commercial, industrial, and residential). Inter-class similarity refers to buildings and other classes having similar features on remote sensing imagery. For instance, on optical imagery, some buildings share akin colors with paved roads, whereas on SAR imagery large storage tanks can have radar-reflective properties similar to some buildings. Furthermore, the precision of building extraction from remote sensing images is hindered by complex background interference and the absence of relevant sensor information (such as illumination conditions, shadows, and shooting angle).

In the past decades, numerous approaches have been proposed to extract buildings from remote sensing images. Early efforts have relied on heuristic feature design procedures, which combine different spatial, spectral, or ancillary information for the construction of building hypotheses. Nevertheless, feature engineering makes it difficult to achieve scalable robust, and generic solutions. Recently, the field of remote sensing image interpretation has seen significant advancements thanks to deep learning techniques. These methods leverage convolutional neural networks (CNNs), known for their superior feature learning capability from raw data [15], recently

This work is jointly supported by the Excellence Strategy of the Federal Government and the Länder through TUM Innovation Network EarthCare, by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (grant number: 01DD20001), by German Federal Ministry for Economic Affairs and Climate Action in the framework of the "national center of excellence ML4Earth" (grant number: 50EE2201C) and by the Munich Center for Machine Learning.

Corresponding author: Xiao Xiang Zhu.

Q. Li, L. Mou, Y, Sun, Y. Hua, and X.X. Zhu are with Data Science in Earth Observation, Technische Universität München (TUM), 80333 Munich, Germany (e-mails: qingyu.li@tum.de; lichao.mou@tum.de; yao.sun@tum.de; yuansheng.hua@tum.de; xiaoxiang.zhu@tum.de)

X.X. Zhu is also with the Munich Center for Machine Learning

Y. Shi is with the School of Engineering and Design, Technische Universität München (TUM), 80333 Munich, Germany (e-mail: yilei.shi@tum.de)

becoming a popular strategy.

To the best of our knowledge, there are two review articles about building extraction from optical imagery [16] [17] in the existing literature. However, they ignore SAR imagery, which can also contribute to this task. SAR imagery can offer data irrespective of time or weather conditions. This capability makes SAR data particularly valuable for application after natural disasters (e.g., earthquake [18] and tsunami [19]) and war conflict [20] and for investigations in areas frequently obscured by clouds [21]. Moreover, these two studies mainly concentrate on building footprint generation. Accompanied by progress in remote sensing technology and data processing strategies, a series of new research tasks have also emerged, e.g., building type classification and roof superstructure segmentation. These tasks aim at extracting geometric structures and semantic attributes of buildings, whereas they have rarely been summarized and discussed. Therefore, a timely overview is essential to summarize works related to these new tasks. In all, a comprehensive and systematic review concerning the building extraction from both optical imagery and SAR imagery has not yet been conducted in the existing literature.

This study mainly focuses on geometric structures and semantic attributes of buildings that can be extracted from remote sensing imagery. Note that the aspects (e.g., building topology) related to building information modeling (BIM) [22] are out of scope in this study. Our research aims to comprehensively review the major tasks within the remote sensing field that exhibit correlations with building extraction, i.e., building footprint generation, building facade segmentation, roof segment and superstructure segmentation, building height retrieval, building type classification, building change detection, and annotation data correction. We conducted a literature search on peer-reviewed scholarly publications that primarily originate from mainstream journals or conferences within the field of remote sensing. Through an in-depth analysis, we identified the related publications and categorized them with respect to corresponding tasks. The primary scientific progress highlighted in the literature is first summarized. Then, some benchmark remote sensing imagery datasets for these tasks are introduced. Furthermore, challenges and outlooks toward future research are presented. Finally, the main applications of geometrical structures and semantic attributes of buildings are discussed.

II. Review of the Major Tasks Involved in Building Extraction

A. Building Footprint Generation

To initiate our exploration into building extraction from remote sensing imagery, we commence with the fundamental task of building footprint generation. This foundational step lays the groundwork for subsequent analyses by establishing the spatial extent of structures. The building footprint is a two-dimensional (2D) visual representation of a building, describing its exact location, size, and shape in the ground [23] [24] [25]. As illustrated in Fig. 1, three representation types (i.e., mask, boundary, and corner) are usually utilized to represent the building footprint. Fig. 2 shows the building footprint map (mask) corresponding to optical and SAR imagery in the same region.

Early efforts to generate building footprints from remote sensing images have three main types: 1) geometrical primitive-based, 2) over-segmentation-based, and 3) classifier-based methods. In the first type, geometric primitives (e.g., building corners [24] and edges [26] [27] [28] [29]) are first extracted and subsequently assembled into enclosed polygons corresponding to individual buildings. In the second type, different segments –so-called super-pixels– are obtained from the partition of an image to delineate building regions. For instance, some commonly used over-segmentation techniques are clustering [30] [31], graph model [32] [33], active contour model [34] [35], and watershed segmentation [36] [37]. The third type mainly consists of two stages: hand-crafted feature extraction and classification. Features from each pixel are extracted and subsequently fed into classifiers that can determine its label. Classifier-based methods utilize machine learning models (e.g., support vector machine [38]) to distinguish buildings from non-building objects [39] [40] [41] [42]. Note that optical imagery encompasses another method: the index-based method. This method devises an index by taking into account the contrast and brightness of buildings and then utilizes an empirical threshold to extract buildings. Specifically, morphological building index (MBI) [43] and its improved versions [44] are commonly used indices.

In the past decades, a significant number of deep learning-based approaches have been proposed, and they have significantly outperformed traditional methods in both efficiency and accuracy. Based on the visual cues they utilize, these methods can be categorized into three groups: 1) corner-based, 2) boundary-based, and 3) mask-based methods. On optical imagery, buildings usually show distinct traits such as straight lines and sharp corners, inspiring some scholars to leverage these traits as prominent and differentiable features with which to extract buildings based on the former two methods. The advancement of keypoint detection networks has further propelled corner-based methods. PolygonRNN [45] is an advantageous approach that comprises a CNN and a recurrent neural network (RNN). The CNN is responsible for extracting corner points, and the RNN then connects these points to create closed polygonal representations. PolyMapper [46] incorporates the Feature Pyramid Network (FPN) [47] into PolygonRNN [45], eliminating the necessity for bounding box annotations. Considering the difficulty in the CNN-RNN training, graph convolutional network (GCN) combined with CNN recently become a more popular strategy in this field [48] [49]. To reduce vertex redundancy in the CNN-GCN paradigm, a transformer [50] -based approach, PolyBuilding [51] is proposed to learn building corner points from remote sensing images. To generate building footprints, boundary-based methods directly learn building boundaries in an end-to-end manner. Some works [52] [53] use semantic segmentation networks to learn building boundaries. To refine the boundaries of individual buildings, other works exploit instance segmentation networks (e.g., Mask R-CNN [54]) for building boundary learning [55] [56]. To obtain sharp building boundaries, some studies exploit the active contour model

(a) (b) (c)

Fig. 1. (a) Mask. (b) Boundary. (c) Corner points of the corresponding building footprint.

(a) (b) (c)

Fig. 2. (a) Optical imagery. (b) SAR imagery. (c) The corresponding building footprints (mask). [59]

(ACM) where parameterizations are learned by an end-to-end network [57] [58]. However, ACM-based methods are tailored for extracting a single building instance from a cropped input image. Thus, their initialization depends on external methods not integrated into an end-to-end learning process.

Most methods for this task learn masks of buildings from optical and SAR images. Their primary objective is to address pixel-level labeling challenges. More specifically, these approaches employ semantic segmentation networks to assign each pixel within the image its relevant label, namely either "building" or "non-building". In the following, we introduce these methods according to their addressed issues.

Buildings exhibit considerable variability within the same class, such as differences in size, which poses challenges for this task. This limitation stems from the fact that the efficacy of semantic segmentation networks is constrained when dealing with extremely small or large buildings. Owing to the restricted receptive field, large buildings often show fragmented and incomplete shapes, whereas many small buildings might be overlooked. Many approaches have been introduced to extract buildings at multiple scales from both optical and SAR images. The majority of research concentrates on aggregating multi-scale information [60] [61] [62]. Some studies concentrate on multi-scale feature extraction [63] [13] [64], while others devise dedicated architectures, e.g., Siamese network [65] [66] and multi-task learning network [67] [68].

On optical imagery, buildings commonly exhibit straight lines and sharp corners. However, the inherent translational and spatial invariance properties of CNNs can result in the loss of intricate information necessary for precise localization. This often leads to inaccurate and irregular building boundaries. A range of methods have been introduced to maintain the geometrical details of buildings. Improved output representation-based methods devise various output representations capable of encoding geometrical details concerning buildings, e.g., signed distance transform (SDT) [69] [70], frame field [71], and attraction field representation [72] [73]. Compared to other output representations, attraction field representation can preserve more detailed structures for complicated buildings [72]. Geometric priors of buildings are not evident on optical imagery with a relatively low spatial resolution. Thus, adversarial training-based approaches and graph model-based methods can be adopted for these optical images. Adversarial training-based approaches harness generative adversarial networks (GANs), comprising a generator and a discriminator [74] [75]. Graph models, which facilitate the representation of pixel interactions, can also be employed. Graph model-based methods have integrated graph models in end-to-end network learning frameworks [76] [77].

When preparing the training data, manually annotating buildings requires more effort compared to annotating woodlands, water bodies, and roads [78]. Thus, different strategies have been designed to diminish the requirement for extensive pixel-level annotations and compensate for the limited supervisory information. Weakly-supervised methods construct models through learning with weak supervision. In addition to pixel-level labels, weakly-supervised approaches still need weaker labels, including point labels [79], bounding boxes [80] [81], and image-level labels [82] [83]. However, weakly-supervised methods neglect the opportunity to leverage extensive unlabeled data. A study [78] explores pseudo-labeling, where a model is initially trained using a small set of labeled data to create pseudo-segmentation maps for unlabeled samples. Consistency training-based approaches enforce prediction consistency by assigning diverse perturbations to the input [84] [85], which are more efficient to implement than the other methods. Domain adaptation is aimed at transferring knowledge from a source domain to a target domain, mitigating domain shift. In this context, the source domain dataset consists of ample annotated samples, whereas the target domain dataset has no labeled instances. Domain adaptation-based approaches aim to enhance CNNs' performance on the target domain by leveraging the source domain dataset and aligning data distribution between the two domains. This helps to mitigate the scarcity of supervisory information in the target domain. Domain shift can be addressed at different levels. For instance, [86] seeks to address the domain-shift problem by only aligning the image distribution, whereas [87] tackles domain adaptation at both the image and feature levels.

Rapid and accurate generation of building footprint maps holds critical significance for disaster emergency response, military reconnaissance, and loss assessment. Some lightweight networks have been developed to realize a balance between computational costs and accuracy by designing specific network architectures. [88] devises a compressing module to reduce feature channels, [89] reduces the count of convolution kernels in its network, and [90] incorporates atrous convolutions [91], thereby diminishing the training parameters.

Fig. 3. Off-nadir optical imagery with building facades extracted by the method described in [92].

B. Building Facade Segmentation

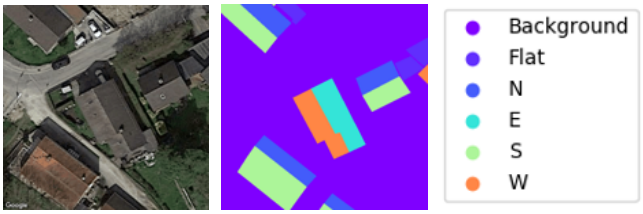
With the building footprints delineated, our focus shifts to a more detailed examination of structures through building facade segmentation. This task delves into the exterior face of buildings, contributing essential information for a more comprehensive understanding of their architectural characteristics. For SAR imagery, there are no studies focusing on extracting building facades. This is due to the side-looking geometry of SAR: building areas refer to roofs and facades, making it difficult to extract sole facade information [93]. For optical imagery, the building facade is usually invisible at the nadir angle. Thus, off-nadir imagery is the primary type of data source to provide beneficial information for building facade segmentation (see Fig. 3).

Early studies in segmenting building facades from off-nadir optical imagery have two main types: 1) geometrical primitive-based, and 2) index-based approaches. The first type extracts geometric primitives (e.g., building corners and edges), which are grouped to form a building facade by applying spatial constraints [92] [94]. In the second type, the index is devised by considering the spatial features of the facade, and then an empirical threshold is applied to extract facade regions [95].

Recently, a deep learning network [96] has been proposed to learn building facades directly from off-nadir imagery, and this information is combined with other elements (e.g., footprint) for 3D building reconstruction.

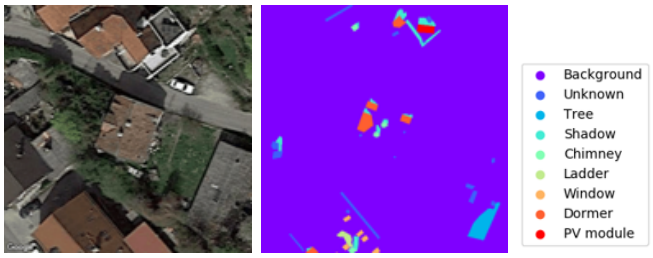
C. Roof Segment and Superstructure Segmentation

Now, we proceed to roof segment and superstructure segmentation, advancing our analysis to the uppermost regions of buildings. This phase enriches our understanding by capturing the roof structures. Each planar roof segment (c.f. Fig. 4 (b)) of the building usually has a specific orientation. Moreover, roofs usually contain some structures (c.f. Fig. 5 (b)), e.g., chimneys and windows, which are generally named roof superstructures. Very-high-resolution optical images provide a valuable resource for roof segment and superstructure segmentation, as the details of roof segments and superstructures are visible.



(a) (b) (c)

Fig. 4. (a) Optical imagery. (b) Roof segment map. (c) Roof segment classes (legend).



(a) (b) (c)

Fig. 5. (a) Optical imagery. (b) Roof superstructure map. (c) Roof superstructure classes (legend).

Roof segment segmentation aims to extract individual roof planar segments. One early work [97] relies on a line detection algorithm to detect roof ridges and gutters, and then roof planar segments (which face in various orientations) of the building can be deduced. Recently, semantic segmentation networks have been implemented to directly learn roof segments from aerial imagery [98] [99].

Roof superstructure segmentation focuses on segmenting different superstructures on the roof. To extract roof superstructure, early efforts [97] [100] utilize either contour detection [101] or watershed segmentation [102], while recent studies [99] [103] use semantic segmentation networks.

D. Building Height Retrieval

Ascending to the three-dimensional realm, our attention turns towards building height retrieval. This critical task augments our knowledge by providing insights into the vertical dimension of structures. Building height retrieval involves addressing two problems: 1) delineating building footprints, and 2) estimating building heights. Fig. 6 shows the building height maps (pixel-wise) retrieved from optical imagery and Fig. 7 illustrates the building height maps (instance-wise) obtained from SAR imagery.

Traditional approaches first extract building footprints and subsequently model the height. For optical and SAR imagery, most of these methods utilize geometrical primitives or the shadow information as primary indicators [106] [107] [108] [109]. Meta information of the sensor (e.g., the sun-earth relative position) is also needed for height estimation. For SAR imagery, its side-looking imaging geometry introduces different types of geometric distortion, which lead to difficulties in image interpretation. In this regard, simulation-based

(a)

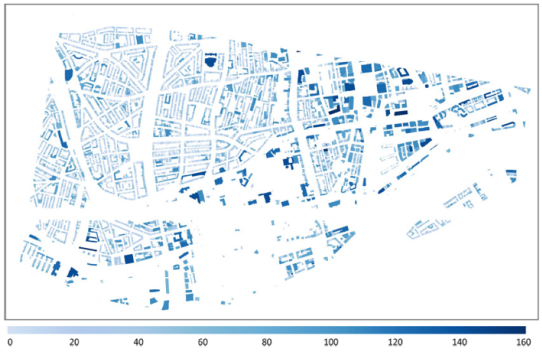
(b)

Fig. 6. (a) Optical imagery. (b) Building height map obtained by the method described in [104].

methods are devised to iteratively simulate SAR images by making a hypothesis of geometric and radiometric properties [110] [111]. Afterward, the target building height is gradually obtained by minimizing the disparity between real and simulated data.

Recent advances in deep learning-based methods have made it possible to directly learn height maps and semantic masks from remote sensing imagery via a multi-task network. In this manner, the efficacy of both sub-tasks can be enhanced through a concurrent optimization procedure. The integration of the 3D centripetal shift representation and decoupling module in [104] yields superior results on near-nadir optical images when compared to other competitors [112] [113]. [96] devises a specific network for off-nadir optical imagery where building facades are also partially visible. For SAR imagery, buildings show special geometric characteristics induced by the SAR view geometry. Thus, two main types of methods are utilized to retrieve building heights. The first type considers the building footprint as preliminary input to estimate the instance-wise building height from a bounding box regression network [105]. The second type exploits semantic segmentation networks to learn building regions [93] [114]. This is because building regions correspond to both roof and layover areas on SAR

(a)



(b)

Fig. 7. (a) SAR imagery. (b) Building height map obtained by the method described in [105].

imagery, and building heights can be estimated from their layover lengths.

E. Building Type Classification

Extending our analysis beyond geometric attributes, we delve into building type classification that interprets the semantic attributes of individual buildings according to their geometry or functions. For instance, buildings can be classified into different roof types, e.g., gable, flat, and hip, or different function types, e.g., industrial, commercial, and residential. Very-high-resolution optical imagery provides the potential for building type classification, as finer building structures can be observed. Fig. 8 (b) and (d) illustrate the roof geometry types and building function types of individual buildings, respectively.

In traditional methods [115] [116], buildings are first segmented and then their types are distinguished by using extracted features.

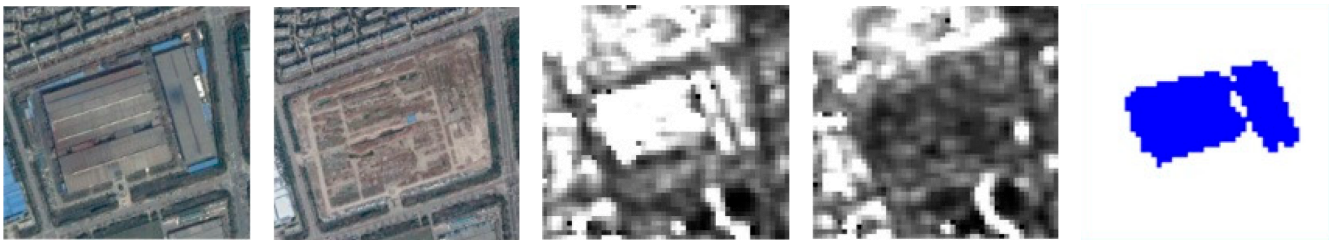
In recent years, deep learning techniques have been exploited to identify building types directly from optical imagery. Owing to the lack of pixel-level annotation data, early studies [117] can only assign each image patch with a label of the corresponding building type. Nowadays, researchers [118] [119] focus on pixel-level-type classification of individual buildings.

F. Building Change Detection

Acknowledging the dynamic nature of urban environments, we introduce building change detection. This task aims to

(a) (b) (c) (d) (e)

Fig. 8. (a) Optical imagery. (b) Roof geometry type map. (c) Roof geometry type classes (legend). (d) Building function type map. (e) Building function classes (legend).



(a) (b) (c) (d) (e)

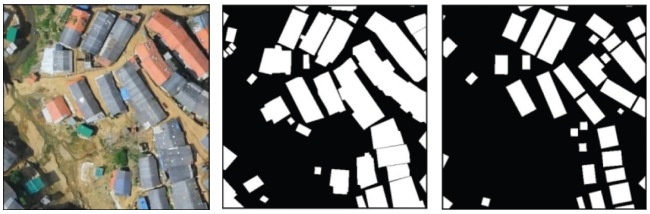
Fig. 9. (a) Pre-change optical imagery. (b) Post-change optical imagery. (c) Pre-change SAR imagery. (d) Post-change SAR imagery. (e) Changed building masks. [120]

identify changes in buildings in bi-temporal or multi-temporal remote sensing imagery that are captured from identical geographic regions. Fig. 9 shows the corresponding changed building mask between pre-change and post-change remote sensing imagery in the same region. Specifically, the change types usually refer to newly constructed or demolished buildings [121] and building damage [122]. In the existing literature, there are two main strategies for building change detection. One solution is based on change detection algorithms, while the other solution is to first extract buildings in the post-change remote sensing imagery and then identify changes by comparison with the pre-change building maps. In this paper, we introduce the literature related to the first solution.

Tradition methods usually consist of two steps [123]: feature extraction and change detection. Multiple features (e.g., spectral, textural, geometrical properties) of buildings need to be engineered to explore the type of changes. For optical imagery, MBI [43] is a commonly used feature, and its variation can be used to carry out building change analysis [124] [125]. For SAR imagery, the double bounce line [126] or the properties of backscattering [127] are detected for monitoring changed buildings. To generate difference images (DI) for further change analysis, three types of indicators are usually utilized: algebra-, transform-, and classifier-based methods. Change vector analysis (CVA), image ratio, and image differencing are commonly used algebraic-based methods. Principal component analysis (PCA), which emphasizes change information in the transformed feature space, is a widely used transform-based approach. In classifier-based methods, change detection can also be realized by exploiting the classifiers that assign pixel-level labels of “change” or “non-change”.

Recently, a number of deep learning-based approaches have achieved impressive performance. However, the potential of existing approaches is usually limited by two factors. First, buildings show various sizes and shapes, which makes it difficult to extract representative features of buildings with different sizes and shapes. Second, the similarity between buildings and other objects as well as the complexity of the background may also lead to mistaken identification. To address the aforementioned issues, many methods are proposed to enhance their capability of feature extraction. Most studies introduce attention mechanisms [128] [129] [130] [131] that select the most discernible features. Multi-scale pyramid structures [132] [133] can also be implemented to extract multi-scale features by increasing the reception field. Some special modules, such as feature space alignment module [134], feature difference enhancement module [135] [136], and context extraction module [137] are also proposed to enlarge the interclass disparity in the feature space.

Deep networks usually require the same number of ground reference labels for pre- and post-change. However, the annotation of changed/unchanged building labels is time-consuming and laborious [121]. Two strategies are usually exploited to compensate for the limited supervisory information. One is generative adversarial training, which can synthesize new labeled samples to expand the training sets [138]. Nevertheless, methods based on adversarial training face a considerable risk of model collapse attributed to the imbalance between adversarial networks. In this regard, semi-supervised learning is more efficient for implementation and can be exploited to improve the model performance by leveraging a considerable number of unlabeled samples [139] [140].



(a) (b) (c)

Fig. 10. (a) Optical imagery. (b) True labels. (c) Labels from OpenStreetMap. [141]

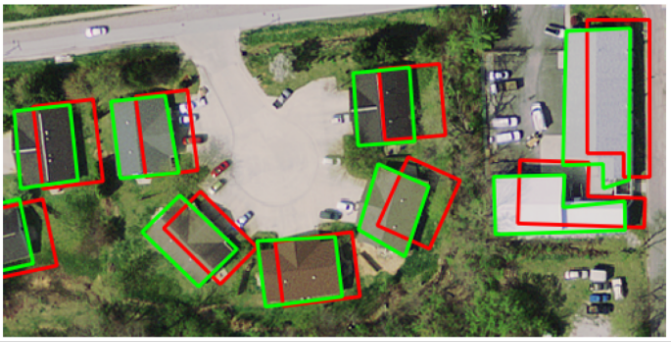


Fig. 11. Aerial imagery with alignment results obtained by the method described in [142].

G. Annotation Data Correction

Recognizing the significance of precisely labeled training data in deep learning or machine learning applications, we now turn our attention to annotation data correction. In fact, data annotation is a time-consuming process and requires expertise. Fortunately, community-based organizations or companies have provided open cadastral maps (e.g., OpenStreetMap). However, these datasets also have two limitations [143] [144] [145]. One limitation is incorrectness (c.f. Fig. 10), where the labels from open cadastral maps differ from the ones in the real world [141] [146]. For example, owing to the time difference between the two data sources, a newly constructed building might be missing, while a demolished building exists in the open cadastral maps. Moreover, the outlines of buildings on open cadastral maps are sometimes much simplified. The other limitation is misalignment (see Fig. 11), where annotated buildings are rotated and translated from their position in the remote sensing imagery [142] [147]. This is due to two factors: 1) errors from the projections of two data sources, and 2) errors from the annotators. If these open data are used as training samples, the noise in class labels will impair the model performance.

To deal with both issues, existing studies involve two main strategies: 1) noise modeling and 2) data cleansing. The first strategy approximates noise transition matrices [141] or devises robust loss functions [148]. However, estimating the noise transition matrices poses a significant challenge, and loss function-based methods suffer from the accumulation of errors [149]. Data cleansing methods, in essence, adhere to a simple yet intuitive concept: the removal of noisy data and training exclusively with the cleaner subset [150][151].

Most existing works concentrate on the alignment of optical imagery and cadastral maps. Cross-correlation-based methods assume that the estimated alignment location refers to the maximum value of the cross-correlation [69]. However, conducting the cross-correlation is a time-consuming process. In energy minimization-based approaches, the alignment problem is solved by designing and minimizing an energy function [143] [152]. Nevertheless, the algorithm for energy minimization encompasses a considerable number of parameters. CNN-based methods propose novel networks to address the misalignment, i.e., displacement field learning [142] [153] [154], probability transition modular [155], and robust loss function [156]. A notable benefit of employing CNN-based approaches lies in their better generalizability.

III. Dataset

With the available computational resources like graphics processing units (GPU), deep learning methods have the capacity to automatically extract information from a large volume of remote sensing imagery. In this regard, some benchmark datasets (see Table I) have been proposed to extract geometrical structures or semantic attributes of buildings.

For building footprint generation, a considerable number of benchmark datasets are available. However, for other tasks (e.g., building facade segmentation), only limited available benchmark datasets are available. This might be due to the amount of effort needed to acquire corresponding labels. Compared to other tasks, annotating building footprints is now much easier since different label sources have become available [17], e.g., OpenStreetMap.

In terms of sensor type, optical imagery is dominant in data sources, and only a few benchmarks provide SAR imagery. This is due to two factors. First, for the SAR sensor, its side-looking geometry leads to difficulties in data interpretation. Thus, most researchers prefer to use optical imagery that is easier to interpret. Second, the number of SAR sensors is much smaller than that of optical sensors, which means that only a limited number of SAR products are available for the whole community.

The spatial resolution of remote sensing images in most datasets is very high (i.e., ranging from centimeter level to decimeter level). However, for the SpaceNet 7 dataset, the spatial resolution is relatively coarse (i.e., 4 m), which introduces more challenges in building footprint generation, as it is difficult to identify individual buildings on such spatial resolution.

The spatial coverage of some benchmark datasets is limited to one specific city or country. Since deep networks focus on learning location-specific building patterns, the model's ability to generalize is restricted when exploiting such datasets. Intra-class variation of buildings is evident across different geolocations. On the one hand, the appearances of urban settlements (which can be densely or sparsely populated) vary across different continents. On the other hand, buildings

TABLE I
REPRESENTATIVE BENCHMARK DATASETS FOR DIFFERENT TASKS.

Task	Name	Sensor	Spatial resolution	Coverage	Website
Building footprint generation	ISPRS-Potsdam	Optical	0.05m	Potsdam, Germany	https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx
	ISPRS-Vaihingen	Optical	0.09m	Vaihingen, Germany	https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx
	Massachusetts building	Optical	1m	Boston, USA	http://www.cs.toronto.edu/~vmnih/data/
	WHU building-aerial	Optical	0.3m	Christchurch, New Zealand	http://gpcv.whu.edu.cn/data/building_dataset.html
	WHU building-satellite	Optical	0.3-2.5m	Multiple cities around the world	http://gpcv.whu.edu.cn/data/building_dataset.html
	Inria aerial image labeling	Optical	0.3m	Multiple cities in Austria and the USA	https://project.inria.fr/aerialimagelabeling/
	CrowdAI	Optical	0.3m	Multiple cities around the world	https://www.crowdai.org/challenges/mapping-challenge
	SpaceNet	Optical	0.5m	Rio de Janeiro, Brazil	https://spacenet.ai/spacenet-buildings-dataset-v1/
	SpaceNet2	Optical	0.3m	Las Vegas, USA; Paris, France; Shanghai, China; Khartoum, Sudan	https://spacenet.ai/spacenet-buildings-dataset-v2/
	SpaceNet4	Optical	0.3m	Atlanta, USA	https://spacenet.ai/off-nadir-building-detection/
	SpaceNet6	Optical and SAR	Optical: 0.5-2m; SAR: 0.5m	Rotterdam, The Netherlands	https://spacenet.ai/sn6-challenge/
	SpaceNet7	Optical	4m	Multiple cities around the world	https://spacenet.ai/sn6-challenge/
	GaoFen-3 Building	SAR	1m	Multiple cities around the world	https://doi.org/10.1109/JSTARS.2021.3085122
Building facade segmentation	3D reconstruction	Optical	-	Multiple cities in China	https://liweijia.github.io/projects/building_3d/
Building height retrieval	ISPRS-Potsdam	Optical	0.05m	Potsdam, Germany	https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx
	ISPRS-Vaihingen	Optical	0.09m	Vaihingen, Germany	https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx
	USSOCOM Urban 3D	Optical	0.5m	Jacksonville and Tampa, USA	https://spacenet.ai/the-ussocom-urban-3d-competition/
	3D reconstruction	Optical	-	Multiple cities in China	https://liweijia.github.io/projects/building_3d/
Roof segment and superstructure segmentation	DeepRoof	Optical	-	Multiple cities in the USA	https://traces.cs.umass.edu/index.php/Smart/Smart/
	RID	Optical	0.1m	Wartenburg, Germany	https://github.com/TUMFTM/RID
Building type classification	Urban Building Classification	Optical	0.5-0.8m	Beijing, China; Munich, Germany	https://github.com/AICyberTeam/UBC-dataset
	DFC23	Optical and SAR	Optical: 0.5-0.8m; SAR: 1m	Multiple cities around the world	https://ieee-dataport.org/competitions/2023-ieee-grss-data-fusion-contest-large-scale-fine-grained-building-classification
Building change detection	LEVIR CD	Optical	0.5m	Texas, USA	https://justchenhao.github.io/LEVIR/
	WHU Building Change Detection	Optical	0.2m	Christchurch, New Zealand	http://gpcv.whu.edu.cn/data/building_dataset.html
	S2Looking	Optical	0.5-0.8m	Multiple cities around the world	https://github.com/S2Looking/Dataset
	SI-BU	Optical	0.5-0.8m	Guiyang, China	https://github.com/liaochengcsu/BCE-Net
	BANDON	Optical	0.6m	Multiple cities in China	https://github.com/fitzpchao/BANDON

come in a wide variety of shapes and colors. Therefore, the benchmark datasets that have wider spatial coverage and a more diverse building pattern are more popular. This is because they can help improve the generalizability of deep networks.

For all benchmark datasets, the most commonly used metrics to evaluate algorithms are precision, recall, F1 score, and Intersection over Union (IoU). In terms of different tasks or goals, new metrics will be considered to provide a comprehensive evaluation. For instance, root mean square error (RMSE) and mean absolute error (MAE) are metrics for the benchmark datasets related to building height retrieval. To provide instance-level evaluation, the standard MS COCO measures [157] including average precision (AP, averaged over IoU thresholds), or AP at different scales will be exploited. For the evaluation of the quality of the predicted boundaries, boundary F-score (BoundF) [158] and polygon similarity [159]

will also be taken into account.

Given the rapidly evolving nature of the field, numerous resources, such as the website "Papers with Code" (https://paperswithcode.com/) and project webpages related to the benchmark datasets, provide detailed and up-to-date information on the quantitative performance of different methods. For instance, both https://paperswithcode.com/sota/semantic-segmentation-on-inria-aerial-image/ and https://project.inria.fr/aerialimagelabeling/leaderboard/ provide the comparison of the performance of different methods on the Inria aerial image labeling dataset.

IV. Perspectives and Insights

A. Challenges and Future Directions

In this section, the challenges of building extraction from remote sensing imagery are summarized. Moreover, possible future directions are also discussed.

Polygonization: Polygonization refers to mapping building corners. For the tasks of building footprint generation and building height retrieval, polygonization should be taken into account. This is because in geographic information system (GIS), building footprints are usually stored as vector formats where building shapes are characterized as building corner points. Early efforts [160] perform the polygonization on predicted semantic masks, and they exploit post-processing steps (e.g., Douglas-Peucker algorithm [161]) to acquire an abstract version of the building shape. Recently, some deep learning-based networks that can directly learn building corner points from remote sensing imagery have become more favored. However, there are several challenges arising from these methods. First, building corners are not distinct on remote sensing imagery with relatively low resolution (e.g., Planet satellite imagery with 3 m/pixel). Thus, when applying these methods to such imagery, results might not be satisfactory. Second, the current strategies for corner point connection (e.g., manually defined rules [162] and graph model [49]) cannot deal with complex shapes (e.g., buildings with holes).

Multimodal data fusion: Multimodal data denotes the data collected by various sensors, and the synergistic utilization of multimodal data empowers the network for the acquisition of more details. For example, optical sensors capture spectral attributes of objects, SAR remains unaffected by weather conditions, and LiDAR can acquire precise geometrical information. A common application for building change detection is to assess information on building damage after an earthquake, and multimodal data can contribute. For instance, pre-event optical and post-event SAR imagery is compared to detect the destroyed buildings [163]. In [164], bi-temporal optical images and post-event LiDAR data are used to extract building damage. A primary challenge emerges in determining the "where" and "how" of fusing multimodal data for specific tasks [165] [166]. "How" denotes fusion strategies to fully exploit the distinct data, while 'where' refers to the level of fusion, encompassing three categories: data-, feature-, and decision-level. The other main issue is the registration of multi-modal data, as geometrical registration accuracy will affect image fusion results. Moreover, different fusion levels might have different sensitivities to registration errors [167].

Domain shift: For all tasks discussed in section II, the generalization capability of deep neural networks is of great concern for large-scale applications. For instance, deep networks tend to yield unsatisfactory outcomes when directly applying a model trained on one dataset (source domain) to another dataset (target domain) [168]. In other words, the transferring capability of the trained model is restricted owing to the domain shift between the target domain and the source domain. An example is in large-scale building footprint generation [169], the model which is trained with samples collected from European cities performs badly on test instances in the African cities. Domain gaps arise from several factors. Firstly, the appearances of urban settlements (which can be densely or sparsely populated) are varied across different continents [70]. Secondly, the intra-class variation of buildings is evident, e.g., buildings have various shapes and colors. Thirdly, disparities in the process of data acquisition (such as illumination conditions and atmospheric effects) might cause various radiometries of remote sensing images [170]. Domain adaptation and domain generalization can be helpful in tackling the domain shift problem. Some strategies aim to learn representations that are invariant to domains. Specifically, domain alignment strategies can be designed to minimize the divergence of distributions between target and source domains [171]. Self-supervised learning can also be explored to capture generic representation [172]. Other strategies attempt to improve the generalizability of models by avoiding overfitting issues. For instance, to simulate the domain shift, various types of data augmentation approaches are devised, including image-, model-, and feature-based augmentations [171].

B. Potential Applications of Geometrical Structures and Semantic Attributes of Buildings

The geometrical structures and semantic attributes of buildings provide valuable insights for many practical applications at both micro and macro scales. In this study, several examples are provided, including, 1) environmental and socioeconomic analysis, 2) disaster risk management, and 3) high-resolution population map production.

Environmental and socioeconomic analysis: Urbanization involves the construction of buildings on former non-urban land. Rapid urbanization can lead to detrimental consequences, e.g., the spread of epidemics, air and water pollution, and resource depletion. For instance, morphological parameters and landscape metrics of buildings are derived from investigating their correlation with the thermal environment [173]. Carbon dioxide emission [174] can be allocated to individual buildings with respect to attributes (e.g., type, area, and height). The analysis of the relationship between pedestrian-level wind velocity and building density facilitates a better understanding of urban ventilation [175]. The urban living environment, such as the building density in the community, has also been proven to be associated with the health of residents [176]. Moreover, the geometrical features of buildings contribute to the estimation of energy consumption [177] and solar energy potential [178].

Disaster risk management: For disaster risk management, the assessment of vulnerability and risk to natural hazards is an

essential process. Hazard refers to environmental phenomena that potentially cause detrimental effects on both humans and infrastructure. Different types of hazards–including landslide, tsunami, drought, earthquake, flood, and volcanic ash –can lead to building damage [179]. When a hazard occurs, we can identify the buildings that are situated within vulnerable regions. Moreover, evaluating the vulnerability of buildings also aids practitioners and stakeholders by helping improve the decision-making process. Specifically, parameters associated with the properties of buildings (e.g., height, shape, orientation, and accessibility) are derived to quantify the vulnerability of buildings [180].

High-resolution population map production: Population maps refer to population distributions and dynamics, offering insights for diverse applications such as comprehending interactions between humans and the environment, and assessing populations at risk. Nonetheless, population data frequently lags behind or remains absent in certain regions. Considering that there is a high correlation between population and buildings, geometrical structures and semantic attributes of buildings can be harnessed to generate detailed population maps [181] [182].

V. Conclusion

Buildings are indispensable objects in the urban environment and play an essential role in urban planning and monitoring. Remote sensing imagery provides excellent potential for the detailed interpretation of buildings. Many methods have been proposed for extracting geometrical structures and semantic attributes of buildings from optical and SAR imagery. Therefore, we present a comprehensive review of both early efforts and recent advances in relation to building extraction on optical and SAR imagery. We summarize six main categories of studies in terms of their extracted building characteristics, including building footprint generation, building facade segmentation, roof segment and superstructure segmentation, building height retrieval, building type classification, and building change detection. Moreover, we also survey the methods aimed at annotation data correction. Furthermore, the corresponding benchmark datasets of these six categories are described. Finally, we discuss the challenges of the current approaches and introduce promising applications for the extracted geometrical structures and semantic attributes of buildings. Although much information about buildings can be acquired by the existing methods, new efforts for developing and improving current approaches should continue to be a high research priority. With the accumulation of a wide range of remote sensing data, more diverse types of information are of interest. How to handle and fully explore these data is becoming a new challenge for the research community, but this also opens new opportunities to gain a deep understanding of buildings.

References

[1] United Nations, "Sustainable development goal 11: Make cities inclusive, safe, resilient and sustainable," https://www.un.org/sustainabledevelopment/cities/, accessed: 2021-12-16.

[2] X. Huang, Y. Cao, and J. Li, "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*, vol. 244, p. 111802, 2020.

[3] A. Rapoport, *Human aspects of urban form: towards a man—environment approach to urban form and design*. Elsevier, 2016.

[4] H. Guo, Q. Shi, A. Marinoni, B. Du, and L. Zhang, "Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images," *Remote Sensing of Environment*, vol. 264, p. 112589, 2021.

[5] M. Alahmadi, P. Atkinson, and D. Martin, "Estimating the spatial distribution of the population of riyadh, saudi arabia using remotely sensed built land cover and height data," *Computers, Environment and Urban Systems*, vol. 41, pp. 167–176, 2013.

[6] F. Biljecki, K. Arroyo Ohori, H. Ledoux, R. Peters, and J. Stoter, "Population estimation using a 3d city model: A multi-scale country-wide study in the netherlands," *PloS one*, vol. 11, no. 6, p. e0156808, 2016.

[7] R. Borck, "Will skyscrapers save the planet? building height limits and urban greenhouse gas emissions," *Regional Science and Urban Economics*, vol. 58, pp. 13–25, 2016.

[8] M. M. Stojiljković, M. G. Ignjatović, and G. D. Vučković, "Greenhouse gases emission assessment in residential sector through buildings simulations and operation optimization," *Energy*, vol. 92, pp. 420–434, 2015.

[9] E. Resch, R. A. Bohne, T. Kvamsdal, and J. Lohne, "Impact of urban density and building height on energy use in cities," *Energy Procedia*, vol. 96, pp. 800–814, 2016.

[10] T. Hong, Y. Chen, S. H. Lee, and M. A. Piette, "CityBES: A web-based platform to support city-scale building energy efficiency," *Urban Computing*, vol. 14, p. 2016, 2016.

[11] B. Oshri, A. Hu, P. Adelson, X. Chen, P. Dupas, J. Weinstein, M. Burke, D. Lobell, and S. Ermon, "Infrastructure quality assessment in Africa using satellite imagery and deep learning," in *ACM SIGKDD*, 2018, pp. 616–625.

[12] W. Kang, Y. Xiang, F. Wang, and H. You, "EU-Net: An efficient fully convolutional network for building extraction from optical remote sensing images," *Remote Sensing*, vol. 11, no. 23, p. 2813, 2019.

[13] Q. Zhu, C. Liao, H. Hu, X. Mei, and H. Li, "MAP-Net: Multiple attending path neural network for building footprint extraction from remote sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[14] W. Yang, X. Yin, H. Song, Y. Liu, and X. Xu, "Extraction of built-up areas from fully polarimetric SAR imagery via PU learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1207–1216, 2013.

[15] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.

[16] L. Luo, P. Li, and X. Yan, "Deep learning-based building extraction from remote sensing images: A comprehensive review," *Energies*, vol. 14, no. 23, p. 7982, 2021.

[17] J. Li, X. Huang, L. Tu, T. Zhang, and L. Wang, "A review of building detection from very high resolution optical remote sensing images," *GIScience & Remote Sensing*, vol. 59, no. 1, pp. 1199–1225, 2022.

[18] T.-L. Wang and Y.-Q. Jin, "Postearthquake building damage assessment using multi-mutual information from pre-event optical image and postevent SAR image," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 452–456, 2011.

[19] Y. Endo, B. Adriano, E. Mas, and S. Koshimura, "New insights into multiclass damage classification of tsunami-induced building damage from SAR images," *Remote Sensing*, vol. 10, no. 12, p. 2059, 2018.

[20] F. Fakhri and I. Gkanatsios, "Integration of sentinel-1 and sentinel-2 data for change detection: A case study in a war conflict area of Mosul city," *Remote Sensing Applications: Society and Environment*, vol. 22, p. 100505, 2021.

[21] B. Huang, Y. Li, X. Han, Y. Cui, W. Li, and R. Li, "Cloud removal from optical satellite imagery with SAR imagery using sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1046–1050, 2015.

[22] S. Azhar, M. Khalfan, and T. Maqsood, "Building information modeling (bim): now and beyond," *Australasian Journal of Construction Economics and Building, The*, vol. 12, no. 4, pp. 15–28, 2012.

[23] O. Wang, S. K. Lodha, and D. P. Helmbold, "A bayesian approach to building footprint extraction from aerial lidar data," in *3DPVT*. IEEE, 2006, pp. 192–199.

[24] M. Cote and P. Saeedi, "Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 313–328, 2012.

[25] Q. Li, "Deep learning for building footprint generation from optical imagery," Ph.D. dissertation, Technische Universität München, 2022.

[26] S. Cui, Q. Yan, and P. Reinartz, "Complex building description and extraction based on Hough transformation and cycle detection," *Remote Sensing Letters*, vol. 3, no. 2, pp. 151–159, 2012.

[27] J. Wang, X. Yang, X. Qin, X. Ye, and Q. Qin, "An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 3, pp. 487–491, 2014.

[28] F. Xu and Y.-Q. Jin, "Automatic reconstruction of building objects from multiaspect meter-resolution sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 7, pp. 2336–2353, 2007.

[29] A. Ferro, D. Brunner, and L. Bruzzone, "Automatic detection and reconstruction of building radar footprints from single VHR SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 935–952, 2012.

[30] F. Cellier, H. Oriot, and J.-M. Nicolas, "Introduction of the mean shift algorithm in sar imagery: Application to shadow extraction for building reconstruction," in *Proceedings of the Earsel 3D Remote Sensing Workshop, Porto, Portugal*. Citeseer, 2005, pp. 6–11.

[31] B. Liu, K. Tang, and J. Liang, "A bottom-up/top-down hybrid algorithm for model-based building detection in single very high resolution SAR image," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 6, pp. 926–930, 2017.

[32] W. He, M. Jäger, A. Reigber, and O. Hellwich, "Building extraction from polarimetric SAR data using mean shift and conditional random fields," in *EUSAR*. VDE, 2008, pp. 1–4.

[33] I. Grinias, C. Panagiotakis, and G. Tziritas, "MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 122, pp. 145–166, 2016.

[34] S. Ahmady, H. Ebadi, M. V. Zouj, and H. A. Moghaddam, "Automatic building extraction from high resolution aerial images using active contour model," *ISPRS Archives*, vol. 37, pp. 453–456, 2008.

[35] R. Hill, C. Moate, and D. Blacknell, "Estimating building dimensions from synthetic aperture radar image sequences," *IET Radar, Sonar & Navigation*, vol. 2, no. 3, pp. 189–199, 2008.

[36] L. Zhao, X. Zhou, and G. Kuang, "Building detection from urban sar image using building characteristics and contextual information," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–16, 2013.

[37] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.

[38] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT*, 1992, pp. 144–152.

[39] J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 3, pp. 236–248, 2007.

[40] F. Dornaika, A. Moujahid, A. Bosaghzadeh, Y. El Merabet, and Y. Ruichek, "Object classification using hybrid holistic descriptors: Application to building detection in aerial orthophotos," *Polibits*, vol. 51, pp. 11–17, 2015.

[41] L. Xue, X. Yang, and Z. Cao, "Building extraction of sar images using morphological attribute profiles," in *Communications, Signal Processing, and Systems*. Springer, 2012, pp. 13–21.

[42] Y. Zhang, C. Wang, X. Chen, and S. Su, "Support vector machine approach to identifying buildings using multi-temporal ALOS/PALSAR data," *International Journal of Remote Sensing*, vol. 32, no. 22, pp. 7163–7177, 2011.

[43] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 77, no. 7, pp. 721–732, 2011.

[44] Q. Bi, K. Qin, H. Zhang, Y. Zhang, Z. Li, and K. Xu, "A multi-scale filtering building index for building extraction in very high-resolution satellite imagery," *Remote Sensing*, vol. 11, no. 5, p. 482, 2019.

[45] L. Castrejon, K. Kundu, R. Urtasun, and S. Fidler, "Annotating object instances with a polygon-rnn," in *CVPR*, 2017, pp. 5230–5238.

[46] Z. Li, J. D. Wegner, and A. Lucchi, "Topological map extraction from overhead images," in *ICCV*, 2019, pp. 1715–1724.

[47] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 2117–2125.

[48] W. Zhao, C. Persello, and A. Stein, "Extracting planar roof structures from very high resolution images using graph neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 187, pp. 34–45, 2022.

[49] S. Zorzi, S. Bazrafkan, S. Habenschuss, and F. Fraundorfer, "Polyworld: Polygonal building extraction with graph neural networks in satellite images," in *CVPR*, 2022, pp. 1848–1857.

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[51] Y. Hu, Z. Wang, Z. Huang, and Y. Liu, "Polybuilding: Polygon transformer for building extraction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 199, pp. 15–27, 2023.

[52] T. Lu, D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang, "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network," *Remote Sensing*, vol. 10, no. 9, p. 1496, 2018.

[53] G. Wu, Z. Guo, X. Shi, Q. Chen, Y. Xu, R. Shibasaki, and X. Shao, "A boundary regulated network for accurate roof segmentation and outline extraction," *Remote Sensing*, vol. 10, no. 8, p. 1195, 2018.

[54] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *ICCV*, 2017, pp. 2961–2969.

[55] A. Hu, L. Wu, S. Chen, Y. Xu, H. Wang, and Z. Xie, "Boundary shape-preserving model for building mapping from high-resolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[56] W. Li, W. Zhao, J. Yu, J. Zheng, C. He, H. Fu, and D. Lin, "Joint semantic–geometric learning for polygonal building segmentation from high-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 201, pp. 26–37, 2023.

[57] D. Marcos, D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, and R. Urtasun, "Learning deep structured active contours end-to-end," in *CVPR*, 2018, pp. 8877–8885.

[58] D. Cheng, R. Liao, S. Fidler, and R. Urtasun, "DARNet: Deep active ray network for building segmentation," in *CVPR*, 2019, pp. 7431–7439.

[59] J. Xia, N. Yokoya, B. Adriano, L. Zhang, G. Li, and Z. Wang, "A benchmark high-resolution gaofen-3 sar dataset for building semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5950–5963, 2021.

[60] G. Wu, X. Shao, Z. Guo, Q. Chen, W. Yuan, X. Shi, Y. Xu, and R. Shibasaki, "Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks," *Remote Sensing*, vol. 10, no. 3, p. 407, 2018.

[61] S. Ji, S. Wei, and M. Lu, "A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery," *International Journal of Remote Sensing*, vol. 40, no. 9, pp. 3308–3322, 2019.

[62] S. Wei, S. Ji, and M. Lu, "Toward automatic building footprint delineation from aerial images using CNN and regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2178–2189, 2019.

[63] W. Deng, Q. Shi, and J. Li, "Attention-gate-based encoder–decoder network for automatical building extraction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2611–2620, 2021.

[64] H. Jing, X. Sun, Z. Wang, K. Chen, W. Diao, and K. Fu, "Fine building segmentation in high-resolution sar images via selective pyramid dilated network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6608–6623, 2021.

[65] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 574–586, 2018.

[66] D. Zhou, G. Wang, G. He, R. Yin, T. Long, Z. Zhang, S. Chen, and B. Luo, "A large-scale mapping scheme for urban building from Gaofen-2 images using deep learning and hierarchical approach," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 530–11 545, 2021.

[67] H. Guo, Q. Shi, B. Du, L. Zhang, D. Wang, and H. Ding, "Scene-driven multitask parallel attention network for building extraction in high-resolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4287–4306, 2020.

[68] Z. Zhang, W. Guo, W. Yu, and W. Yu, "Multi-task fully convolutional networks for building segmentation on sar image," *The Journal of Engineering*, vol. 2019, no. 20, pp. 7074–7077, 2019.

[69] J. Yuan, "Learning building extraction in aerial scenes with convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2793–2798, 2017.

[70] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multi-task learning for segmentation of building footprints with deep neural networks," in *ICIP*. IEEE, 2019, pp. 1480–1484.

[71] N. Girard, D. Smirnov, J. Solomon, and Y. Tarabalka, "Polygonal building extraction by frame field learning," in *CVPR*, 2021, pp. 5891–5900.

[72] Q. Li, L. Mou, Y. Hua, Y. Shi, and X. X. Zhu, "Building footprint generation through convolutional neural networks with attraction field representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2021.

[73] B. Xu, J. Xu, N. Xue, and G.-S. Xia, "Hisup: Accurate polygonal mapping of buildings in satellite imagery with hierarchical supervision," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 198, pp. 284–296, 2023.

[74] X. Li, X. Yao, and Y. Fang, "Building-A-Nets: robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3680–3687, 2018.

[75] S. Zorzi and F. Fraundorfer, "Regularization of building boundaries in satellite images using adversarial and regularized losses," in *IGARSS*. IEEE, 2019, pp. 5140–5143.

[76] Y. Shi, Q. Li, and X. Zhu, "Building footprint extraction with graph convolutional network," in *IGARSS*. IEEE, 2019, pp. 5136–5139.

[77] Q. Li, Y. Shi, X. Huang, and X. X. Zhu, "Building footprint generation by integrating convolution neural network with feature pairwise conditional random field (FPCRF)," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[78] L. Xia, X. Zhang, J. Zhang, H. Yang, and T. Chen, "Building extraction from very-high-resolution remote sensing images using semi-supervised semantic edge detection," *Remote Sensing*, vol. 13, no. 11, p. 2187, 2021.

[79] J.-H. Lee, C. Kim, and S. Sull, "Weakly supervised segmentation of small buildings with point labels," in *ICCV*, 2021, pp. 7406–7415.

[80] M. U. Rafique and N. Jacobs, "Weakly supervised building segmentation from aerial images," in *IGARSS*. IEEE, 2019, pp. 3955–3958.

[81] D. Zheng, S. Li, F. Fang, J. Zhang, Y. Feng, B. Wan, and Y. Liu, "Utilizing bounding box annotations for weakly supervised building extraction from remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[82] J. Chen, F. He, Y. Zhang, G. Sun, and M. Deng, "SPMF-Net: Weakly supervised building segmentation by combining superpixel pooling and multi-scale feature fusion," *Remote Sensing*, vol. 12, no. 6, p. 1049, 2020.

[83] Z. Li, X. Zhang, P. Xiao, and Z. Zheng, "On the effectiveness of weakly supervised semantic segmentation for building extraction from high-resolution remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3266–3281, 2021.

[84] J. Wang, C. HQ Ding, S. Chen, C. He, and B. Luo, "Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label," *Remote Sensing*, vol. 12, no. 21, p. 3603, 2020.

[85] E. Lee, S. Jeong, J. Kim, and K. Sohn, "Semantic equalization learning for semi-supervised sar building segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[86] X. Li, M. Luo, S. Ji, L. Zhang, and M. Lu, "Evaluating generative adversarial networks based image-level domain transfer for multi-source remote sensing image segmentation and object detection," *International Journal of Remote Sensing*, vol. 41, no. 19, pp. 7343–7367, 2020.

[87] L. Shi, Z. Wang, B. Pan, and Z. Shi, "An end-to-end network for remote sensing imagery semantic segmentation via joint pixel-and representation-level domain adaptation," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[88] H. Liu, J. Luo, B. Huang, X. Hu, Y. Sun, Y. Yang, N. Xu, and N. Zhou, "DE-Net: Deep encoding network for building extraction from high-resolution remote sensing imagery," *Remote Sensing*, vol. 11, no. 20, p. 2380, 2019.

[89] M. Chen, J. Wu, L. Liu, W. Zhao, F. Tian, Q. Shen, B. Zhao, and R. Du, "DR-Net: An improved network for building extraction from high resolution remote sensing image," *Remote Sensing*, vol. 13, no. 2, p. 294, 2021.

[90] X. Wang, L. Cavigelli, M. Eggimann, M. Magno, and L. Benini, "Hr-sar-net: A deep neural network for urban scene segmentation from high-resolution sar data," in *SAS*. IEEE, 2020, pp. 1–6.

[91] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[92] J. Liu and Y. Liu, "Local regularity-driven city-scale facade detection from aerial images," in *CVPR*, 2014, pp. 3778–3785.

[93] Y. Sun, Y. Hua, L. Mou, and X. X. Zhu, "CG-Net: Conditional GIS-aware network for individual building segmentation in vhr sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[94] X. Yang, X. Qin, J. Wang, J. Wang, X. Ye, and Q. Qin, "Building façade recognition using oblique aerial images," *Remote Sensing*, vol. 7, no. 8, pp. 10 562–10 588, 2015.

[95] M. Kakooei, Y. Baleghi, and M. Amani, "Adaptive thresholding for detecting building facades with or without openings in single-view oblique remote sensing images," *Journal of Applied Remote Sensing*, vol. 15, no. 3, p. 036511, 2021.

[96] W. Li, L. Meng, J. Wang, C. He, G.-S. Xia, and D. Lin, "3D building reconstruction from monocular remote sensing images," in *ICCV*, 2021, pp. 12 548–12 557.

[97] K. Mainzer, S. Killinger, R. McKenna, and W. Fichtner, "Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques," *Solar Energy*, vol. 155, pp. 561–573, 2017.

[98] S. Lee, S. Iyengar, M. Feng, P. Shenoy, and S. Maji, "Deeproof: A data-driven approach for solar potential estimation using rooftop imagery," in *ACM SIGKDD*, 2019, pp. 2105–2113.

[99] S. Krapf, L. Bogenrieder, F. Netzler, G. Balke, and M. Lienkamp, "RID—roof information dataset for computer vision-based photovoltaic potential assessment," *Remote Sensing*, vol. 14, no. 10, p. 2299, 2022.

[100] Y. E. Merabet, C. Meurie, Y. Ruichek, A. Sbihi, and R. Touahni, "Building roof segmentation from aerial images using a line-and region-based watershed segmentation technique," *Sensors*, vol. 15, no. 2, pp. 3172–3203, 2015.

[101] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.

[102] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 06, pp. 583–598, 1991.

[103] S. Krapf, B. Willenborg, K. Knoll, M. Bruhse, and T. H. Kolbe, "Deep learning for semantic 3d city model extension: Modeling roof superstructures using aerial images for solar potential analysis." *ISPRS Annals*, vol. 10, 2022.

[104] Q. Li, L. Mou, Y. Hua, Y. Shi, S. Chen, Y. Sun, and X. X. Zhu, "3DCentripetalNet: Building height retrieval from monocular remote sensing imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 120, p. 103311, 2023.

[105] Y. Sun, L. Mou, Y. Wang, S. Montazeri, and X. X. Zhu, "Large-scale building height retrieval from single SAR imagery based on bounding box regression networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 184, pp. 79–95, 2022.

[106] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, 2012.

[107] M. Izadi and P. Saeedi, "Three-dimensional polygonal building model estimation from single satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2254–2272, 2011.

[108] G. L. Laprade and E. Leonardo, "Elevations from radar imagery," *Photogrammetric Engineering*, 1969.

[109] M. Jahangir, D. Blacknell, C. Moate, and R. Hill, "Extracting information from shadows in SAR imagery," in *ICMV*. IEEE, 2007, pp. 107–112.

[110] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, "Building height retrieval from vhr sar imagery based on an iterative simulation and matching technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1487–1504, 2009.

[111] Z. Wang, L. Jiang, L. Lin, and W. Yu, "Building height estimation from high resolution sar imagery via model-based geometrical structure prediction," *Progress In Electromagnetics Research M*, vol. 41, pp. 11–24, 2015.

[112] S. Srivastava, M. Volpi, and D. Tuia, "Joint height estimation and semantic labeling of monocular aerial images with CNNs," in *IGARSS*. IEEE, 2017, pp. 5173–5176.

[113] J. Mahmud, T. Price, A. Bapat, and J.-M. Frahm, "Boundary-aware 3D building reconstruction from a single overhead image," in *CVPR*, 2020, pp. 441–451.

[114] J. Chen, X. Qiu, C. Ding, and Y. Wu, "CVCMFF Net: Complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[115] S. Du, F. Zhang, and X. Zhang, "Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 107–119, 2015.

[116] J. Xie and J. Zhou, "Classification of urban building type from high spatial resolution remote sensing imagery using extended MRS and soft BP network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3515–3528, 2017.

[117] T. Partovi, F. Fraundorfer, S. Azimi, D. Marmanis, and P. Reinartz, "Roof type selection based on patch-based classsification using deep learning for high resolution satellite imagery," *ISPRS Archives*, vol. 42, no. W1, pp. 653–657, 2017.

[118] X. Huang, L. Ren, C. Liu, Y. Wang, H. Yu, M. Schmitt, R. Hänsch, X. Sun, H. Huang, and H. Mayer, "Urban building classification (UBC)-A dataset for individual building detection and classification from satellite imagery," in *CVPR*, 2022, pp. 1413–1421.

[119] N. Skuppin, E. J. Hoffmann, Y. Shi, and X. X. Zhu, "Building type classification with incomplete labels," in *IGARSS*. IEEE, 2022, pp. 5844–5847.

[120] L. Li, C. Wang, H. Zhang, B. Zhang, and F. Wu, "Urban building change detection in SAR images using combined differential image and residual u-net network," *Remote Sensing*, vol. 11, no. 9, p. 1091, 2019.

[121] C. Liao, H. Hu, X. Yuan, H. Li, C. Liu, C. Liu, G. Fu, Y. Ding, and Q. Zhu, "Bce-net: Reliable building footprints change extraction based on historical map and up-to-date images using contrastive learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 201, pp. 138–152, 2023.

[122] J. Ge, H. Tang, N. Yang, and Y. Hu, "Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 105–128, 2023.

[123] D. Wen, X. Huang, F. Bovolo, J. Li, X. Ke, A. Zhang, and J. A. Benediktsson, "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 68–101, 2021.

[124] X. Huang, L. Zhang, and T. Zhu, "Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 105–115, 2013.

[125] Y. Tang, X. Huang, and L. Zhang, "Fault-tolerant building change detection from urban high-resolution remote sensing imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1060–1064, 2013.

[126] P. T. Brett and R. Guida, "Earthquake damage detection in urban areas using curvilinear features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4877–4884, 2013.

[127] C. Marin, F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution SAR images," *IEEE transactions on geoscience and remote sensing*, vol. 53, no. 5, pp. 2664–2682, 2014.

[128] L. Song, M. Xia, J. Jin, M. Qian, and Y. Zhang, "SUACDNet: Attentional change detection network based on siamese U-shaped structure," *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, p. 102597, 2021.

[129] D. Wang, X. Chen, M. Jiang, S. Du, B. Xu, and J. Wang, "ADS-Net: An attention-based deeply supervised network for remote sensing image change detection," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, p. 102348, 2021.

[130] Q. Ding, Z. Shao, X. Huang, and O. Altan, "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*, vol. 105, p. 102591, 2021.

[131] L. Pang, F. Zhang, L. Li, Q. Huang, Y. Jiao, and Y. Shao, "Assessing buildings damage from multi-temporal sar images fusion using semantic change detection," in *IGARSS*. IEEE, 2022, pp. 6292–6295.

[132] J. Li, C. Wang, H. Zhang, F. Wu, L. Li, and L. Gong, "Automatic extraction of built-up areas for cities in china from gf-3 images based on improved residual u-net network," in *IGARSS*. IEEE, 2020, pp. 4399–4402.

[133] T. Liu, M. Gong, D. Lu, Q. Zhang, H. Zheng, F. Jiang, and M. Zhang, "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*, vol. 60, pp. 1–17, 2021.

[134] Y. Zhang, M. Deng, F. He, Y. Guo, G. Sun, and J. Chen, "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*, vol. 14, pp. 8125–8134, 2021.

[135] J. Zheng, Y. Tian, C. Yuan, K. Yin, F. Zhang, F. Chen, and Q. Chen, "MDESNet: Multitask difference-enhanced siamese network for building change detection in high-resolution remote sensing images," *Remote Sensing*, vol. 14, no. 15, p. 3775, 2022.

[136] Z. Chen, Y. Zhou, B. Wang, X. Xu, N. He, S. Jin, and S. Jin, "Egde-net: A building change detection method for high-resolution remote sensing imagery based on edge guidance and differential enhancement," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 191, pp. 203–222, 2022.

[137] F. Zhou, C. Xu, R. Hang, R. Zhang, and Q. Liu, "Mining joint intra- and inter-image context for remote sensing change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[138] H. Chen, W. Li, and Z. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[139] W. G. C. Bandara and V. M. Patel, "Revisiting consistency regularization for semi-supervised change detection in remote sensing images," *arXiv preprint arXiv:2204.08454*, 2022.

[140] C. Sun, J. Wu, H. Chen, and C. Du, "SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using siamese networks with graph attention," *Remote Sensing*, vol. 14, no. 12, p. 2801, 2022.

[141] N. Ahmed, R. M. Rahman, M. S. G. Adnan, and B. Ahmed, "Dense prediction of label noise for learning building extraction from aerial drone imagery," *International Journal of Remote Sensing*, vol. 42, no. 23, pp. 8906–8929, 2021.

[142] N. Girard, G. Charpiat, and Y. Tarabalka, "Noisy supervision for correcting misaligned cadaster maps without perfect ground truth data," in *IGARSS*. IEEE, 2019, pp. 10 103–10 106.

[143] J. E. Vargas-Muñoz, S. Lobry, A. X. Falcão, and D. Tuia, "Correcting rural building annotations in openstreetmap using convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 283–293, 2019.

[144] S. Fobi, T. Conlon, J. Taneja, and V. Modi, "Learning to segment from misaligned and partial labels," in *ACM SIGCAS*, 2020, pp. 286–290.

[145] C. Ayala, R. Sesma, C. Aranda, and M. Galar, "A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery," *Remote Sensing*, vol. 13, no. 16, p. 3135, 2021.

[146] N. Ahmed and R. M. Rahman, "Label noise tolerance of deep semantic segmentation networks for extracting buildings in ultra-high-resolution aerial images of semi-built environments," *Geocarto International*, vol. 37, no. 25, pp. 8062–8079, 2022.

[147] Y. Zhang, W. Li, W. Gong, Z. Wang, and J. Sun, "An improved boundary-aware perceptual loss for building extraction from VHR images," *Remote Sensing*, vol. 12, no. 7, p. 1195, 2020.

[148] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *ICML*, 2012, pp. 567–574.

[149] Z. Sun, F. Shen, D. Huang, Q. Wang, X. Shu, Y. Yao, and J. Tang, "Pnp: Robust learning from noisy labels by probabilistic noise prediction," in *CVPR*, 2022, pp. 5311–5320.

[150] C. M. Gevaert, C. Persello, S. O. Elberink, G. Vosselman, and R. Sliuzas, "Context-based filtering of noisy labels for automatic basemap

updating from UAV data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2731–2741, 2017.

[151] H. Song, L. Yang, and J. Jung, "Self-filtered learning for semantic segmentation of buildings in remote sensing imagery with noisy labels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1113–1129, 2023.

[152] D. Bulatov, "Alignment of building footprints using quasi-nadir aerial photography," in *SCIA*. Springer, 2019, pp. 361–373.

[153] A. Zampieri, G. Charpiat, N. Girard, and Y. Tarabalka, "Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing," in *ECCV*, 2018, pp. 657–673.

[154] N. Girard, G. Charpiat, and Y. Tarabalka, "Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning," in *ACCV*. Springer, 2018, pp. 675–690.

[155] Z. Zhang, W. Guo, M. Li, and W. Yu, "GIS-supervised building extraction with label noise-adaptive fully convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 12, pp. 2135–2139, 2020.

[156] H. Chen, W. Xie, A. Vedaldi, and A. Zisserman, "AutoCorrect: Deep inductive alignment of noisy geometric annotations," *arXiv preprint arXiv:1908.05263*, 2019.

[157] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*. Springer, 2014, pp. 740–755.

[158] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *CVPR*, 2016, pp. 724–732.

[159] S. Wang, M. Bai, G. Mattyus, H. Chu, W. Luo, B. Yang, J. Liang, J. Cheverie, S. Fidler, and R. Urtasun, "Torontocity: Seeing the world with a million eyes," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3009–3017.

[160] K. Zhao, J. Kang, J. Jung, and G. Sohn, "Building extraction from satellite images using mask R-CNN with building boundary regularization," in *CVPR Workshops*, 2018, pp. 247–251.

[161] D. H. Douglas and T. K. Peucker, "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*, vol. 10, no. 2, pp. 112–122, 1973.

[162] S. Zorzi, K. Bittner, and F. Fraundorfer, "Machine-learned regularization and polygonization of building segmentation masks," in *ICPR*. IEEE, 2021, pp. 3098–3105.

[163] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using vhr optical and sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, 2010.

[164] X. Wang and P. Li, "Extraction of urban building damage using spectral, height and corner information from vhr satellite images and airborne lidar data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 322–336, 2020.

[165] H. Hosseinpour, F. Samadzadegan, and F. D. Javan, "Cmgfnet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images," *ISPRS journal of photogrammetry and remote sensing*, vol. 184, pp. 96–115, 2022.

[166] X. Li, G. Zhang, H. Cui, S. Hou, Y. Chen, Z. Li, H. Li, and H. Wang, "Progressive fusion learning: A multimodal joint segmentation framework for building extraction from optical and sar images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 178–191, 2023.

[167] W. Wu, Z. Shao, X. Huang, J. Teng, S. Guo, and D. Li, "Quantifying the sensitivity of SAR and optical images three-level fusions in land cover classification to registration errors," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102868, 2022.

[168] X. Yao, Y. Wang, Y. Wu, and Z. Liang, "Weakly-supervised domain adaptation with adversarial entropy for building segmentation in cross-domain aerial imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8407–8418, 2021.

[169] Q. Li, L. Mou, Y. Hua, Y. Shi, and X. X. Zhu, "Crossgeonet: A framework for building footprint generation of label-scarce geographical regions," *International Journal of Applied Earth Observation and Geoinformation*, vol. 111, p. 102824, 2022.

[170] O. Tasar, S. Happy, Y. Tarabalka, and P. Alliez, "Colormapgan: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7178–7193, 2020.

[171] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[172] Y. Wang, C. M. Albrecht, N. Ait Ali Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geoscience and Remote Sensing Magazine*, 2022.

[173] H. Lu, F. Li, G. Yang, and W. Sun, "Multi-scale impacts of 2D/3D urban building pattern in intra-annual thermal environment of hangzhou, china," *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, p. 102558, 2021.

[174] R. Cong, M. Saito, R. Hirata, A. Ito, and S. Maksyutov, "Visualization on fossil-fuel carbon dioxide (co2) emissions from buildings in tokyo metropolis." *ISPRS Annals*, vol. 4, no. 4, 2018.

[175] T. Kubota, M. Miura, Y. Tominaga, and A. Mochida, "Wind tunnel tests on the relationship between building density and pedestrian-level wind velocity: Development of guidelines for realizing acceptable wind environment in residential neighborhoods," *Building and Environment*, vol. 43, no. 10, pp. 1699–1708, 2008.

[176] Y. Zhang, N. Chen, W. Du, Y. Li, and X. Zheng, "Multi-source sensor based urban habitat and resident health sensing: A case study of wuhan, china," *Building and Environment*, vol. 198, p. 107883, 2021.

[177] G. V. Fracastoro and M. Serraino, "A methodology for assessing the energy performance of large scale building stocks and possible applications," *Energy and Buildings*, vol. 43, no. 4, pp. 844–852, 2011.

[178] Q. Li, S. Krapf, Y. Shi, and X. X. Zhu, "SolarNet: A convolutional neural network-based framework for rooftop solar potential estimation from aerial imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103098, 2023.

[179] C. J. Van Westen, "Remote sensing and gis for natural hazards assessment and disaster risk management," *Treatise on geomorphology*, vol. 3, pp. 259–298, 2013.

[180] M. Mück, H. Taubenböck, J. Post, S. Wegscheider, G. Strunz, S. Sumaryono, and F. Ismail, "Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data," *Natural hazards*, vol. 68, no. 1, pp. 97–114, 2013.

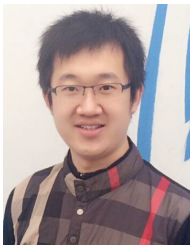
[181] T. G. Tiecke, X. Liu, A. Zhang, A. Gros, N. Li, G. Yetman, T. Kilic, S. Murray, B. Blankespoor, E. B. Prydz *et al.*, "Mapping the world population one building at a time," *arXiv preprint arXiv:1712.05839*, 2017.

[182] G. Boo, E. Darin, D. R. Leasure, C. A. Dooley, H. R. Chamberlain, A. N. Lázár, K. Tschirhart, C. Sinai, N. A. Hoff, T. Fuller *et al.*, "High-resolution population estimation using household survey data and building footprints," *Nature Communications*, vol. 13, no. 1, pp. 1–10, 2022.

Qingyu Li (S'21) received the bachelor's degree in Remote Sensing Science and Technology from Wuhan University, Wuhan, China, in 2015, and the master's degree in Earth Oriented Space Science and Technology (ESPACE) from Technische Universität München (TUM), Munich, Germany, in 2018, and the master's degree in Photogrammetry and Remote Sensing from Wuhan University, Wuhan, China, in 2019, and the doctor of engineering (Dr.-Ing.) degree from TUM, Munich, Germany, in 2022.

She is currently a postdoctoral researcher with the chair of Data Science in Earth Observation at TUM, Munich, Germany. From 2019 to 2022, she was a Research Associate at the Remote Sensing Technology Institute (IMF) of the German Aerospace Center (DLR).

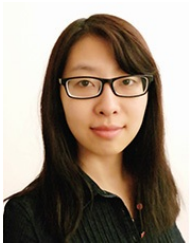
Her research interests include remote sensing data processing and interpretation, artificial intelligence, remote sensing applications, and urban analysis.



Lichao Mou received the Bachelor's degree in automation from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2012, the Master's degree in signal and information processing from the University of Chinese Academy of Sciences (UCAS), China, in 2015, and the Dr.-Ing. degree from the Technical University of Munich (TUM), Munich, Germany, in 2020.

He is currently a Guest Professor at the Munich AI Future Lab AI4EO, TUM and the Head of Visual Learning and Reasoning team at the Department "EO Data Science", Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany. Since 2019, he is a Research Scientist at DLR-IMF and an AI Consultant for the Helmholtz Artificial Intelligence Cooperation Unit (HAICU). In 2015 he spent six months at the Computer Vision Group at the University of Freiburg in Germany. In 2019 he was a Visiting Researcher with the Cambridge Image Analysis Group (CIA), University of Cambridge, UK.

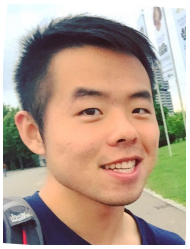
He was the recipient of the first place in the 2016 IEEE GRSS Data Fusion Contest and finalists for the Best Student Paper Award at the 2017 Joint Urban Remote Sensing Event and 2019 Joint Urban Remote Sensing Event.



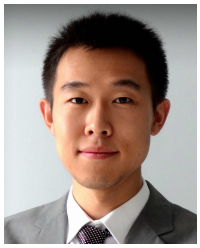
Yao Sun received the Bachelor's degree in cartography and geo-information system from Wuhan University, Wuhan, China, in 2012, the Master's degree in Earth Oriented Space Science and Technology (ESPACE) from Technical University of Munich (TUM), Munich, Germany, in 2016, and her doctor of engineering (Dr.-Ing.) degree from Technical University of Munich (TUM), Munich, Germany, in 2021.

She is currently a Postdoc Researcher with the chair of Data science in earth observation at the Technical University of Munich (TUM), Munich, Germany. From 2016 to 2021, she was a Research Scientist at the Remote Sensing Technology Institute (IMF) of the German Aerospace Center (DLR).

Her main research interests are building information extraction, post-disaster damage estimation, computer vision, and deep learning, especially leveraging remote sensing data and open geo-information sources for urban environment analysis.



Yuansheng Hua (S'18) received the bachelor's degree in Remote Sensing Science and Technology from Wuhan University, Wuhan, China, in 2014, and double master's degrees in Earth Oriented Space Science and Technology (ESPACE) and Photogrammetry and remote sensing from the Technical University of Munich (TUM), Munich, Germany, and Wuhan University, Wuhan, China, in 2018 and 2019, respectively. He received the Ph.D. degree from the Technical University of Munich (TUM), Munich, Germany in 2022. In 2019, he was a visiting researcher with the Wageningen University & Research, Wageningen, Netherlands. Currently, he is an assistant professor in the School of Civil and Traffic Engineering, Shenzhen University. His research interests include remote sensing, computer vision, and deep learning, especially their applications in remote sensing.



Yilei Shi (M'18) received the Dipl.-Ing degree in mechanical engineering and Dr.-Ing degree in signal processing from the Technische Universität München (TUM), Munich, Germany, in 2010 and 2019, respectively. In April and May 2019, he was a guest scientist with the department of applied mathematics and theoretical physics, University of Cambridge, United Kingdom. He is currently a senior scientist with the School of Engineering and Design, TUM.

His research interests include fast solver and parallel computing for large-scale problems, high performance computing and computational intelligence, advanced methods on SAR and InSAR processing, machine learning and deep learning for variety of data sources, such as SAR, optical images, and medical images, and PDE-related numerical modeling and computing.

Xiao Xiang Zhu (S'10–M'12–SM'14–F'21) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her "Habilitation" in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is the Chair Professor for Data Science in Earth Observation at Technical University of Munich (TUM) and was the founding Head of the Department "EO Data Science" at the Remote Sensing Technology Institute, German Aerospace Center (DLR). Since May 2020, she is the PI and director of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond", Munich, Germany. Since October 2020, she also serves as a Director of the Munich Data Science Institute (MDSI), TUM. From 2019 to 2022, Zhu has been a co-coordinator of the Munich Data Science Research School (www.mu-ds.de) and the head of the Helmholtz Artificial Intelligence – Research Field "Aeronautics, Space and Transport". Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan and University of California, Los Angeles, United States in 2009, 2014, 2015 and 2016, respectively. She is currently a visiting AI professor at ESA's Phi-lab, Frascati, Italy. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with their applications in tackling societal grand challenges, e.g. Global Urbanization, UN's SDGs and Climate Change.

Dr. Zhu has been a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the scientific advisory board in several research organizations, among others the German Research Center for Geosciences (GFZ, 2020-2023) and Potsdam Institute for Climate Impact Research (PIK). She is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing, Pattern Recognition and served as the area editor responsible for special issues of IEEE Signal Processing Magazine (2021-2023). She is a Fellow of IEEE.