ISPRS Journal of Photogrammetry and Remote Sensing 187 (2022) 306–327

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives

Yansheng Li [a,*], Bo Dang [a], Yongjun Zhang [a], Zhenhong Du [b]

[a] School of Remote Sensing and Information Engineering, Wuhan University, China
[b] School of Earth Sciences, Zhejiang University, China

ARTICLE INFO

Keywords:
Water body classification
High-resolution
Optical remote sensing (RS) image
Deep learning (DL)

ABSTRACT

Water body classification from high-resolution optical remote sensing (RS) images, aiming at classifying whether each pixel of the image is water or not, has become a hot issue in the area of RS and has extensive practical applications in a variety of fields. Numerous existing methods have drawn broad attention and achieved remarkable advancements, meanwhile, serious challenges and potential opportunities also exist, which deserves in thinking and discussing deeply. By taking into account the comprehensive survey is still lacking, through the compilation of approximately 200 papers, this paper summarizes and analyzes the achievements, and discusses the perspectives of future research directions. Specifically, we first analyze 5 challenges according to the characteristics of water bodies in high-resolution optical RS imagery, and 5 corresponding significant opportunities combined with advanced deep learning techniques are discussed to respond mentioned challenges. Then, we divide the existing methods into several groups in light of their core ideas and introduce them chiefly. In addition, some practical applications and publicly open benchmarks are listed intuitively. 10 and 9 representative methods are implemented on two widely used datasets to assess their performance, respectively. To facilitate the qualitative and quantitative comparison in the research avenue, the two benchmarks employed in the comparative experiments and links to other relevant datasets and open-source codes will be summarized and released in https://github.com/Jack-bo1220/Benchmarks-for-Water-Body-Extraction-from-HRORS-Imagery. Finally, we discuss a range of promising research directions to provide some references and inspiration for the following research. The studies of our paper, including the existing methods, challenges, opportunities, derived applications, and future research directions, provide a fuller understanding of water body classification from high-resolution optical remote sensing imagery.

1. Introduction

With the great advances of remote sensing (RS) sensors, computer science, and other technologies, RS has ushered in a climax of development, and the volume of that is growing at high speed. Numerous sensors, particularly high-resolution optical satellites and unmanned aerial vehicles (UAVs), become significant platforms for earth observation (McCabe et al., 2017; Li et al., 2012). Since 1972, the first satellite called ERTS-1 applied to survey and research the earth's surface was launched by NASA, which marks the beginning of the task of monitoring the land cover on Earth's surface from the angle of view of space (Huang et al., 2018). Subsequently, a large number of satellites were launched for various purposes, including Sentinel, SPOT, Gaofen, and WorldView series. Nanosatellites (e.g., GRUS-1A, SkySat series, and Flock series), especially CubeSats that use a standard size and form factor, have the advantages of short development cycle, low cost, and the ability to carry a variety of RS sensors, and have achieved rapid development. According to the differences of the sensors' imaging band, they can be divided into optical RS sensors and synthetic aperture radar (SAR) sensors. Further, taking into account the local coverage, optical RS sensors are categorized into 3 kinds in terms of spatial resolution—coarse-resolution (>200 m), medium-resolution (5–200 m), and high-resolution (<5 m) (Huang et al., 2018). In recent years, as a novel observation platform, UAV has been extensively used in specific tasks of RS (Adão et al., 2017; Nex and Remondino, 2014). Due to its portability, operation flexibility, and the ability to provide images with the high spatial resolution, it is becoming a new generation of sensors to supplement the conventional RS technology. Relying on the advancements

* Corresponding author.
E-mail addresses: yansheng.li@whu.edu.cn (Y. Li), bodang@whu.edu.cn (B. Dang), zhangyj@whu.edu.cn (Y. Zhang), duzhenhong@zju.edu.cn (Z. Du).

https://doi.org/10.1016/j.isprsjprs.2022.03.013
Received 29 November 2021; Received in revised form 15 March 2022; Accepted 17 March 2022
Available online 25 March 2022
0924-2716/© 2022 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

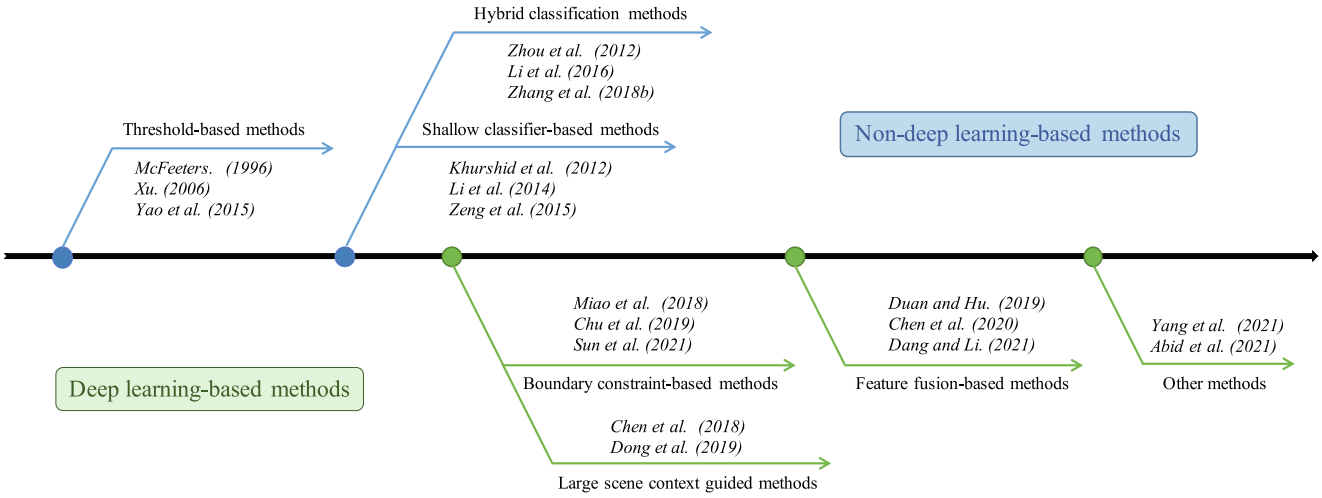

Fig. 1. The road map of existing methods for water body classification from high-resolution optical RS imagery. (For each method category, some representative work is selected for display.).

of the above sensor techniques, the high-resolution optical RS imagery obtained from high-resolution optical satellites and high-resolution UAVs are the focus of researchers' attention and are frequently used in various tasks in the area of RS.

Water bodies, as one of the fundamental elements of the earth, are not only essential to the natural ecological cycle, but closely geared to human life as well, such as health, irrigation, electric power generation, and so on (Vörösmarty et al., 2000). Water bodies mainly include rivers, canals, ponds, lakes, and seas. It should be noted that for water bodies that change with seasonal climate, for example, seasonal rivers that are covered by water in the rainy season but sand in the summer. When images are acquired during the rainy season, they are naturally identified as water bodies. Thanks to the earth observation capability of the abovementioned sensors and advanced image processing techniques, it is possible to extract water bodies from RS imagery. For optical RS images, the methods based on handcrafted features appear and demonstrate fine performance (McFeeters, 1996; Xu, 2006; Acharya et al., 2016). The wide application of the deep learning (DL) algorithm makes it further developed (Li et al., 2021c; Sun et al., 2021). For SAR images, there are also a considerable number of studies that achieved great success. For example, gray-level co-occurrence matrix (GLCM) and support vector machine (SVM) (Lv et al., 2010), Graph Cut model (Bao et al., 2021), homogeneity response (Sghaier et al., 2016), PA-UNet (Li et al., 2021b), cascaded fully-convolutional network (CFCN) (Zhang et al., 2020b), and superpixel segmentation (Pappas et al., 2020). Compared with the speckle noise interference of SAR data and the fuzziness of lower-resolution optical data which is difficult to obtain detailed information and is not conducive to accurately distinguishing small-sized land cover classification, high-resolution optical RS images have absolute advantages of observing and classifying the land use categories. Consequently, raising the spatial resolution of optical satellites has always been, without doubt, one of the most critical tasks in the field of RS, as shown in Tables A1 and A2.

In the background of increasing global water scarcity, water body classification from high-resolution optical RS imagery can be widely used in water resources assessment, environmental protection, urban planning, etc, which makes it attach importance in the RS community (Nath and Deb, 2010). Although there are still many challenges to be tackled in water body classification from SAR images and coarse-resolution optical RS images, it is worth mentioning that high-resolution optical RS images, which are the focus of this article, are increasingly utilized in the task of water body classification owing to extensive data sources, high enough spatial resolution.

Currently, a large number of water body classification methods have emerged and the accuracy of that is gradually increasing. To better organize and introduce the existing approaches, and explore the feature research, in this article, they are divided into two categories in light of the core idea of algorithms: non-DL-based methods and DL-based methods. Specifically, the former can be separated into three component subcategories: threshold-based methods, shallow classifier-based methods, and hybrid classification methods. Generally speaking, threshold-based methods usually formulate the discriminative water body index that relies on the different spectral response in various bands and make an attempt to combine them effectively. However, the second is to excavate the spatial features of images, such as edge, shape, and so on. In addition, unsupervised classification algorithms (e.g., K-means clustering (Likas et al., 2003), MRF model (Dubes et al., 1990), and decision tree (Friedl and Brodley, 1997)) or low-level machine learning models (Huang et al., 2015; Li et al., 2021a) are embedded. As its name implies, the third is to combine the above two categories to form an overall water body segmentation workflow. Owing to the powerful capability of deep neural networks mining the deeper-level abstract features of images, DL-based methods have achieved better performance, and have gradually become the mainstream approaches of water body classification from high-resolution optical RS imagery. Facing different motivations, the optimization schemes of these methods can be systematically summarized into four categories. In order to integrate more multi-scale or spatial-spectral feature information to adapt to the variability of water bodies, feature fusion-based methods emerged as time requires, and some studies (Duan and Hu, 2019; Chen et al., 2020; Sun et al., 2021) as milestones demonstrate great potential in this research direction. Boundary constraint-based methods are also common water body classification projects, mainly including well-designed boundary constraint loss functions (Miao et al., 2018), boundary refinement modules (Chen et al., 2020; Cui et al., 2020), and post-processing optimization (Sun et al., 2021; Feng et al., 2018; Li et al., 2019b). The methods driven by large scene context (Chen et al., 2018b; Dong et al., 2019) and cross-domain migration approaches (Yang et al., 2021; Abid et al., 2021) should not be ignored in this field. The road map of existing methods for water body classification from high-resolution optical RS imagery is shown in Fig. 1 clearly and intuitively.

To sum up, it is not hard to see that non-DL-based methods are usually specially formulated for certain imagery from specific sensors, which leads to their poor transferability and low efficiency of water body classification. Under the background of the RS big data era (Li et al., 2021h; Chi et al., 2016), non-DL-based methods cannot meet the needs of practical application. Therefore, it is urgent to extract water from high-resolution optical RS images accurately and efficiently.

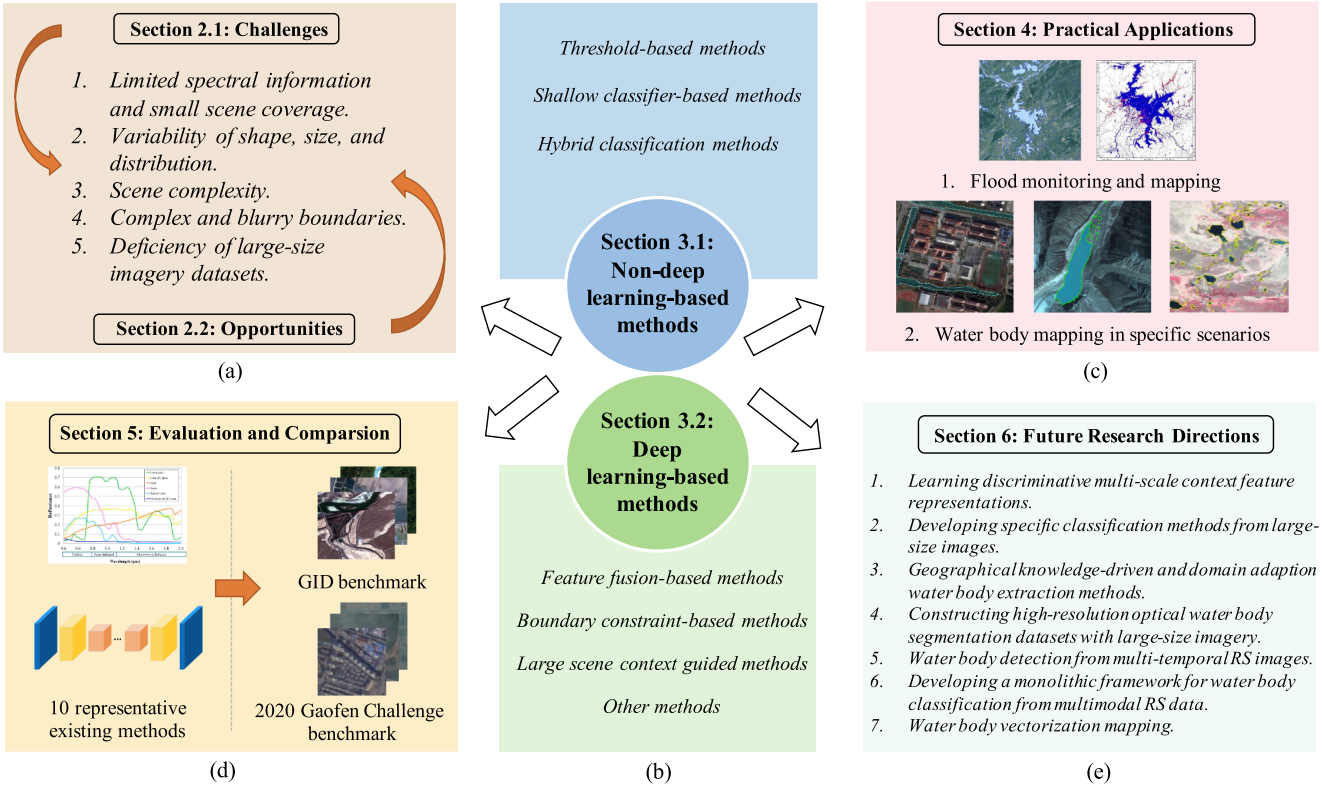

Fig. 2. An outline of this paper, including (a) challenges and opportunities, (b) existing methods, (c) practical applications (The pictures in the second and third columns of the second row of this component are downloaded from (Qayyum et al., 2020; Tian et al., 2017) and edited, respectively), (d) evaluation and comparison (The picture in the first row and first column of this component is downloaded from (Huang et al., 2018)), and (e) future research directions.

Fortunately, relying on the great advances of DL technology in image processing and visual recognition, the bottleneck of automatic water body classification has been broken through. A large number of explorations on DL technology appeared, which made considerable achievements and breakthroughs (Krizhevsky et al., 2012; He et al., 2016; Xie et al., 2017; Zhang et al., 2020a). Especially for semantic segmentation, fully connected network (FCN) (Long et al., 2015), PSPNet (Zhao et al., 2017), DeepLab series (Chen et al., 2017a; Chen et al., 2017b; Chen et al., 2018), High-resolution network (HRNet) (Wang et al., 2020b), Swin Transformer (Liu et al., 2021), and other advanced DL network architectures and training strategies continue to emerge, promoting the continuous improvement of classification performance. Motivated by DL technology, DL-based methods, as alternative ways to extract water bodies from high-resolution optical RS imagery, have made many attempts from different perspectives, which provide good inspiration for the follow-up improvement research. More exactly, there are four contributions for existing DL-based methods. First, fusing the high- and low-level features from different layers to enrich the representation of the final feature map, or aggregating contextual information at multi-scale receptive fields is to adapt the variable shape, size of water bodies and make full use of the multi-scale characteristic of water bodies. Then, through the boundary constraint loss, boundary refinement units, and post-processing algorithms, the complex details of boundaries can be preserved to the greatest extent. Last, some large scene context-guided methods and domain adaptive methods appeared gradually, which is beneficial for extracting water bodies from large-size scene imagery and images with the complex scene or distribution, respectively.

Although lots of achievements have been obtained in water body classification from high-resolution optical RS imagery, many tricky challenges still remain. However, a thorough review of water body classification from high-resolution optical RS imagery is still lacking, which motivates us to systematically analyze the challenges and investigate the existing methods. Different from the existing relevant review articles (Govender et al., 2007; Nath and Deb, 2010; Haibo et al., 2011; Jawak et al., 2015; Musa et al., 2015; Huang et al., 2018; Wang and Xie, 2018; Acharya et al., 2018; Shen et al., 2019; Bijeesh and Narasimhamurthy, 2020)(as shown in Table A3), our paper focuses on the developing methods, practical application, challenges, and opportunities of water body classification from high-resolution optical RS imagery. Through detailed theoretical analysis and intuitive experimental comparison, our paper aims to summarize the achievements and perspectives in this field and fill the gap of a systematic review in this field. To the best of our knowledge, about 47 methods have been disclosed, and they have made remarkable advancements. At present, the methods based on DL technique occupy the mainstream gradually. Then, 10 and 9 representative methods are reproduced and compared on two public benchmarks respectively, which will help potential readers quickly study and compare, and facilitate engineers and technicians to rapidly lock the most appropriate method by viewing this paper. Furthermore, we summarize some practical applications. Last but not least, combined with the current progress in the field of DL, future development opportunities and research directions are discussed, which is conducive to providing research ideas for young researchers.

The rest of this paper is organized as follows. Challenges and opportunities are discussed in Section 2. In Section 3, existing methods are categorized and introduced. Some practical applications are discussed in Section 4, and then, Section 5 mainly focuses on the publicly open datasets and applied some milestones to compare and discuss their performance. In Section 6, we conclude the possible future research directions that can provide inspiration for potential readers. Ultimately, we conclude this paper in Section 7. In addition, Fig. 2 intuitively displays the outline of this paper.

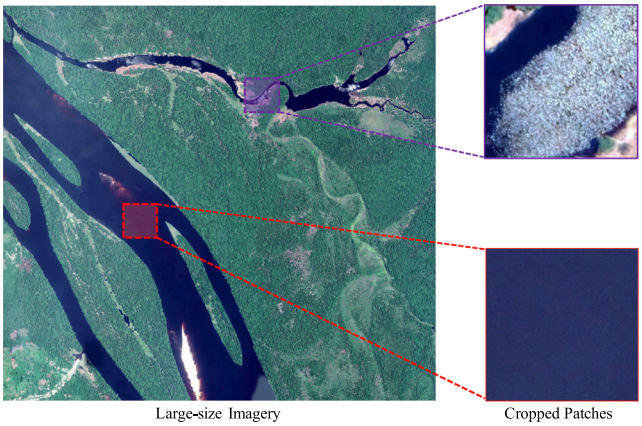

Fig. 3. Large-size imagery and corresponding cropped patches.

2. Water body classification from high-resolution optical remote sensing imagery: Challenges and opportunities

The use of remote sensing for automatic water body classification, as one of the significant tasks in the field of intelligent interpretation and monitoring of natural resources, still has many problems and challenges to be tackled. In the same way, a number of advanced artificial intelligence (AI) approaches, especially DL methods, not only bring broader development opportunities but also promote advanced algorithms to succeed in practical applications. In this section, we first systematically discuss the main challenges of water body classification from high-resolution optical RS imagery according to the characteristic of relevant images of water bodies. And then, aiming at several challenges we concluded, many cutting-edge DL approaches that can assist to solve these problems are analyzed one by one.

2.1. Challenges of water body classification from high-resolution optical remote sensing imagery

Compared with SAR data, despite high-resolution optical RS imagery being easy to be influenced by the natural climate and having no ability to provide 24 h and all-weather RS data, they still have incomparable advantages. Panchromatic and visible images with high-spatial resolution provide clear spatial texture information for water bodies. On the basis of that, multispectral data (e.g., WorldView-2 data) has the advantage of integrating spectral information and can present the intrinsical characteristics of water bodies. In other words, due to the different characteristics of receiving and radiating electromagnetic waves, different types of ground objects show different spectral curves.

In recent years, with the progress of high-resolution RS imaging technology, the texture structure of the obtained image is finer, the geometry of the ground object is clearer, and the detail difference between the visual and the natural image has become smaller and smaller (Cheng et al., 2020). Therefore, in addition to the existence of traditional water body classification methods, many advanced Computer Vision (CV) algorithms have been successfully transferred to water body classification from high-resolution optical RS images. However, it still needs to be noted that there has not yet been an algorithm that can achieve the target of that with satisfactory accuracy. Intuitively, the challenges include the following:

(1) **Limited spectral information and small scene coverage.** In terms of the preeminence of NDWI (McFeeters, 1996), automated water extraction index (AWEI) (Feyisa et al., 2014), water index created with linear discriminant analysis (WI2015) (Fisher et al., 2016), it demonstrates the significance and desirability of enough spectral bands, especially a series of infrared bands.Unfortunately, apart from the WorldView-3 satellite, other common optical high-resolution sensors have no capability of supplying much band information such as short-wave infrared band (SWIR), which makes most threshold-based methods ineffective. Additionally, it is easy to understand that with the increase of spatial resolution, the geographical area covered by each tile is narrower when the size of patches is fixed. Currently, mature DL algorithms prefer to crop large-size samples into small-size (e.g., 256 × 256, 512 × 512) tiles and then use them to train models. There are two major reasons for this. On one hand, the limited computational power and GPU memory cannot meet the setting of training large-size samples directly. On the other hand, the receptive field of neurons is finite, and small-size tiles are conducive to the model to learn abstract features better. However, different from natural imagery, RS imagery contains more complex geoscience knowledge and interdependent ground objects. Especially, river, lake, and so on are geographically continuous, but the small scene coverage undermines the application of continuity, as shown in Fig. 3.

(2) **Variability of shape, size, and distribution.** Water bodies include surface rivers, ponds, lakes, oceans, and so on. Their

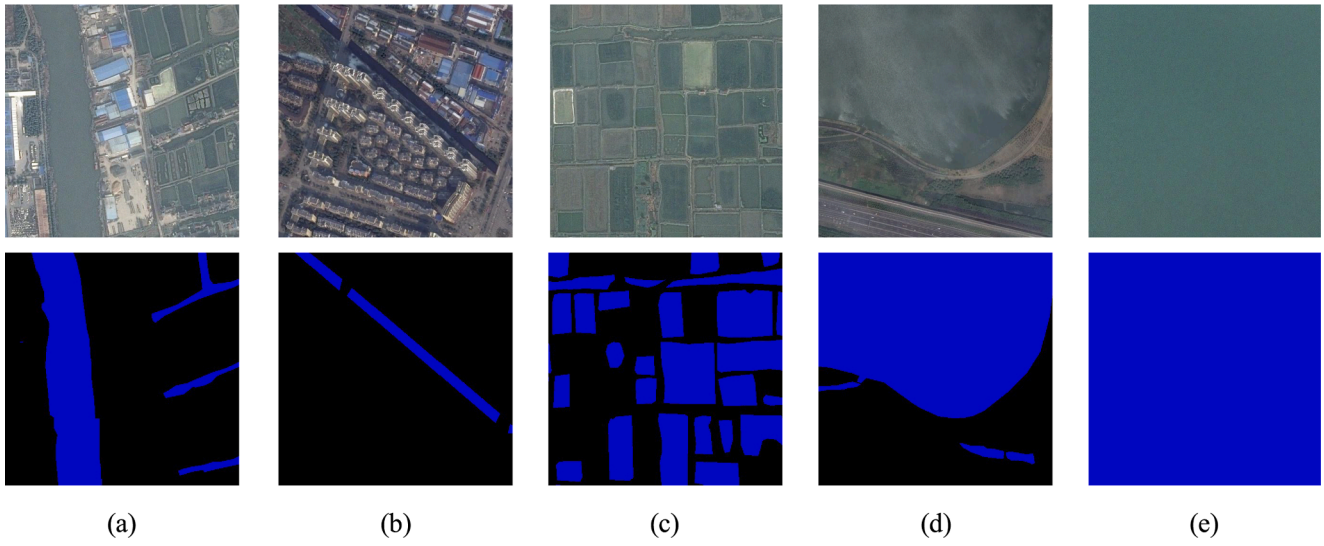

Fig. 4. Water bodies with different scale characteristics, including (a) rivers, (b) canals, (c) ponds, (d) lakes, and (e) seas.

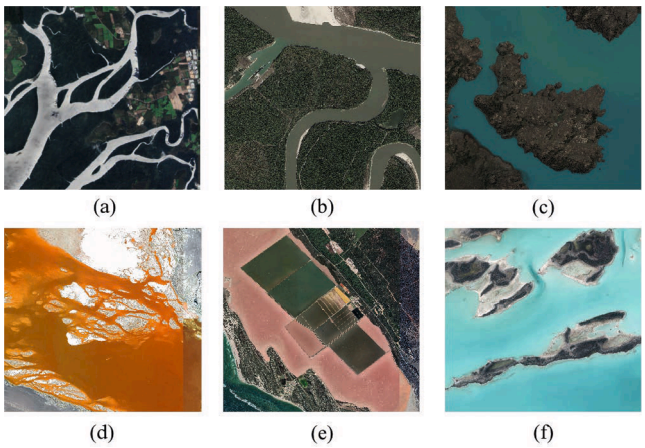

Fig. 5. Water body images from different areas of the world captured by DigitalGlobe and downloaded through Google Earth software (spatial resolution: 0.3 m). Data source areas: (a) Thailand; (b) Mississippi River, U.S.A.; (c) Bertalano Lake, Argentina; (d) Colorado River, Bolivia; (e) Hillier Lake, Australia; (f) Bahamas.

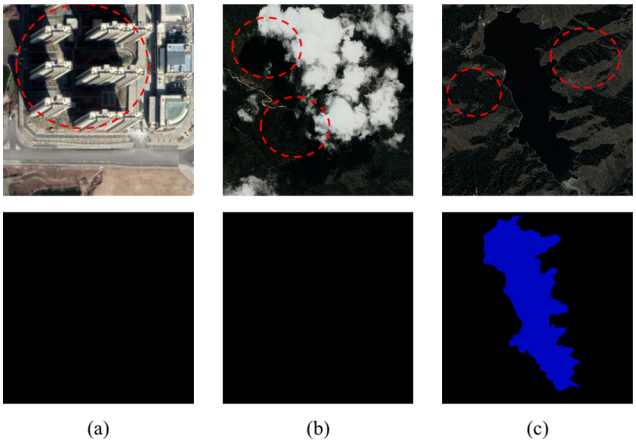

Fig. 6. Samples disturbed by (a) building shadows, (b) cloud shadows, and (c) mountain shadows. The red dashed circle indicates the shadow coverage area that is prone to misclassification.

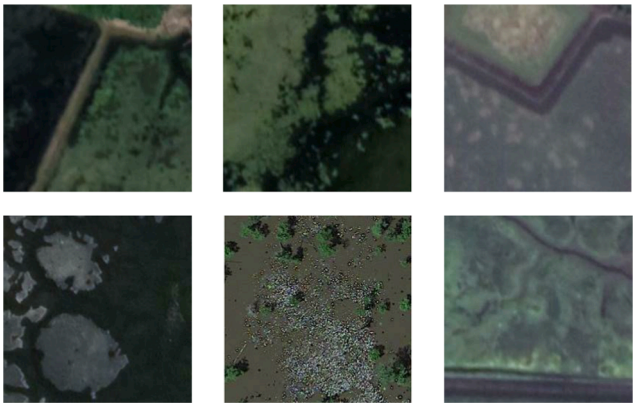

Fig. 7. Polluted black and smelly water bodies.

shapes are various. Therefore, if classification methods only depend on morphological algorithms, it is difficult to define water bodies with a unified paradigm. In addition, water bodies can be roughly divided into large-size ones (e.g., lakes and trunk streams) and small ones (e.g., ponds and tributaries). These multi-scale characteristics present many challenges to the scale-generalization of existing approaches. In Fig. 4, the above-mentioned viewpoints are demonstrated visually. It is also important to note that the impact of distribution differences should not be ignored. For example, images obtained from divergent regions or sensors have great differences in illumination, color tone, texture, and appearance, as shown in Fig. 5.

(3) **Scene complexity.** The spectral characteristics of some water bodies are similar to those of shadow on the optical RS image, which is particularly easy to cause false classification of water bodies, as shown in Fig. 6. Therefore, the elimination of mountain, building, and cloud shadow is a necessary aspect to improve the accuracy of water body information classification. To solve this problem, scholars have conducted extensive experiments via using thinning segmentation (Gao et al., 2016), multiband spectral relationship (Xu, 2006), and object-oriented methods (Xu et al., 2010), the interference of shadow is eliminated to a certain extent. Sun glint (i.e., specular reflection of light from water surfaces) is another important factor impacting the accuracy of water body mapping. The reason is that in the sun glint area of a satellite image, smooth ocean water becomes a silvery mirror, while rougher surface waters appear dark. More importantly, the deteriorating environment and increasingly serious pollution have gradually aroused the concern of human beings, which also has a negative impact on water body extraction. Water body is mainly contaminated by water eutrophication, toxic organic matter and heavy metals. Polluted waters usually represent black and smelly characteristics, as shown in Fig. 7, in which water eutrophication seriously affects the edge extraction. The principal reason is that a series of water eutrophication products such as cyanobacteria and the red tide will show vague information like vegetation on RS images, which makes it laborious to classify the water or land boundary. Similarly, macrophytes often cover parts of water bodies, making accurate classification more difficult.

(4) **Complex and blurry boundaries.** Boundary optimization, as a hot research issue, has been widely exploited in both natural image vision (Lafferty et al., 2001; Borse et al., 2021; Zhu et al., 2021; Chen et al., 2019b) and RS image interpretation (Bokhovkin and Burnaev, 2019; Sun et al., 2020; Nong et al., 2021; Zhang et al., 2020e). However, the boundaries of water bodies are too intricate and various to be maintained well during the process of downsample and upsample. Intuitively, in the comparative display of Fig. 8, the boundaries of buildings are usually regular and straight, the reason is that buildings in a certain area usually have similar styles. However, this regional similarity is not tenable in the fine extraction of water body boundary.

(5) **Deficiency of large-size image datasets.** Up to now, we count that a total of 10 publicly available benchmarks or datasets can be applied for the evaluation and supervision learning of water body classification from high-resolution optical RS images, which will be introduced in detail later. It should be noted that the current dedicated datasets are only proposed by the 2020 Gaofen Challenge (Sun et al., 2021), and the large-size fine classification datasets are also scarce except from the Gaofen image dataset (GID) (Tong et al., 2018) to the best of our knowledge.

2.2. Opportunities of water body classification from high-resolution optical remote sensing imagery

Opportunities are often accompanied by challenges. Learning transferable DL models for land-use pixel-level, object-level, and scene-level classification is relatively mature (Cheng et al., 2020). Propelled by the powerful computer capabilities of deep networks, massive novel algorithms have been proposed and conducted to various practical

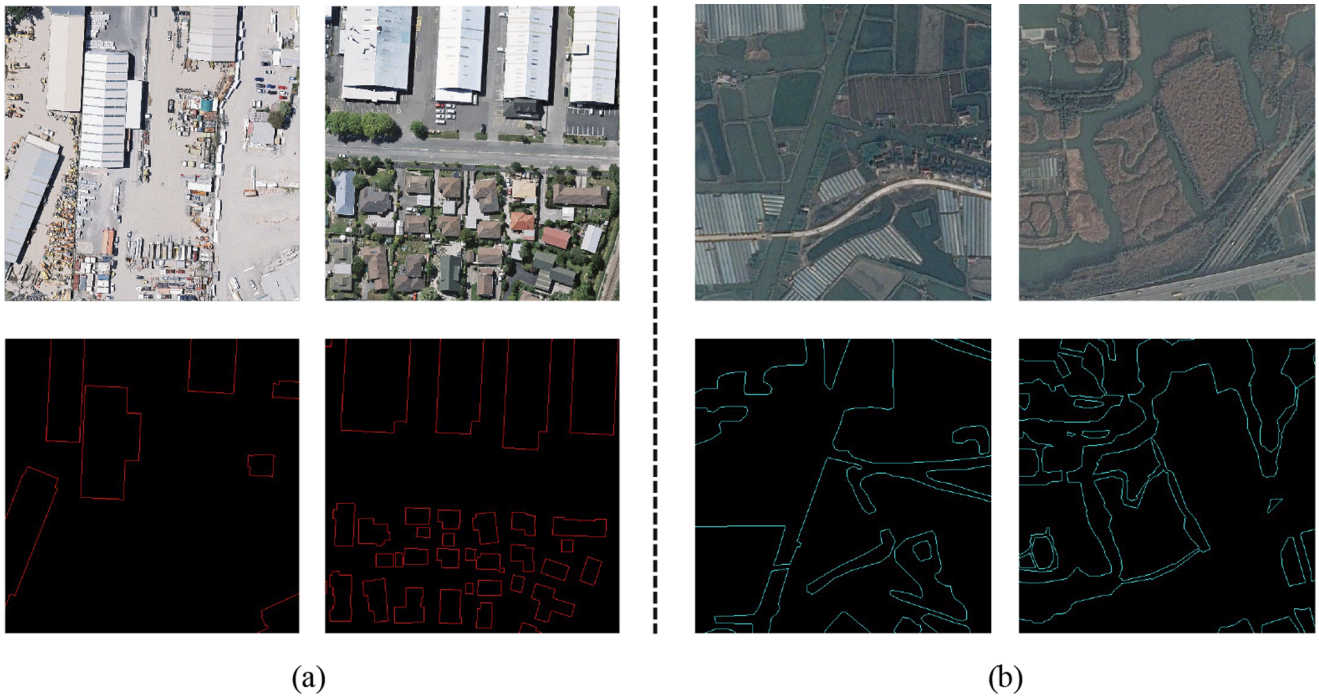

Fig. 8. Comparison of (a) buildings' boundary and (b) water bodies' boundary in high-resolution optical RS images. The first column is the raw images, and the second column is the boundary masks of the corresponding buildings and water bodies. (The images of buildings are from the WHU Aerial imagery dataset (Ji et al., 2018), and the boundary masks were made by. skimage tool.).

applications, such as Image Style Transfer (Liu et al., 2019; Zhao, 2020), Adversarial Attack, and Defense (Yuan et al., 2019; Zhou et al., 2020; Qiu et al., 2019), Salient Detection (Borji et al., 2019; Wang et al., 2021b), and so on. Furthermore, In the other tasks of RS, many methods driven by DL have drawn remarkable attention and made great achievements, which are worthy of reference. For example, positive fruit from cloud shadow removal (Zhang et al., 2020d; Bermudez et al., 2018), cross-domain segmentation (Li et al., 2021i; Yan et al., 2021), and weakly-supervised land-cover mapping (Schmitt et al., 2020; Wang et al., 2020c) is likely to be beneficial for water body classification. In this section, we systematically survey approximately 60 promising algorithms for sorting out potential opportunities clearly. To help potential readers better comprehend, we discussed the opportunities point by point in terms of the abovementioned challenges.

(1) **To cope with the first challenge: Limited spectral information and small scene coverage.** As we all know, high-spatial resolution and high-spectral resolution seem to be an inherent contradiction. Therefore, under the background of scanty spectral information, some manually designed or calculated features are embedded into original spectral channels as auxiliary information. Liu et al. (2017) applied handcrafted features like NDVI and nDSM to logistic regression, which is readable and active for segmentation. Similar works (Xu et al., 2018; Pan et al., 2018; Sherrah, 2016; Du et al., 2019) were presented to demonstrate the potential of the idea of this auxiliary channel. Another way to do this is to construct an extra branch to learn relative knowledge. For example, Ma et al. (2020) delved into a structure-preserving framework to super resolution, which integrates the gradient branch and super-resolution branch to fuse multi-level representations. It is conducive to better capturing geometric relationships via supervising the image-space and gradient-space. The enhancement of image resolution has brought great benefits to daily life and scientific research. At present, some methods are applicable to resolution even for ultra-high resolution images larger than 4 K. For example, Cheng et al. (2020) proposed a plug-and-play segmentation framework named CascadePSP that refines the large-size images and shows the high-quality refinement capability, even without extra finetuning. Similarly in the field of RS, land cover classification from very-high resolution (VHR) has raised increasing interest recently. Ding et al. (2020) provided a two-stage training strategy to break through the limitation of cropped image patches that only contain partial context information. Based on this core idea, a Wider-Context Network (WiCNet) (Ding et al., 2021) was devised to enhance the performance further. In particular, a novel Context Transformer was created and assessed to learn the correlations among various areas. Li et al. (2021d) introduced a novel pipeline called contextual semantics refinement network that is enabling to leverage local and context mask comprehensively and refine mask contours to obtain final refined high-resolution labels. The experimental conclusion declares that the segmentation accuracy can be effectively improved by combining the small, medium, and large context sizes. Motivated by the synergy of global and local branches, Collaborative Global–Local Network (GLNet) was designed by Chen et al. (2019a), it mainly focuses on fusing multi-branches feature maps and exploiting the contextual interdependence from inputs. Three large-size datasets from RS and medical image challenges are applied to verify the effectiveness and universality of their approaches. Furthermore, they also demonstrate that the global scene information is essential for semantic segmentation.

(2) **To cope with the second challenge: Variability of shape, size, and distribution.** How to accurately and automatically detect small water bodies has always been a research hotspot in environmental monitoring and protection. Coincidentally, small objects segmentation is also significant. In (Takikawa et al., 2019), a two-stream architecture named Gated-SCNN was created to pay more attention to shape information as a separate stream, which was proven the effectiveness of inferencing thinner and smaller objects accurately. Multi-scale feature fusion strategy is the most common to deal with the recognition of changeable objects.

Pyramid Pooling Module (PPM) learned from PSPNet (Zhao et al., 2017), as a classical multi-scale fusion block, was used or modified in visual recognition or intelligent interpretation. Furthermore, He et al. (2019) came up with Adaptive Pyramid Context Network (APCNet) that especially considers global-guided local affinity to generate powerful multi-scale and global contexts. Propelled by vision transformer (Dosovitskiy et al., 2020), CrossViT (Chen et al., 2021a) was formulated to learn multi-scale feature representations via a dual-branch transformer block and cross-attention fusion strategy for vision applications, which is superior to other methods. Expanding the receptive field is another available strategy that has great promise. Based on the ASPP module, Multi Receptive Field Module (MRFM) (Yuan et al., 2020a) was proposed to replace the traditional backbone to capture the multi-receptive field. To model the long-range dependencies better and explore rich contextual representative, the strip loss (Hou et al., 2020) was released. It should be mentioned that making better use of object-context information is also of significance for multi-scale ground objects like water body. OCRNet (Yuan et al., 2020b) showed the powerful ability to mine object-context representation. The above schemes are worthy of reference for extracting water bodies with various shapes and sizes, however, distribution discrepancy is expected to be eliminated by GAN, if it is roughly equivalent to the image style difference. In (Zhu et al., 2017), CycleGAN was proposed for image-to-image translation, and obtained promising performance. Although pair patches are not available, it is also possible to generate transferred samples. Because there are no sufficient constraints such as labels, the phenomenon of confused classification is frequent among transferred images, which is immensely harmful for subsequent semantic segmentation. To address this issue, a variety of consistency training (Melas-Kyriazi and Manrai, 2021;Kim and Byun, 2020; Li et al., 2021i) were introduced to constrain the training of GAN.

(3) **To cope with the third challenge: Scene complexity.** For resolving shadow occlusion, high-quality shadow detection algorithms and cloud removal algorithms provide another potential solution. In the past several years, these algorithms are also gradually mature enough to be leveraged in practical applications. Zheng et al. (2019) proposed a discriminative framework to learn distraction-aware features that can boost the performance of shadow detection. A weakly-supervised training scheme was designed by Le and Samaras (2021), which was beneficial for shadow removal when the shadow-free images are enabled to be available. Facing the various and unknown shadow patterns, Fu et al. (2021) delved into a robust overexposure fusion approach that is superior in recovering the background content. In the field of RS, Chen et al. (2019c) and Li et al. (2019a) conducted their experiments in different ways, and both gained a pretty good performance, respectively. Adversarial Attack and Defense, as an exploratory study direction, is considered to be of great importance for information safety. However, another point of thought, can the disturbance of adversarial attack simulate the color tone disturbance caused by water pollution? Currently, not only many related approaches have emerged in the CV (Cho et al., 2020; Xie et al., 2017), but also they have been transferred to RS image interpretation (Xu et al., 2020).

(4) **To cope with the fourth challenge: Complex and blurry boundaries.** How to accurately extract or refine boundaries of objects has attracted the interest of scholars from multi-domain. For example, in the assignment of salient object detection, BASNet (Qin et al., 2019) and Contour loss (Chen et al., 2019d) was proposed to make an attempt to refine boundary through designing a predict refinement module and formulating object contours-guided perception loss function, respectively. In the mission of segmentation, the shape stream from Gated-SCNN (Takikawa et al., 2019) only take the advantage of relevant boundary information, which is beneficial for retaining the fine boundaries. Chen et al. (2019b) analyzed the reason for the loss of fine boundaries, which is mainly that manual labels via annotation tools cannot meet the requirements. Then, BANet was formulated to recover the fine boundaries like thin hairs, including semantic branch, boundary feature mining branch, and fusion unit. In the field of medical image processing, Kervadec et al. (2019) presented a well-designed boundary loss that calculates the distance metric on the space of shape to alleviate the problem of high-unbalanced scenarios and enhance the accuracy of segmentation. Besides, driven by boundary metric, a novel boundary loss was introduced by Bokhovkin and Burnaev (2019), which applies the pixel-wise max pooling operation to define the boundary, and calculate the precision and recall combined with ground truth. Extensive experiments demonstrate that this boundary loss can be successfully transplanted to RS binary classification. Furthermore, boundary refinement also plays a key role in building extraction from RS imagery, therefore, many works (Jin et al., 2021; Jung et al., 2021; Zhao et al., 2018) have emerged to tackle this issue. The abovementioned methods afford water body classification lessons that merit attention, however, it should be noted that relevant boundary refinement algorithms should be devised around the uniqueness of water bodies.

(5) **To cope with the fifth challenge: Deficiency of large-size image datasets.** It is undeniable that constructing large datasets is necessary for training sufficiently robust deep network models. However, gathering and annotating the class labels are expensive and time-consuming, which is difficult to build a large-size scene dataset in the short term. Fortunately, some semi-, self- and unsupervised approaches use fewer labeled data and incorporate unlabeled data into training, which is possible to reach a similar even equal performance (Schmarje et al., 2020). For instance, the student–teacher method (Tarvainen and Valpola, 2017) was proposed to replace label predictions with average model weights for semi-supervised learning. Pseudo labels (Lee et al., 2013) and MixUp (Zhang et al., 2018a) are classical approaches to boost the performance under the condition of a small amount of labeled data. Self-supervised learning (SSL), as one of the most prevailed research directions, utilize a pretext task to learn representations for downstream tasks. Komodakis and Gidaris (2018) created a pretext task that recognizes the rotation angle to learn high-level image features, and its state-of-the-art (SOTA) performance demonstrates that it can stand comparison quite effectively by supervised learning. Then, SimSiam architecture (Chen and He, 2021) was reported to emphasize the vitalness of Siamese networks in representative learning. Additionally, some weakly-supervised land cover mapping, object detection, and scene classification methods (Li et al., 2020; Wang et al., 2020c; Cheng et al., 2013; Cheng and Han, 2016; Robinson et al., 2020; Yao et al., 2016; Li et al., 2018; Li et al., 2021j) and benchmarks (Schmitt et al., 2019; Yokoya et al., 2020) were conducted to deal with relevant circumstances.

3. Methods for water body classification from high-resolution optical remote sensing imagery

In the past decades, water body classification from high-resolution optical RS images has attracted the attention of many researchers due to its wide and essential application in human being society. In the early stage of development, a number of approaches based on handcrafted features, especially water body indexes, have been proposed. Yet, considering that the limitation of spectral information of high-spatial optical imagery, shallow classifier tools that pay more attention to spatial texture information are also applied in this mission. Propelled by DL and thanks to the availability of massive labeled data and powerful

computational resources, DL-based methods have emerged and occupied the mainstream in this area due to their puissant high-level semantic features exploiting capability and the SOTA accuracy. In this section, we survey and analyze about 45 existing methods to explore their core contributions. Especially, to understand better, we divide them into two categories and seven subcategories and introduce them separately at length.

3.1. Non-deep learning-based methods

(1) Threshold-based methods: The definition of threshold-based method is to select the appropriate spectral band to build the model according to the spectral characteristic curve of water body, and carry out the classification method of water body extraction according to certain threshold determination rules. In 1996, NDWI (McFeeters, 1996), as a pioneer of water body index, was proposed for the first time and achieved great success in assessing water resources. The core principle of that is using the low reflectance of NIR by water, and combined Green and NIR bands. And then Yao et al. (2015) developed the High Resolution Water Index (HRWI) that only uses the visible and NIR channels. Its performance in ZY-3 images demonstrates that it can achieve a high accuracy under different conditions. Similarly, Wu et al. (2018) released the Two-Step Urban Water Index (TSUWI), which is easy to popularize and apply to most common high-resolution optical imagery that has only visible and NIR bands. Due to the circumscription of spectral channels, some water body indexes cannot be applied to most high-resolution RS images containing RGB and NIR channels. However, they are suitable for specific high-resolution RS sensors, which still has good application value. The NDWI was modified to form the MNDWI (Xu, 2006) which was modified to boost the robustness under the building noise disturbance. In particular, it should be emphasized that the middle-infrared red (MIR) band that is needed for calculating the MNDWI is not common in images from high-resolution optical sensors, but some research surveyed that it is successful to replace that with Band-7 of WorldView-2. Because WorldView-2 includes 9 bands, which makes the threshold-based methods that are extremely dependent on spectral information work, several exclusive water body indexes appeared. For example, Jawak and Luis (2015) introduced some fresh modified NDWI and evaluated them with other target extraction methods to prove their effectiveness. Incorporated a complex NDWI and morphological shadow index (MSI), NDWI-MSI (Xie et al., 2016) was proposed to further give prominence to water bodies and simultaneously suppress shadow areas. Besides, it made full use of the rich information of 8 bands of WorldView-2, and extensive experiments showed its advances. In terms of the red edge (RE) band that was provided specially by RapidEye, Klemenjak et al. (2012) came up with RE-NDWI that applied RE band rather than NIR. Draw support from the fact that the red edge band is very sensitive to the change of chlorophyll, RE-NDWI can better distinguish water from vegetation and soil.

With the continuous development of threshold-based methods, their utilization of spectral information is more reasonable. However, they depend on the spectral information and ignore the spatial information of the images, which is easy to lead to the problems of significant errors in boundary classification and terrible detection accuracy of the transition region.

(2) Shallow classifier-based methods: SVM, as one of the most classical shallow classifiers, was highly applied and compared in many works (Yao et al., 2015; Wu et al., 2018; Zhang et al., 2018b; Li et al., 2014; Sui et al., 2013). Generally, the texture information of the samples was trained by SVM, and then the decision function was obtained to distinguish the water body. In addition, Khurshid and Khan (2012) formulated the two-stage algorithms that contain computing approximate masks and refining detailed masks. Experiments in a SPOT-5 image prove it is an efficient way to enhance the accuracy via multiple iterations. In (Zeng et al., 2015), a framework based on Natural-Rule-Based-Connection (NRBC) was proposed to connect the patch pair and segment an intact river mask. Simultaneously, the river centerlines were extracted to verify river continuity. Zhang et al. (2010) devised a workflow to extract water body, including level set segmentation and fast matching algorithm. Markov model was used to refine initial masks in (Qi et al., 2019), and it is an interesting and significant work for water body classification from panchromatic imagery.

Most shallow classifier-based methods are derived from traditional machine learning and have obtained good performance under specific data conditions. However, they usually need to design classifier functions, which are sophisticated and difficult to transfer to other scenes.

(3) Hybrid classification methods: In general, threshold-based methods mainly focus on spectral knowledge, and shallow classifier-based methods leverage more spatial features such as texture. Therefore, combing them to comprehensively analyze the characteristics of water bodies is more stable at a certain cost. In 2012, Zhou et al. (2012) proposed an adaptive framework to extract urban water. It uses NDWI to roughly classify the binary class of water body and background, and construct the segmented buffer with automatic multi-scale buffer length to extend the margin of water body. Finally, in light of the spatial adjacency and similarity principle to enhance the segmentation performance. Li et al. (2014) introduced a multi-hierarchies method that combines spectral and shape features. They adopted the modified Statistical Region Merging (SRM) approaches to image segmentation, merging original channels and two indexes to boost the robustness. MRF model and Graph cut algorithm were integrated with NDWI to improve the accuracy of water body classification, and significant performance on GF-1 images proved its appropriateness under complex surroundings (Li et al., 2016). On the basis of detecting urban water bodies with NDWI, Yang et al. (2017) also used an object-oriented shadow detection method to remove shadow noise, and finally obtained the optimized mask. Zhang et al. (2018b) delved into an unsupervised approach that applies pixel region index (PRI) and NDWI to calculate the preliminary mask, and then further processes by K-means clustering and merges them.

Hybrid classification methods combine the advantages of the above two approaches. They can not only effectively utilize the rich spatial and texture information of high-resolution optical RS images, but also unite the spectral features. However, their implementation process is generally relatively complicated.

3.2. Deep learning-based methods

(1) Feature fusion-based methods: Under a certain geographic coverage, the water area in the image contains rivers, lakes, ponds, and other various types of water bodies that have different shapes and sizes, demonstrating different scales characteristics, as shown in Fig. 4. These water bodies with different scales bring certain limitations and difficulties to the training and inference of general DL models. Therefore, the multi-scale features of water bodies are not only the hardship of extracting them, but also the key point to boost the classification accuracy. Inspired by the multi-scale features fusion algorithm in the area of CV, many researchers explored various integration strategies to take the advantage of multi-scale features. Concretely, Li et al. (2019b) introduced a hierarchical expansion splitting approach to get the multi-scale scene dataset and provided multi-scale data instead of a single-scale sample as input. In the decoder architecture, multi-scale features were fused with weights. The multi-scale features supervision strategy was conducted by MSR-Net (Duan and Hu, 2019), and it can leverage multi-scale information from different layers to refine final results. Inspired by ASPP (Chen et al., 2018), Guo et al. (2020) presented a multi-scale convolutional neural network (MWEN) that expands the receptive fields via atrous convolution to capture the features with various scales. Likewise, Expanding the receptive fields was adopted by some other methods (Weng et al., 2020; Yu et al., 2021; Wang et al., 2020; Cui et al., 2020), and relevant multi-scale features from distinct dilation rates convolution were integrated flexibly. In (Chen et al., 2020), a multi-

scale learning module was created to capture multi-scale contextual information through upsampling and downsampling input data and fusing their score maps. Recently, motivated by PPM (Zhao et al., 2017) and spatial pyramid pooling (SPP) (He et al., 2015), Dang and Li (2021) devised MSResNet that can seize the scale-invariant feature and fuse the multi-scale contextual information by integrating high- and low-level feature maps. Kang et al. (2021) proposed a multi-scale context feature extractor module and used Res2Net to tackle the problem of shape and size variations of water bodies. Aside from multi-scale fusion strategy, many other features fusion methods have emerged. For example, high-resolution and low-resolution feature maps were combined to gain dense and accurate water body localization (Fu et al., 2018). Li et al. (2021c) designed an LFC module to integrate spatial and spectral features, which is an important part of the overall architecture. Zhang et al. (2021) presented a semantic feature fusion module (DSFF) to focus on global feature information, which is effective to enhance the semantic consistency between encoder and decoder. In addition, designing diverse features or channels as input data is another significant way. Li et al. (2021f) and Yuan et al. (2021) embedded artificial features and multi-channel features into the original RGB channel as novel input data, and fuse them adaptively, respectively.

As the most common DL-based methods, feature fusion-based methods mainly optimize the network structure according to the multi-scale features of water bodies. However, under the condition of certain DL model parameters, the more efficient multi-scale feature fusion modules are still developing.

(2) Boundary constraint-based methods: At present, the existing methods can be divided into three categories: boundary constraint loss functions, boundary refinement modules, and post-processing optimization. Conditional random field (CRF) (Krähenbühl and Koltun, 2011), as an effective post-processing algorithm, attracted the attention and interest of researchers and was utilized in many works. Intuitively, taking each pixel or superpixel as a node and the relationship between pixels or superpixels as an edge constitutes a conditional random field, which can make full use of the spatial neighborhood information of labeled images and observed images to effectively obtain spatial context information. Extensive experiments (Sun et al., 2021; Feng et al., 2018; Li et al., 2019b; Chu et al., 2019) showed that it is a significant tool for boundary refinement. For specific loss functions, Miao et al. (2018) proposed a novel loss named Edges Weighting Loss (EWLoss) to compute the edge weight contribution to constrain the network to catch accurate boundaries of water bodies. In addition, Cui et al. (2020) surveyed the boundary-sensitive feature maps and embedded the Squeeze-and-excitation (SE) module (Hu et al., 2018) to improve the boundaries' representation. Chen et al. (2020) presented spatial-spectral convolution (SS-Conv) to construct a boundary refinement unit like a residual block, which can generate refined boundary score maps.

Although the experimental results of existing boundary constraint-based methods show that they can improve the accuracy of boundary detection, most of them are universal boundary post-processing optimization methods and boundary loss functions. Enhancement algorithms in the light of the meandering boundaries of water bodies are scarce.

(3) Large scene context guided methods: Some researchers found that the importance of capturing context information from large scene RS images to maintain the continuity of water bodies. Chen et al. (2018b) applied a super-pixel segmentation algorithm to crop the large-size images into small patches rather than cropping them in sequence. The advantage is that it can retain the whole rivers and lakes to the maximum, even without consuming huge computing resources. Dong et al. (2019) released a sub-neighbor system constraint algorithm to generate a more coherent water body mask. Experimental results on a large scene dataset proved that it is superior among relevant methods.

Compared with feature fusion-based methods and boundary constraint-based methods, only a small number of literature about large scene context guided methods have been reported so far. Nevertheless, it is a promising research direction to accurately identify water bodies from large scene RS images.

(4) Other methods: Domain gap is a common problem in land cover classification, and it also exists in water body extraction. Yang et al. (2021) discovered this issue and proposed an unsupervised content-adaptive water-body extraction framework (UCWater) and content-consistent matching sample selecting strategy to alleviate the domain shift. Its progressiveness is that it not only boosts the classification accuracy of cross-domain samples, but also reduces the consumption of manual annotation. In (Abid et al., 2021), unsupervised Curriculum Learning (UCL) was introduced to classify water body without relying on any labeled data. Simply put, a pre-trained CNN was applied to extract features, and then clustered and selected them into water or non-water, fine-tuned the CNN with selected samples, repeated the above steps until the model converges.

4. Practical applications of water body classification from high-resolution optical remote sensing

Against the backdrop of the rapid advance of various types of sensors, high-resolution optical imaging sensors are undoubtedly unique among them because they can provide 'clear' imagery with a high spatial resolution that not only retains the most complete details but also improves the visual impression of images, which is significant for both visual and intelligent interpretation. Throughout the development of satellite and UAV techniques in the past few decades, it is not difficult to find that sensors with higher spatial resolution have always been the hot topic in the field of the RS community. Furthermore, water is a precious resource for human survival. With the surge of the global population and the rapid economical development of the human community, the demand for water resources is increasing. Meanwhile, unreasonable use and uncontrollable water pollution lead to a huge waste of water resources, further leading to a perplexing ecological deterioration. Consequently, water body classification from high-resolution optical RS imagery will have broader applications. In this section, we discuss some prevailed applications and consider the advantages and shortages of the existing methods in these practical applications.

4.1. Flood monitoring and mapping

Flood event is one of the most frequently occurring natural disasters in the world, especially under the background of drastic global climate change, it has gradually become a hot research field (Ward et al., 2014; Lamovec et al., 2013; Byun et al., 2015). Flood monitoring mainly focuses on obtaining basic information about flood ranges and land cover changes quickly. In (Huang et al., 2008), the destructiveness of flood disasters is discussed. Compared with any other natural disaster, flood hazard is more destructive, which can cause huge casualties and economic losses. In addition, because floods destroy key substances such as power, water supply, and transportation, rapid and even real-time monitoring of flood extent can help decision-makers put forward effective disaster management strategies and rescue plans.

In fact, due to the advantages of numerous available spectral bands and short revisit periods, multi-temporal images obtained from sensors with low spatial resolution but high spectral resolution have been widely applied in flood monitoring (Byun et al., 2015; Ticehurst et al., 2014). In addition, considering SAR has the unique capability of capturing images regardless of the weather condition, their images have also been commonly applied in flood monitoring(Grimaldi et al., 2020).

It must be acknowledged that due to the limitations of high-resolution optical satellites, their images are not the first and optimum choice in the general mission of flood monitoring. Nevertheless, related research has made some achievements. Byun et al. (2015) proposed a novel method based on VHR image fusion, and extensive experiments were conducted to assess the effectiveness. Malinowski et al. (2015) employed the multispectral images that were captured from WorldView-

2 to tackle the problem of detecting and mapping inundation on complex areas. Thanks to the advances of UAV technology, some inundation monitoring methods based on aerial images from UAVs (Munawar et al., 2021; Koutalakis et al., 2020; Munawar et al., 2021) have emerged and drawn remarkable attention, and UAVs are considered to be the next generation of favorable monitoring tools due to their cost-effectiveness, ability to fly at lower altitudes, and ability to enter a dangerous area (Gebrehiwot et al., 2019).

Different from rapid or real-time inundation monitoring, flood mapping is mainly employed for disaster assessment after flood hazards and helps for post-disaster reconstruction work, which emphasizes obtaining high-precision extraction maps. Although low-resolution optical data and SAR data can also be applied for mapping (Thomas et al., 2011; Dhara et al., 2020; Benoudjit and Guida, 2019), high-resolution optical RS images are more broadly employed for local flood mapping because of their comprehensibility and high quality. In (Hashemi-Beni and Gebrehiwot, 2021), a novel method based on FCN and region growth (RG) algorithm was presented to map the inundation extent from UAV data. The preliminary results were predicted by FCN, and then, DEM and water level information were employed for RG algorithm to identify the floods underneath vegetation canopy. Gebrehiwot et al. (2019) surveyed the effectiveness of CNN approaches to flood mapping accurately. Driven by discriminative texture features, Feng et al. (2015) applied a shallow random forest classifier to make a satisfactory urban inundation map and provide a case study.

4.2. Water body mapping in specific scenarios

Water body mapping in some specific scenarios is another momentous application, which needs to fully consider the peculiarity of landscape elements and the characteristics of water bodies contained therein.

As an important part of the urban ecosystem, the urban water body is of great significance in urban environmental quality monitoring and the urban heat island effect. Moreover, the change of urban water bodies also has a great impact on human life, such as floods and waterlogging in the city (Fletcher et al., 2013). More importantly, high-resolution optical RS imagery can better demonstrate small water bodies such as ponds and narrow rivers, which is significant for urban water resources investigation. Therefore, it is necessary to understand the distribution and change of urban water bodies via high-resolution optical RS imagery timely and accurately. Considering that the shadow influence of high-rise buildings or plants is more remarkable, some studies proposed a modified water body index to reduce the adverse influence of shadow as much as possible (Xie et al., 2016; Wu et al., 2018; Yang et al., 2017). For example, Wu et al. (2018) combined an Urban Water Index (UWI) and an Urban Shadow Index (USI) to improve the performance of urban water body mapping. Depending on the advanced DL model, Chen et al. (2018b) took into account the complexity of China's urban surface water network and employed ZY-3 and GF-2 images from various cities including divergent urban water bodies as experimental data to verify the effectiveness of their proposed DL network architecture. The experimental results of threshold-based and traditional machine learning methods demonstrated that in the face of relatively complex urban surface water bodies, DL-based methods have more advantages than other methods. Similarly, based on the WorldView-3 images and GF-2 images of Beijing, Song et al. (2020) verified the identification ability of the improved Mask-RCNN for urban water bodies. Compared with other methods, it can significantly reduce the confusion with the shadow of buildings and vegetation.

The conventional field surveys to map glacial lakes are time-consuming, laborious, costly, and even dangerous, especially in alpine countries (Mitkari et al., 2017). However, RS data with wide sources and easy access become the most practical source for mapping glacial lakes. Using high-resolution RS images to map glacial lakes is considerably valuable for studying the impact of climate change and the mitigation and risk assessment of a Glacial Lake Outburst Flood (GLOF) (Shugar et al., 2020). In 2014, Jawak and Luis (2014) designed four customized NDWI to map the glacier lakes of Larsemann Hills in the Antarctica region using WorldView-2 pan-sharpened images. Unfortunately, their proposed method was not superior to the target extraction method, but the experimental results proved that it was feasible to use high-resolution optical data to extract glacial lakes. In (Mitkari et al., 2017), a novel object-based image analysis (OBIA) approach was applied to the case study of mapping the small supraglacial lakes (SGLs) of Gangotri glacier (Uttarakhand Himalayas) from the high-spatial-resolution data of the LISS-IV sensor (spatial resolution: 5 m). Compared with the object-based NDWI and other methods, it is observed that the accuracy of mapping is significantly enhanced. Qayyum et al. (2020) focused on utilizing PlanetScope imagery with 4 bands to explore the application of DL technology in glacier lake mapping and change monitoring for accuracy evaluation. The superior experiment performance confirmed the great potential of high-resolution optical images captured by CubeSats in this field.

Thermokarst lakes, topographic depressions formed by the melting of ice-rich permafrost or a large amount of ground ice, are easy to cause a series of thorny environmental problems such as groundwater level decline, vegetation degradation, and land desertification due to erosion and expansion (Kokelj and Jorgenson, 2013; Tian et al., 2017). RS data is an effective tool to map the distribution of thermokarst lakes. Massive optical satellites have been widely employed in the change monitoring task of thermokarst lakes in the Qinghai Tibet Plateau (QTP) (Zhang et al., 2015; Mergili et al., 2013; Song et al., 2013). However, because of the limitation of coarse spatial resolution, they cannot identify massive small thermal-karst lakes. Therefore, the utilization of high-resolution optical RS images is a potentially effective way to obtain their accurate distribution. Tian et al. (2017) delved into the thermokarst lake shorelines extraction approach based on Nonlocal Active Contours (NLAC) model and NDWI. The shoreline vector map corroborated that their method was robust and superior in mapping the shorelines from the GF-2 images, and was feasible for detecting thermokarst lakes of the QTP. Huang et al. (2018) exploited high-resolution UAV images to verify an automatic extraction method of thermokarst lakes based on DL techniques. The experimental results confirmed that it can be extended to other thermokarst lakes and larger areas in the future by collecting corresponding training data.

5. Experimental results and discussion

5.1. Survey on publicly open benchmarks

Benchmark datasets, as powerful partners in the implementation and development of DL algorithms, are essential in the era of RS big data. With the increase of RS sensors and the reduction of the cost of manual annotation, a large number of land cover classification benchmarks appear, which is of great benefit to enhance the accuracy of water body classification from high-resolution optical RS imagery. In the past several years, many researchers and organizations have released numerous benchmarks in which images are captured from various sensors for training or evaluation. For example, DeepWaterMap (Isikdogan et al., 2019), Deltares Aqua Monitor (Donchyts et al., 2016), Global surface water (Pekel et al., 2016), Earth surface water knowledge base (ESWKB) (Luo et al., 2021), 2020 Gaofen challenge water body segmentation dataset (Sun et al., 2021), DeepGlobe Land Cover Classification Challenge benchmark (Demir et al., 2018), GID (Tong et al., 2018), and so on were proposed to facilitate this area forward. In particular, this paper only discusses the publicly open benchmarks that consist of high-resolution optical RS images, including Zurich Summer (Volpi and Ferrari, 2015), WHDLD (Shao et al., 2018; Shao et al., 2020), DLRSD (Chaudhuri et al., 2017), DeepGlobe (Demir et al., 2018), DroneDeploy (Nicholas et al., xxxx), EORSSD (Zhang et al., 2020c), Landcover_ai (Boguszewski et al., 2021), LoveDA (Wang et al., 2021a), GID

Table 1
10 publicly available datasets for water body classification from high-resolution optical remote sensing imagery.

Datasets	Total image number	Image bands number	Image size (pixels)	Spatial resolution	Data sources	Year
Zurich Summer	20	4	600~1600 × 600~1600	0.62 m	QuickBird	2015
WHDLD	4940	3	256 × 256	2 m	GF-1 & ZY-3	2018
DLRSD	2100	3	256 × 256	0.3 m	UC Merced archive	2018
DeepGlobe	803	3	2448 × 2448	0.5 m	DigitalGlobe	2018
GID	10/150	3/4	7200 × 6800	4 m	GF-2	2018
DroneDeploy	55	3	6000 × 6000	0.1 m	Aerial images	2019
EORSSD	2000	3	500 × 500	-	Google Earth	2020
Landcover_ai	41	3	9000 × 9500/ 4200 × 4700	0.25 m/0.5 m	Public Geodetic Resource	2020
2020 Gaofen Challenge dataset	1000/2500	3	492~2000 × 492~2000	1 to 4 m	GF-2	2020
LoveDA	5987	3	1024 × 1024	0.3 m	Google Earth	2021

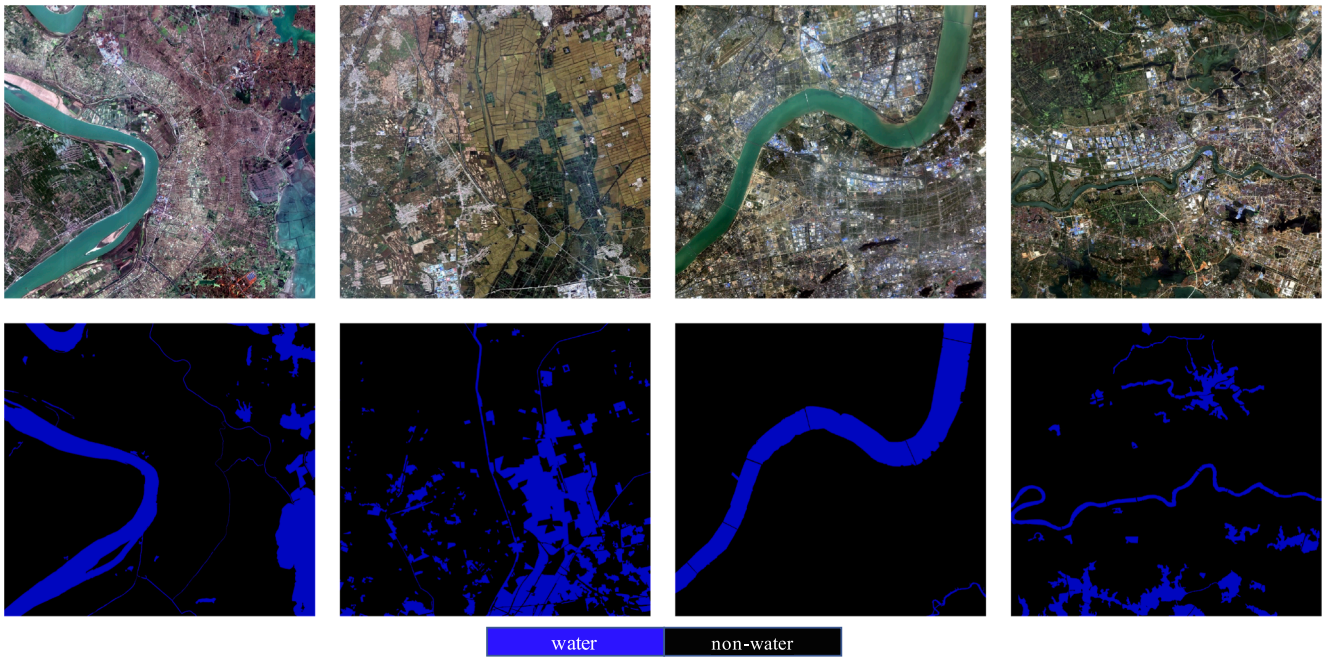

Fig. 9. Raw images and ground truths from the GID dataset.

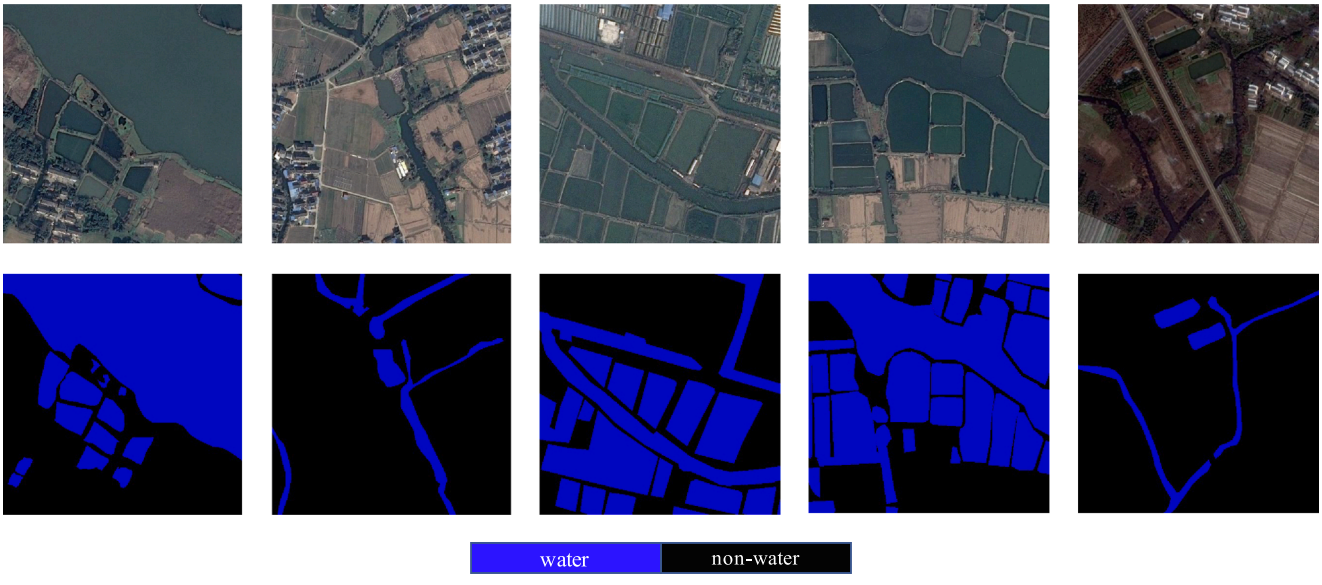

Fig. 10. Raw images and ground truths from the 2020 Gaofen challenge water body segmentation dataset.

Table 2
Quantitative accuracy comparison of 10 water body classification methods on the GID benchmark. (repro means that the experimental results are reproduced by us according to the description in reference.)

Methods	Year	Publication	Backbone	OA(%)	MIoU(%)	FWIoU(%)
NDWI (McFeeters, 1996)	1996	IJRS	-	91.19	83.34	83.82
Feng et al. (2018) w/o Regional Restriction (repro.)	2018	IEEE GRSL	-	96.51	93.01	93.23
Chu et al. (2019) (repro.)	2019	IGARSS 2019	ResNet-34	96.36	92.78	92.99
Li et al. (2019b) (repro.)	2019	IEEE Access	ResNet-101	96.42	92.89	93.10
MWEN (Guo et al., 2020) (repro.)	2020	MDPI ISPRS	-	96.75	93.53	93.72
SR-SegNet (Weng et al., 2020) (repro.)	2020	MDPI ISPRS	ResNet-50	96.32	92.72	92.93
MSCENet (Kang et al., 2021) (repro.)	2021	JAG	Res2Net-50	96.67	93.38	93.58
MECNet (Zhang et al., 2021)	2021	Remote Sensing	-	96.27	92.62	92.83
Dang and Li (2021)	2021	Remote Sensing	ResNet-34	97.27	94.53	94.70
The Wu Da Ti Shui Gao Fen Team (Sun et al., 2021) (repro.)	2021	IEEE JSTARS	ResNet-34	96.73	93.49	93.68

Table 3
Quantitative accuracy comparison of 9 water body classification methods on the 2020 Gaofen challenge water body segmentation benchmark. (repro means that the experimental results are reproduced by us according to the description in reference.)

Methods	Year	Publication	Backbone	OA(%)	MIoU(%)	FWIoU(%)
Feng et al. (2018) w/o Regional Restriction (repro.)	2018	IEEE GRSL	-	94.95	85.53	90.65
Chu et al. (2019) (repro.)	2019	IGARSS 2019	ResNet-34	94.46	84.19	89.79
Li et al. (2019b) (repro.)	2019	IEEE Access	ResNet-101	93.48	81.85	88.17
MWEN (Guo et al., 2020) (repro.)	2020	MDPI ISPRS	-	95.18	85.95	90.99
SR-SegNet (Weng et al., 2020) (repro.)	2020	MDPI ISPRS	ResNet-50	94.46	84.17	89.78
MSCENet (Kang et al., 2021) (repro.)	2021	JAG	Res2Net-50	94.87	85.16	90.46
MECNet (Zhang et al., 2021)	2021	Remote Sensing	-	94.66	84.96	90.21
Dang and Li (2021)	2021	Remote Sensing	ResNet-34	95.17	85.81	90.94
The Wu Da Ti Shui Gao Fen Team (Sun et al., 2021) (repro.)	2021	IEEE JSTARS	ResNet-34	94.92	84.92	90.43

(Tong et al., 2018), and 2020 Gaofen challenge water-body segmentation dataset (Sun et al., 2021). The relevant details are listed in Table 1. Among them, the GID (Tong et al., 2018) and the 2020 Gaofen challenge water body segmentation dataset (Sun et al., 2021) are two representative benchmarks that were introduced in detail and used for evaluating the performance of various methods later.

5.1.1. Gaofen Image Dataset (GID)

The GID was proposed in 2018 by a group from Wuhan University and contains 5 or 15 classes originally. In total, this dataset includes 10 fine land cover images and annotations, and 150 large-scale images and annotations. They are all captured from the GF-2 satellite and the pixel resolution is 4 m. To the best of our knowledge, the dataset is broadly employed for land cover classification in the area of RS due to the characteristics of images and satisfactory manual labels. To better meet the requirements of water body extraction, we redefined the classification standard, that is, rivers, lakes, and ponds are divided into water, while other classes are classified as non-water. Some samples are displayed in Fig. 9. In our later experiments, we select representative samples and crop them to small samples with the size of 256 × 256, of which the total images number is 19500. 60% are used for training, 20% and 20% are used for validation and testing, respectively. In the case where the sample size is not particularly large, in order to reduce the information leakage and obtain a more accurate response model. Such a proportion of divided datasets seems to be more reasonable.

5.1.2. 2020 Gaofen challenge water body segmentation dataset

The 2020 Gaofen challenge water body segmentation dataset was released by the 2020 Gaofen Challenge committee, which is the current only specific high-resolution optical benchmark for water body classification. The benchmark contains 2500 RGB images from the GF-2 satellite, of which the pixel resolution is ranging from 1 to 4 m. Noted that only 1000 samples with the size of 492 × 492 can be publicly available, of which the pixel resolution is 1 m. Fig. 10. lists the samples from the dataset. When we conduct experiments, the train and test ratio are the same as GID.

5.2. Performance comparison and discussion

5.2.1. Evaluation metrics

Three commonly used evaluation metrics: the overall accuracy (OA), the mean intersection over union (MIoU), and the frequency weighted intersection over union (FWIoU) are applied to assess the 10 existing milestone works on the GID dataset and the 9 existing milestone works on the 2020 Gaofen challenge water body segmentation dataset. The confusion matrix between the water body masks and ground truths is calculated, consisting of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Then, the abovementioned criteria can be denoted as Eqs. (1)–(3):

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$MIoU = \frac{1}{n+1}\sum_{i=0}^{n}\frac{TP_i}{FN_i + FP_i + TP_i} \tag{2}$$

$$FWIoU = \frac{1}{n+1}\sum_{i=0}^{n}\left(\frac{TP_i}{TP_i + TN_i + FN_i}\cdot\frac{TP_i + FN_i}{TP_i + FP_i + TN_i + FN_i}\right) \tag{3}$$

5.2.2. Performance comparison and experiment setting

In the past few years, numerous DL-based approaches have been proposed. In this paper, we select 9 representative DL-based methods and compare their performance on two open benchmarks. Among them, 2 of them released their source code publicly, and 7 of them were reimplemented by us. In addition, we also select the most representative non-DL-based method (i.e., NDWI) to compare with other approaches on the GID benchmark. Tables 2 and 3, and Fig. 11 and Fig. 12 demonstrate the water body classification accuracy comparison of 10 methods on the GID dataset and 9 DL-based methods on the 2020 Gaofen challenge water body segmentation dataset, respectively.

For a fair comparison, all DL experiments were implemented using Pytorch platform (Paszke et al., 2019) on a single NVIDIA GeForce RTX 3090 GPU, and the networks were optimized with the Adam optimizer (Kingma and Ba, 2014). The Adam optimizer uses the exponential decay

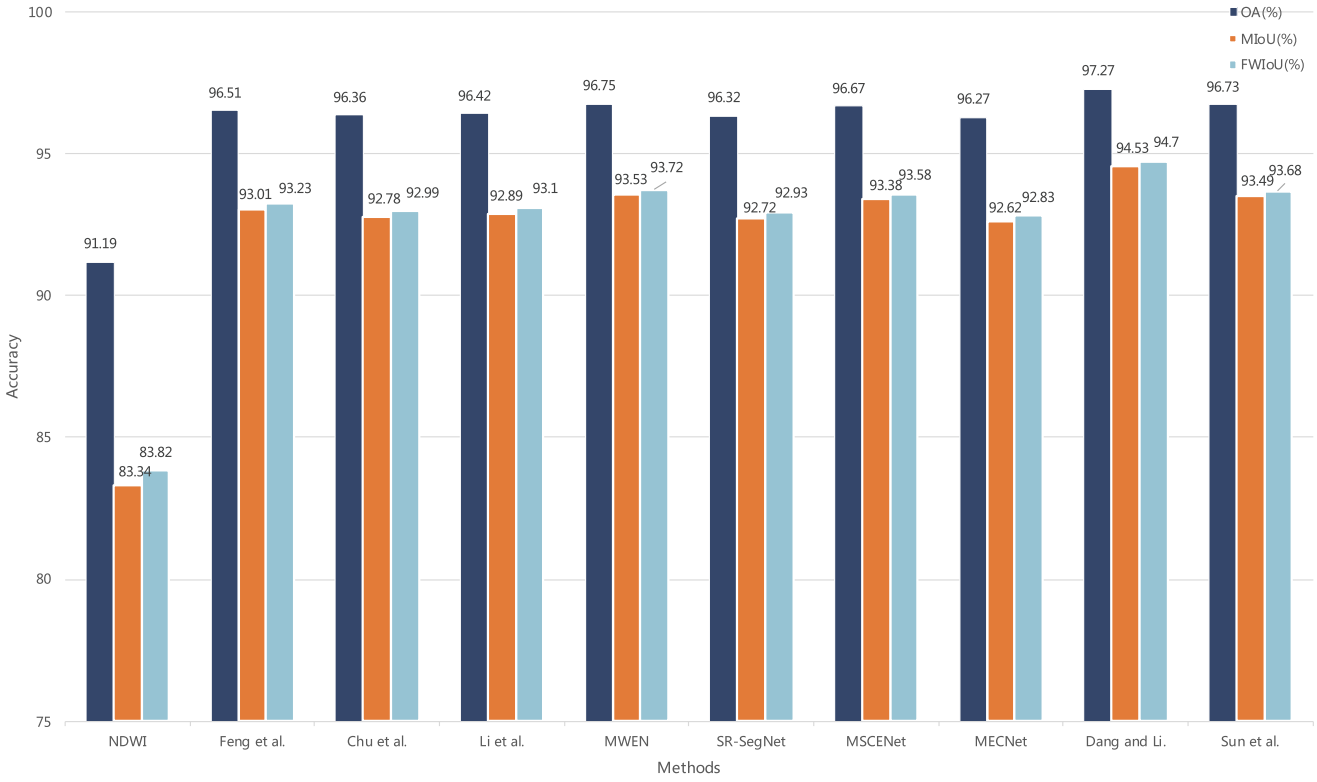

Fig. 11. The histogram of quantitative accuracy comparison of 10 water body classification methods on the GID benchmark.

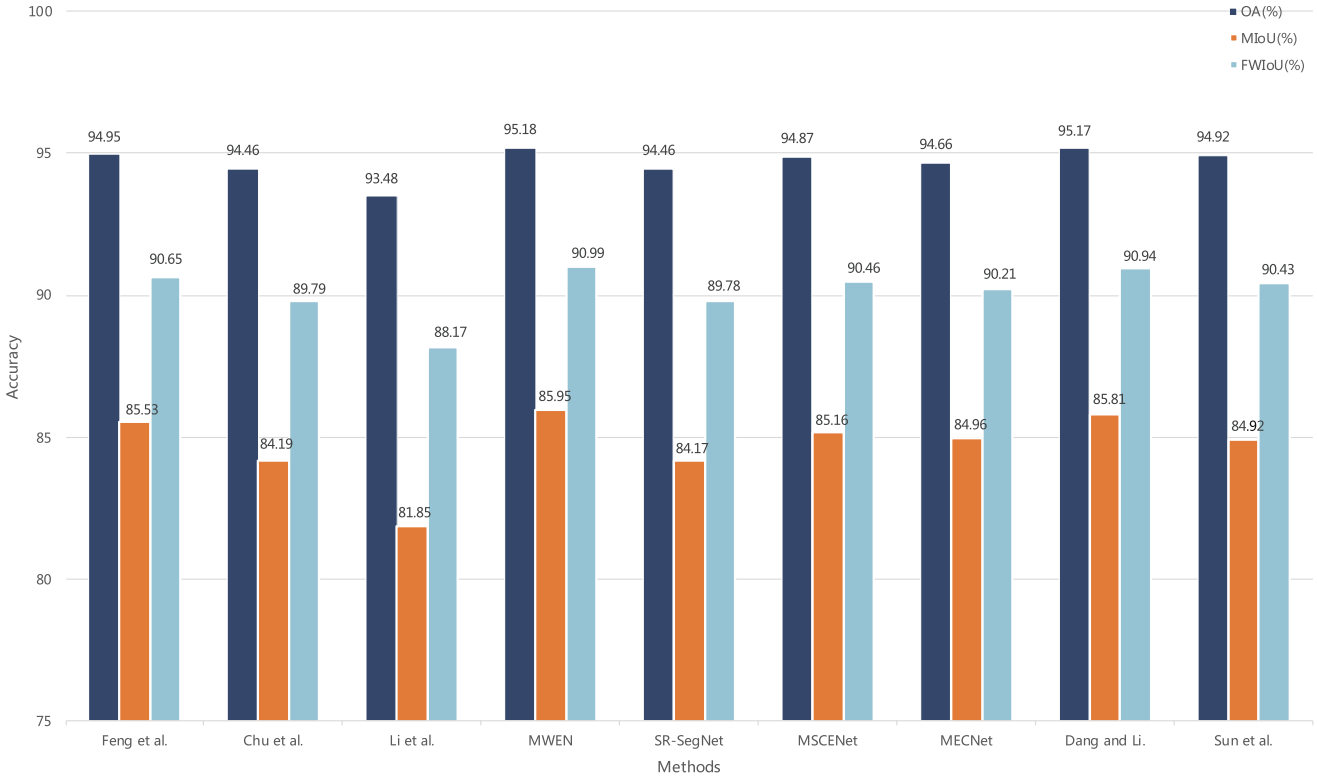

Fig. 12. The histogram of quantitative accuracy comparison of 9 water body classification methods on the 2020 Gaofen challenge water body segmentation benchmark.

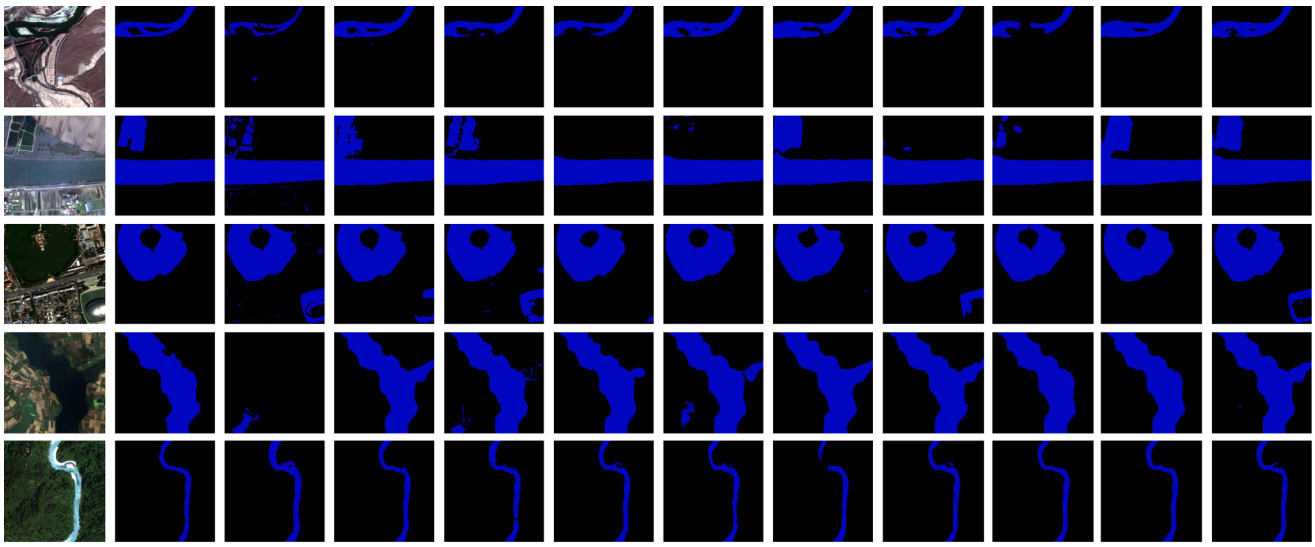

Fig. 13. Visible water body classification of the GID dataset using the codes of the previously published researches or the codes that we reproduce according to the description of reference. (a,b) are raw images and ground truth, respectively. The result of NDWI (McFeeters, 1996) are shown in c. (d,e) are the results of (Feng et al., 2018) and the results of (Chu et al., 2019), respectively. The results of (Li et al., 2019b) and the results of MWEN (Guo et al., 2020) are shown in (f,g), respectively. (h) The results of SR-SegNet (Weng et al., 2020), and (i) the results of MSCENet (Kang et al., 2021). The results of MECNet (Zhang et al., 2021) are displayed in (j). (k, l) are the results of Dang and Li (2021) and the results of (Sun et al., 2021.), respectively.

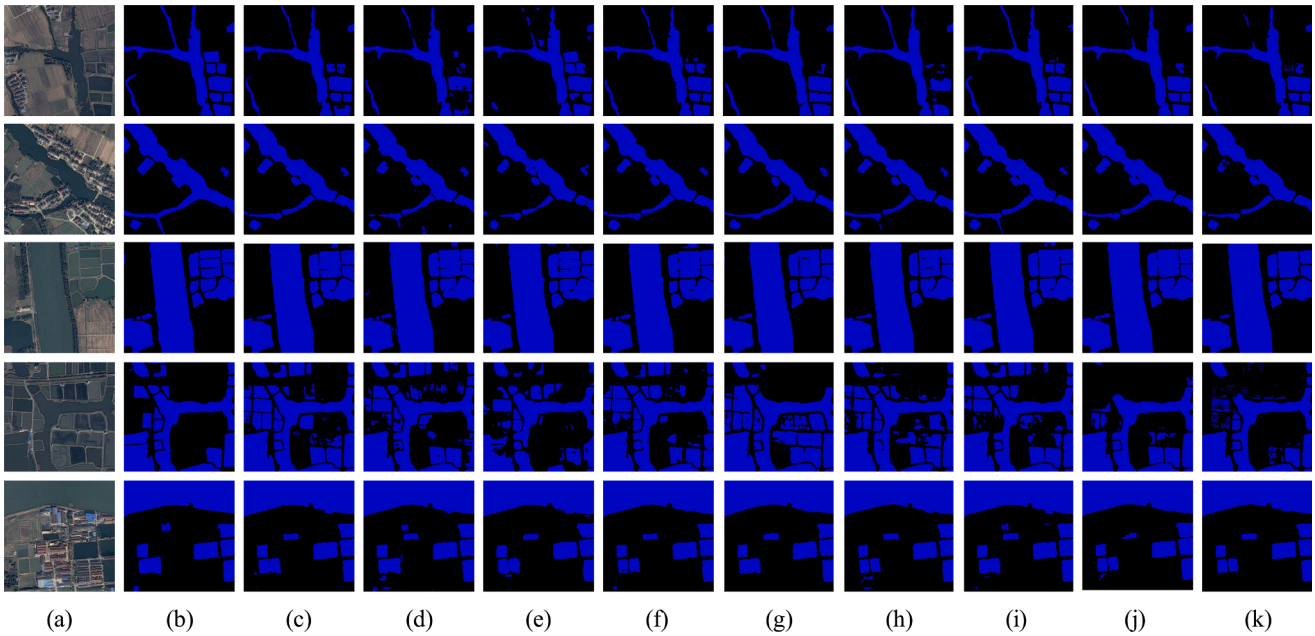

Fig. 14. Visible water body classification of the 2020 Gaofen challenge water body segmentation dataset using the codes of the previously published researches or the codes that we reproduce according to the description of reference. (a,b) are raw images and ground truth, respectively. (c,d) are the results of (Feng et al., 2018) and the results of (Chu et al., 2019), respectively. The results of (Li et al., 2019b) and the results of MWEN (Guo et al., 2020) are shown in (e,f), respectively. (g) The results of SR-SegNet (Weng et al., 2020), and (h) the results of MSCENet (Kang et al., 2021). The results of MECNet (Zhang et al., 2021) are displayed in (i). (j,k) are the results of (Dang and Li, 2021) and the results of (Sun et al., 2021.), respectively.

rate with a coefficient of 0.9 to control the weight distribution (momentum and current gradient), and used the exponential decay rate with a coefficient of 0.999 to control the effect of the square of the previous gradient. The initial learning rate was set to 0.0001 and was reduced during the training process. The batch-size was set to 4, and the maximum epoch was 100. We finally adopted binary cross-entropy loss for training and picked the epoch that achieved the highest MIoU on the validation tile. In particular, after continuous experimental attempts, we finally determined that the optimal segmentation threshold of NDWI is 0.332. (i.e., when the calculated value is greater than 0.332, it is recognized as a water body, otherwise it is a non-water body).

5.2.3. Discussion

In this section, we compare 9 DL-based learning methods, of which 5 belong to feature fusion-based methods, while the other methods are classified as boundary constraint-based methods. we also add a non-DL-based methods on the GID benchmark. Note that because the principles of approaches employed in the experiments and discussions are all introduced in Section 3, we do not repeat them here. The quantitative results on the two public benchmarks are shown in Tables 2 and 3 and Fig. 11 and Fig. 12 in detail. Furthermore, Fig. 13 and Fig. 14 display some samples of the water body classification results and their ground truths for the GID dataset and the 2020 Gaofen Challenge water body segmentation dataset, respectively.

It is obvious from the comparison in Table 2 and Fig. 11 that the result of NDWI is significantly inferior to that of other DL methods. The third column in Fig. 13 also shows obvious classification errors. In fact, most non-DL-based methods cannot stably adapt to abundant classification tasks. Furthermore, it can be seen intuitively from Fig. 11 and Fig. 13 that Dang and Li (2021) is superior to the other nine methods on the GID dataset, it shows that SSL algorithms and intervention of prior knowledge can improve the performance of the network, which confirms the opportunity mentioned in Section 2.2(5. However, it is worth noting that MWEN (Weng et al., 2020) achieved the best performance on another benchmark. It demonstrates that the existing DL-based methods are very dependent on specific datasets, and cannot guarantee to obtain enough stable performance on data with large style differences. As we described in the challenges faced in Section 2.1(2) and (5), the distribution of existing datasets varies greatly. At present, there is no benchmark that can cover most of the world's water conditions for training sufficiently stable models, which seems to be one of the potential research directions in the future.

Another significant problem that can be seen from Fig. 13 and Fig. 14 is the randomness in the classification of various DL algorithms. As a matter of fact, DL models often have thousands of parameters that change in each epoch of training, thereby affecting the validation results of each epoch. When the DL model training converges, the overall inference results are stable, but for the extraction results of a certain image, this is not necessarily the best classification result. This is why one algorithm is detecting water in one area and not detecting in other areas in Fig. 13 and Fig. 14. Moreover, this problem can be attributed to the instability and insufficient interpretability of deep network. It is not only an overt problem but also a research hotspot in the field of DL (Ovadia et al., 2019; Wang et al., 2020a).

In addition, it can be found from Tables 2 and 3 that the method based on feature fusion is better than the method based on boundary constraints as a whole. It illustrates that the exploration of adaptive fusion through multi-scale or other different features has become mature, but how to more effectively use or recover boundary information is still a tough problem, as we mentioned in Section 2.1(4. From the conclusion of (Sun et al., 2021), we can understand that CRF and other common post-processing algorithms play a negative role in the recognition of small water bodies, so more methods dedicated to water boundary restoration need to be further developed.

6. Future research directions

Due to the broad applications and numerous unresolved problems, water body classification from high-resolution optical RS imagery is still a significant and challenging mission of RS images intelligent interpretation. Propelled by the great advancement of relevant DL algorithms and RS land cover classification benchmarks, the accuracy of water body extraction is enhancing gradually. Although many meaningful achievements have been proposed, there still exist a lot of challenges and gaps that need to be further studied, as described in Section 2. Therefore, by summarizing the existing algorithms and publicly open benchmarks, this section discusses several potential future research directions of water body extraction, hoping to provide some inspiration for potential readers.

(1) **Learning discriminative multi-scale context feature representations.** Rivers, lakes, ponds, and other various types of water bodies have different shapes and sizes, demonstrating different scales characteristics. Therefore, how to capture multi-scale features has been a vital issue. Despite most existing methods based on DL techniques focus on improving the capability of multi-scale representation or mining object-context information, the fact is that it still has a great room to upgrade for further research. As mentioned in Section 2.2, some advanced techniques were exploited to strengthen multi-scale features or make more effective use of spatial context information. In the future, there is much work to be done in this direction.

(2) **Developing specific classification methods from large-size images.** As the increase of spatial resolution, the geographical area covered by a tile is narrower, which seems to be an inherent contradiction and hampers the performance of water body classification. Some researches (Cheng et al., 2020; Ding et al., 2020; Ding et al., 2021; Li et al., 2021d; Chen et al., 2019a) about large-size scene segmentation in other tasks have emerged in recent years, such as exploiting wider windows' information and fusing global and local features. However, these existing developments have just started and are still far from simulating the intelligent perception of human vision. Currently, the mainstream methods tend to crop large-size images into small patches. In the future, how to use the limited GPU memory to model the spatial relative position consistency of large-scale area effectively? How to eliminate the negative edge noise during the inference?

(3) **Geographical knowledge-driven and domain adaption water body classification methods.** The images from existing open benchmarks are generally concentrated in some specific areas, such as Wuhan, Suzhou, and Zurich. However, water bodies contained in these areas can not completely cover the characteristics of global water bodies, which is a disadvantage for global water resources monitoring. From the algorithm aspect, introducing geographical knowledge representation or relevant knowledge graph to supervise or pre-train networks is a good choice. To the best of our knowledge, some researches (Li et al., 2022a; Li et al., 2021g; Li et al., 2021e; Ouyang and Li, 2021; Li et al., 2021k) have been done in the past several years, which can provide inspiration for detecting water body. Among them, Li et al. (2022a) innovatively delved into a collaboratively boosting framework (CBF) to obtain better RS classification performance by combining knowledge-guided reasoning algorithm and data-driven deep learning model. Further, a new remote sensing knowledge graph (RSKG) was released, which greatly promotes the application of geographical knowledge in the field of RS intelligence interpretation (Li et al., 2021g). Another valuable option is to explore algorithms inspired by domain adaption technology, as described in Section 2.2(2), GAN and its improved version (Zhu et al., 2017; Melas-Kyriazi and Manrai, 2021; Kim and Byun, 2020; Li et al., 2021i) have made great advancements

in the problem of domain shift. Consequently, it is worthy of developing relevant approaches for water body classification.

(4) **Constructing high-resolution optical water body segmentation datasets with large-size imagery.** As mentioned above, existing high-resolution optical datasets are not enough to provide global characteristics of various water bodies, and images with small sizes destroy the continuity of water bodies. From the data aspect, building a novel larger-size scene dataset that contains massive water body objects from all regions of the world is meaningful and necessary.

(5) **Water body detection from multi-temporal RS images.** As one of the essential branches of intelligent interpretation of RS images, automatic water body classification from high-resolution optical images can be employed in many practical application scenarios. Considering this, analyzing and monitoring water bodies from the same area at different times is worthwhile and necessary for flood monitoring and mapping of water resources change. Multi-temporal water body change monitoring, as an extended task of water body classification, can be directly served for major projects such as the emergency rescue of natural disasters and evaluation of ecological environment.

(6) **Developing a monolithic framework for water body classification from multimodal RS data.** In this paper, water body classification from high-resolution optical images is discussed chiefly, yet in fact, hyperspectral images, SAR images (or even future high-resolution SAR based products such as Surface Water and Ocean Topography (SWOT) (Morrow et al., 2019)), and even Laser Radar (Lidar) data are also employed for water body detection because of their respective imaging characteristics. Considering that each sensor has its own limitations, modeling a robust framework based on multimodal images is significant for further development in the future. In the past several years, some multimodal images fusion strategies (Audebert et al., 2018; Tuia et al., 2014; Ferreira et al., 2016; Li et al., 2022b) and alignment and matching approaches (Tuia et al., 2016; Hughes et al., 2020) have emerged. For example, as mentioned in (Mahdianpari et al., 2021), multiple spectral, contextual, and elevation features are extracted from multispectral data and Lidar data respectively, and then combined into a multi-feature stack. Among them, morphometric topographic features (such as slope and curvature) extracted from Lidar data can also be added to the multi-feature stack. When classifying water bodies on a global scale, it seems inevitable that the elevation changes significantly in different scenes. In this case, the segmentation scene is divided by Digital Elevation Model (DEM) (Acharya et al., 2018), or the elevation is used as additional supervision information. However, how to build a multimodal framework is still a challenging need to be tackled in the future.

(7) **Water body vectorization mapping.** Intuitively, water body vectorization mapping consists of two independent pipelines: water body classification and water body vectorization. The former has been relatively mature, while the latter still has great room for improvement. Up to now, existing methods have extracted water body areas and saved them as raster data, which is inconvenient for spatial analysis and relevant information queries. Furthermore, pixel-level classification can not meet the high-quality requirements of high-precision projects, consequently, studying the methods of end-to-end or automatic water body vectorization is promising. Currently, some researches about road and building vectorization (Chen et al., 2021b; Wei and Ji, 2021; Abdollahi et al., 2021) have been already done, such as creating an end-to-end vectorization mapping framework, using graph convolutional networks (GCN) to vector maps, and tracking pixels. In a word, it is valuable to exploit water body vectorization approaches.

7. Conclusions

Because the spectral range of water bodies contained in the high-resolution images is relatively narrow. They have a variety of shapes and sizes, and the distribution scene is wide and complex. Moreover, their boundaries are generally winding, and there is a lack of data sets that can be used for supervised training. Therefore, extracting the water body from high-resolution optical RS images still remains great challenges. Fortunately, profit by the rapid development of related CV techniques (e.g., effective multi-scale feature fusion approaches, object-contextual representations, boundary optimization strategies, weakly supervised learning, and so on.) and relevant RS technologies (e.g., multi-modal feature fusion methods, cloud and shadow removal approaches, natural resource mapping techniques, and so on), The abovementioned challenges seem to be expected to be solved in the future.

In the past decades, water body classification from high-resolution optical RS images has attracted great attention and obtained major advancement. Especially after the great achievements of DL techniques in the intelligent interpretation of RS images, articles on novel water body classification algorithms have increased rapidly. Considering that there is no relevant survey to summarize the existing methods, in this paper, we first analyze the current challenges in light of the features of water bodies in high-resolution optical RS images. Meanwhile, in conjunction with the progressive DL-based approaches in response to these challenges, we discussed the corresponding potential opportunities one by one. Then, we surveyed two kinds of existing methods and introduced them chiefly. Next, we listed some practical applications and demonstrated some publicly open benchmarks. Specially, we selected two representative benchmarks that were employed for evaluating 10 typical approaches, and discussed their performance. Finally, we discussed a range of promising opportunities for providing some references for follow-up research.

CRediT authorship contribution statement

Yansheng Li: Methodology, Funding acquisition, Investigation, Writing - original draft. **Bo Dang:** Data processing, Writing - original draft, Methodology, Validation. **Yongjun Zhang:** Writing - review & editing, Supervision. **Zhenhong Du:** Writing - review & editing, Supervision.

Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/Jack-bo1220/Benchmarks-for-Water-Body-Extraction-from-HRORS-Imagery.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0505003; the National Natural Science Foundation of China under Grant 41971284.

Appendix A

See Tables A1, A2, and A3.

Table A1
The main parameters of some high-resolution optical sensors.

Remote sensor	Spatial resolution (m)	Radiometric resolution (bit)	Number of bands	Spectral range (micron)	Revisit time (days)	Data acquisition permission	Launch year
IKONOS	1, 4	11	5	Pan:0.45 to 0.9 Blue:0.45 to 0.53 Green:0.52 to 0.61 Red:0.64 to 0.72 NIR:0.76 to 0.86	1.5 to 2.9	paid	1999
QuickBird	0.61 to 0.72, 2.44 to 2.88	11	5	Pan:0.405 to 1.053 Blue:0.430 to 0.545 Green:0.466 to 0.62 Red:0.59 to 0.71 NIR:0.715 to 0.918	1 to 6	paid	2001
SPOT-5	2.5, 10, 20	-	5	Pan:0.49 to 0.69 B1:0.49 to 0.61 B2:0.61 to 0.68 B3:0.78 to 0.89 B4:1.58 to 1.78	26	paid	2002
WorldView-1	0.5	11	1	Pan:0.4 to 0.9	1.7	paid	2007
GeoEye-1	0.41, 1.65	11	5	Pan:0.45 to 0.80 Blue:0.45 to 0.51 Green:0.51 to 0.58 Red:0.655 to 0.69 NIR:0.78 to 0.92	less than 3	paid	2008
WorldView-2	0.46, 1.85	11	9	Pan:0.45 to 0.80 CA:0.40 to 0.45 Blue:0.45 to 0.51 Green:0.51 to 0.58 Yellow:0.585 to 0.625 Red:0.63 to 0.69 RE:0.705 to 0.745 NIR:0.77 to 0.895 NIR:0.86 to 1.04	1.1	paid	2009
Cartosat-2B	0.8	-	1	Pan:0.45 to 0.85	2/4	paid	2010
KOMPSAT-3	0.7, 2.8	14	5	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.60 Red:0.63 to 0.69 NIR:0.76 to 0.90	3	paid	2012
ZY-3	2.1, 5.8	-	5	Pan:0.45 to 0.80 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.77 to 0.89	3 to 5	paid	2012
SPOT-6/7	1.5, 6.0	12	5	Pan:0.45 to 0.745 Blue:0.45 to 0.52 Green:0.53 to 0.59 Red:0.625 to 0.695 NIR:0.76 to 0.89	3	paid	2012/ 2014
GF-1	2, 8/16	-	5, 4	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.77 to 0.89	4	partial free	2013
GF-2	0.8, 3.2	-	5	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.770 to 0.89	5	paid	2014
WorldView-3	0.31, 1.24, 3.7	11/14	17	Pan:0.45 to 0.80 MS:0.40 to 2.365	1	paid	2014

Table A2
The main parameters of some high-resolution optical sensors.

Remote sensor	Spatial resolution (m)	Radiometric resolution (bit)	Number of bands	Spectral range (micron)	Revisit time (days)	Data acquisition permission	Launch year
Flock-1, −1b, −1c	3, 5	-	4	Blue:0.42 to 0.53 Green:0.50 to 0.59 Red:0.61 to 0.70 NIR:0.76 to 0.86	1 to 2	paid	2014
DMC-3	1, 4	-	5	Pan:0.45 to 1.00 Blue:0.45 to 0.52 Green:0.52 to 0.60 Red:0.60 to 0.68 NIR:0.73 to 1.30	1	paid	2015
WorldView-4	0.31, 1.24	11	5	Pan:0.45 to 0.80 Blue:0.45 to 0.51 Green:0.51 to 0.58 Red:0.655 to 0.69 NIR:0.78 to 0.92	1	paid	2016
SuperView-1	0.5, 2	11	5	Pan:0.45 to 0.89 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.77 to 0.89	1	paid	2016/2018
GF-6	2, 8/16	12	5, 8	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.77 to 0.89	4	partial free	2018
GRUS-1A	2.5, 5	-	6	Pan:0.45 to 0.90 Blue:0.45 to 0.505 Green:0.515 to 0.585 Red:0.62 to 0.685 RE:0.705 to 0.745 NIR:0.77 to 0.90	1	paid	2018
GF-7	0.8, 2.6	-	5	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.59 Red:0.63 to 0.69 NIR:0.77 to 0.89	-	paid	2019
Jilin-1 GF −02A, 02B, 03A	1, 4.2	-	5	Pan:0.45 to 0.70 Blue:0.45 to 0.51 Green:0.51 to 0.58 Red:0.63 to 0.69 NIR:0.77 to 0.895	3.3	paid	2019
KEOSat	3	-	4	-	1	paid	2019
HYDRA-1, 2	4, 35	-	5	-	-	paid	2020
1HOPSat	1	-	8	-	0.04	paid	2020
SkySat-1 to −18	0.86, 1	12	5	Pan:0.45 to 0.90 Blue:0.45 to 0.515 Green:0.515 to 0.595 Red:0.605 to 0.695 NIR:0.74 to 0.90	0.5	paid	2013–2020
KOMPSAT-7	0.3, 1.2, 4	-	6	Pan:0.45 to 0.90 Blue:0.45 to 0.52 Green:0.52 to 0.60 Red:0.63 to 0.69 NIR:0.76 to 0.90 MWIR:3.00 to 5.00	-	paid	2021

Table A3
Summary of existing reviews of water body classification.

No.	Review title	Year	Publication	Main content
1	A review of hyperspectral remote sensing and its applicationin vegetation and water resource studies (Govender et al., 2007)	2007	Water Sa	Survey the application of hyperspectral imagery in water resource studies.
2	Water-body area extraction from high resolution satellite images-an introduction, review, and comparison(Nath and Deb, 2010)	2010	IJIP	Review early methods applied for water body extraction using satellite remote sensing.
3	Water body extraction methods study based on RS and GIS(Haibo et al., 2011)	2011	PROENV	Analysis unsupervised classification, supervised classification, single-band threshold, inter spectrum relation method and water index method.
4	A review on extraction of lakes from remotely sensed optical satellite data with a special focus on cryospheric lakes(Jawak et al., 2015)	2015	ARS	Reviewing methods, technologies, and satellite sensors employed for the extraction of cryospheric lakes from satellite imagery.
5	A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation (Musa et al., 2015)	2015	HESS	Providing a review of applications and limitations of satellite remote sensing in surface water modelling, mapping and parameter estimation.
6	Detecting, extracting, and monitoring surface water from space using optical sensors: A review(Huang et al., 2018)	2018	REV GEOPHYS	Reviewing the current status and challenges of detecting, extracting, and monitoring surface water using optical remote sensing in the last decade.
7	A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management(Wang and Xie, 2018)	2018	Water	Providing an analyze on some applications of remote sensing and GIS in water resources and flood risk management.
8	Evaluation of water indices for surface water extraction in a Landsat 8 scene of Nepal(Acharya et al., 2018)	2018	Sensors	Comparison of three water body index methods using Landsat 8 data.
9	Inundation extent mapping by synthetic aperture radar: A review(Shen et al., 2019)	2019	Remote Sensing	Reviewing algorithms, strengths, and limitations of flood inundation mapping using SAR data.
10	Surface water detection and delineation using remote sensing images: A review of methods and algorithms(Bijeesh and Narasimhamurthy, 2020)	2020	SWAM	Review of techniques, methods, algorithms and the sensors/satellites for surface water body detection.
11	Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives	2022	ISPRS JPRS	A systematic review of the latest and advanced approaches, achievements, and the perspectives of future research directions in water body classification from high-resolution optical remote sensing imagery.

References

Abdollahi, A., Pradhan, B., Alamri, A., 2021. Roadvecnet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up. GIScience Remote Sens. 1–24.

Abid, N., Shahzad, M., Malik, M.I., Schwanecke, U., Ulges, A., Kovács, G., Shafait, F., 2021. Ucl: Unsupervised curriculum learning for water body classification from remote sensing imagery. Int. J. Appl. Earth Obs. Geoinf. 105, 102568.

Acharya, T.D., Lee, D.H., Yang, I.T., Lee, J.K., 2016. Identification of water bodies in a landsat 8 oli image using a j48 decision tree. Sensors 16, 1075.

Acharya, T.D., Subedi, A., Lee, D.H., 2018. Evaluation of water indices for surface water extraction in a landsat 8 scene of Nepal. Sensors 18, 2580.

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., Sousa, J.J., 2017. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. Remote Sens. 9, 1110.

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond rgb: Very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogram. Remote Sens. 140, 20–32.

Bao, L., Lv, X., Yao, J., 2021. Water extraction in sar images using features analysis and dual-threshold graph cut model. Remote Sens. 13, 3465.

Benoudjit, A., Guida, R., 2019. A novel fully automated mapping of the flood extent on sar images using a supervised classifier. Remote Sens. 11, 779.

Bermudez, J., Happ, P., Oliveira, D., Feitosa, R., 2018. Sar to optical image synthesis for cloud removal with generative adversarial networks. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci., 4.

Bijeesh, T., Narasimhamurthy, K., 2020. Surface water detection and delineation using remote sensing images: A review of methods and algorithms. Sustain. Water Resour. Manage. 6, 1–23.

Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Dziedzic, T., Zambrzycka, A., 2021. Landcover. ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1102–1110.

Bokhovkin, A., Burnaev, E., 2019. Boundary loss for remote sensing imagery semantic segmentation. In: International Symposium on Neural Networks, Springer. pp. 388–401.

Borji, A., Cheng, M.M., Hou, Q., Jiang, H., Li, J., 2019. Salient object detection: A survey. Comput. Visual Media 5, 117–150.

Borse, S., Wang, Y., Zhang, Y., Porikli, F., 2021. Inverseform: A loss function for structured boundary-aware segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5901–5911.

Byun, Y., Han, Y., Chae, T., 2015. Image fusion-based change detection for flood extent extraction using bi-temporal very high-resolution satellite images. Remote Sens. 7, 10347–10363.

Chaudhuri, B., Demir, B., Chaudhuri, S., Bruzzone, L., 2017. Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method. IEEE Trans. Geosci. Remote Sens. 56, 1144–1158.

Chen, C.F., Fan, Q., Panda, R., 2021a. Crossvit: Cross-attention multi-scale vision transformer for image classification. arXiv preprint arXiv:2103.14899.

Chen, D., Zhong, Y., Zheng, Z., Ma, A., Lu, X., 2021b. Urban road mapping based on an end-to-end road vectorization mapping network framework. ISPRS J. Photogram. Remote Sens. 178, 345–365.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

Chen, W., Jiang, Z., Wang, Z., Cui, K., Qian, X., 2019a. Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8924–8933.

Chen, X., He, K., 2021. Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15750–15758.

Chen, X., Qi, D., Shen, J., 2019b. Boundary-aware network for fast and high-accuracy portrait segmentation. arXiv preprint arXiv:1901.03814.

Chen, Y., Fan, R., Yang, X., Wang, J., Latif, A., 2018b. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. Water 10, 585.

Chen, Y., Tang, L., Kan, Z., Bilal, M., Li, Q., 2020. A novel water body extraction neural network (wbe-nn) for optical high-resolution multispectral imagery. J. Hydrol. 588, 125092.

Chen, Y., Tang, L., Yang, X., Fan, R., Bilal, M., Li, Q., 2019c. Thick clouds removal from multitemporal zy-3 satellite images using deep learning. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 13, 143–153.

Chen, Z., Zhou, H., Xie, X., Lai, J., 2019d. Contour loss: Boundary-aware learning for salient object segmentation. arXiv preprint arXiv:1908.01975.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogram. Remote Sens. 117, 11–28.

Cheng, G., Han, J., Guo, L., Qian, X., Zhou, P., Yao, X., Hu, X., 2013. Object detection in remote sensing imagery using a discriminatively trained mixture model. ISPRS J. Photogram. Remote Sens. 85, 32–43.

Cheng, G., Xie, X., Han, J., Guo, L., Xia, G.S., 2020. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 13, 3735–3756.

Cheng, H.K., Chung, J., Tai, Y.W., Tang, C.K., 2020b. Cascadepsp: toward class-agnostic and very high-resolution segmentation via global and local refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8890–8899.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. Proc. IEEE 104, 2207–2219.

Cho, S., Jun, T.J., Oh, B., Kim, D., 2020. Dapas: Denoising autoencoder to prevent adversarial attack in semantic segmentation. In: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8.

Chu, Z., Tian, T., Feng, R., Wang, L., 2019. Sea-land segmentation with res-unet and fully connected crf. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 3840–3843.

Cui, B., Jing, W., Huang, L., Li, Z., Lu, Y., 2020. Sanet: A sea–land segmentation network via adaptive multiscale feature learning. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 116–126.

Dang, B., Li, Y., 2021. Msresnet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery. Remote Sens. 13, 3122.
Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 172–181.
Dhara, S., Dang, T., Parial, K., Lu, X.X., 2020. Accounting for uncertainty and reconstruction of flooding patterns based on multi-satellite imagery and support vector machine technique: A case study of can tho city, Vietnam. Water 12, 1543.
Ding, L., Lin, D., Lin, S., Zhang, J., Cui, X., Wang, Y., Tang, H., Bruzzone, L., 2021. Looking outside the window: Wide-context transformer for the semantic segmentation of high-resolution remote sensing images. arXiv preprint arXiv: 2106.15754.
Ding, L., Zhang, J., Bruzzone, L., 2020. Semantic segmentation of large-size vhr remote sensing images using a two-stage multiscale training architecture. IEEE Trans. Geosci. Remote Sens. 58, 5367–5376.
Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., Van De Giesen, N., 2016. Earth's surface water change over the past 30 years. Nat. Clim. Change 6, 810–813.
Dong, S., Pang, L., Zhuang, Y., Liu, W., Yang, Z., Long, T., 2019. Optical remote sensing water-land segmentation representation based on proposed sns-cnn network. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 3895–3898.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929.
Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. ISPRS J. Photogram. Remote Sens. 158, 63–75.
Duan, L., Hu, X., 2019. Multiscale refinement network for water-body segmentation in high-resolution satellite imagery. IEEE Geosci. Remote Sens. Lett. 17, 686–690.
Dubes, R.C., Jain, A.K., Nadabar, S.G., Chen, C.C., 1990. Mrf model-based algorithms for image segmentation. In: [1990] Proceedings. 10th International Conference on Pattern Recognition, IEEE. pp. 808–814.
Feng, Q., Liu, J., Gong, J., 2015. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier–a case of yuyao, china. Water 7, 1437–1455.
Feng, W., Sui, H., Huang, W., Xu, C., An, K., 2018. Water body extraction from very high-resolution remote sensing imagery using deep u-net and a superpixel-based conditional random field model. IEEE Geosci. Remote Sens. Lett. 16, 618–622.
Ferreira, E., de Albuquerque Araújo, A., Dos Santos, J.A., 2016. A boosting-based approach for remote sensing multimodal image classification. In: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, pp. 416–423.
Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated water extraction index: A new technique for surface water mapping using landsat imagery. Remote Sens. Environ. 140, 23–35.
Fisher, A., Flood, N., Danaher, T., 2016. Comparing landsat water index methods for automated water classification in eastern australia. Remote Sens. Environ. 175, 167–182.
Fletcher, T.D., Andrieu, H., Hamel, P., 2013. Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. Adv. Water Resour. 51, 261–279.
Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. 61, 399–409.
Fu, K., Lu, W., Diao, W., Yan, M., Sun, H., Zhang, Y., Sun, X., 2018. Wsf-net: Weakly supervised feature-fusion network for binary segmentation in remote sensing image. Remote Sens. 10, 1970.
Fu, L., Zhou, C., Guo, Q., Juefei-Xu, F., Yu, H., Feng, W., Liu, Y., Wang, S., 2021. Auto-exposure fusion for single-image shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10571–10580.
Gao, H., Wang, L., Jing, L., Xu, J., 2016. An effective modified water extraction method for landsat-8 oli imagery of mountainous plateau regions. In: IOP conference series: earth and environmental science, IOP Publishing. p. 012010.
Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., Langan, T.E., 2019. Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data. Sensors 19, 1486.
Govender, M., Chetty, K., Bulcock, H., 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water Sa 33, 145–151.
Grimaldi, S., Xu, J., Li, Y., Pauwels, V.R., Walker, J.P., 2020. Flood mapping under vegetation using single sar acquisitions. Remote Sens. Environ. 237, 111582.
Guo, H., He, G., Jiang, W., Yin, R., Yan, L., Leng, W., 2020. A multi-scale water extraction convolutional neural network (mwen) method for gaofen-1 remote sensing images. ISPRS Int. J. Geo-Inform. 9, 189.
Haibo, Y., Zongmin, W., Hongling, Z., Yu, G., 2011. Water body extraction methods study based on rs and gis. Proc. Environ. Sci. 10, 2619–2624.
Hashemi-Beni, L., Gebrehiwot, A.A., 2021. Flood extent mapping: an integrated method using deep learning and region growing using uav optical data. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 2127–2135.
He, J., Deng, Z., Zhou, L., Wang, Y., Qiao, Y., 2019. Adaptive pyramid context network for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7519–7528.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Hou, Q., Zhang, L., Cheng, M.M., Feng, J., 2020. Strip pooling: Rethinking spatial pooling for scene parsing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4003–4012.
Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.
Huang, C., Chen, Y., Zhang, S., Wu, J., 2018. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. Rev. Geophys. 56, 333–360.
Huang, L., Liu, L., Jiang, L., Zhang, T., 2018. Automatic mapping of thermokarst landforms from remote sensing images using deep learning: A case study in the northeastern tibetan plateau. Remote Sens. 10, 2067.
Huang, X., Tan, H., Zhou, J., Yang, T., Benjamin, A., Wen, S.W., Li, S., Liu, A., Li, X., Fen, S., et al., 2008. Flood hazard in hunan province of china: an economic loss analysis. Nat. Hazards 47, 65–73.
Huang, X., Xie, C., Fang, X., Zhang, L., 2015. Combining pixel-and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 8, 2097–2110.
Hughes, L.H., Marcos, D., Lobry, S., Tuia, D., Schmitt, M., 2020. A deep learning framework for matching of sar and optical imagery. ISPRS J. Photogram. Remote Sens. 169, 166–179.
Isikdogan, L.F., Bovik, A., Passalacqua, P., 2019. Seeing through the clouds with deepwatermap. IEEE Geosci. Remote Sens. Lett. 17, 1662–1666.
Jawak, S., Luis, A., 2015. A rapid extraction of water body features from antarctic coastal oasis using very high-resolution satellite remote sensing data. Aquatic Proc. 4, 125–132.
Jawak, S.D., Kulkarni, K., Luis, A.J., et al., 2015. A review on extraction of lakes from remotely sensed optical satellite data with a special focus on cryospheric lakes. Adv. Remote Sens. 4, 196.
Jawak, S.D., Luis, A.J., 2014. A semiautomatic extraction of antarctic lake features using worldview-2 imagery. Photogram. Eng. Remote Sens. 80, 939–952.
Ji, S., Wei, S., Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Trans. Geosci. Remote Sens. 57, 574–586.
Jin, Y., Xu, W., Zhang, C., Luo, X., Jia, H., 2021. Boundary-aware refined network for automatic building extraction in very high-resolution urban aerial images. Remote Sens. 13, 692.
Jung, H., Choi, H.S., Kang, M., 2021. Boundary enhancement semantic segmentation for building extraction from remote sensed image. IEEE Trans. Geosci. Remote Sens.
Kang, J., Guan, H., Peng, D., Chen, Z., 2021. Multi-scale context extractor network for water-body extraction from high-resolution optical remotely sensed images. Int. J. Appl. Earth Obs. Geoinf. 103, 102499.
Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, I.B., 2019. Boundary loss for highly unbalanced segmentation. in: International conference on medical imaging with deep learning, PMLR. pp. 285–296.
Khurshid, M.H., Khan, M.F., 2012. River extraction from high resolution satellite images. in: 2012 5th International Congress on Image and Signal Processing, IEEE. pp. 697–700.
Kim, M., Byun, H., 2020. Learning texture invariant representation for domain adaptation of semantic segmentation. in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12975–12984.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Klemenjak, S., Waske, B., Valero, S., Chanussot, J., 2012. Unsupervised river detection in rapideye data. in: 2012 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 6860–6863.
Kokelj, S.V., Jorgenson, M., 2013. Advances in thermokarst research. Permafrost Periglac. Process. 24, 108–119.
Komodakis, N., Gidaris, S., 2018. Unsupervised representation learning by predicting image rotations. in: International Conference on Learning Representations (ICLR).
Koutalakis, P., Tzoraki, O., Gkiatas, G., Zaimes, G.N., 2020. Using uav to capture and record torrent bed and banks, flood debris, and riparian areas. Drones 4, 77.
Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected crfs with gaussian edge potentials. Adv. Neural Inform. Process. Syst. 24, 109–117.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. 25, 1097–1105.
Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
Lamovec, P., Velkanovski, T., Mikos, M., Osir, K., 2013. Detecting flooded areas with machine learning techniques: case study of the selška sora river flash flood in september 2007. J. Appl. Remote Sens. 7, 073564.
Le, H., Samaras, D., 2021. Physics-based shadow image decomposition for shadow removal. IEEE Trans. Pattern Anal. Mach. Intell. 1–1.
Lee, D.H., et al., 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML, p. 896.
Li, A., Fan, M., Qin, G., Xu, Y., Wang, H., 2021a. Comparative analysis of machine learning algorithms in automatic identification and extraction of water boundaries. Appl. Sci. 11, 10062.
Li, B., Zhang, H., Xu, F., 2014. Water extraction in high resolution remote sensing image based on hierarchical spectrum and shape features. In: IOP Conference Series: Earth and Environmental Science, IOP Publishing. p. 012123.

Li, D., Tong, Q., Li, R., Gong, J., Zhang, L., 2012. Current issues in high-resolution earth observation technology. Sci. China Earth Sci. 55, 1043–1051.
Li, J., Wang, C., Xu, L., Wu, F., Zhang, H., Zhang, B., 2021b. Multitemporal water extraction of dongting lake and poyang lake based on an automatic water extraction and dynamic monitoring framework. Remote Sens. 13, 865.
Li, M., Wu, P., Wang, B., Park, H., Yang, H., Wu, Y., 2021c. A deep learning method of water body extraction from high resolution remote sensing images with multisensors. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 3120–3132.
Li, Q., Yang, W., Liu, W., Yu, Y., He, S., 2021d. From contexts to locality: Ultra-high resolution image segmentation via locality-aware contextual correlation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7252–7261.
Li, S., Wang, S., Zheng, Z., Wan, D., Feng, J., 2016. A new algorithm for water information extraction from high resolution remote sensing imagery. In: 2016 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 4359–4363.
Li, W., Chen, K., Chen, H., Shi, Z., 2021e. Geographical knowledge-driven representation learning for remote sensing images. IEEE Trans. Geosci. Remote Sens.
Li, W., Li, Y., Gong, J., Feng, Q., Zhou, J., Sun, J., Shi, C., Hu, W., 2021f. Urban water extraction with uav high-resolution remote sensing data based on an improved u-net model. Remote Sens. 13, 3165.
Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., Tan, Y., 2020. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. Remote Sens. Environ. 250, 112045.
Li, Y., Kong, D., Zhang, Y., Tan, Y., Chen, L., 2021g. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification. ISPRS J. Photogram. Remote Sens. 179, 145–158.
Li, Y., Ma, J., Zhang, Y., 2021h. Image retrieval from remote sensing big data: A survey. Inform. Fusion 67, 94–115.
Li, Y., Ouyang, S., Zhang, Y., 2022a. Combining deep learning and ontology reasoning for remote sensing image semantic segmentation. Knowl.-Based Syst. 243, 108469.
Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., Li, H., 2021i. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. ISPRS J. Photogram. Remote Sens. 175, 20–33.
Li, Y., Zhang, Y., Huang, X., Yuille, A.L., 2018. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. ISPRS J. Photogram. Remote Sens. 146, 182–196.
Li, Y., Zhang, Y., Zhu, Z., 2021j. Error-tolerant deep learning for remote sensing image scene classification. IEEE Trans. Cybernet. 51, 1756–1768.
Li, Y., Zhou, Y., Zhang, Y., Zhong, L., Wang, J., Chen, J., 2022b. Dkdfn: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. ISPRS J. Photogram. Remote Sens. 186, 170–189.
Li, Y., Zhu, Z., Yu, J.G., Zhang, Y., 2021k. Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification. IEEE Trans. Geosci. Remote Sens. 59, 10590–10603.
Li, Z., Shen, H., Cheng, Q., Li, W., Zhang, L., 2019a. Thick cloud removal in high-resolution satellite images using stepwise radiometric adjustment and residual correction. Remote Sens. 11, 1925.
Li, Z., Wang, R., Zhang, W., Hu, F., Meng, L., 2019b. Multiscale features supported deeplabv3+ optimization scheme for accurate water semantic segmentation. IEEE Access 7, 155787–155804.
Likas, A., Vlassis, N., Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern Recogn. 36, 451–461.
Liu, L., Xi, Z., Ji, R., Ma, W., 2019. Advanced deep learning techniques for image style transfer: A survey. Signal Process.: Image Commun. 78, 465–470.
Liu, Y., Piramanayagam, S., Monteiro, S.T., Saber, E., 2017. Dense semantic labeling of very-high-resolution aerial imagery and lidar with fully-convolutional neural networks and higher-order crfs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 76–85.
Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
Luo, X., Tong, X., Hu, Z., 2021. An applicable and automatic method for earth surface water mapping based on multispectral images. Int. J. Appl. Earth Obs. Geoinf. 103, 102472.
Lv, W., Yu, Q., Yu, W., 2010. Water extraction in sar images using glcm and support vector machine. In: IEEE 10th international conference on signal processing proceedings.
Ma, C., Rao, Y., Cheng, Y., Chen, C., Lu, J., Zhou, J., 2020. Structure-preserving super resolution with gradient guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7769–7778.
Mahdianpari, M., Granger, J.E., Mohammadimanesh, F., Warren, S., Puestow, T., Salehi, B., Brisco, B., 2021. Smart solutions for smart cities: Urban wetland mapping using very-high resolution satellite imagery and airborne lidar data in the city of st. john's, nl, canada. J. Environ. Manage. 280, 111676.
Malinowski, R., Groom, G., Schwanghart, W., Heckrath, G., 2015. Detection and delineation of localized flooding from worldview-2 multispectral data. Remote Sens. 7, 14853–14875.
McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E., Franz, T.E., et al., 2017. The future of earth observation in hydrology. Hydrol. Earth Syst. Sci. 21, 3879–3914.
McFeeters, S.K., 1996. The use of the normalized difference water index (ndwi) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.
Melas-Kyriazi, L., Manrai, A.K., 2021. Pixmatch: Unsupervised domain adaptation via pixelwise consistency training. in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12435–12445.
Mergili, M., Müller, J.P., Schneider, J.F., 2013. Spatio-temporal development of high-mountain lakes in the headwaters of the amu darya river (central asia). Global Planet. Change 107, 13–24.
Miao, Z., Fu, K., Sun, H., Sun, X., Yan, M., 2018. Automatic water-body segmentation from high-resolution satellite images via deep networks. IEEE Geosci. Remote Sens. Lett. 15, 602–606.
Mitkari, K.V., Arora, M.K., Tiwari, R.K., 2017. Extraction of glacial lakes in gangotri glacier using object-based image analysis. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 10, 5275–5283.
Morrow, R., Fu, L.L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., d'Ovidio, F., Farrar, J.T., Gille, S.T., Lapeyre, G., et al., 2019. Global observations of fine-scale ocean surface topography with the surface water and ocean topography (swot) mission. Front. Mar. Sci. 6, 232.
Munawar, H.S., Ullah, F., Qayyum, S., Heravi, A., 2021. Application of deep learning on uav-based aerial images for flood detection. Smart Cities 4, 1220–1242.
Munawar, H.S., Ullah, F., Qayyum, S., Khan, S.I., Mojtahedi, M., 2021. Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection. Sustainability 13, 7547.
Musa, Z., Popescu, I., Mynett, A., 2015. A review of applications of satellite sar, optical, altimetry and dem data for surface water modelling, mapping and parameter estimation. Hydrol. Earth Syst. Sci. 19, 3755–3769.
Nath, R.K., Deb, S., 2010. Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. Int. J. Image Process. (IJIP) 3, 265–384.
Nex, F., Remondino, F., 2014. Uav for 3d mapping applications: a review. Appl. Geomat. 6, 1–15.
Nicholas, P., Stacey, S., Tom, H., Dronedeploy segmentation benchmark. https://github.com/dronedeploy/dd-ml-segmentation-benchmark.
Nong, Z., Su, X., Liu, Y., Zhan, Z., Yuan, Q., 2021. Boundary-aware dual stream network for vhr remote sensing images semantic segmentation. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.
Ouyang, S., Li, Y., 2021. Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. Remote Sens. 13, 119.
Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., Snoek, J., 2019. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. Adv. Neural Inform. Process. Syst. 32.
Pan, X., Gao, L., Marinoni, A., Zhang, B., Yang, F., Gamba, P., 2018. Semantic labeling of high resolution aerial imagery and lidar data with fine segmentation network. Remote Sens. 10, 743.
Pappas, O.A., Anantrasirichai, N., Achim, A.M., Adams, B.A., 2020. River planform extraction from high-resolution sar images via generalized gamma distribution superpixel classification. IEEE Trans. Geosci. Remote Sens. 59, 3942–3955.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inform. Process. Syst. 32, 8026–8037.
Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422.
Qayyum, N., Ghuffar, S., Ahmad, H.M., Yousaf, A., Shahid, I., 2020. Glacial lakes mapping using multi satellite planetscope imagery and deep learning. ISPRS Int. J. Geo-Inform. 9, 560.
Qi, B., Zhuang, Y., Chen, H., Dong, S., Li, L., 2019. Fusion feature multi-scale pooling for water body extraction from optical panchromatic images. Remote Sens. 11, 245.
Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019. Basnet: Boundary-aware salient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7479–7489.
Qiu, S., Liu, Q., Zhou, S., Wu, C., 2019. Review of artificial intelligence adversarial attack and defense technologies. Appl. Sci. 9, 909.
Robinson, C., Malkin, K., Hu, L., Dilkina, B., Jojic, N., 2020. Weakly supervised semantic segmentation in the 2020 ieee grss data fusion contest. In: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 7046–7049.
Schmarje, L., Santarossa, M., Schröder, S.M., Koch, R., 2020. A survey on semi-, self-and unsupervised techniques in image classification. arXiv preprint arXiv:2002.08721.
Schmitt, M., Hughes, L., Qiu, C., Zhu, X., 2019. Sen12ms-a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci. 42, 153–160.
Schmitt, M., Prexl, J., Ebel, P., Liebel, L., Zhu, X.X., 2020. Weakly supervised semantic segmentation of satellite images for land cover mapping–challenges and opportunities. arXiv preprint arXiv:2002.08254.
Sghaier, M.O., Foucher, S., Lepage, R., 2016. River extraction from high-resolution sar images combining a structural feature set and mathematical morphology. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 10, 1025–1038.
Shao, Z., Yang, K., Zhou, W., 2018. Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset. Remote Sens. 10, 964.
Shao, Z., Zhou, W., Deng, X., Zhang, M., Cheng, Q., 2020. Multilabel remote sensing image retrieval based on fully convolutional network. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 13, 318–328.

Shen, X., Wang, D., Mao, K., Anagnostou, E., Hong, Y., 2019. Inundation extent mapping by synthetic aperture radar: A review. Remote Sens. 11, 879.
Sherrah, J., 2016. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. arXiv preprint arXiv:1606.02585.
Shugar, D.H., Burr, A., Haritashya, U.K., Kargel, J.S., Watson, C.S., Kennedy, M.C., Bevington, A.R., Betts, R.A., Harrison, S., Strattman, K., 2020. Rapid worldwide growth of glacial lakes since 1990. Nat. Clim. Change 10, 939–945.
Song, C., Huang, B., Ke, L., 2013. Modeling and analysis of lake water storage changes on the tibetan plateau using multi-mission satellite data. Remote Sens. Environ. 135, 25–35.
Song, S., Liu, J., Liu, Y., Feng, G., Han, H., Yao, Y., Du, M., 2020. Intelligent object recognition of urban water bodies based on deep learning for multi-source and multi-temporal high spatial resolution remote sensing imagery. Sensors 20, 397.
Sui, H., Chen, G., Hua, L., 2013. An automatic integrated image segmentation, registration and change detection method for water-body extraction using hsr images and gis data. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. 7, W2.
Sun, X., Shi, A., Huang, H., Mayer, H., 2020. BasΘ{4} net: Boundary-aware semi-supervised semantic segmentation network for very high resolution remote sensing images. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 13, 5398–5413.
Sun, X., Wang, P., Yan, Z., Diao, W., Lu, X., Yang, Z., Zhang, Y., Xiang, D., Yan, C., Guo, J., et al., 2021. Automated high-resolution earth observation image interpretation: Outcome of the 2020 gaofen challenge. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 8922–8940.
Takikawa, T., Acuna, D., Jampani, V., Fidler, S., 2019. Gated-scnn: Gated shape cnns for semantic segmentation. in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5229–5238.
Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv preprint arXiv:1703.01780.
Thomas, R.F., Kingsford, R.T., Lu, Y., Hunter, S.J., 2011. Landsat mapping of annual inundation (1979–2006) of the macquarie marshes in semi-arid australia. Int. J. Remote Sens. 32, 4545–4569.
Tian, B., Li, Z., Zhang, M., Huang, L., Qiu, Y., Li, Z., Tang, P., 2017. Mapping thermokarst lakes on the qinghai–tibet plateau using nonlocal active contours in chinese gaofen-2 multispectral imagery. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 10, 1687–1700.
Ticehurst, C., Guerschman, J.P., Chen, Y., 2014. The strengths and limitations in using the daily modis open water likelihood algorithm for identifying flood events. Remote Sens. 6, 11791–11809.
Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2018. Learning transferable deep models for land-use classification with high-resolution remote sensing images. arXiv preprint arXiv:1807.05713.
Tuia, D., Marcos, D., Camps-Valls, G., 2016. Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. ISPRS J. Photogram. Remote Sens. 120, 1–12.
Tuia, D., Volpi, M., Trolliet, M., Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. IEEE Trans. Geosci. Remote Sens. 52, 7708–7720.
Volpi, M., Ferrari, V., 2015. Semantic segmentation of urban scenes by learning local class interactions. in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–9.
Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. science 289, 284–288.
Wang, H., Wu, X., Huang, Z., Xing, E.P., 2020a. High-frequency component helps explain the generalization of convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8684–8694.
Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020b. Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43, 3349–3364.
Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y., 2021a. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2). URL https://openreview.net/forum?id=bLBIbVaGDu.
Wang, S., Chen, W., Xie, S.M., Azzari, G., Lobell, D.B., 2020c. Weakly supervised deep learning for segmentation of remote sensing imagery. Remote Sens. 12, 207.
Wang, W., Lai, Q., Fu, H., Shen, J., Ling, H., Yang, R., 2021b. Salient object detection in the deep learning era: An in-depth survey. In: IEEE Trans. Pattern Anal. Mach. Intell..
Wang, X., Xie, H., 2018. A review on applications of remote sensing and geographic information systems (gis) in water resources and flood risk management.
Wang, Z., Gao, X., Zhang, Y., Zhao, G., 2020. Mslwenet: A novel deep learning network for lake water body extraction of google remote sensing images. Remote Sens. 12, 4140.
Ward, D.P., Petty, A., Setterfield, S.A., Douglas, M.M., Ferdinands, K., Hamilton, S.K., Phinn, S., 2014. Floodplain inundation and vegetation dynamics in the alligator rivers region (kakadu) of northern Australia assessed using optical and radar remote sensing. Remote Sens. Environ. 147, 43–55.
Wei, S., Ji, S., 2021. Graph convolutional networks for the automated production of building vector maps from aerial images. IEEE Trans. Geosci. Remote Sens.
Weng, L., Xu, Y., Xia, M., Zhang, Y., Liu, J., Xu, Y., 2020. Water areas segmentation from remote sensing images using a separable residual segnet network. ISPRS Int. J. Geo-Inform. 9, 256.
Wu, W., Li, Q., Zhang, Y., Du, X., Wang, H., 2018. Two-step urban water index (tsuwi): A new technique for high-resolution mapping of urban surface water. Remote Sens. 10, 1704.
Xie, C., Huang, X., Zeng, W., Fang, X., 2016. A novel water index for urban high-resolution eight-band worldview-2 imagery. Int. J. Digital Earth 9, 925–941.
Xie, C., Wang, J., Zhang, Z., Zhou, Y., Xie, L., Yuille, A., 2017a. Adversarial examples for semantic segmentation and object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1369–1378.
Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017b. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492–1500.
Xu, H., 2006. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.
Xu, T., Tan, Z., Yan, X., et al., 2010. Extraction techniques of urban water bodies based on object-oriented. Geospatial Inform. 8, 64–66.
Xu, Y., Du, B., Zhang, L., 2020. Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses. IEEE Trans. Geosci. Remote Sens. 59, 1604–1617.
Xu, Y., Wu, L., Xie, Z., Chen, Z., 2018. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sens. 10, 144.
Yan, L., Fan, B., Xiang, S., Pan, C., 2021. Cmt: Cross mean teacher unsupervised domain adaptation for vhr image semantic segmentation. In: IEEE Geosci. Remote Sens. Lett..
Yang, F., Guo, J., Tan, H., Wang, J., 2017. Automated extraction of urban water bodies from zy-3 multi-spectral imagery. Water 9, 144.
Yang, J., Zhu, Q., Lv, J., Guan, Q., 2021. Ucwater: Unsupervised content-adaptive water-body extraction framework for high-resolution satellite imagery. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE. pp. 2767–2770.
Yao, F., Wang, C., Dong, D., Luo, J., Shen, Z., Yang, K., 2015. High-resolution mapping of urban surface water using zy-3 multi-spectral imagery. Remote Sens. 7, 12336–12355.
Yao, X., Han, J., Cheng, G., Qian, X., Guo, L., 2016. Semantic annotation of high-resolution satellite images via weakly supervised learning. IEEE Trans. Geosci. Remote Sens. 54, 3660–3671.
Yokoya, N., Ghamisi, P., Hänsch, R., Schmitt, M., 2020. 2020 ieee grss data fusion contest: Global land cover mapping with weak supervision [technical committees]. IEEE Geosci. Remote Sens. Magaz. 8, 154–157.
Yu, Y., Yao, Y., Guan, H., Li, D., Liu, Z., Wang, L., Yu, C., Xiao, S., Wang, W., Chang, L., 2021. A self-attention capsule feature pyramid network for water body extraction from remote sensing imagery. Int. J. Remote Sens. 42, 1801–1822.
Yuan, J., Deng, Z., Wang, S., Luo, Z., 2020a. Multi receptive field network for semantic segmentation. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 1883–1892.
Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., Fang, H., 2021. Deep-learning-based multispectral satellite image segmentation for water body detection. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 7422–7434.
Yuan, X., He, P., Zhu, Q., Li, X., 2019. Adversarial examples: Attacks and defenses for deep learning. IEEE Trans. Neural Networks Learn. Syst. 30, 2805–2824.
Yuan, Y., Chen, X., Wang, J., 2020b. Object-contextual representations for semantic segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16, Springer. pp. 173–190.
Zeng, C., Bird, S., Luce, J.J., Wang, J., 2015. A natural-rule-based-connection (nrbc) method for river network extraction from high-resolution imagery. Remote Sens. 7, 14055–14078.
Zhang, G., Yao, T., Xie, H., Wang, W., Yang, W., 2015. An inventory of glacial lakes in the third pole region and their changes in response to global warming. Global Planet. Change 131, 148–157.
Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2018a. mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations.
Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., et al., 2020a. Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955.
Zhang, J., Xing, M., Sun, G.C., Chen, J., Li, M., Hu, Y., Bao, Z., 2020b. Water body detection in high-resolution sar images with cascaded fully-convolutional network and variable focal loss. IEEE Trans. Geosci. Remote Sens. 59, 316–332.
Zhang, Q., Cong, R., Li, C., Cheng, M.M., Fang, Y., Cao, X., Zhao, Y., Kwong, S., 2020c. Dense attention fluid network for salient object detection in optical remote sensing images. IEEE Trans. Image Process. 30, 1305–1317.
Zhang, Q., Yuan, Q., Li, J., Li, Z., Shen, H., Zhang, L., 2020d. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. ISPRS J. Photogram. Remote Sens. 162, 148–160.
Zhang, Y., Li, W., Gong, W., Wang, Z., Sun, J., 2020e. An improved boundary-aware perceptual loss for building extraction from vhr images. Remote Sens. 12, 1195.
Zhang, Y., Liu, X., Zhang, Y., Ling, X., Huang, X., 2018b. Automatic and unsupervised water body extraction based on spectral-spatial features using gf-1 satellite imagery. IEEE Geosci. Remote Sens. Lett. 16, 927–931.
Zhang, Y., Liu, Z., Zhang, J., Yan, H., 2010. Water extraction from high resolution satellite image based on the fast matching level set method. In: International Conference on Geo-spatial Solutions for Emergency Management, Citeseer. p. C4.
Zhang, Z., Lu, M., Ji, S., Yu, H., Nie, C., 2021. Rich cnn features for water-body segmentation from very high resolution aerial and satellite imagery. Remote Sens. 13, 1912.
Zhao, C., 2020. A survey on image style transfer approaches using deep learning, in: Journal of Physics: Conference Series, IOP Publishing. p. 012129.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890.
Zhao, K., Kang, J., Jung, J., Sohn, G., 2018. Building extraction from satellite images using mask r-cnn with building boundary regularization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 247–251.

Zheng, Q., Qiao, X., Cao, Y., Lau, R.W., 2019. Distraction-aware shadow detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5167–5176.
Zhou, M., Niu, Z., Wang, L., Zhang, Q., Hua, G., 2020. Adversarial ranking attack and defense. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16, Springer. pp. 781–799.
Zhou, Y., Luo, J., Shen, Z., Cheng, X., Hu, X., 2012. Adaptive extraction of water in urban areas based on local iteration using high-resolution multi-spectral image. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, Ieee. pp. 6024–6027.
Zhu, F., Zhu, Y., Zhang, L., Wu, C., Fu, Y., Li, M., 2021. A unified efficient pyramid transformer for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2667–2677.
Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.