

A High-Resolution Multi-Temporal Remote Sensing Dataset for Levee-like Feature Segmentation in Arid Regions

Osman Ilniyaz ^{3,4}, Qingwu Hu ⁵, Hao Lu ^{1,2} and Kaisar Ahmat ^{1,2,*}

¹ Laboratory of Energy Carbon Neutrality, School of Electrical Engineering, Xinjiang University, Urumqi 830047, China; luhao@xju.edu.cn

² Engineering Research Center of Northwest Energy Carbon Neutrality, Ministry of Education, Xinjiang University, Urumqi 830047, China

³ Turpan Academy of Museology, Turpan 838000, China; osmanjan99@163.com

⁴ Institute of Archaeology, Academia Turfanica, Turpan 838000, China

⁵ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China; huqw@whu.edu.cn

* Correspondence: kaiser@xju.edu.cn; Tel.: +86-156-2383-7233

Abstract

Levee-like features are critical for water regulation in arid regions, but their automated extraction from remote sensing imagery remains challenging due to the scarcity of high-resolution labeled datasets. This data descriptor introduces a high-resolution remote sensing image dataset for semantic segmentation of levee-like features. The dataset covers 11 regions across Xinjiang and Gansu Province in northwestern China. It includes 459 single-phase base images with a spatial resolution of 0.50 m, as well as multi-temporal images of the same regions captured at different times. All annotations were manually drawn in polygon mode using the LabelMe tool and converted into YOLO format label files. The dataset adopts a strict strategy to prevent data leakage: first, training, validation and test sets are divided based on single-phase images, and then multi-temporal images are allocated to the corresponding data subsets according to their spatial locations. The dataset has been publicly released on the ScienceDB platform under the CC BY 4.0 license. YOLO and U-Net segmentation experiments on the test set achieved promising results, demonstrating its usability for levee-like feature segmentation. This dataset can provide fundamental data support for research on levee-like feature extraction, remote sensing change detection, and cross-region model transfer learning.

Dataset: <https://doi.org/10.57760/sciencedb.34604>.

Dataset License: CC-BY 4.0

Keywords: levee; semantic segmentation; YOLO; deep learning; data leakage; multi-temporal dataset



Academic Editor: Tianwen Zhang

Received: 28 April 2026

Revised: 11 June 2026

Accepted: 12 June 2026

Published: 16 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Summary

Arid zones cover about 41% of the world's land area and support the livelihood and production activities of more than 2 billion people [1]. In China, arid areas are mainly distributed in northern parts such as Xinjiang, Gansu, western parts of Inner Mongolia etc. [2], among which the southern and eastern Xinjiang region and western parts of Gansu Province are typical arid areas. These areas have low precipitation, high evaporation, and extreme water scarcity, and agricultural production and human survival are highly

dependent on artificial irrigation systems [3,4]. An artificial levee is a linear structure built by humans, which is mainly used for flood control, water storage, and irrigation, protecting the residents and providing agricultural and industrial production within the dike by intercepting water flow [5,6]. They are usually built by piling soil, sand, or rocks on a cleared, level surface. Levee-like features, including artificial irrigation canal embankments, silt dams, reservoir embankments, natural riverbank uplifts, and ancient water conservancy sites, are the core components of water resource regulation and conservancy facilities in arid areas. Accurate identification and monitoring of the spatial distribution, morphological characteristics and dynamic changes of levee-like features have important scientific value and practical significance for water resource management, agricultural production security, disaster prevention and mitigation, and cultural heritage protection in arid areas.

However, the identification and monitoring of levee-like features in arid areas face numerous challenges. The arid regions are vast, and traditional ground surveys are inefficient and costly, making it difficult to support large-scale periodic monitoring. Furthermore, they are slender in shape, scattered in distribution, and spectrally similar to surrounding features, making them easily confused with natural gullies, roads, and other elements in remote sensing images, which increases the difficulty of automatic identification [7]. In addition, some hydraulic sites have been subjected to wind and sand erosion, their shapes damaged or even buried, further increasing the challenges of remote sensing identification [8].

In recent years, with the rapid development of deep learning technology in the field of remote sensing image processing, object detection [9,10] and semantic segmentation [11,12] methods based on convolutional neural networks have made significant progress in tasks such as feature recognition [13] and change detection [14]. The foundation of modern remote sensing image segmentation was laid by fully convolutional networks (FCNs) [15], which enabled end-to-end pixel-wise prediction. Subsequently, encoder–decoder architectures such as U-Net [16] and DeepLab series [17,18] became widely adopted for various remote sensing segmentation tasks. These advances have spurred a number of studies on levee detection and monitoring using various remote sensing data sources. Some researchers have leveraged medium-resolution satellite data and DEMs for large-scale levee mapping; for example, Khanh et al. (2025) used 10 m LiDAR DEMs to detect river levees across five countries, identifying approximately 49,000 km of levees [19], while Sasaki et al. (2023) developed a DEM-based morphological parameter approach for levee detection in Japan [20]. Others have focused on deformation monitoring using InSAR techniques, such as Ghorbani et al. (2025) with Sentinel-1 data [21] and Ghosh et al. (2026) with Q-PS InSAR [22], the latter also noting the resolution limitations for detecting localized defects. A third group has pursued high-accuracy extraction using advanced models or fine-resolution data, such as Ai et al. (2025) who proposed an attention-enhanced UNet for paddy field levee extraction from Gaofen-1 imagery [23]. Lee et al. (2023) developed a mobile LiDAR point cloud filtering algorithm achieving 95.9% classification accuracy [24], while Su et al. (2024) designed a UAV-borne multi-sensor system for levee hazard detection [25]. Collectively, these studies demonstrate the growing potential of remote sensing and deep learning for levee-related tasks.

Despite these advances, a critical limitation pervades most existing studies: they rely on data with spatial resolutions that are too coarse to accurately capture the fine morphology of levee-like features. The majority of the above-cited works use data at 10 m or coarser resolutions (e.g., Sentinel, Gaofen, Landsat), which are fundamentally insufficient for delineating slender structures that are often only a few meters wide, especially those that are partially buried or have indistinct boundaries after centuries of wind and sand erosion. Even the studies that achieve high accuracy using LiDAR or UAV data [24,25] are limited in spatial coverage and not easily scalable across vast arid regions. What is urgently

needed is a high-resolution, large-coverage, publicly available dataset with pixel-level annotations specifically designed for levee-like features in arid areas, yet such a dataset has remained absent.

This absence is part of a broader gap in the remote sensing data landscape. Existing public segmentation datasets, such as SpaceNet [26] and xBD [27], focus on urban features (buildings, roads) and disaster damage assessment, respectively, but are limited to urban scenes. DeepGlobe [28] and LoveDA [29] provide land cover annotations, yet neither covers arid regions nor includes levee-like features (such as artificial irrigation canal embankments, silt dams, ancient water conservancy sites, etc.). Moreover, most existing datasets contain only single-phase imagery. Given that deep learning methods are data-driven and their performance is highly dependent on large-scale, high-quality labeled datasets [30], this data gap has become a major bottleneck for advancing levee-like feature extraction and monitoring in arid regions.

To fill the above data gaps, a high-resolution remote sensing dataset for segmentation of levee-like features was constructed. The dataset covers a total of 11 locations in Xinjiang and Gansu Province of China, all of which are typical arid or extreme arid areas. The spatial resolution of remote sensing images is 0.5 m, including 459 single-phase base images and their corresponding annotations, as well as multi-temporal images of the same area (totaling 1550 images). All annotations were drawn in polygon mode using the LabelMe tool and converted into YOLO format label files (.txt) that can be used directly for deep learning model training. In terms of data construction methodology, this study adopts a strict data leakage prevention strategy: first, the training set (about 80%), validation set (about 10%), and test set (about 10%) are divided based on the single-phase base images, and then the multi-temporal images are assigned to the corresponding data subsets according to their spatial location attribution. This division ensures that different temporal images in the same geographic region are not scattered between training and test sets, thus avoiding performance overestimation caused by spatial autocorrelation. Based on this dataset, we conducted segmentation baseline experiments using YOLO and U-Net models, where both achieved competitive Precision on the test set, demonstrating the feasibility of using the dataset for segmentation tasks.

This dataset is publicly released on the ScienceDB platform (DOI: 10.57760/sciencedb.34604) under the CC BY 4.0 license. To the best of our knowledge, this is the first publicly available dataset that combines (1) 0.5 m high-resolution imagery, (2) pixel-level annotations of levee-like features, and (3) coverage of arid regions in northwestern China with multi-temporal imagery.

Beyond segmentation, the multi-temporal nature of this dataset supports change detection (e.g., levee erosion or vegetation encroachment), heritage monitoring (e.g., ancient water conservancy sites), irrigation management, and cross-region transfer learning. We propose three benchmark tasks for future research: semantic segmentation, binary change detection using multi-temporal pairs, and domain adaptation across different regions.

2. Data Description

In this dataset, levee-like features are defined as elongated raised structures extending in straight or curved lines, with a width-to-length ratio typically less than 0.2, obvious shadow or texture differences compared to surrounding terrain, and significant relative elevation. Specific types include artificial embankments, silt dams, ancient irrigation canal embankments, and man-made elevations. During annotation, boundaries are traced along clearly visible lines between the feature and the ground. For areas partly blurred by vegetation or shadows, annotators apply reasonable inference based on contextual continuity. To distinguish levee-like features from visually similar landforms, annotators

use the following criteria: roads typically have uniform texture and lack an elevation shadow; natural ravines are depressions (not raised) with dendritic drainage patterns. For ambiguous or partially buried features, continuity with well-preserved levee segments serves as a key cue. All areas not belonging to levee-like features (bare soil, Gobi, water bodies, desert vegetation, roads, buildings, etc.) are treated as background. All annotations were reviewed by a senior expert to ensure consistency.

2.1. Study Area and Image Data Acquisition

This dataset covers the southern to eastern Xinjiang, and Jiuquan City of Gansu Province, China, totaling 11 locations in 6 districts (Figure 1). These areas are all typical arid regions, with surface coverage mainly consisting of bare soil, Gobi, and desert vegetation, while levee-like features (such as artificial channel embankments, silt dams, and ancient water conservancy sites) are distributed across different geomorphic units. The distribution of training image data in the study area is given in Table 1.

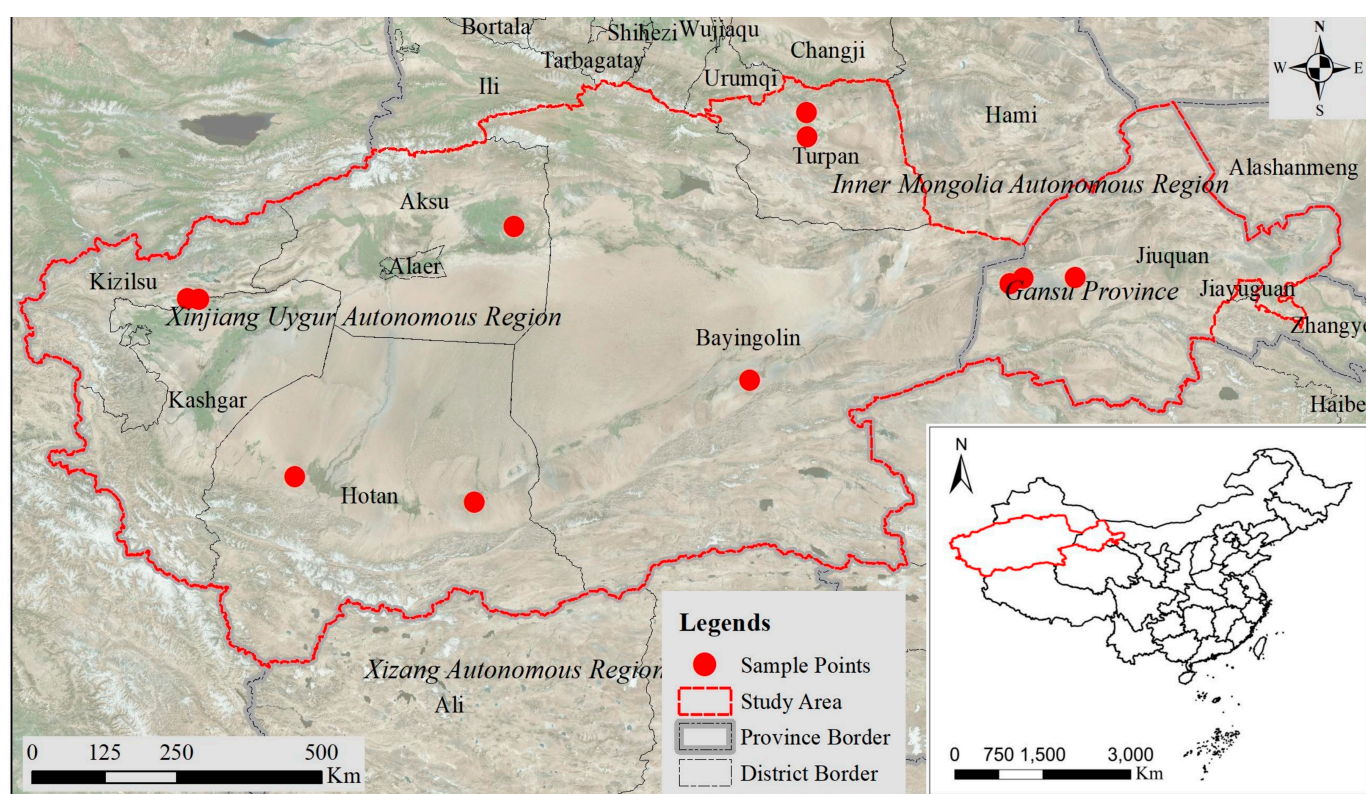


Figure 1. Location of study area and distribution of sample points.

Table 1. Distribution of images across study areas (derived from high-resolution satellite imagery via Google Earth).

No.	District	Number of			Captured Dates of Images
		Locations	Base Images	All Images	
1	Kashgar	2	121	542	2019, 2021, 2010, 2011, 2024, 2025
2	Aksu	1	168	343	2003, 2013, 2024
3	Hotan	2	36	150	2013, 2021, 2011, 2016, 2019, 2020, 2025
4	Bayingolin	1	21	109	2014, 2010, 2011, 2012, 2013, 2025
5	Turpan	2	18	62	2020, 2019, 2013, 2018, 2019, 2025
6	Jiuquan	3	75	344	2006, 2010, 2012, 2013, 2014, 2015, 2018, 2019, 2020, 2021, 2025
	Total	11	459	1550	

The selection of sampling regions followed three criteria: (1) the region should cover primarily bare soil landscapes typical of arid areas, including oasis edges, river terraces, piedmont alluvial fans, and deserts; (2) ensure that levee-like features are distributed across different landscape backgrounds, including man-made irrigation canal embankments and water conservancy sites; and (3) meet the fine annotation requirements of 0.5 m spatial resolution. The remote sensing images are sourced from high-resolution satellite imagery with an original spatial resolution of about 0.46 m, and all images were resampled to a uniform 0.5 m resolution. For each location, two types of images were collected as described in Table 2.

Table 2. Description and naming methods of single and multi-temporal images.

Type	Description	Naming Rules and Example
Single-phase base images	For each location, select one reference image as the initial annotation object	[region][no.][id].png Dunhuang1_1.png
Multi-temporal images	Images collected at different times for the same location, used for temporal analysis and annotation correction	[region][no.][id]([acquisition_date(YYYYMMDD)]).png Dunhuang1_1(20120907).png

2.2. Dataset Structure

The folder structure follows the standard YOLO format for segmentation tasks, allowing direct use with YOLOv8-seg and similar models. It is shown in Figure 2. The root “Levee Dataset” directory contains “images” and “labels” folders, “data.yaml”, “metadata.csv”, “license”, “readme”, “labelme_to_yolo.ipynb” and “yolo_to_unet.ipynb” files.

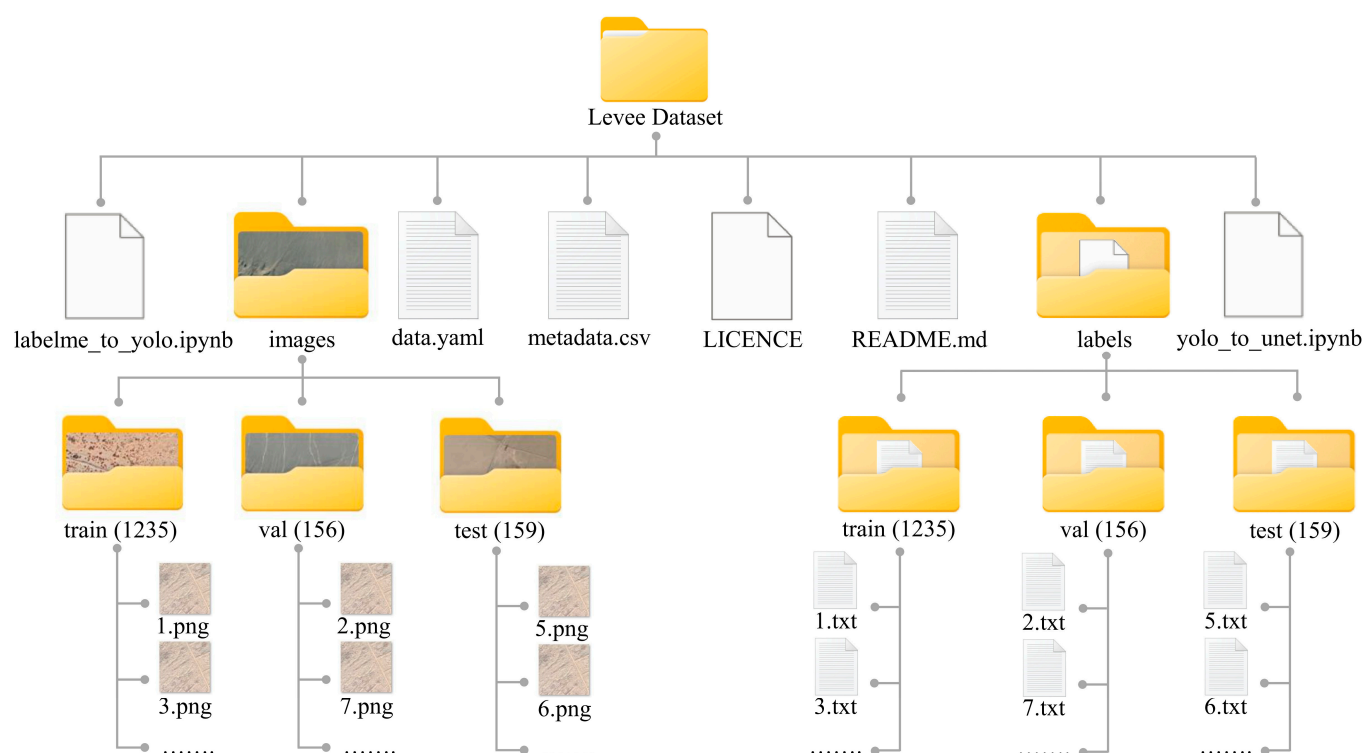


Figure 2. Folder structure of the dataset. The number of images/labels in each subset is indicated in parentheses.

The “data.yaml” file records the number of classes, class names and file directories of training, validation and test folders. The “images” folder contains the subfolders with training, validation and test image data, while the “labels” folder contains the subfolder

with corresponding training, validation and test annotation data. The metadata file records the property of each file in the images folder and its corresponding annotation files in the labels folder, including the file ID, file names, and path of images and annotations, locations, acquisition date, base image ID, image width, height, pixel size (in meter), etc. The License file records the copyright information of the dataset. The readme file describes the dataset including the dataset title with the version number, authors, description of data, data file structure, methodologies, file format, and citation and contact information. The “labelme_to_yolo.ipynb” and “yolo_to_unet.ipynb” files are the scripts used to convert the annotations from labelme JSON format to YOLO txt format and from YOLO txt format to U-Net format, respectively.

2.3. Data Split

This dataset uses a spatial location-based data split strategy. First, the original 459 single-phase base images were divided into a training set (about 80%), a validation set (about 10%), and a test set (about 10%). Then multi-temporal images are assigned to the same subset as their corresponding base images based on their spatial location. This strategy ensures that images of different times from the same geographic area do not appear in different subsets, avoiding performance overestimation caused by spatial autocorrelation.

2.4. Annotation Format

The annotation files of this dataset used the YOLO segmentation format, with the file extension .txt, and have the same name as the corresponding image files. In the txt format annotation files, each line represents a levee-like feature, with the first number being the class identifier (uniformly 0 in this dataset), followed by pairs of numbers, each representing the normalized coordinates of a polygon vertex (x and y values ranging from 0 to 1, divided by the image width and height, respectively). For example, a line with the content “0 0.00198 0.29468 0.34112 0.13091 0.36740 0.18530 0.0 0.34566” represents a quadrilateral polygon with four vertices, belonging to the levee-like feature class.

The original annotations of this dataset were made using the LabelMe tool in polygon mode and saved in JSON format. A custom script was then used to convert the polygon coordinates in the JSON files into YOLO segmentation format TXT files for direct use in models that support segmentation.

2.5. Data Availability

The dataset has been publicly released on the ScienceDB platform (DOI: 10.57760/sciencedb.34604) under the CC BY 4.0 license. Users can access and download it via the following link: <https://doi.org/10.57760/sciencedb.34604>.

The list of files included in the ScienceDB archive are given in Table 3. The annotations (label files, metadata, and scripts) are released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, permitting unrestricted non-commercial use with appropriate attribution. The source imagery was accessed via Google Earth, which permits non-commercial research use. Redistribution of the imagery in this dataset is permitted for non-commercial research purposes. Any commercial use of the dataset would require separate licensing from the original imagery providers. Users must provide attribution to Google Earth and the original imagery providers when publishing results. Redistribution of the raw imagery outside of this dataset package is not permitted.

The accompanying metadata file (metadata.csv) contains detailed attributes for each image (region, acquisition date, image and label paths, number of annotations, etc.), allowing users to filter and customize subsets of the data (Table 4). Due to cropping and conversion to PNG format, the released dataset images do not retain embedded georefer-

encing information. However, all images share a uniform spatial resolution (0.50 m) and projection (WGS_1984_UTM_Zone_45N) as described in Section 3.1.

Table 3. Files included in the ScienceDB archive.

File/Directory	Description	Size
images/	Contains all images (train/val/test subfolders)	425.10 MB
labels/	Contains all YOLO format label files (train/val/test subfolders)	1.97 MB
data.yaml	Dataset configuration file for YOLO	4.00 KB
metadata.csv	Metadata for all images	287.28 KB
README.md	Dataset description and usage instructions	2.94 KB
LICENSE	License information (CC BY 4.0)	0.35 KB
labelme_to_yolo.ipynb	Annotation conversion script	11.60 KB
yolo_to_unet.ipynb	Annotation conversion script	17.70 KB

Table 4. Metadata schema for the metadata.csv file.

Column Name	Description	Example
file_id	Unique identifier for each image	1, 2, 3, ...
split	Dataset split assignment (train/val/test)	train, val, test
image_filename	Original image file name	Dunhuang1_1.png
label_filename	Corresponding label file name	Dunhuang1_1.txt
image_path	Relative path to image file	images/train/Dunhuang1_1.png
label_path	Relative path to label file	labels/train/Dunhuang1_1.txt
region	Geographic region name	Dunhuang
acquisition_date	Image acquisition date (YYYYMMDD)	20120907
base_image_id	For multi-temporal images, links to corresponding single-phase base image	Dunhuang
image_width	Image width in pixels	513
image_height	Image height in pixels	513
pixel_size_m	Spatial resolution in meters	0.50
num_objects	Number of annotated levee-like features in the image	3
classes	Class identifiers present in the image	0
annotation_source	Annotation method	labelme_converted_to_yolo

The dataset contains 4792 annotated levee-like features. Due to the elongated nature of these features, object sizes vary considerably depending on the length of the levee segment captured in each image. The dataset contains a single class (levee-like features with a class ID of 0).

3. Methods

3.1. Image Acquisition and Preprocessing

The remote sensing images used in this dataset were sourced from high-resolution satellite imagery available through the Google Earth service, which integrates data from multiple commercial providers, including Maxar (e.g., WorldView series, GeoEye) and CNES/Airbus (e.g., SPOT series, Pleiades). The imagery is provided as RGB (three-band) files. Original multispectral bands and radiometric resolution information are not retained, as Google Earth only distributes RGB composites. The original spatial resolution is about 0.46 m. All images were obtained already orthorectified by the original data providers. The images cover the southern and eastern regions of Xinjiang and the western region of Gansu Province, China, comprising a total of 11 locations (Figure 1). All locations are situated in typical arid or extremely arid regions.

The image selection followed these criteria: cloud cover below 5% to ensure clear visibility of features, priority given to images covering various seasons to capture seasonal variation of features, and images free from obvious geometric distortion or striping noise.

All images undergo the following preprocessing steps before annotation:

(1) Projection: Each image was projected to the WGS_1984_UTM_Zone_45N coordinate system to ensure spatial consistency across all samples.

(2) Resampling: Although the original spatial resolution of the source imagery varies by sensor (approximately 0.46 m), all images were uniformly resampled to a standard spatial resolution of 0.50 m using the Nearest Neighbor (NEAREST) method. This resampling algorithm was chosen because it preserves the original pixel values without introducing new spectral information, which is critical for maintaining the integrity of the land cover features.

(3) Cropping: Images were cropped according to the distribution range of levee-like features to remove irrelevant areas and reduce data volume.

(4) Format conversion: The images, originally in TIFF format, were converted to PNG format to facilitate reading by common annotation tools (LabelMe) and deep learning libraries.

For each location, two types of images were collected: (1) single-phase base images as the primary annotation object, and (2) multi-temporal images of the same area at different time periods.

3.2. Annotation Process

This dataset was annotated using the LabelMe tool (MIT open-source license) in polygon mode, which allows precise outlining of the irregular boundaries of levee-like features. The annotation work is divided in two stages. The first stage involves annotating 459 single-phase base images and saving them in LabelMe JSON format. The second stage involves annotating and correcting multi-temporal images (Figure 3). Annotators referenced the single-phase base image annotations and compared the land features in multi-temporal images with those in the baseline: if features have not changed, the original annotation is retained; if features' shapes have changed, the boundaries are adjusted; newly appeared features are annotated; and disappeared features are removed. All adjustments of multi-temporal images are finally reviewed by a senior expert. For partially obscured features, a simple rule was applied: if the obscured portion was short and continuity could be inferred from visible neighboring segments, the feature was annotated; if the obscured portion was long and boundaries could not be confidently determined, the feature was excluded. This ensures label quality while maintaining reasonable completeness. After annotation, a custom Python script is used to convert the LabelMe JSON format into YOLO segmentation TXT files. During conversion, the polygon vertex coordinates in the JSON are first read, normalized, and then written into the TXT file, with each line formatted as "class_ID x1 y1 x2 y2 . . . xn yn", and the class ID is uniformly set to 0.

3.3. Data Splitting Strategy

To prevent data leakage, this dataset uses a spatial location-based split strategy. First, the 459 single-phase base images were randomly divided into training (80%), validation (10%), and test (10%) sets using random shuffling (without a fixed random seed). However, the resulting train, validation and test split is fixed in the released dataset. All users who download the dataset from ScienceDB will receive the same split used in our experiments. All multi-temporal images (1091 images) were then assigned to the same subset as their corresponding base images based on spatial location. The final subset sizes are summarized in Table 5.

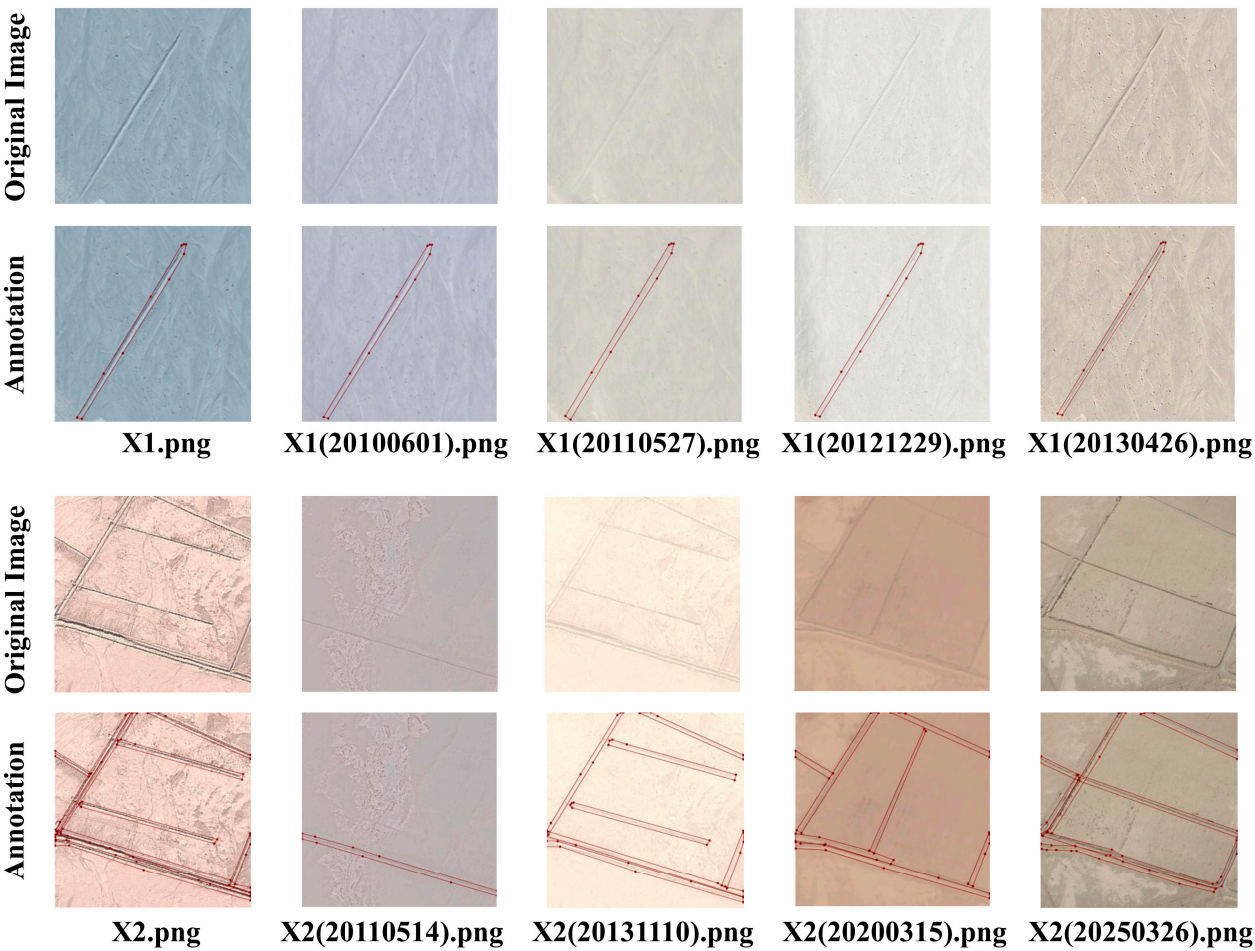


Figure 3. Example of image annotation for base (first column) and multi-temporal images (second to fifth columns). Temporal changes visible in this example include variations in shadow intensity, differences in levee boundary clarity, and the addition or loss of levee segments over time.

Table 5. Final subset sizes of base images and multi-temporal images.

Subset	Base Images	Multi-Temporal Images	Total
Training	367	868	1235
Validation	45	111	156
Test	47	112	159
Total	459	1091	1550

We confirm that there is no spatial overlap between subsets. Because the split was performed at the location level (all images from the same geographic area are assigned to the same subset), images from the same physical location never appear in different subsets. This eliminates spatial autocorrelation and prevents data leakage that could otherwise lead to overestimated model performance. For example, as shown in Figure 4a, if multi-temporal images from the same geographic location were split across training, validation and test sets, it will cause spatial autocorrelation leakage. The model may learn location-specific features rather than generalizable levee morphology. Figure 4b showed the correct approach, where all images from the same geographic location are assigned to the same subset. This ensures that the test set only contains images from locations not seen during training, providing a realistic evaluation of model generalization. Table 6 shows the distribution of images by region across the three subsets.

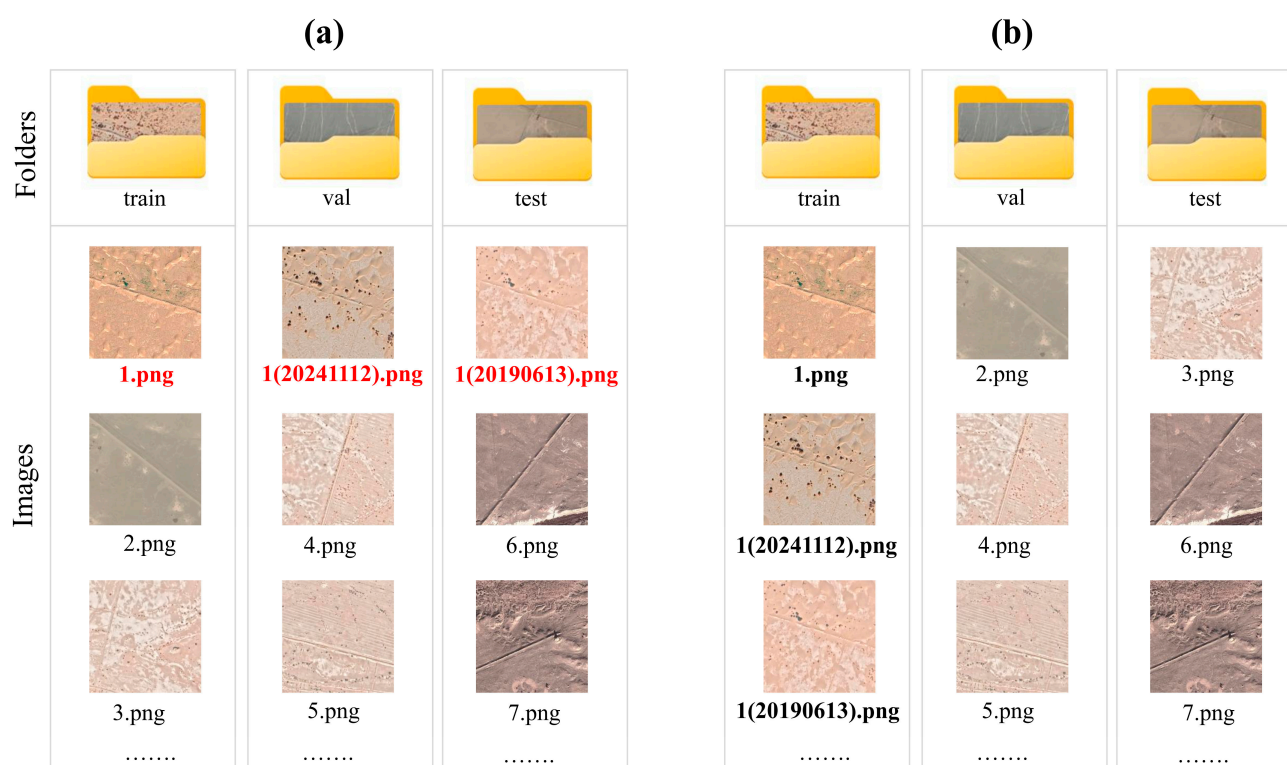


Figure 4. Schematic illustration of data leakage prevention strategy. (a) Incorrect: same location across training and test sets. (b) Correct: each location assigned to a single subset.

Table 6. Distribution of images by region across training, validation, and test subsets.

Region	Training	Validation	Test	Total
Kashgar	444	63	35	542
Aksu	264	35	44	343
Hotan	115	9	26	150
Bayingolin	75	22	12	109
Turpan	62	0	0	62
Jiuquan	275	27	42	344
Total	1235	156	159	1550

3.4. Quality Control

To ensure the accuracy and consistency of the labeled data, multiple rounds of manual inspection and correction were conducted. After all annotations were completed using LabelMe, each image was reviewed by a second annotator to check for missing annotations, incorrect boundaries, or annotation errors. Any issues found were corrected promptly. For multi-temporal images, annotators compared them against the corresponding single-phase base images to ensure that unchanged features retained consistent annotations, while newly appeared or disappeared features were properly updated. Quality control was performed by a senior expert who reviewed all of the annotations. Formal inter-annotator agreement statistics were not calculated, as the process was designed as sequential verification rather than parallel independent annotation.

3.5. Model Evaluation

To validate the annotation quality and usability of this dataset, we conducted segmentation experiments using the YOLOv8s-seg and U-Net models. U-Net is a typical fully convolutional neural network (FCN) with an encoder–decoder structure, specifically designed for image segmentation tasks. It is named for its symmetrical “U”-shaped

architecture and skip connections and was first proposed in 2015 for biomedical image segmentation [16]. YOLOv8 achieves a favorable balance between architectural design and training efficiency and has become one of the mainstream benchmark models in object detection and segmentation tasks [31].

For the YOLO model, the standard YOLOv8s-seg model was downloaded from the website of Ultralytics (www.ultralytics.com, version 8.4.56). For the U-Net baseline, we adopted the standard architecture with four encoder and four decoder blocks, each containing two 3×3 convolutional layers. Skip connections were used to preserve spatial details. The model was trained using a combination of BCE (Binary Cross-Entropy) and Dice loss. The division of the training, validation, and test sets follows the spatial location-based split strategy described in Section 3.3. The training hyperparameters for experiments are given in Table 7.

Table 7. Training hyperparameters for baseline experiments.

Parameter	YOLOv8s-Seg	U-Net
Input size	512 × 512 pixels	512 × 512 pixels
Optimizer	SGD	Adam
Initial learning rate	0.01	0.0001
Learning rate scheduler	Cosine annealing	ReduceLROnPlateau (factor = 0.5, patience = 10)
Batch size	16	4
Epochs	1000	200
Weight decay	0.0005	0.0001
Momentum	0.937	-
Data augmentation	Mosaic, flip, mixup, scale, HSV	Mosaic, flip, mixup, scale, HSV
Confidence/IoU threshold	Conf = 0.5, IoU = 0.5 (mAP@0.5)	Segmentation threshold = 0.5
Hardware	NVIDIA RTX 4060 Laptop GPU (8 GB)	NVIDIA RTX 4060 Laptop GPU (8 GB)
Dependencies	Python 3.10.20, PyTorch 2.5.0 + cu118, OpenCV 4.13.0	Python 3.10.20, PyTorch 2.5.0 + cu118, OpenCV 4.13.0

For YOLOv8s-seg, instance-level metrics were reported such as Precision, Recall, mAP@0.5 (IoU threshold = 0.5), and mAP@0.5:0.95 (averaged over IoU thresholds from 0.5 to 0.95 with step 0.05). For U-Net, pixel-level metrics were reported such as Precision, Recall, F1-score, and IoU, computed directly from the binary segmentation masks using a threshold of 0.5. Training was conducted separately on the single-phase base dataset and the multi-temporal dataset.

The evaluation results for YOLOv8s-seg and U-Net are summarized in Table 8 and Table 9, respectively. As shown in Table 8, YOLOv8s-seg achieved robust performance on both the base dataset (459 images) and the full multi-temporal dataset (1550 images). On the base dataset, the model attained a test Precision of 0.800, Recall of 0.708, mAP@0.5 of 0.750, and F1-score of 0.751. When evaluated on the full multi-temporal dataset, the performance remained competitive, with a test Precision of 0.804, mAP@0.5 of 0.710, and F1-score of 0.698. The slightly lower performance on the multi-temporal dataset is expected, as the increased data diversity and temporal variations introduce additional challenges for the model.

Table 9 presents the results for the U-Net baseline. On the base dataset, U-Net achieved strong performance with a test Precision of 0.837, Recall of 0.853, IoU of 0.735, and F1-score of 0.844. These results indicate that U-Net is highly effective at capturing levee-like features, particularly in terms of Recall (0.853), suggesting a low rate of false negatives. On the multi-temporal dataset, U-Net maintained solid performance with a test Precision of 0.757,

Recall of 0.774, IoU of 0.622, and F1-score of 0.759. The decrease in IoU from 0.735 to 0.622 reflects the increased complexity of the multi-temporal data.

Table 8. Evaluation results of Yolov8 model.

Dataset	Evaluation	P	R	mAP50	mAP50–95	F1-Score
Base (459)	validation	0.809	0.686	0.752	0.400	0.742
	test	0.800	0.708	0.750	0.411	0.751
Multi-temporal (1550)	validation	0.808	0.677	0.747	0.416	0.737
	test	0.804	0.617	0.710	0.384	0.698

Table 9. Evaluation results for U-Net model.

Dataset	Evaluation	P	R	IoU	F1-Score
Base (459)	validation	0.862	0.765	0.679	0.805
	test	0.837	0.853	0.735	0.844
Multi-temporal (1550)	validation	0.788	0.777	0.638	0.775
	test	0.757	0.774	0.622	0.759

When comparing the two models on the multi-temporal test set, U-Net achieved a higher F1-score (0.759) than YOLOv8s-seg (0.698), while YOLOv8s-seg achieved a higher Precision (0.804 vs. 0.757). This indicates that YOLOv8s-seg is more conservative in its predictions (fewer false positives), whereas U-Net is more sensitive in detecting levee-like features (fewer false negatives). The complementary strengths of the two architectures suggest that ensemble or fusion approaches could potentially yield further improvements.

The competitive performance of both models on the base and multi-temporal datasets suggests that the dataset is suitable for deep learning-based segmentation. As expected, both models showed slightly lower performance on the multi-temporal dataset compared to the single-phase base dataset. This is attributable to the greater variability in image acquisition conditions, seasonal changes, and potential land cover changes captured in the multi-temporal imagery.

The observed differences between YOLOv8s-seg and U-Net reflect their architectural differences. YOLOv8s-seg, as an instance segmentation model, is designed to distinguish individual levee instances, which may lead to more precise boundary localization but potentially higher false negatives when features are fragmented or partially buried. U-Net, as a semantic segmentation model, performs pixel-wise classification without distinguishing individual instances, which may result in higher Recall but lower Precision when features are ambiguous. These complementary characteristics suggest that the dataset is suitable for evaluating both types of segmentation approaches.

3.6. Limitations

Several limitations of this dataset should be noted. First, only levee-like features visible on bare soil were annotated; features obscured by vegetation, shadows, or sand cover were excluded, potentially introducing omission bias. This exclusion may limit model generalizability to areas where such obscurations are common. Users should be aware of this bias and consider mitigation strategies, such as collecting additional samples from obscured areas or using occlusion-aware data augmentation. Second, the sampling regions are limited to Xinjiang and Gansu Province, China, and transferability to other arid regions requires further validation. Third, the temporal coverage is uneven across locations, which may hinder systematic change detection studies. Fourth, manual annotation inevitably involves subjective interpretation, particularly for degraded features. Finally, no explicit

negative samples (e.g., roads or natural ravines) are provided, limiting the evaluation of model discriminability. Furthermore, no formal inter-annotator agreement metrics (e.g., Cohen's kappa or IoU overlap) were calculated for this dataset. As described in Section 3.4, quality control was performed through sequential verification rather than parallel independent annotation. Users should be aware of this limitation when using the dataset for tasks requiring precise inter-annotator consistency benchmarks.

Accordingly, future work will focus on four main directions: (1) extending the dataset to other arid regions to improve geographic representativeness; (2) providing explicit negative samples and change annotations to support discriminative model training and supervised change detection; (3) exploring semi-supervised learning approaches to reduce annotation costs; and (4) organizing benchmarking challenges to establish standardized evaluation protocols. The dataset will be maintained and updated on the ScienceDB platform, and community contributions are welcome.

4. User Notes

This dataset was released on the ScienceDB platform in the form of a ZIP archive. After downloading and extracting, the complete folder structure can be obtained. The root directory of the dataset contains two main folders: `images/` and `labels/`, which store the images and the corresponding YOLO segmentation format annotation files, respectively. Each folder is further divided into three subsets: `train/`, `val/`, and `test/`. Users can directly use this directory structure to load the data without reorganizing it.

The annotation files of this dataset use the YOLO segmentation format (each line in the format `class_id x1 y1 x2 y2 ... xn yn`), suitable for YOLO series models that support segmentation, such as YOLOv8-seg, YOLOv9-seg, and YOLOv10-seg. For other semantic segmentation models (such as U-Net, DeepLabV3, etc.), users need to convert the polygon annotations to binary masks. The python script for converting the LabelMe JSON format to YOLO txt format and from YOLO txt format to U-Net format are given in the archive. Note that annotations were made only for levee-like features visible on bare soil surfaces; features obscured by vegetation or shadows were not annotated.

The `metadata.csv` file records detailed information for each image, including fields such as file identifier (`file_id`), subset (split), region name (`region`), acquisition date (`acquisition_date`), whether it is multi-temporal (`is_multi_temporal`), corresponding base image identifier (`base_image_id`), image dimensions (`image_width`, `image_height`), number of annotated objects (`num_objects`), and others. Users can filter the desired data subset based on these fields.

When using this dataset, please note the following points: (1) The naming of all images and annotation files follows the same rules, i.e., single-phase base images are named as "`region_identifier_number.png`" (e.g., `Dunhuang1_1.png`), and multi-temporal images append the acquisition date in parentheses after the base name (e.g., `Dunhuang1_1(20120907).png`). Annotation files have the same name as their corresponding image files with the extension `.txt`. (2) This dataset contains only one class (levee-like features, class ID 0), so users do not need to configure an additional class file. (3) Since features in arid regions exhibit visual variations in different seasons, it is recommended to pay attention to seasonal matching or use domain adaptation methods when applying across time series. (4) If using this dataset for model training, it is recommended to adopt the same data splitting strategy as in this work (i.e., splitting by spatial location) to ensure comparable results.

5. Conclusions

This data descriptor presents a high-resolution, multi-temporal remote sensing dataset for semantic segmentation of levee-like features in arid regions of northwestern China. The

dataset contains 459 single-phase base images with pixel-level annotations and 1091 multi-temporal images across 11 locations in Xinjiang and Gansu Province. Key contributions include (1) the first publicly available dataset specifically targeting levee-like features in arid regions; (2) a spatial location-based splitting strategy to prevent data leakage; and (3) multi-temporal coverage supporting change detection research.

Baseline experiments using YOLOv8s-seg and U-Net demonstrated the dataset's usability, achieving F1-scores of 0.698 and 0.759, respectively, on the multi-temporal test set. The dataset is publicly released on ScienceDB under the CC BY 4.0 license.

Future work includes (1) extending the dataset to other arid regions to improve geographic representativeness; (2) providing explicit negative samples and change annotations; (3) exploring semi-supervised learning to reduce annotation costs; and (4) organizing benchmark challenges to establish standardized evaluation protocols. The dataset will be maintained and updated on the ScienceDB platform, and community contributions are welcome.

Author Contributions: Conceptualization, O.I., Q.H. and K.A.; methodology, O.I. and Q.H.; software, O.I.; validation, Q.H. and H.L.; formal analysis, O.I., Q.H., H.L. and K.A.; investigation, O.I. and K.A.; resources, K.A.; data curation, O.I.; writing—original draft preparation, O.I.; writing—review and editing, K.A. and H.L.; visualization, O.I.; supervision, Q.H. and H.L.; project administration, K.A.; funding acquisition, K.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Xinjiang Tianchi Doctoral Project (CN), grant number 5105250180U, Innovation Program of the Electrical Engineering Postdoctoral Research Station, Xinjiang University, grant number XJU-DQGCBSHLDZ-2025005. The APC was funded by the Xinjiang Tianchi Doctoral Project (CN), grant number 5105250180U.

Data Availability Statement: The original data presented in the study are openly available in the ScienceDB platform at <https://doi.org/10.57760/sciencedb.34604>.

Acknowledgments: During the preparation of this manuscript, the authors used DeepSeek API (DeepSeek-V4-Flash) for the purposes of grammatical refinement and language polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
DEM	Digital Elevation Model
InSAR	Interferometric Synthetic Aperture Radar
IoU	Intersection over Union
LiDAR	Light Detection and Ranging
mAP	Mean Average Precision
RMSE	Root Mean Square Error
YOLO	You Only Look Once

References

1. Gaur, M.K.; Squires, V.R. (Eds.) Geographic Extent and Characteristics of the World's Arid Zones and Their Peoples. In *Climate Variability Impacts on Land Use and Livelihoods in Drylands*; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–20.
2. Wang, Y.; Yu, Q.; Avirmed, B.; Zhao, J.; Sun, W.; Liu, Y.; Cui, H.; Khishigjargal, M.; Oyuntuya, S.; Dambadarjaa, N.; et al. The response of ecosystem services to ecological spatial network patterns in China's arid and semi-arid regions. *Ecol. Indic.* **2025**, *172*, 113300. [[CrossRef](#)]

3. Wang, S.; Zhang, Q.; Yue, P.; Wang, J. Effects of evapotranspiration and precipitation on dryness/wetness changes in China. *Theor. Appl. Climatol.* **2020**, *142*, 1027–1038.
4. Liu, X.; Tang, Q.; Zhao, Y.; Wang, P. Persistent Water Scarcity Due To High Irrigation Demand in Arid China: A Case Study in the North Slope of the Tianshan Mountains. *Earth's Future* **2024**, *12*, e2024EF005070.
5. Tobin, G.A. The levee love affair: A stormy relationship? *JAWRA J. Am. Water Resour. Assoc.* **1995**, *31*, 359–367. [[CrossRef](#)]
6. Knox, R.L.; Wohl, E.E.; Morrison, R.R. Levees don't protect, they disconnect: A critical review of how artificial levees impact floodplain functions. *Sci. Total Environ.* **2022**, *837*, 155773.
7. de Mattos Teixeira, C.A.; Alves de Araujo, T.M.; Cardoso, E.; Costantin Filho, M.A.; Weyl Costa, J.; Lisboa Frances, C.R. Aerial image segmentation of embankment dams based on multispectral remote sensing: A case study in the Belo Monte Hydroelectric Complex, Pará, Brazil. *PeerJ Comput. Sci.* **2025**, *11*, e2917. [[CrossRef](#)]
8. Ali, L.; Tanaka, N. Experimental Investigation of Levee Erosion during Overflow and Infiltration with Varied Hydraulic Conductivities of Levee and Foundation Properties in Saturated Conditions. *GeoHazards* **2023**, *4*, 286–301. [[CrossRef](#)]
9. Gui, S.; Song, S.; Qin, R.; Tang, Y. Remote Sensing Object Detection in the Deep Learning Era—A Review. *Remote Sens.* **2024**, *16*, 327. [[CrossRef](#)]
10. Amjoud, A.B.; Amrouch, M. Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review. *IEEE Access* **2023**, *11*, 35479–35516. [[CrossRef](#)]
11. Yang, S.; Luo, L.; Li, Q.; Chen, Y.; Wu, L.; Wang, X. Auto-identification of linear archaeological traces of the Great Wall in northwest China using improved DeepLabv3+ from very high-resolution aerial imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *113*, 102995.
12. Hossain, M.D.; Chen, D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 115–134. [[CrossRef](#)]
13. He, L.-H.; Zhou, Y.-Z.; Liu, L.; Cao, W.; Ma, J.-H. Research on object detection and recognition in remote sensing images based on YOLOv11. *Sci. Rep.* **2025**, *15*, 14032. [[CrossRef](#)]
14. Liu, T.; Yang, L.; Lunga, D. Change detection using deep learning approach with object-based image analysis. *Remote Sens. Environ.* **2021**, *256*, 112308. [[CrossRef](#)]
15. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer International Publishing: Cham, Switzerland, 2015.
17. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
18. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of Computer Vision—ECCV 2018*; Springer International Publishing: Cham, Switzerland, 2018.
19. Khanh, D.N.; Tsumura, Y.; Sasaki, O.; Shiraishi, K.; Akimoto, D.; Yamazaki, D.; Zhao, G.; Hirabayashi, Y. Mapping the World's River Levees: A Hyper-Resolution Levee Database Based on Digital Elevation Models. *Geophys. Res. Lett.* **2025**, *52*, e2024GL114121. [[CrossRef](#)]
20. Sasaki, O.; Tsumura, Y.; Yamada, M.; Hirabayashi, Y. Automatic levee detection using a high-resolution DEM—Case study in Kinu river basin, Japan. *Hydrol. Res. Lett.* **2023**, *17*, 9–14. [[CrossRef](#)]
21. Zahra, G.; Ali, K.; Yasser, M.; Vroman Ethan, T. An InSAR-Based Remote Sensing Approach for Monitoring Levee Deformation? A Case Study of Lake Tholocco, Alabama. In *Geotechnical Frontiers 2025*; ASCE: Reston, VA, USA, 2025; pp. 84–93.
22. Debayan, G.; Puneet, B.; Vinaykrishnan, L.; Anand, P. Assessment of Satellite-Based Remote Sensing Methods for Levee Monitoring. In *Proceedings of the Geo-Congress 2026*, Salt Lake City, UT, USA, 9–12 March 2026.
23. Ai, H.; Zhu, X.; Han, Y.; Ma, S.; Wang, Y.; Ma, Y.; Qin, C.; Han, X.; Yang, Y.; Zhang, X. Extraction of Levees from Paddy Fields Based on the SE-CBAM UNet Model and Remote Sensing Images. *Remote Sens.* **2025**, *17*, 1871. [[CrossRef](#)]
24. Lee, J.; Yoo, S.; Kim, C.; Sohn, H.-G. Automatic levee surface extraction from mobile LiDAR data using directional equalization and projection clustering. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103143. [[CrossRef](#)]
25. Su, S.; Yan, L.; Xie, H.; Chen, C.; Zhang, X.; Gao, L.; Zhang, R. Multi-Level Hazard Detection Using a UAV-Mounted Multi-Sensor for Levee Inspection. *Drones* **2024**, *8*, 90. [[CrossRef](#)]
26. Etten, A.V.; Hogan, D.; Manso, J.M.; Shermeyer, J.; Weir, N.; Lewis, R. The Multi-Temporal Urban Development SpaceNet Dataset. In *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 20–25 June 2021.
27. Zhao, Z.; Wang, F.; Chen, S.; Wang, H.; Cheng, G. Deep object segmentation and classification networks for building damage detection using the xBD dataset. *Int. J. Digit. Earth* **2024**, *17*, 2302577. [[CrossRef](#)]

28. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018.
29. Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; Zhong, Y. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Virtual, 6–14 December 2021.
30. Dritsas, E.; Trigka, M. Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. *Remote Sens.* **2025**, *17*, 550. [[CrossRef](#)]
31. Varghese, R.; M., S. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.