Article

Intelligent Structural Health Monitoring and Noncontact Measurement Method of Small Reservoir Dams Using UAV Photogrammetry and Anomaly Detection

Sizeng Zhao [1], Fei Kang [2,*], Lina He [3], Junjie Li [1,2,*], Yiqing Si [4] and Yiping Xu [5]

[1] College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China; zhaosizeng@hhu.edu.cn
[2] School of Hydraulic Engineering, Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China
[3] School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China; hlnyh@hhu.edu.cn
[4] Chongqing Surveying and Design Institute Co. Ltd. of Water Resources, Electric Power and Architecture, Chongqing 401120, China
[5] Chongqing Water Resources and Electric Engineering College, Chongqing 402160, China
* Correspondence: kangfei@dlut.edu.cn (F.K.); lijunjie@dlut.edu.cn (J.L.); Tel.: +86-0411-84708516 (F.K.); Fax: +86-0411-84708501 (J.L.)

Abstract: This study proposes a UAV-based remote measurement method for accurately locating pedestrians and other small targets within small reservoir dams. To address the imprecise coordinate information in reservoir areas after prolonged operations, a transformation method for converting UAV coordinates into the local coordinate system without relying on preset parameters is introduced, accomplished by integrating the Structure from Motion (SfM) algorithm to calculate the transformation parameters. An improved YOLOv8 network is introduced for the high-precision detection of small pedestrian targets, complemented by a laser rangefinder to facilitate accurate 3D locating of targets from varying postures and positions. Furthermore, the integration of a thermal infrared camera facilitates the detection and localization of potential seepage. The experimental validation and application across two real small reservoir dams confirm the accuracy and applicability of the proposed approach, demonstrating the efficiency of the proposed routine UAV surveillance strategy and proving its potential to establish electronic fences and enhance maintenance operations.

Keywords: computer vision; noncontact measurement; small reservoir dam; unmanned aerial vehicle; structural health monitoring; deep learning

1. Introduction

Reservoirs and dams are essential infrastructures for flood protection, water supply storage and distribution, energy generation, and other purposes. Small reservoir dams have consistently represented a substantial portion of hydraulic infrastructures given their construction cost advantages, applicability to certain locations, and convenient access due to proximity [1]. During long-term operation, the safety of the dam is often affected by several factors, such as structural aging and loading [2]. To monitor dam health and plan operational responses accurately, many researchers have conducted numerical simulations or model experiments for dam maintenance analysis [3,4]. However, these methods mainly focus on large hydraulic facilities, and, considering the cost constraints and efficiency, small reservoir dams have received limited attention [5]. Due to the lack of sufficient automated monitoring systems or inspection personnel, numerous small reservoir dams are often considered vulnerable structures in flood defense systems.

The safety of pedestrians in the reservoir dam area is a focal point for managers [6]. However, unlike conventional damage, pedestrians exhibit both spatiotemporal randomness and mobility. Furthermore, in contrast to large-scale hydraulic structures, small

reservoir dams often serve as thoroughfares, leading to frequent appearances or crossings by pedestrians. During routine inspections, it is essential to promptly detect and locate pedestrians, issuing timely warnings when they venture into areas outside the designated pathways. In the event of changes in water levels, the prompt determination of pedestrian presence in areas like river channels becomes imperative. Identifying areas with frequent pedestrian activity allows management to implement targeted measures. Although deploying an extensive array of cameras could accomplish these objectives, the associated construction and maintenance costs are impractical for small reservoir dams. Therefore, efficiently achieving accurate pedestrian detection and three-dimensional (3D) positioning is a crucial measure for ensuring the safety of individuals in reservoir areas. Seepage has a significant impact on the structural safety of small reservoirs, primarily those composed of earth–rock dams [7]. Due to the concealed and spatiotemporal randomness of seepage, the conventional methods struggle to identify it. Recently, several contact measurement methods have been adopted to explore the internal features, including ground-penetrating radar [8], electrical resistivity measurement [9], and other sensing techniques [10]. However, for small reservoir dams, the cost of detection is a significant consideration. More importantly, rapidly and accurately pinpointing the location of leaks in the dam is also challenging. Consequently, an efficient and cost-effective method is needed to address both seepage and pedestrian detection and localization in small reservoir dams.

Recently, computer vision has been recognized as an important method for structural health monitoring, and, with the development of both aerial and terrestrial vehicles to carry the requisite instruments and sensors, computer vision has been widely applied in the fields of damage detection and noncontact measurements [11–13]. Jiang used a single panoramic camera to measure full-field structural deformation and accurately calculate the displacement of specified nodes, proving the practicality of computer vision in noncontact measurements [14]. Xu proposed a visualization method for simulating the seismic dynamic responses of buildings based on aerial photography, and the resulting realistic visualization results provide a basis for decision making during sudden disasters [15].

To obtain images more flexibly and comprehensively, unmanned aerial vehicles (UAVs) are widely adopted and integrated with positioning systems to monitor structural health [16,17]. Narazaki applied a comprehensive inspection method to automatically recognize and localize critical structural components, and the results confirmed the significant potential of image processing methods for efficient post-earthquake structural inspection [18]. Xu proposed a 3D reconstruction method for bridge geometry measurements based on UAV images that improved the computational efficiency without reducing the accuracy [19]. For construction safety, Chen presented a homography-based method for measuring the vibration of a bridge model, and the effectiveness was validated against the fixed-camera-based method [20]. In addition to visible light, infrared thermography can also enable the detection of temperature fields. Zhou combined passive infrared thermography and deep learning to identify embankment leakages automatically; the results indicated the practicality and flexibility of UAVs for seepage detection [21]. Qin conducted research on urban evapotranspiration using UAVs, demonstrating the potential of UAVs for urban environmental planning applications [22]. Zhong also used UAVs to analyze urban temperatures at the microscale and proposed a method for analyzing the spatial variations in urban thermal environments [23]. These studies demonstrate the valuable utility of UAVs in facilitating high-precision noncontact measurements, and the introduction of 3D reconstruction methods enables damage localization and change detection [24–26].

A key challenge for UAV-photogrammetry-based structural health monitoring involves the accurate estimation of the relative pose and position between the image and the target, such as the registration from a 2D image to 3D coordinates [27–29]. Chen fine-tuned the rough pose estimated by photogrammetry and solved the challenging indoor localization problem [30]. Tan presented a simplified coordinate transformation method to transform damage positions in the real world to coordinates in a building information modeling (BIM) model, enabling an efficient method for detecting and locating structural

damage in high buildings [31]. Ren proposed a method to reduce the systematic errors in the vertical axis direction of the 3D reconstruction results based on GPS data [32]. Coordinate transformation is a crucial technique for achieving target localization and measurements. Lasers are often incorporated into noncontact measurement methods and introduce another dimensional constraint [33–35]. Zhuge developed a method for precisely measuring bridge deflection with a laser combined with a UAV [36]. Xu combined laser scanning data to monitor the dimensional accuracy of prefabricated components and enhance the quality management efficiency [37]. In summary, compared with the traditional methods, UAVs equipped with different measurement and sensing devices can efficiently detect and accurately locate multiple types of targets in a noncontact manner. The flexibility and efficiency of UAVs integrated with sensors have been demonstrated in various applications, and they can provide rapid and automated structural assessment as well as route planning in advance [11].

Small reservoir dams possess distinct characteristics. Due to inadequate recognition of the emphasis and limited financial resources, they often lack automated monitoring systems, necessitating UAVs as a supplementary measure. However, the absence of control points and terrain deformation complicate the establishment of a correspondence between UAV positions and the coordinates of small reservoir dams. Consequently, this limitation obstructs the precise 3D localization of observed entities such as pedestrians, which is elaborated upon subsequently:

1. UAVs typically rely on the World Geodetic System-1984 (WGS-84) coordinate system, which uses longitude (*L*), latitude (*B*), and altitude (*H*). This coordinate system is not commonly utilized in dam construction. While conversion parameters are publicly available, within the scope of small reservoirs, the transformation precision may be insufficient for precise positioning.
2. Coordinate transformations typically require simultaneous measurements of the coordinates in both coordinate systems for multiple control points. However, the control points for small reservoir dams that have been in operation for many years may not have been adequately preserved, making it difficult to accurately determine their positions in the dam coordinate system.
3. Both the pedestrian and seepage locations in small reservoirs are highly random, requiring methods capable of accurately determining the 3D positions of targets from 2D images photographed in arbitrary positions and angles, as shown in Figure 1.
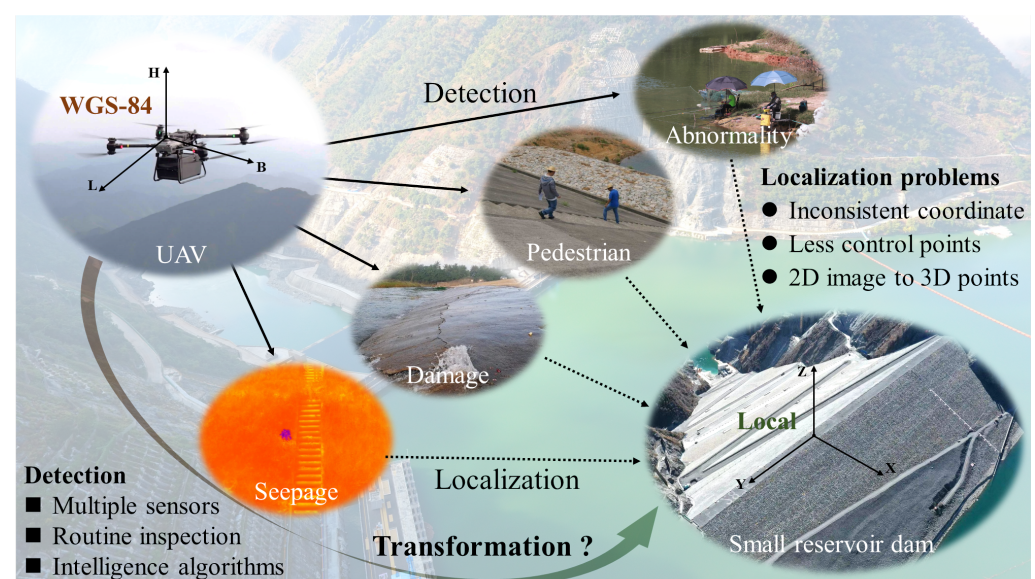

Figure 1. Target 3D localization challenges with UAV inspection.

Moreover, the accurate identification of pedestrians is crucial during UAV inspections, but the small pixel size of pedestrians can lead to poor accuracy using the conventional detection algorithms. Therefore, achieving rapid target localization in the dam coordinate system and accurate pedestrian detection need to be guaranteed during UAV inspection.

This study introduces an effective method for coordinate transformation and non-contact measurement to detect and localize pedestrian targets. UAV coordinates are decomposed into planar coordinates projected onto the Gauss–Krüger plane and altitude coordinates. By incorporating ground control points (GCPs) to establish a local coordinate system, a correspondence between the UAV coordinates and local dam coordinates is proposed using the Structure from Motion (SfM) algorithm. Subsequently, a combination of a four-parameter model and the altitude fitting method is introduced, establishing a rapid and accurate approach for coordinate transformation. Integrated with laser rangefinder measurements, this methodology enables the 3D localization of targets regardless of the photographing position and posture. To enhance pedestrian detection accuracy, an improved YOLOv8 network incorporating attention modules is introduced, bolstering the feature extraction capability for small targets. The precision of the proposed method is validated by applying it to two existing earth–rock dams. By comparing the 3D point cloud generated after the coordinate transformation with the laser scanning results, it accurately identified and positioned various seepage locations and pedestrians. The results of this study provide both the theoretical and technical frameworks required for the routine inspection of small reservoir dams using UAV photogrammetry.

The remainder of this paper is organized as follows. Section 2 describes the theory and method of coordinate transformation. Section 3 discusses the coordinate calculation method based on the SfM algorithm and the detection method for pedestrians and dam seepages. The experimental results and application case studies are analyzed in Section 4. Finally, the conclusions are presented in Section 5.

2. 3D Localization of UAV Targets in the Dam Local Coordinate System

The objective of UAV inspection is to identify and locate different objects, thereby determining the structural locations of these objects based on the coordinate system of the dam. Therefore, an efficient and convenient coordinate transformation and positioning method suitable for small reservoir dams must be developed.

2.1. Coordinate Transformation Methods

Since UAVs use WGS-84 coordinate system while local coordinate system or spatial Cartesian coordinate system is often employed during dam construction, coordinate transformation methods are required to convert between different coordinate systems [38,39]. As shown in Figure 2, WGS-84 coordinate system can be transformed into the spatial Cartesian coordinate system through the seven-parameter method expressed in Equation (1), which is generally employed when the conversion range is large. The UAV coordinate system can then be projected onto the dam planar coordinate system using the reference parameters, as shown in Figure 3a.

$$\mathbf{P}_j = (1 + m_1) \begin{bmatrix} 1 & -\kappa & -\varphi \\ \kappa & 1 & -\omega \\ \varphi & \omega & 1 \end{bmatrix} \mathbf{P}_i + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \tag{1}$$

where $\mathbf{P}_i$ (X_i, Y_i, Z) and $\mathbf{P}_j$ (X_j, Y_j, Z_j) are the 3D coordinate points matched in the two coordinate systems; m_1 indicates a scale factor; (ω, φ, κ) and (ΔX, ΔY, ΔZ) represent three Euler angle rotation parameters and three translation parameters, respectively. It can be observed that a minimum of three high-precision control points with two sets of coordinates are required for computation when employing the seven-parameter method. Furthermore,

when calculating plane and altitude conversions separately, the four-parameter method can be adopted, as depicted in Figure 3b and described as

$$\mathbf{p}_j = (1 + m_2) \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{p}_i + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (2)$$

where $\mathbf{p}_i$ (x_i, y_i) and $\mathbf{p}_j$ (x_j, y_j) represent the planar coordinate points; m_2 also indicates a scale factor; θ denotes the rotation angle of the coordinate axis. Meanwhile, altitude h_j and h_i can be transformed based on the altitude anomaly Δh, as shown in Figure 3c and expressed as

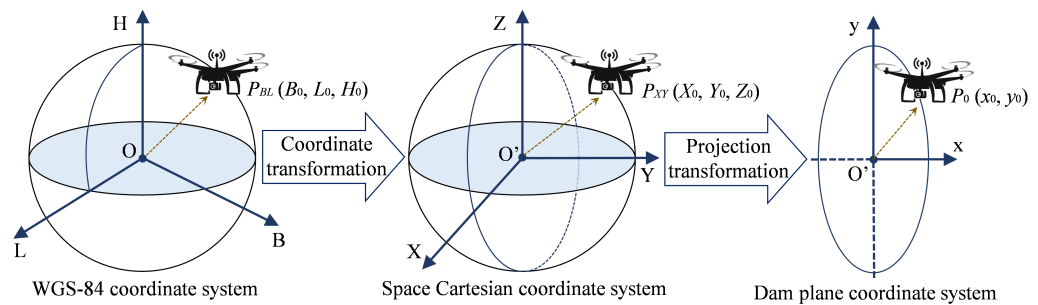$$h_j = h_i - \Delta h. \quad (3)$$

Figure 2. Conventional coordinate transformation process between UAV and dam coordinate systems.
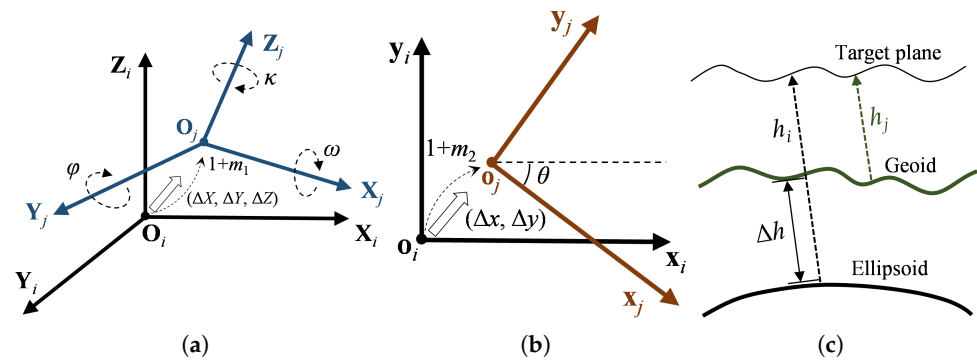

Figure 3. Geometric models for coordinate transformation in different dimensions. (**a**) Seven-parameter model. (**b**) Four-parameter model. (**c**) Altitude conversion.

The altitude anomaly Δh is a variable that varies from one geographic location to another due to the non-uniform nature of Earth's surface. Then, coordinate transformation can be achieved when there are a sufficient number of control points with precision coordinates in both coordinate systems.

However, despite the efficiency of the seven-parameter method in achieving coordinate transformation, it presents challenges when the center of the local coordinate system is established at the dam crest, which is significantly distant from the center of the reference ellipsoid. In such scenarios, the rotation (ω, φ, κ) and translation (ΔX, ΔY, ΔZ) parameters tend to become relatively large when using the seven-parameter method. This can lead to a singularity in the coefficient matrix when attempting to solve for these parameters using the least-squares method, resulting in suboptimal and biased solutions, along with substantial errors post-conversion [40]. Simultaneously, in the conventional altitude transformation method, there is a reliance on altitude anomaly Δh. However, as the planar position undergoes changes, the height difference between the reference ellipsoid and the geoid also varies. This renders the fixed altitude anomaly Δh inadequate for accurately representing altitude transformation in the reservoir area. Hence, it becomes imperative to explore

coordinate transformation methods and frameworks tailored to the distinctive environment of small reservoir dams.

2.2. SfM-Based UAV to Dam Coordinate Registration

While UAVs are equipped with positioning capabilities, the absence of a sufficient number of high-precision control points results in the lack of a reference for conventional coordinate transformation methods. The SfM algorithm can calculate the spatial relationship between a UAV and target points through multiple images in a noncontact manner, thereby generating a 3D point cloud or a real scene model [18,41]. The coordinates of a point P (X, Y, Z) in space can be calculated by multiple cameras using the principle of triangulation:

$$s_n \begin{bmatrix} u_n \\ v_n \\ 1 \end{bmatrix} = \mathbf{K}_{3\times4}[\mathbf{R}|\mathbf{t}]_{4\times4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{84} \tag{4}$$

where $\mathbf{u}_n$ (u_n, v_n) is the projection of point P in image n; $\mathbf{K}$ and $[\mathbf{R}|\mathbf{t}]$ are the intrinsic matrix and extrinsic matrix, respectively; and s_n is a scale factor. To calculate the relationship between images, the pixel points $\mathbf{u}_n$ should be constrained by the fundamental matrix $\mathbf{F}$ with $\mathbf{u}_n^T \mathbf{F} \mathbf{u}_{n-1} = 0$ according to the epipolar constraint, and further decomposed into $\mathbf{u}_n^T \mathbf{K}^{-T} \mathbf{E} \mathbf{K}^{-1} \mathbf{u}_{n-1} = 0$ to resolve the essential matrix $\mathbf{E}$ when $\mathbf{K}$ is known. Finally, bundle adjustment is employed to optimize the camera parameters and positions by minimizing the reprojection error [42]:

$$g(C, X) = \sum_{n=1}^{all} \sum_{m=1}^{all} \omega_{nm} \|\mathbf{u}_{nm} - \mathbf{project}(\mathbf{C}_n, P_m)\|^2 \tag{5}$$

where $\|\mathbf{u}_{nm}\text{-}\mathbf{project}(\mathbf{C}_n, P_m)\|$ represents the difference between the pixel point m from image n and the projection position of space point P, $\mathbf{C}_n$ includes several parameters; $\omega_{nm} = 1$ indicates the point is observed. With a sufficient number of images captured by a UAV, the 3D coordinates of the target point can be calculated.

However, the 3D points generated by the SfM algorithm have no corresponding spatial relationship with respect to the location of the small reservoir dam, and the lack of a sufficient number of high-precision control points hinders the application of traditional coordinate transformation methods [33], as shown in Figure 4a. This paper introduces the ground control points (GCPs) associated with dam local coordinate system to establish the foundation for transformation. GCPs are usually measured by a total station or other equipment to obtain the 3D coordinates in a specific coordinate system. Corresponding to each UAV image capturing a GCP with location (X_{UAV}, Y_{UAV}, Z_{UAV}), the projection point (u, v) and GCP position (X_{GCP}, Y_{GCP}, Z_{GCP}) are on a single line and expressed as

$$\begin{cases} u - u_0 + k_u = -f\dfrac{a_1(X_{GCP} - X_{UAV}) + b_1(Y_{GCP} - Y_{UAV}) + c_1(Z_{GCP} - Z_{UAV})}{a_2(X_{GCP} - X_{UAV}) + b_2(Y_{GCP} - Y_{UAV}) + c_2(Z_{GCP} - Z_{UAV})} = -f\dfrac{\overline{X}}{\overline{Z}} \\ v - v_0 + k_v = -f\dfrac{a_3(X_{GCP} - X_{UAV}) + b_3(Y_{GCP} - Y_{UAV}) + c_3(Z_{GCP} - Z_{UAV})}{a_2(X_{GCP} - X_{UAV}) + b_2(Y_{GCP} - Y_{UAV}) + c_2(Z_{GCP} - Z_{UAV})} = -f\dfrac{\overline{Y}}{\overline{Z}} \end{cases} \tag{6}$$

where (u_0, v_0) is the center point of image; (k_u, k_v) denotes the correction values for system errors such as distortion; f represents focal length; and $\mathbf{R} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$ is the direction cosine for the three external orientation angles (ω, φ, κ). Meanwhile, given the offset between the camera photographic center and the GPS position, there exists a displacement described as $\Delta \mathbf{t} = [\Delta X, \Delta Y, \Delta Z]^T$, as shown in Figure 4b. When using UAVs to capture

several GCPs, the internal and external parameters can be calculated based on the SfM algorithm, and Equation (6) simplifies to

$$
\begin{cases}
u = (u_0 - f\dfrac{\overline{X}}{\overline{Z}} - k_u) + D_X \\[2mm]
v = (v_0 - f\dfrac{\overline{Y}}{\overline{Z}} - k_v) + D_Y
\end{cases}
\tag{7}
$$

where D_X and D_Y are the correction factors for the true values during iterative process. Therefore, the first-order term resulting from the Taylor series expansion of Equation (7) is expressed as

$$
\begin{bmatrix} u \\ v \end{bmatrix} =
\begin{bmatrix}
\dfrac{\partial u}{\partial X_{UAV}} & \dfrac{\partial u}{\partial Y_{UAV}} & \dfrac{\partial u}{\partial Z_{UAV}} & \dfrac{\partial u}{\partial \omega} & \dfrac{\partial u}{\partial \varphi} & \dfrac{\partial u}{\partial \kappa} \\[2mm]
\dfrac{\partial v}{\partial X_{UAV}} & \dfrac{\partial v}{\partial Y_{UAV}} & \dfrac{\partial v}{\partial Z_{UAV}} & \dfrac{\partial v}{\partial \omega} & \dfrac{\partial v}{\partial \varphi} & \dfrac{\partial v}{\partial \kappa}
\end{bmatrix}
\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \\ \Delta \omega \\ \Delta \varphi \\ \Delta \kappa \end{bmatrix}
-
\begin{bmatrix} u_0 - f\dfrac{\overline{X}}{\overline{Z}} - k_u \\[2mm] v_0 - f\dfrac{\overline{Y}}{\overline{Z}} - k_v \end{bmatrix}.
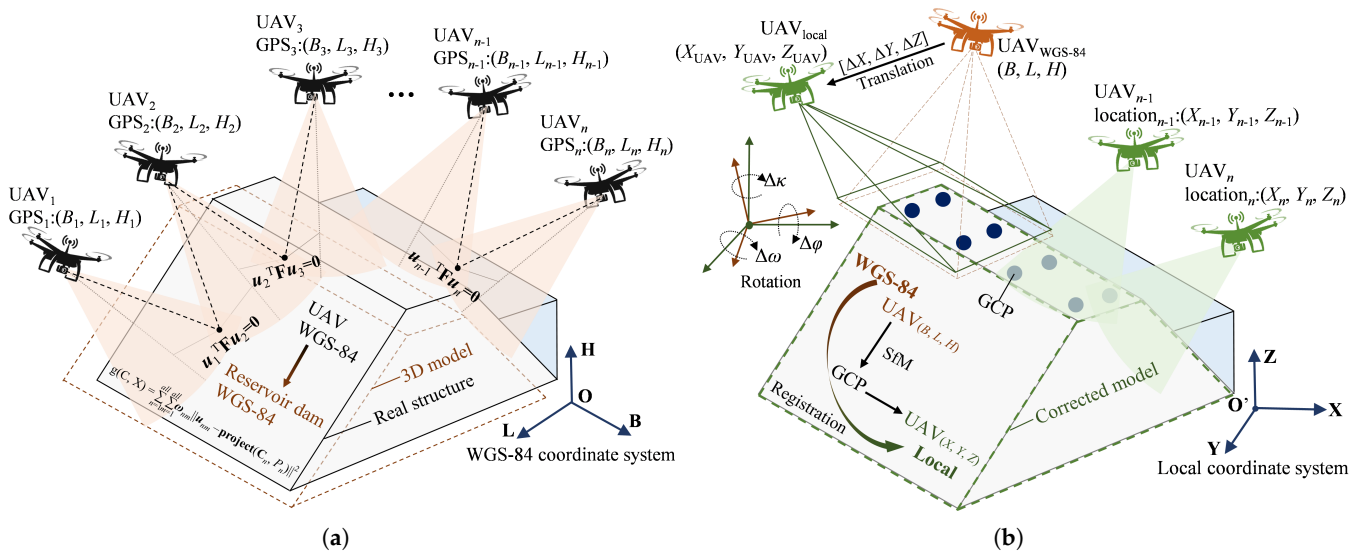\tag{8}
$$

Figure 4. The registration process between UAV and dam coordinates based on the SfM algorithm. (**a**) UAV image-based SfM processes without GCPs. (**b**) UAV coordinates transformed to local coordinate system.

For a massive collection of UAV images, the average of the correction factors can serve as a representative value for the discrepancy between the computed external orientation elements using GCPs and the UAV GPS data. Thus, multiple sets of coordinates (*B*, *L*, *H*) from the UAV internal positioning system can be matched with their corresponding coordinates (*X*, *Y*, *Z*) within the local coordinate system of the small reservoir dam, which provides precise and low-cost spatial referencing.

The correspondence between UAV coordinates and small reservoir dam coordinates can be established by deploying GCPs and utilizing SfM algorithms. Two transformation modes can be considered, including GCP-coordinate-based transformation and UAV-coordinate-based transformation. The GCP-coordinate-based transformation is based on the coordinates of GCPs in both systems. However, GCPs are typically placed at the dam crest, leading to minimal vertical changes, and the GCP-coordinate-based approach might become trapped in a local optimum for vertical transformations, which could result in an increase in overall errors, as shown in Figure 5. Consequently, for a comprehensive coordinate transformation, this paper adopts the UAV-coordinate-based method, which

involves the transformation of the UAV coordinate system at various altitudes and positions into the local coordinate system.

After applying the SfM algorithm, corrections were made to the UAV parameters, and multiple control points with coordinates in two coordinate systems were established. This process laid the foundation for proposing the coordinate transformation method in the subsequent steps.
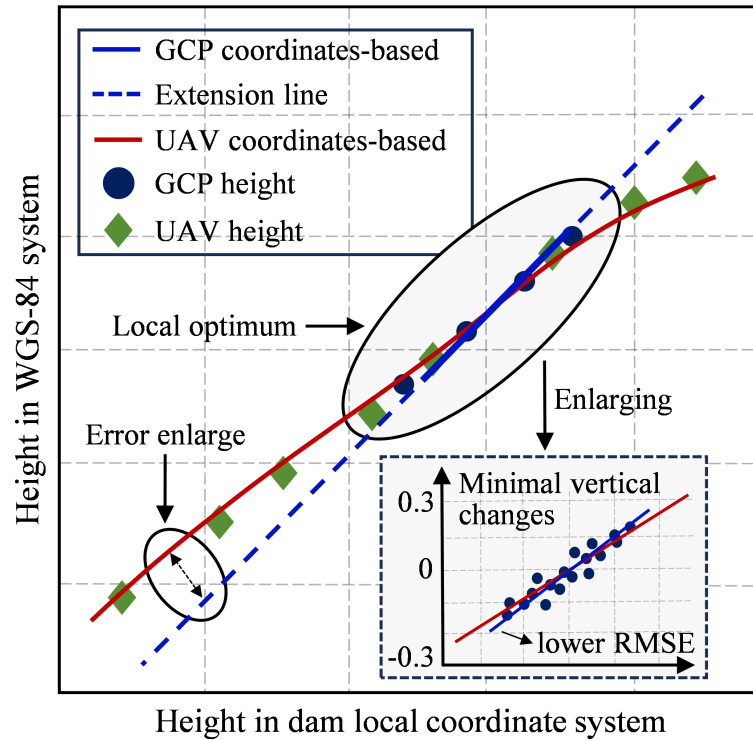

Figure 5. Altitude conversion methods and corresponding error trends.

2.3. Proposed UAV Inspection and Localization Framework

In the distinctive environment of small reservoir dams, precise locating of targets on the surface is paramount based on UAV coordinates and collected data. This section introduces a framework for high-precision noncontact measurement of targets, combining established coordinate transformation formulas with the previously identified challenges.

2.3.1. Projection from WGS-84 to Plan Coordinate

B and L represent spatial positions on the surface of a three-degree sphere and are measured in degrees, in contrast to the typical linear units commonly employed in the local coordinate system. The process of projection onto a plane is introduced to determine the relationship between the UAV system and the local reference [39]. The Gauss–Kruger projection method is widely adopted in topographic surveys given its minimal distortion in terms of length and area. Hence, this study employs the Gauss–Kruger method to project (B_0, L_0) of the UAV into a planar Cartesian coordinate system (x_{proj}, y_{proj}) before transformation:

$$\begin{bmatrix} x_{proj} \\ y_{proj} \end{bmatrix} = f_{proj}(N, e) \begin{bmatrix} B_0 \\ L_0 \end{bmatrix} \tag{9}$$

where N is the radius of the Meridian circle and e is the first eccentricity of the reference ellipsoid. The process is shown in Figure 6b. By using the projection method, the 3D coordinates can be separated into both plane and elevation components. This alignment corresponds to the GCP coordinates measured using a total station on the small reservoir dam.
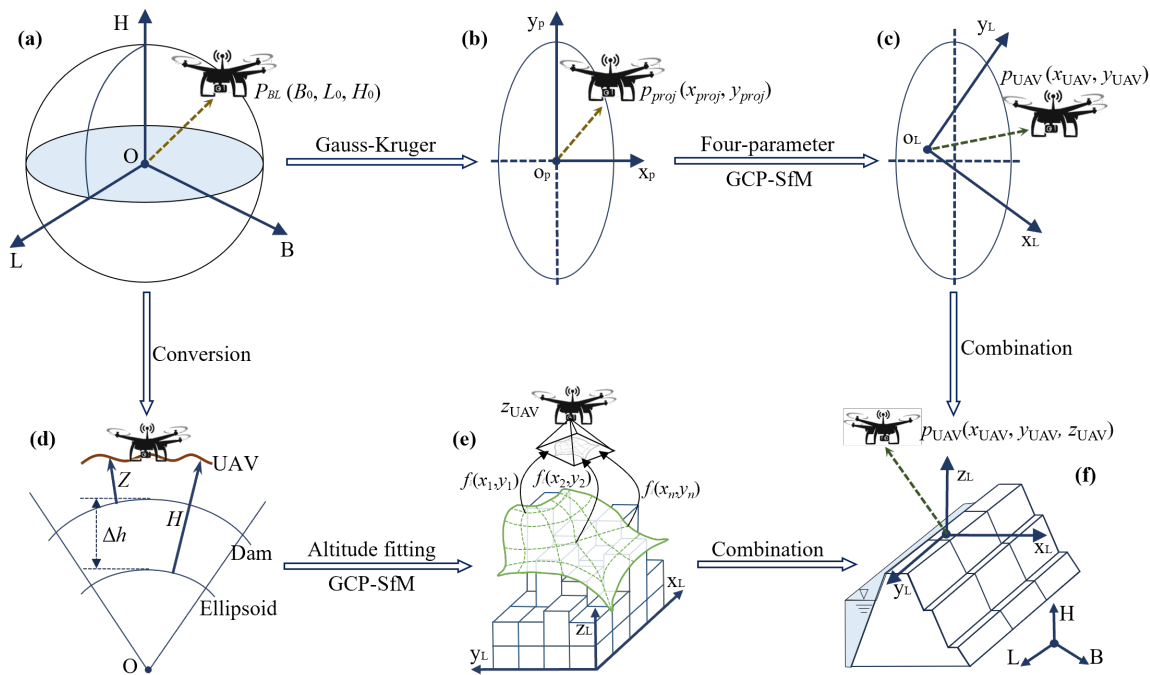
Figure 6. The process of transforming UAV WGS-84 coordinates to dam local coordinates. (**a**) WGS-84 coordinate system. (**b**) Projection to plane. (**c**) Convert to local. (**d**) Altitude transformation. (**e**) Altitude adjustment. (**f**) UAV transformation coordinates.

2.3.2. Transforming UAV Coordinate to Dam Coordinate

After projection conversion, coordinate registration can be achieved by including the UAV altitude information, as described in Section 2.2. Considering image distortion and other influences, based on the four-parameter model in Equation (3), the scale factor m_2 is revised by applying m_x and m_y separately in different directions [31], and the transformation parameters are calculated based on coordinates in both coordinate systems, as shown in Figure 6c:

$$\begin{bmatrix} x_{UAV} \\ y_{UAV} \end{bmatrix} = \begin{bmatrix} 1+m_x & 0 \\ 0 & 1+m_y \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{proj} \\ y_{proj} \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \tag{10}$$

The altitude anomaly mainly consists of two parts: the elevation difference between the geoid height used by the UAV and the normal height and the height difference between the reference surface of the dam local coordinate system and the normal height reference surface. However, considering the influence of the gravity field, the altitude difference between the reference ellipsoid and the geoid varies with the topography, which can have a minor impact on the altitude conversion. Additionally, since GCPs have relatively small changes in the vertical direction, some parameters are required to correct the converted altitudes. Therefore, as shown in Figure 6e, for small reservoir dams, the altitude fitting method is introduced to perform polynomial fitting for the altitude anomaly Δh [43,44], and Equation (3) can be expanded as

$$z_{UAV}(H, x, y) = H + \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 xy + \alpha_4 x^2 + \alpha_5 y^2 \tag{11}$$

where $\alpha_0 \sim \alpha_5$ are the conversion fitting coefficients from WGS-84 altitude H to local altitude Z_{UAV} in the small reservoir dam area. Finally, through the modified four-parameter method and the altitude fitting method, the transformation from UAV WGS-84 to dam local coordinate system can be achieved as shown in Figure 6f. The developed method establishes the correspondence between the UAV and the dam structural position in space, enabling target localization.

2.3.3. Target 3D Localization

After obtaining the coordinates of the UAV in the dam local coordinate system, the position of specific targets within the small reservoir dam can be determined. However, during routine inspections, it is difficult for UAVs to maintain a strictly vertical orientation relative to the target plane, and a spatial angle between UAV and the target is introduced. Furthermore, the target 3D spatial position cannot be solely determined based on 2D projection coordinates (u, v) [27,45,46]. To address this complexity, the use of a laser rangefinder is introduced in this study, adding an additional constraint that establishes the connection between the UAV optical center and a pixel point on the image. The position of target point $(x_{tar}, y_{tar}, z_{tar})$ can be calculated as

$$\begin{bmatrix} x_{tar} \\ y_{tar} \\ z_{tar} \end{bmatrix} = \begin{bmatrix} x_{UAV} \\ y_{UAV} \\ z_{UAV} \end{bmatrix} + d \begin{bmatrix} -\sin\beta\cos\gamma \\ \cos\beta\cos\gamma \\ -\sin\gamma \end{bmatrix} \tag{12}$$

where d refers to the measurement distance of the laser rangefinder; β and γ are the capturing angles, as shown in Figure 7. Meanwhile, as UAV inspections often involve vertical photography, a single capture combined with distance measurements can determine the 3D positions of any pixel point (u_i, v_i) within the dam coordinate system. This approach is particularly advantageous for flat areas such as the dam crest, and the pixel point position can be expressed as

$$\begin{bmatrix} x_{tar-i} \\ y_{tar-i} \end{bmatrix} = \frac{d\lambda}{f} \begin{bmatrix} (u_i - u_0) \times \sin\psi \\ (v_i - v_0) \times \cos\psi \end{bmatrix} + \begin{bmatrix} x_{tar} \\ y_{tar} \end{bmatrix} \tag{13}$$

where λ is the unit pixel length; (u_0, v_0) is the position of pixel center; ψ is the angle between the u-axis of the image and the x-axis of the dam coordinate system. Finally, the position of the target point within the dam coordinate system can be rapidly measured using the UAV integrated with a rangefinder.
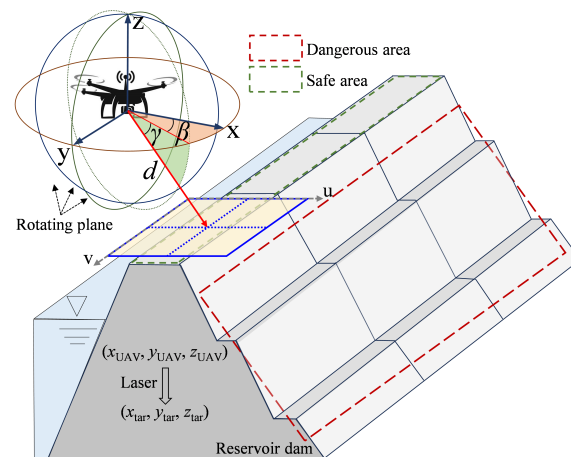

Figure 7. Target 3D localization using a UAV integrated with a laser rangefinder.

Overall, when the calculation of conversion parameters is inaccurate, the proposal is to use a combination of plane and altitude transformations to improve accuracy. For altitude transformations, altitude fitting method can be effective in addressing inaccuracies. A method for target localization based on ranging results during noncontact measurements is established. The approach proposed in this section enables accurate positioning of target points through UAV photogrammetry.

3. Small Pedestrian Target Intelligent Detection Algorithm

To achieve UAV automatic inspection and 3D localization, in addition to the proposed noncontact measurement method, it is also necessary to implement fast and intelligent pedestrian identification algorithms. However, unlike conventional targets, pedestrians captured by UAV cameras usually occupy a small portion of the pixel and lack facial features, and the relevant datasets are smaller, making target detection very challenging for conventional methods. Therefore, it is necessary to develop pedestrian detection methods tailor-made for UAV inspections of small reservoir dams.

3.1. YOLOv8 Detection Algorithm Basic Architecture

The YOLO network is one of the most widely adopted object detection algorithms for UAV image detection, and it is suitable for deployment on UAV platforms to achieve real-time detection [47,48]. YOLOv8, like YOLOv5, was developed and continues to be updated by Ultralytics and provides about notable enhancements in detection speed and accuracy [49]. The input image with size of (w, h, c) is down-sampled five times in the backbone, resulting in dimensions of ($w/32$, $h/32$, c_{set}). These features are then transferred to the neck structure, which is primarily employed for feature fusion and up-sampling. This facilitates the amalgamation of features from different scales, leading to a more comprehensive understanding of the target. The detection part is the head structure, which can classify the target, predict the bounding box, generate three scale detection results, and finally describe the position of the target in image.

YOLOv8 is composed of a similar network architecture to YOLOv5, making it possible to directly compare their performance when using similar parameters. This also enables an effective evaluation of how newly added modules enhance the network's capabilities. Furthermore, compared with YOLOv5, several improvements are adopted in YOLOv8. A C2f module is used to replace the C3 module to obtain more abundant gradient flow information while maintaining a lightweight architecture. At the end of backbone, the spatial pyramid pooling fast (SPPF) module is retained from original YOLOv5 to ensure running speed during the fusion of global and local features. The feature pyramid network (FPN) and path aggregation network (PAN) are also introduced in the neck, but one convolution operation is removed to achieve a lightweight model. The head is designed as a decoupled structure, which separates the prediction of target classification and bounding box regression into different branches. The advantage of this structure lies in simplifying the training and optimization process of the network, enhancing the model trainability and inference efficiency. In the context of object detection, the model directly learns to predict object locations and scales without relying on predefined anchors. With these improvements, YOLOv8 achieves higher detection speed and accuracy, making it more suitable for real-time detection in UAV routine inspections [50].

3.2. Proposed Pedestrian Detection Network

To enhance the detection of small pedestrians lacking distinct facial features, some improvements should be implemented in the YOLOv8 model. In this study, the ConvNeXt v2 [51,52] module and ODConv (Omni-Dimensional Dynamic Convolution) [53] module are introduced to enhance feature extraction capabilities. The network architecture is modified to incorporate a tiny detection head, and the Wise-IoU loss function is added to comprehensively improve the network ability in detecting small pedestrians. The specific details are described as follows.

3.2.1. Additional Modules

The ConvNeXt module is a pure convolutional neural network that performs better in terms of accuracy than a transformer and is widely adopted by various detection tasks [54]. ConvNeXt v2 is proposed by adding a global response normalization (GRN) step to increase the feature contrast between channels and enhance performance. Effective detection of small targets typically requires a model with a good receptive field and strong

representation capabilities for small-sized objects. The improvements in ConvNeXt v2 primarily focus on the depth structure of the network, which may contribute to improving the receptive field and enhancing the detection capabilities for small targets. Based on ConvNeXt, a global response normalization (GRN) step was added to ConvNeXt v2, which can increase the feature contrast between channels and enhance performance. The architecture of ConvNeXt v2 is shown in Figure 8. In contrast to conventional bottleneck structures, a 7×7 depthwise convolution is adopted in the first step, followed by layer normalization. Then, features are passed into convolutional layers and a GRN block applied on each patch. The module ultimately concludes with the convolutional layer and drop path. ConvNeXt can maintain the efficiency of convolution without complex modules, and the weights of small objects are increased to make the network focus more on training.
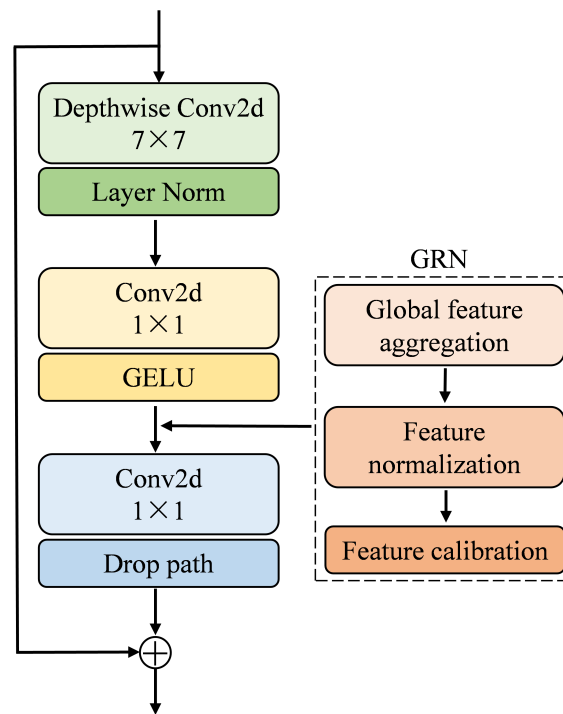

Figure 8. The ConvNeXt v2 module architecture.

This study also modifies the convolutional part of YOLOv8. Considering the characteristics of low resolution and limited information content, small targets are susceptible to being overlooked or falsely detected during the detection process, thus requiring the digging of deeper image features. ODConv adopts a parallel strategy and multi-dimensional attention mechanism of the convolutional kernel space, which provides an advantage when dealing with low-resolution images that contain limited information, thus obtaining full-dimensional convolutional kernel attention, as shown in Figure 9. This allows the network to better understand target features and improve detection accuracy for small targets. Specifically, in addition to the conventional assignment of dynamic attributes to convolution through the number of kernels (α_{wi}), attention is also introduced in the spatial dimension (α_{si}), input channel dimension (α_{ci}), and output channel dimension (α_{fi}), which can be expressed as follows:

$$f_{out} = f_{in} \sum \left(\alpha_{wi} \bullet \alpha_{fi} \bullet \alpha_{ci} \bullet \alpha_{si} \bullet W_i \right) \tag{14}$$

where W_i is the convolutional kernel i; f_{in} and f_{out} denote the input and output features, respectively. The four types of attentions are complementary to each other, providing rich contextual clues and significantly enhancing the feature extraction capability of con-

volutional operations; therefore, it is introduced into the detection network for small pedestrian targets.
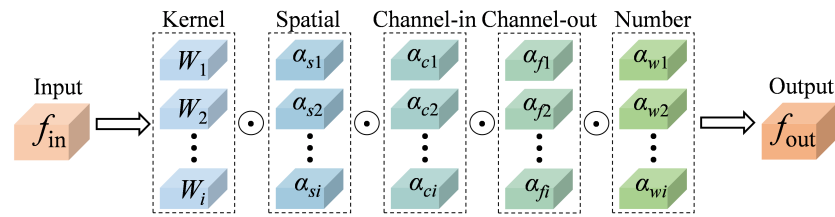

Figure 9. Schematic representation of the ODConv architecture.

3.2.2. Architecture of the Proposed Network

Although the YOLOv8 network performs well on most datasets, specific modifications are needed to further improve detection accuracy, particularly when dealing with small objects that occupy a low number of pixels in images captured by UAVs. The most widely adopted and effective method is the addition of another detection head specifically for tiny objects, preventing their features from being overwhelmed by continuous convolution operations [55–57]. Meanwhile, given the imperative to preserve the real-time detection performance of the network, this study refrained from adding too many complex modules to every layer in the backbone and neck.

The architecture of the proposed network is shown in Figure 10. In the shallow layers of the backbone, the conventional convolution and C2f modules are replaced by ODConv and ConvNeXt to enhance the detection of small targets in large images. In the neck structure, layers 15, 18, and 21 are replaced with ConvNeXt to improve the fusion of shallow and deep features, making the small head and the newly added tiny head more sensitive to small targets. The proposed network is primarily designed for small targets in UAV images, and it enhances the ability to extract features from shallow layers without significantly increasing the number of parameters, making it more suitable for real-time detection in UAV inspections.
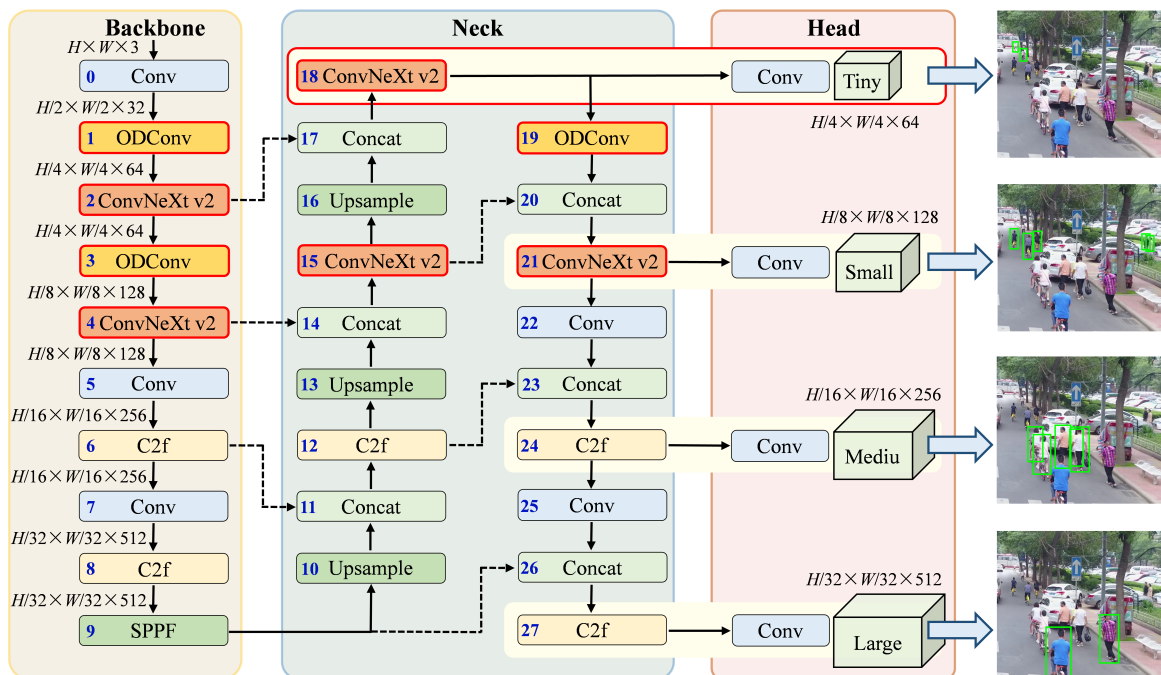

Figure 10. Architecture of the proposed small pedestrian detection network with a tiny head. The main modified part is enclosed in the red box. Modules 0 to 9 are the backbone of the network, while modules 10 to 27 are the Neck structure of the network.

3.2.3. Loss Function

The loss function is a crucial component in the training process of deep learning models. The loss function reduces gradients so that the model parameters can be upgraded. An appropriate loss function can significantly enhance the detection accuracy of a network. YOLOv8 adopts CIoU (Complete Intersection over Union) to calculate the regression loss of bounding boxes. However, for pedestrian targets captured in UAV images, where the pixel proportion is relatively small, CIoU does not consider the balance between hard and easy samples, leading to the inability to learn some useful information [58].

To address this issue, a modified loss function called WIoU (Wise Intersection over Union) is introduced [59]. WIoU v3 incorporates a dynamic non-monotonic mechanism and includes a well-designed gradient gain allocation strategy. This strategy reduces the occurrence of large gradients or harmful gradients in extreme samples. The loss method focuses more on samples of moderate quality, thereby improving the network model generalization ability and overall performance, expressed as

$$L_{\text{WIoUv3}} = \frac{\beta_{\text{v3}}}{\delta_{\text{v3}} \alpha_{\text{v3}}^{\beta_{\text{v3}} - \delta_{\text{v3}}}} \exp \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} \right) (1 - IoU) \tag{15}$$

where α_{v3}, β_{v3}, δ_{v3} are the parameters indicating the quality and dynamically optimizing the anchor boxes; (W_g, H_g) is the smallest size of the bounding box; and (x, y) and (x_{gt}, y_{gt}) are the center of the predicted box and truth box, respectively. The addition of WIoU v3 into the network training process allows the network to focus more on recognizing and challenging small targets, and dynamically optimizes the loss weights, thereby improving the detection performance.

3.3. Evaluation Metrics

To evaluate the performance of the developed network, the mean average precision (*mAP*) is calculated and represents the average precision when the *IoU* threshold is 50%. Moreover, some indexes including *Precision*, *Recall*, and *F*1 are introduced and expressed as

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$F1 = \frac{2 Precision \times Recall}{Precision + Recall} \tag{18}$$

where *TP* refers to the true positives that have been correctly predicted; *FP* represents the true negative sample; *FN* indicates the number of incorrect predictions.

3.4. Coordinate Transformation and Target Localization Procedure

First, GCPs are established and measured at the dam crest, and then images are captured by the UAV camera. The SfM algorithm is subsequently applied to create a correspondence between the UAV planar projected coordinates and altitudes within the local dam coordinate system. The developed method is then introduced to calculate the parameters for the plane and altitude transformation. Finally, small pedestrians are detected by the improved YOLOv8 algorithm and located, with the laser rangefinder ascertaining their presence within potentially hazardous areas. The entire procedure effectively translates 2D pixels into 3D coordinates through coordinate transformation, target detection, and localization. The main steps are processed on Windows workstation; details are as follows and summarized in Figure 11:

Step 1:   Data acquisition. Gathering GCP coordinates through instruments and capturing small reservoir dam images using UAV photography.

Step 2:   Coordinate registration. Employing the SfM algorithm to align the Gauss–Kruger planar projected coordinates and altitudes with dam local positions.

Step 3:   Coordinate transformation. Calculating the transformation parameters of the four-parameter model and applying the altitude fitting formula using the developed method.

Step 4:   Pedestrian detection. Utilizing the improved YOLOv8 algorithm to detect pedestrians in UAV inspection images.

Step 5:   Target localization. Locating pedestrians within the dam coordinate system by combining captured angles and distances using a laser rangefinder.

Step 6:   Position analysis. Determining if pedestrians are within hazardous areas based on their position in the local coordinate system of the small reservoir dam.
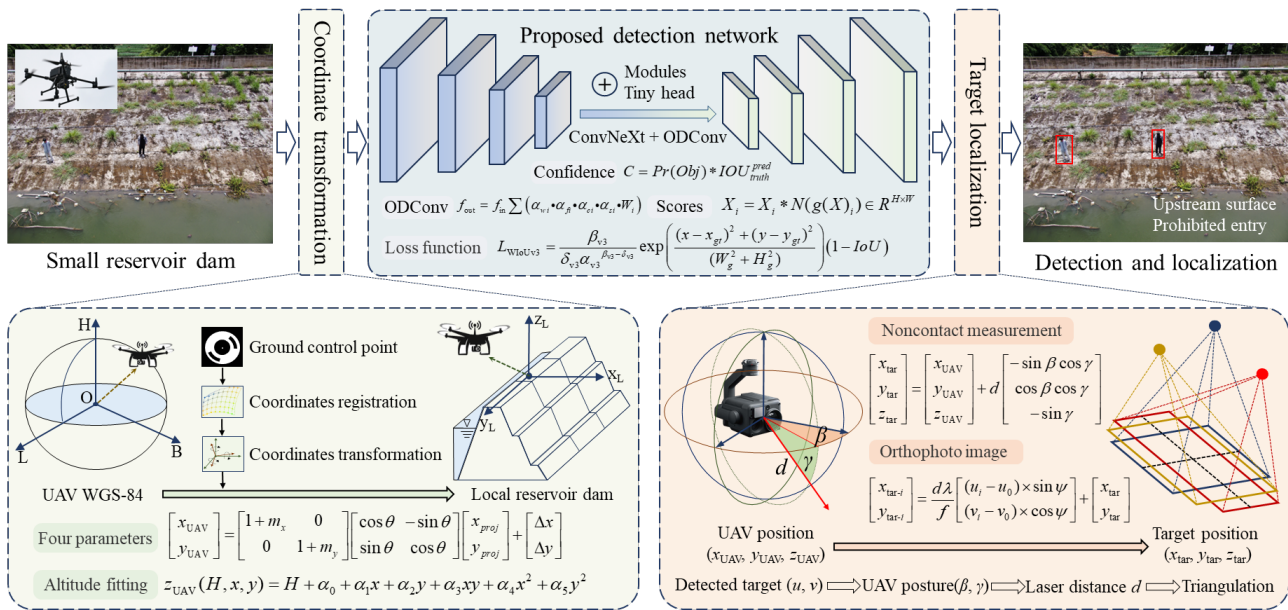

Figure 11. UAV coordinate transformation and detected target 3D localization in a small reservoir dam.

4. Experimental Performance Analysis and Practical Implementation

4.1. Accuracy Analysis of UAV Coordinate Transformation

To analyze and evaluate the impact of the developed coordinate transformation method on the 3D localization accuracy of targets, random point coordinate measurements and analyses were conducted. The images were captured using a DJI M300 RTK UAV, and the SfM algorithm was applied using Agisoft Metashape version 2.0.1 [60]. A workstation running on Windows with an Intel i7-13700F CPU was used, which can integrate its computational power with pre-defined spatial coordinates of the inspection area and produce inspection results directly.

As shown in Figure 12a, six GCPs are placed in the measurement area and their 3D coordinates in the local coordinate system are measured using a total station. A total of 53 images were captured, and, after using the common Gauss–Kruger projection parameters in the transformation library (PROJ), the planar coordinates and altitudes were registered using the SfM algorithm. The parameters described in Section 2.3 could then be calculated. Subsequently, utilizing the transformation parameters and laser rangefinder measurement values, the UAV was used to capture and measure the 3D coordinates of GCPs at various locations. Each GCP was measured nine times from different directions at the same elevation, and these measurements were conducted at three distinct elevations, as shown in Figure 12b. To obtain the most accurate results, professional operators precisely aligned the center of the image with the center of the control point during each measurement. Although this process is time-consuming, it ensures more accurate outcomes. The error

of each measured GCP is shown in Figure 13; the overall in-plane error is generally less than 20 cm, with most values not exceeding 10 cm, and the error remains consistent across the different measurement positions, as in Figure 13a. For altitude, the errors fluctuate greatly, but the average values are less than 10 cm, as in Figure 13b. Some measurement results exhibit larger errors, attributed to significant angles or longer distances during capture, as well as inaccuracies in laser rangefinding. The measurement results in different directions are shown in Table 1, where the errors in the planar directions, x and y, are about 5 cm, while the average error in the altitude direction z is 7.868 cm. This is mainly influenced by the accuracy of the SfM algorithm in calculating the feature points in the altitude direction and the accuracy of the UAV self-positioning. It can be seen that the proposed method can accurately measure the 3D coordinates of the target point in the local coordinate system in a noncontact manner. However, it should be noted that the experiment only used six GCPs in a small area and did not fully exploit the advantages of the altitude fitting method. In a larger-scale measurement process where more GCPs are introduced, the developed method would produce better results.
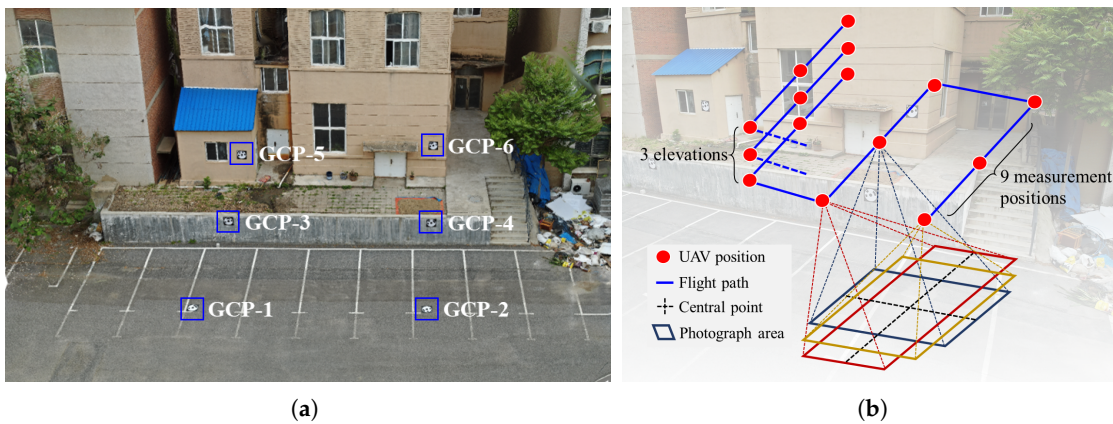

(a) (b)

Figure 12. Evaluation scenario and measurement process of the developed method. (**a**) 3D model of experimental area with 6 GCPs. (**b**) Target measurements in 27 various positions.
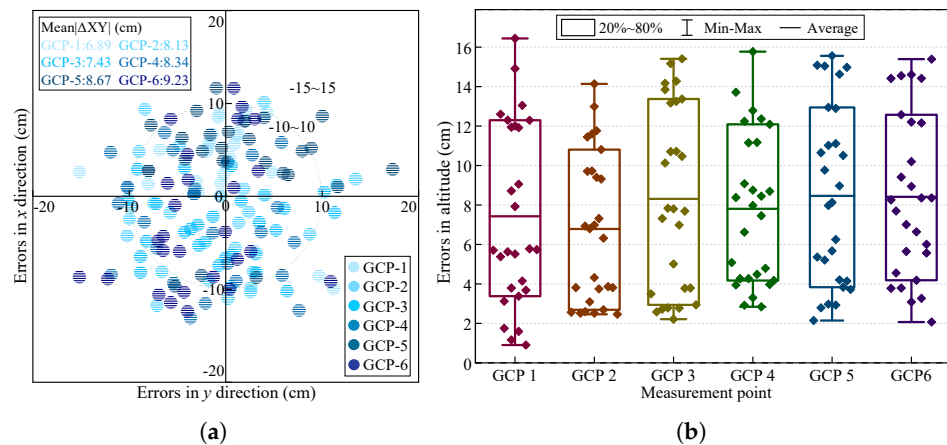

(a) (b)

Figure 13. GCP errors measured from each position. (**a**) Errors in planar measurement. (**b**) Errors in altitude measurement.

Table 1. The measurement errors of the 6 GCPs in each direction (cm).

Point	Mean ΔX	Mean ΔY	Mean ΔZ	Mean ΔXYZ
GCP 1	3.343	5.140	7.425	10.829
GCP 2	5.080	5.128	6.789	11.403
GCP 3	5.010	4.217	8.312	11.767

Table 1. *Cont.*

Point	Mean ΔX	Mean ΔY	Mean ΔZ	Mean ΔXYZ
GCP 4	5.473	5.075	7.807	11.999
GCP 5	5.017	5.943	8.464	12.709
GCP 6	4.679	6.912	8.414	12.999
Total	4.767	5.402	7.868	11.951

4.2. Small Pedestrian Detection Results Based on the Developed Improved YOLOv8

To accurately detect small pedestrian targets on the dam, the developed network was trained and compared with the other networks. A total of 22,400 annotated images were included in the dataset, which were derived from several public datasets, such as WiderPerson [61], VisDrone [62], and COCO [63], and cropped to the appropriate size. The dataset was divided into training, validation, and testing with a ratio of 7:2:1, resulting in 15,680 images for training, 4480 images for validation, and 2240 images for testing. The dataset includes multi-scale targets, ranging from small pedestrians captured from a distance to larger ones photographed up close. The image sizes used during training and testing vary, further enhancing the robustness of the network. Meanwhile, considering the dataset conditions, the training image size was modified to 960 × 960 for improved detection accuracy and better performance. The initial learning rate was set to 0.01, the SGD optimization algorithm was adopted, and the optimizer weight decay was 0.0005. The entire training process was conducted using a Linux system equipped with an RTX 2080 Ti GPU and 128 GB of RAM.

The training process is shown in Figure 14; four other networks are involved, including YOLOv5s, original YOLOv8s, and YOLOv8s with tiny detection heads. All the networks were trained using 200 epochs to ensure convergence. The evaluation results for each trained model are presented in Table 2, and the developed algorithm exhibited the best *F*1 and *mAP* values, reaching up to 70.4 and 78.4%, respectively. These values represent *F*1 and *mAP* improvements of 1.5 and 2.3%, respectively, compared to the original YOLOv8s network. Additionally, with the introduction of the WIoU loss function, the *mAP* value increased by 0.3%. Some detection results are shown in Figure 15, where the developed network excels in accurately identifying pedestrians of various sizes as well as small targets that were ignored by the other models. The developed network can cover targets of all sizes, as shown in Figure 15a, significantly improving the detection accuracy for small targets in Figure 15b–e and reducing the false positives in Figure 15f. These results substantiate the effectiveness of introducing additional modules for enhancing detection capabilities. Moreover, to objectively assess the network generalization ability, the K-Fold Cross-Validation method was utilized, dividing the data of 15,680 images into five parts. The proposed method was then iteratively trained on four parts and validated on the remaining part. The results of the five training iterations are presented in Table 3, with an average *mAP* of 78.2 and *F*1 of 70.5. Additionally, the training results showed no significant differences across the different folds, indicating that the proposed model exhibits a strong level of generalization. To further analyze the robustness of the proposed method, the trained model was utilized to detect the background images, as shown in Figure 16. Although these background images do not contain pedestrians, they feature other disturbances of similar size. The test results confirm that the proposed network did not generate any false detections, further indicating its robust performance.

To evaluate the effectiveness of the proposed method, all the compared networks were kept consistent in terms of the parameters. Additionally, since the Ray Tune method did not significantly improve the network's accuracy during hyperparameter optimization, this study employed the default hyperparameters of YOLOv8. The experimental results are expressed in Table 4; it can be seen from the table that all three modules have a positive impact on the model optimization. With the adoption of ConvNeXt v2, *mAP* increased by 0.9%, and the use of ODConv and a tiny head resulted in a further *mAP* improvement of 1.2%. This is attributed to the fact that ConvNeXt v2 is mainly applied to the backbone

network, contributing more to feature extraction. Additionally, while the tiny head alone only increased *mAP* by 0.3%, it is evident that combining it with ODConv yields better performance. Finally, to analyze the network sensitivity, some layers are replaced with 1×1 convolutional modules. Despite the variations in texture among the different images, the results from a substantial number of validation images indicate that replacing all the layers from zero to six in the backbone significantly hampers the detection of small pedestrian targets within the images.
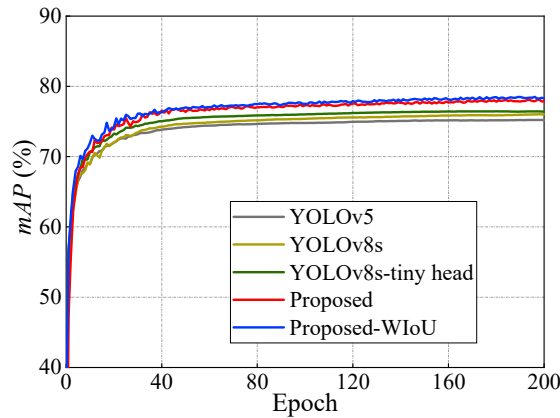

Figure 14. Training process *mAP* curves for the different networks.

Table 2. Comparison of comprehensive evaluation results of different models (%).

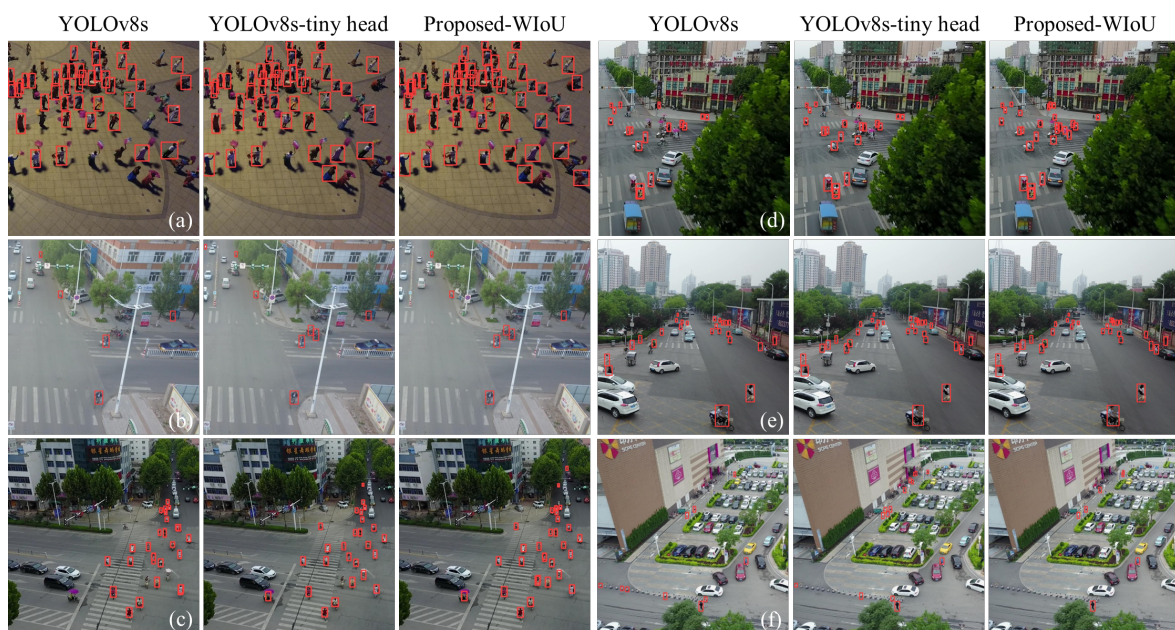| Network	*Precision*	*Recall*	*F1*	*F1* Improved	*mAP*	*mAP* Improved
YOLOv5s	70.6	67.1	68.8	-	75.3	-
YOLOv8s	68.3	69.5	68.9	-	76.1	-
YOLOv8s-tiny head	68.8	69.2	69.0	0.1	76.4	1.3
Proposed	69.7	70.7	70.2	1.3	78.1	2.0
Proposed-WIoU	70.0	70.8	70.4	1.5	78.4	2.3

Figure 15. Small pedestrian detection results of different networks. To visually demonstrate the detection performance, the relevant confidence values are not displayed. Some large target pedestrians appear in (**a**,**b**), while smaller targets appear in (**c**–**f**).

Table 3. Cross-validation results with k = 5 (%).

Evaluation Metrics	Division 1	Division 2	Division 3	Division 4	Division 5	Average
F1	70.6	71.0	70.2	70.3	70.2	70.5
mAP	78.1	78.6	77.9	78.0	78.4	78.2

Figure 16. Examples of added background images.

Table 4. Results of ablation experiments (%).

ConvNeXt v2	ODConv	Tiny Head	*Precision*	*Recall*	*F1*	*mAP*	Speed (ms)
-	-	-	68.3	69.5	68.9	76.1	10.3
√	-	-	68.8	70.1	69.4	77.0	14.4
√	√	-	69.1	70.2	69.6	77.7	14.8
-	√	-	69.0	69.2	69.1	76.7	10.5
-	√	√	68.9	70.3	69.6	77.3	16.0
-	-	√	68.8	69.2	69.0	76.4	15.7
√	√	√	69.7	70.7	70.2	78.1	18.2

Note: √ means adopting the module; - means without the module.

Moreover, to further verify the performance of the proposed network, multiple one-stage and two-stage object detection models were trained in this study, and the results are described in Table 5. In general, the one-stage object detection models exhibit superior performance compared to the two-stage models. On one hand, this superiority is attributed to the effective structures and modules utilized by single-stage algorithms; on the other hand, this might be due to the incorporation of supervised signals regarding relationships directly into the feature maps, which made the learned context information richer.

Table 5. Comparison of comprehensive evaluation results of different models (%).

Category	Network	*Precision*	*Recall*	*F1*	*mAP*	Speed (ms)
One-stage	YOLOv5s	70.6	67.1	68.8	75.3	11.4
	YOLOv8s	68.3	69.5	68.9	76.1	10.3
	Proposed-WIoU	70.0	70.8	70.4	78.4	18.2
	SSD	61.5	62.4	61.9	70.8	13.7
Two-stage	Faster-R-CNN	48.4	52.4	50.3	57.3	47.5
	Mask R-CNN	51.9	57.2	54.4	61.5	62.4
	Cascade R-CNN	60.1	58.7	59.4	64.2	56.7

4.3. Small Reservoir Dam Inspection and Localization Using UAV—Case 1

4.3.1. Analysis of UAV Measurement Accuracy after Coordinate Transformation

The developed method was first applied to a small earth–rock dam, as shown in Figure 17a. The dam has a height of 13 m, and the length of the crest is 75 m. Moreover,

the dam serves as a road that can support traffic. After prolonged operation, the downstream surface of the dam was covered with weeds, and there was a lack of comprehensive monitoring data. To establish the local coordinate system of the dam efficiently and economically, 14 GCPs were set up on the dam crest, and their 3D coordinates were measured using a Leica total station (Figure 17b). The UAV used is a DJI M300 RTK (Figure 17c) equipped with an H20T camera that consists of optical imaging, infrared thermal imaging, and laser ranging capabilities. To enhance the positioning accuracy, the UAV network RTK mode was implemented.

(**a**) (**b**) (**c**)
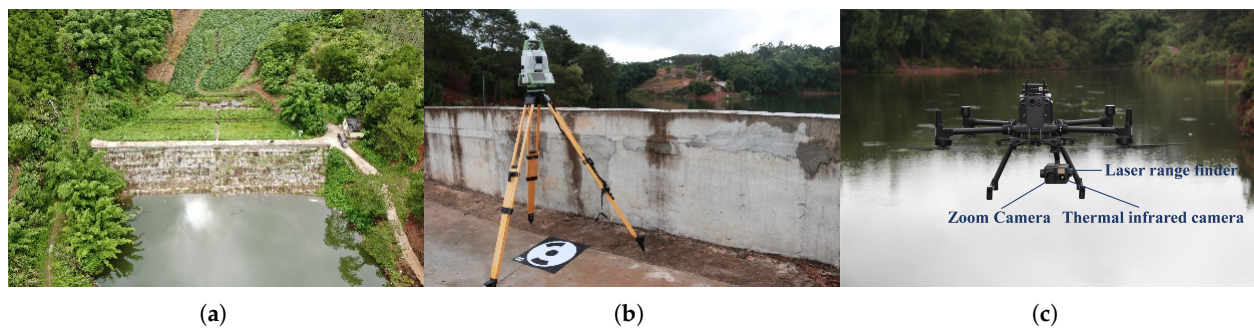
Figure 17. Photos of the small reservoir dam and measurement equipment for Case 1. (**a**) Overall scene of the dam. (**b**) Total station and one GCP. (**c**) M300 RTK UAV and H20T camera.

To adapt to a larger range of UAV inspections, a total of 1215 images were taken to calculate the coordinate transformation parameters, including images from locations far from the dam. The 14 GCPs were then remeasured from 14 random positions, and the results are presented in Table 6. Similar to the experimental results, the errors in the planar direction are smaller than those in the altitude direction, measuring 7.47 cm and 9.49 cm, respectively, verifying the broad applicability of the developed method. However, the measurement error for GCP 4 is significantly high and is attributed to water accumulation on the ground causing significant errors in the laser ranging results.

Table 6. Measurement errors of 14 GCPs on the dam crest for Case 1 (cm).

No	ΔX	ΔY	ΔXY	ΔZ
GCP 1	−3.49	−8.92	9.58	−10.56
GCP 2	−4.77	−0.04	4.77	−6.77
GCP 3	−10.12	0.05	10.12	7.67
GCP 4	15.77	−2.80	16.01	−16.00
GCP 5	−7.63	−6.07	9.75	12.68
GCP 6	−0.28	5.47	5.48	2.82
GCP 7	−7.53	−6.64	10.04	−6.12
GCP 8	−8.18	−9.85	12.80	0.31
GCP 9	−10.02	2.37	10.29	7.67
GCP 10	−2.17	−7.71	8.01	8.70
GCP 11	−3.27	−11.73	12.18	9.96
GCP 12	1.28	−3.28	3.52	−5.22
GCP 13	−3.43	−11.20	11.72	6.66
GCP 14	−8.36	−1.65	8.52	3.42
Mean |Δ|	6.16	5.56	7.47	9.49

A 3D reconstruction of the dam area was created to further evaluate the accuracy of the developed method. The point clouds generated before and after applying coordinate transformation are compared with the results from laser scanning, as shown in Figure 18. To eliminate noise interference, the point cloud data from the trees near the dam were removed. It can be observed that the point cloud generated after coordinate transformation is closer to the laser scanning results, as shown in Figure 18b, with most of the points

having distances within 10 cm. However, there are still some large errors, mainly caused by the poor performance of the SfM algorithm in generating point clouds on the downstream surface covered with weeds. The point cloud comparison results indicate that the coordinate transformation method can improve the localization accuracy and enable a UAV-based routine inspection method.
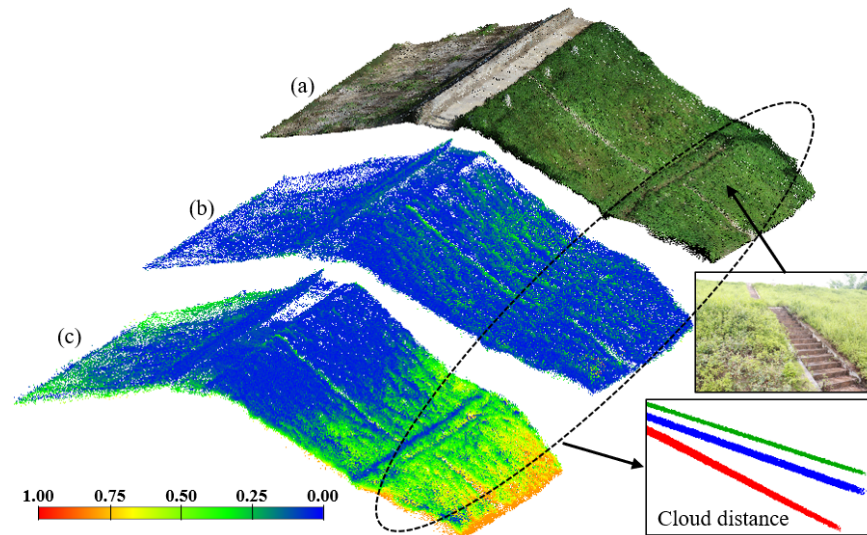

Figure 18. Point clouds generated from UAV images of a laser scan for Case 1. (**a**) Laser scan. (**b**) Transformed points. (**c**) No transformation.

4.3.2. 3D Localization of Pedestrians in Small Reservoir Dams

The developed noncontact method in combination with an improved YOLOv8 detection algorithm enables the analysis of whether pedestrians are present in hazardous areas. Some detection and localization results are shown in Figure 19. Based on the measured coordinates in the dam coordinate system, it is possible to determine the structural positions of pedestrians. For example, if a pedestrian appears near the downstream surface close to the reservoir area, an alert can be triggered. Additionally, virtual electronic fences can be constructed to promptly detect pedestrians in hazardous situations [64]. Compared to manual inspection, the proposed method offers the advantages of efficiency and low cost and is particularly suitable for small reservoir dams without automated monitoring devices.

4.3.3. Potential Seepage Detection and Localization

For small earth–rock dams, seepage is a major cause for alarm. Given that the data collection occurred during the summer, the temperature in the deeper layers of the reservoir is expected to be lower than the ambient temperature. Based on this, the primary criterion for determining seepage occurrence in this study is the manual identification of whether low-temperature areas appear on the downstream face. The UAV camera is equipped with an infrared thermal imaging sensor, capable of detecting the temperature field on the dam surface. A total of 770 infrared images were captured within the dam area, and a 3D model was generated as shown in Figure 20a. It can be observed that there are two anomalous areas on the dam surface, as indicated by the infrared image, where the temperature is lower than the surrounding environment, as shown in Figure 20b,c. However, not all the low-temperature areas can be conclusively identified as seepage. For instance, the temperature at the drainage channels at the bottom of the dam is also low due to the presence of accumulated water in Figure 20d. Combining the proposed method enables the 3D localization of this position. Nevertheless, certain locations may display temperature anomalies, as shown in Figure 20e. By incorporating their positional information, it can be analyzed that these locations do not fall within the dam area, thus avoiding false alarms. Considering that the detection of seepage still requires further analysis of the internal

conditions of the dam, to maintain rigor, in the absence of additional measurements such as osmotic pressure, this study refers to the occurrence of seepage as potential seepage.
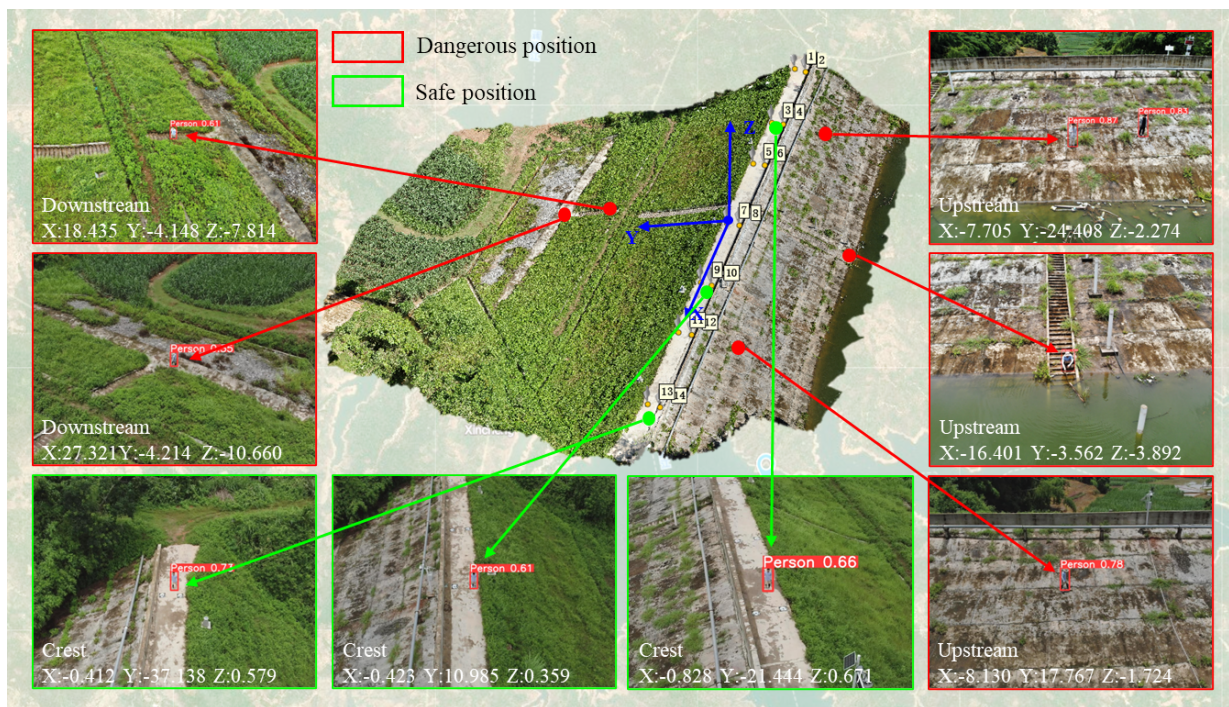

Figure 19. The 3D localization of pedestrians in the dam coordinate system and the analysis of their positions within the dam structure for Case 1.
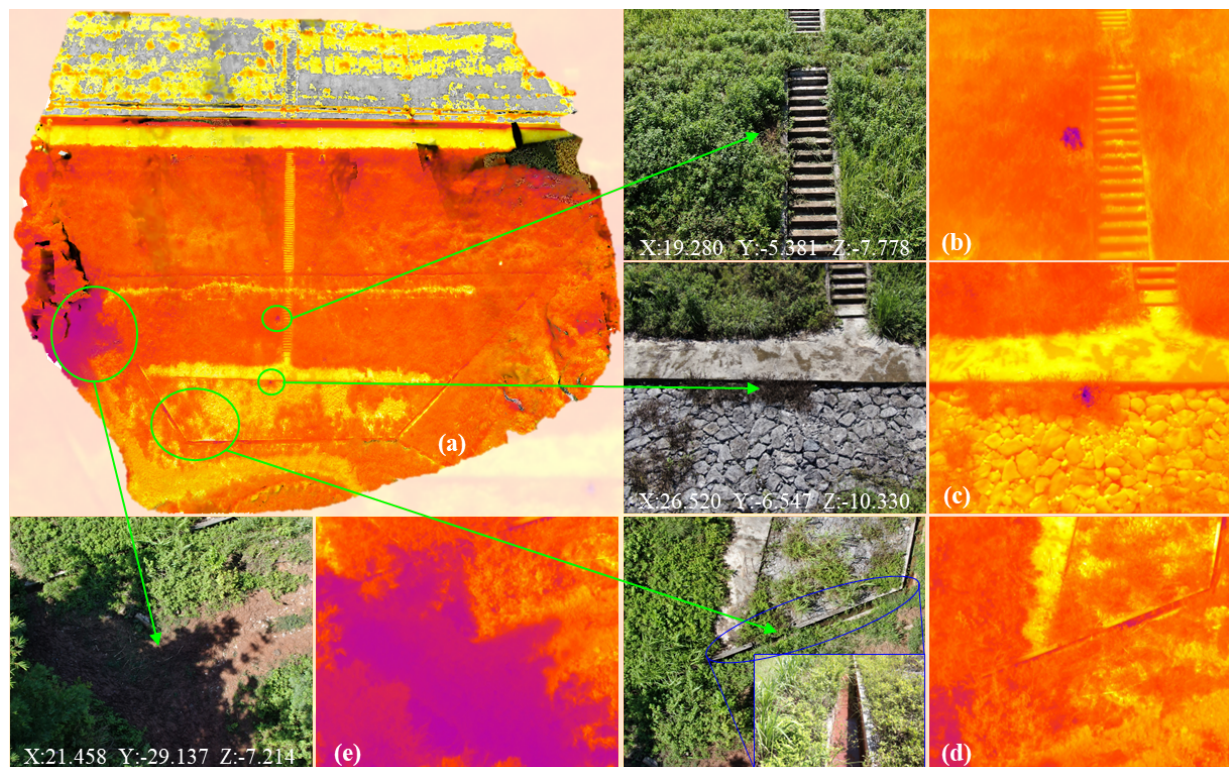

Figure 20. Detection and localization of potential seepage regions based on infrared thermography for Case 1. (**a**) Overall image. (**b**–**e**) Details of visible light and infrared images.

4.4. Small Reservoir Dam Inspection and Localization Using UAV—Case 2

4.4.1. Coordinate Transformation and Measurement Accuracy Analysis

To validate the broad applicability of the proposed method, the coordinate transformation approach was applied to another earth–rock dam. This dam has a maximum height of 21.12 m and a length of 108 m (Figure 21). Similarly, 14 GCPs were placed on the dam crest, utilizing the same UAV and measurement equipment as in Case 1. The origin of the dam local coordinate system was established at the crest.

Figure 21. Photograph of the small reservoir dam for Case 2.

A total of 1688 UAV images were captured for the purposes of coordinate transformation and 3D reconstruction. Following the coordinate transformation, noncontact measurements were carried out on the 14 GCPs from 14 distinct positions. The resulting measurement errors are detailed in Table 7. The average error in the plane is 7.19 cm, while the average error in the altitude direction 8.44 cm, demonstrating a consistent trend where the accuracy of the planar positioning surpasses that of the vertical positioning. The measurement results for Case 2 indicate that the proposed method is applicable to different dam environments.

Table 7. Measurement errors of 14 GCPs on the dam crest for Case 2 (cm).

No	ΔX	ΔY	ΔXY	ΔZ
GCP 15	0.22	−0.01	0.22	3.07
GCP 16	7.59	−4.81	8.98	6.62
GCP 17	12.67	−7.11	14.53	8.59
GCP 18	7.58	−3.94	8.55	8.22
GCP 19	1.51	1.57	2.17	2.50
GCP 10	6.28	−1.90	6.56	14.64
GCP 21	−0.51	−7.40	7.41	4.17
GCP 22	3.31	−2.06	3.89	9.80
GCP 23	−2.88	−2.19	3.62	5.71
GCP 24	−7.20	−1.98	7.47	11.76
GCP 25	−4.43	2.41	5.04	11.02
GCP 26	−11.54	6.84	13.41	14.85
GCP 27	−10.53	1.63	10.66	9.65
GCP 28	−7.84	2.23	8.15	7.56
Mean |Δ|	6.01	3.29	7.19	8.44

The dam laser scanning data are shown in Figure 22a, with the point cloud of trees near the reservoir removed. Using the scan results as a reference, the results between the point clouds generated by the SfM algorithm before and after coordinate transformation can be compared, as shown in Figure 22b,c. Although the model generated directly from the GCPs can depict the shape and structural dimensions, the point cloud generated after the coordinate transformation is noticeably closer to the scanning results. Most of the distances are less than 20 cm, and the errors are distributed relatively evenly from the dam crest to the bottom. In addition, there are some high error points near the crest, mainly caused by noise points such as weeds. Overall, accurate measurement results can be obtained after coordinate transformation for Case 2.
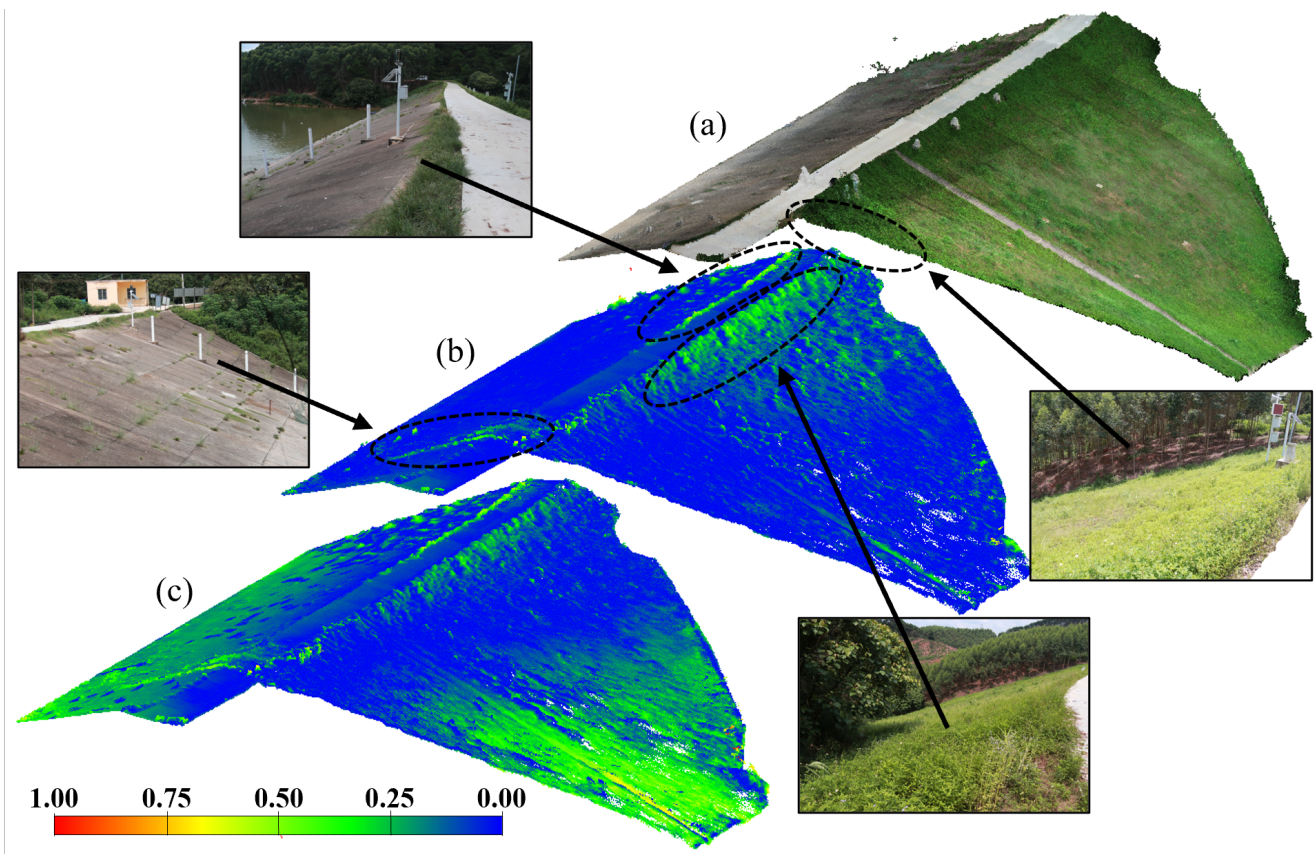
Figure 22. Point clouds generated from UAV images and a laser scan for Case 2. (**a**) Laser scan. (**b**) Transformed points. (**c**) No transformation.

4.4.2. Pedestrian Detection and Localization in Local Coordinate System

Combined with the improved YOLOv8 detection algorithm, several pedestrians at various locations on the surface of the dam can be detected and localized within the local coordinate system and specific structural positions, as shown in Figure 23. Based on the calculated 3D coordinates, it was determined that some pedestrians were present near the reservoir. When the UAV is equipped with a speaker, it can be used to promptly emit an alert. Additionally, since the dam serves a road, pedestrians detected on the crest will not trigger an alarm, thereby establishing a virtual electronic fence. With continuous monitoring, the system can record the prolonged presence of pedestrians, enabling safety management in small dam environments.
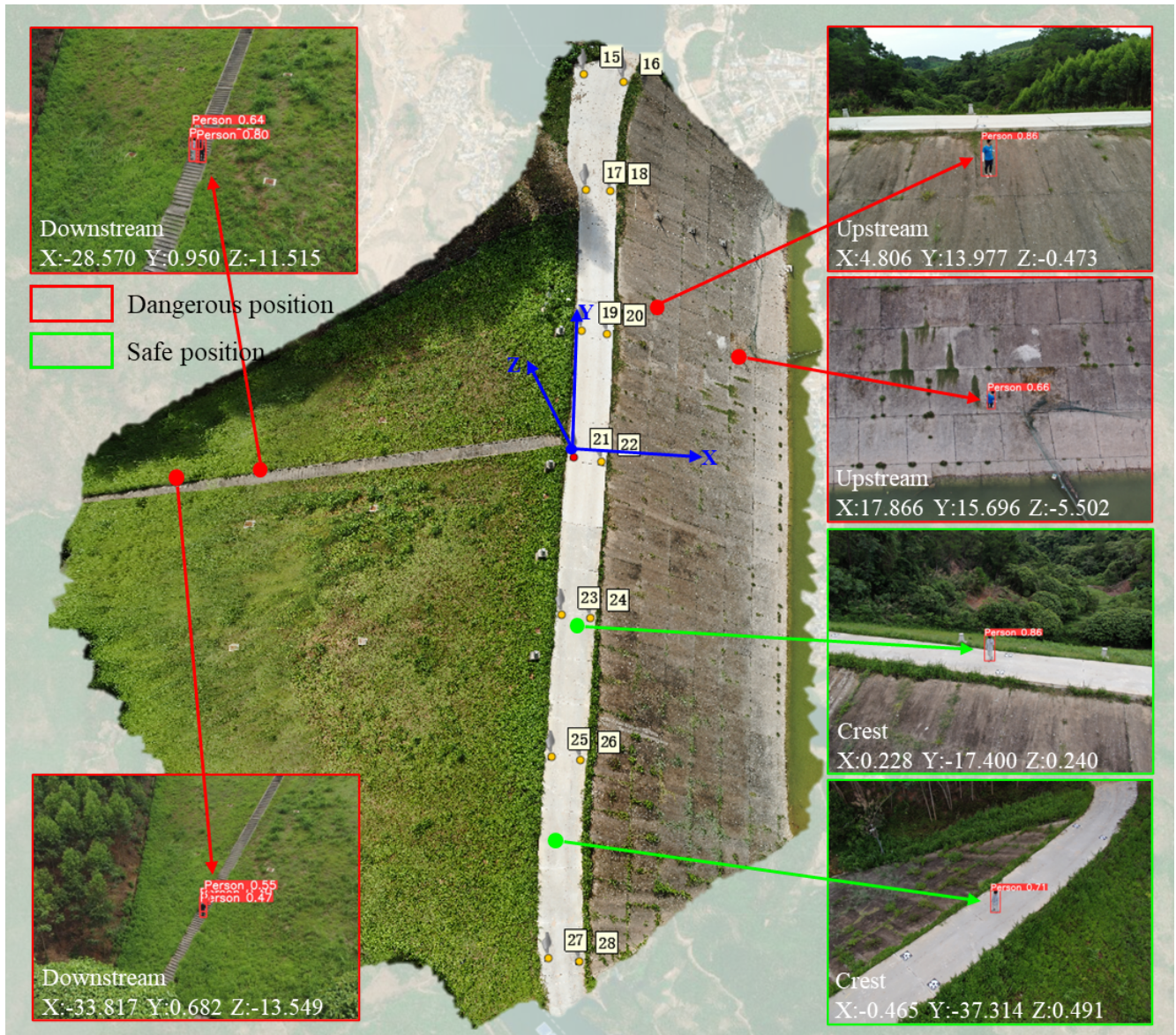
Figure 23. 3D localization of pedestrians in the dam coordinate system and the analysis of their positions within the structure for Case 2.

4.4.3. Potential Seepage Detection Using Infrared Thermography

Infrared thermal images of the dam area were collected by the UAV to assess the surface seepage. A total of 875 infrared images were used to generate the 3D model, as shown in Figure 24a. There are a total of three areas with lower temperatures, which were manually identified as potential seepage points, as shown in Figure 24b–d, all distributed near the bottom of the dam. Comparing the thermal infrared imaging with the water-filled channel reveals similar temperature values, indicating that this area requires special attention. After multiple data collections by the UAV, it is possible to determine whether an area has experienced anomalies by analyzing the temperature or texture variations at the same location at different time periods. These two cases demonstrate the applicability of the proposed method for long-term UAV inspection and localization.
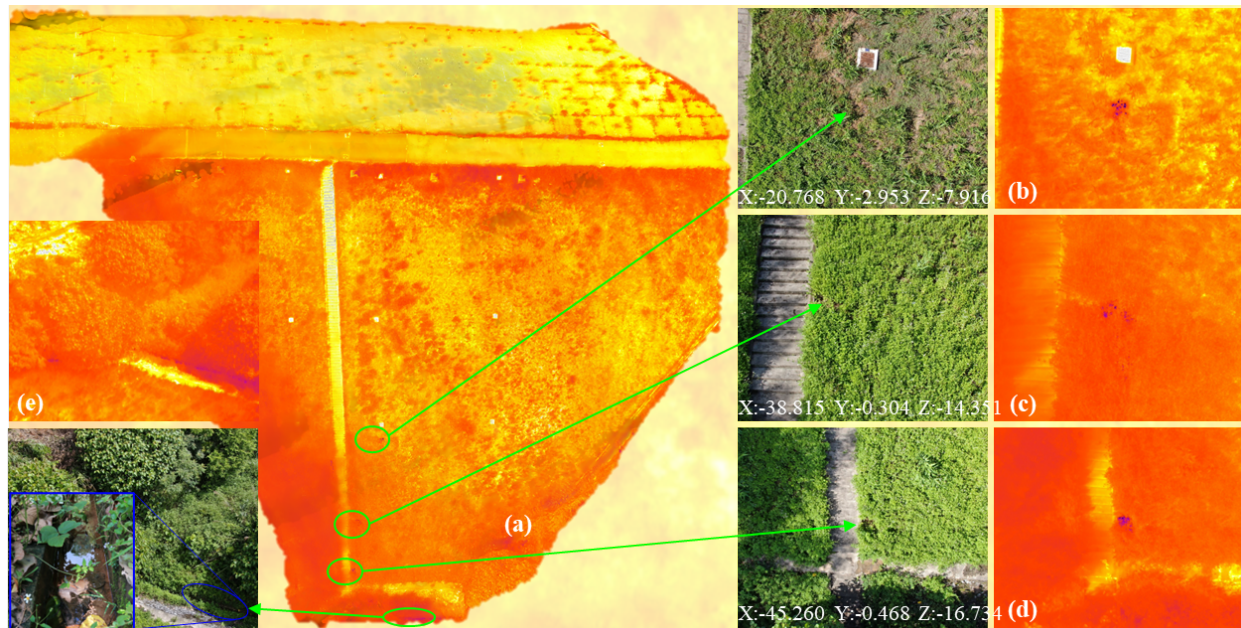
Figure 24. Detection and localization of potential seepage regions based on infrared thermography for Case 2. (**a**) Overall image. (**b–e**) Details of visible light and infrared images.

5. Conclusions

This paper proposed a UAV coordinate transformation and pedestrian localization method designed for small reservoir dams. It effectively addressed the problems associated with the absence of sufficient control points and terrain deformation, which complicate the establishment of a coordinate system due to the absence of geographic information. The precision and applicability of the proposed method were verified by experimental validation and application in two case studies, demonstrating its cost-effectiveness in accurately detecting pedestrians and identifying structural anomalies within small reservoir dams. The major contributions of the study include the following:

1. A coordinate transformation method for converting two coordinate systems without reference points using the SfM algorithm was developed. The UAV 3D coordinates were projected onto a plane using the Gauss–Kruger method and then transformed to the local coordinate system of the small reservoir dam through four-parameter plane transformation and an altitude fitting. The coordinate registration was calculated by the SfM algorithm with the introduction of several GCPs. After the transformation, the dam point cloud generated from the UAV images closely matched the laser scanning results, with errors mostly below 0.2 m and minimal sensitivity to changes in altitude.

2. An improved YOLOv8 algorithm for the accurate detection of small pedestrians in UAV images was also conceived. To detect targets with a limited number of pixels, a tiny head channel was introduced into the original structure. Additionally, ConvNeXt v2 modules and the ODConv algorithm were integrated into the small target detection sections of the backbone, thereby enhancing the feature extraction capability. The *mAP* value of the improved YOLOv8 network can reach up to 78.4%, an improvement of 2.8% compared to the original network. Therefore, the method is more accurate and efficient at detecting small pedestrians in practical scenarios.

3. Combined with laser rangefinding, the developed method enables noncontact 3D measurement of the target by UAV photogrammetry. To convert 2D coordinates to actual 3D coordinates, the capturing distance and camera pose were determined, and the coordinates of the target point were calculated based on the UAV position. The experimental results indicate an overall positioning error of less than 0.2 m, and, in two case studies,

the measured plane and altitude average errors were both less than 0.1 m. Moreover, both pedestrians and potential seepage points can be accurately located.

Compared to the current UAV inspection approaches, the proposed method in this study is suitable for small reservoir dams. From a methodological perspective, unlike 3D reconstruction [25], the proposed method does not require excessive time to capture 3D coordinates in complex structural areas or rely on other structural features for coordinate transformation. The proposed approach only requires a single coordinate transformation, making it more suitable for evaluating the safety of identified individuals in small reservoir dam inspections. In terms of accuracy, the proposed method incorporates a laser rangefinder, effectively converting 2D pixels into 3D coordinates. When relying solely on measurement methods based on the spatial relationship between the image and the structure, the error could be around 1 m. However, in this study, the positioning error was within 0.2 m, which meets the inspection positioning requirements for small reservoir dams. For practicality, the method developed in this research requires minimal additional equipment. The coordinate system of the small reservoir dam does not need to be pre-established; the UAV coordinates can be transformed using ground control point coordinates alone. This is particularly important for meeting the low-cost demands of small reservoir dam inspections. Additionally, the proposed method enables the creation of a virtual electronic fence for small reservoir dams, as illustrated in Figure 25. Thus, compared to the current state-of-the-art methods, the proposed approach is more suitable for small reservoir dams, supporting safety management and guiding disaster prevention strategies [65,66].
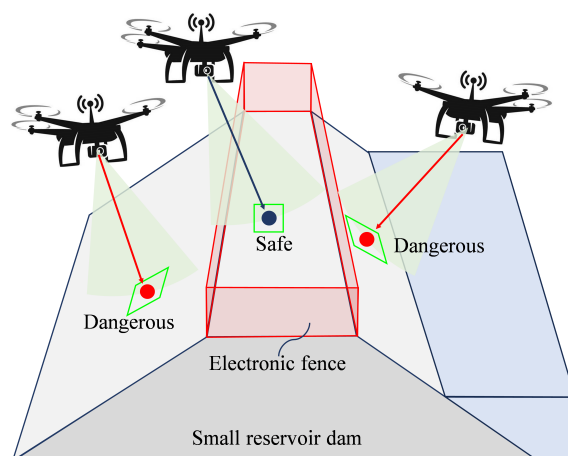

Figure 25. Schematic diagram of electronic fence for small reservoir dams using UAVs.

However, the method presented in this paper still has some limitations. First, the positioning accuracy of the UAV is dictated by satellite signals. Although the study adopted the network RTK mode to enhance the signal strength, fluctuations in positioning can lead to significant measurement errors [67]. Second, the localization results for pedestrians were determined at their midpoint and do not correspond to their positions on the dam surface. This could introduce interference when assessing the critical positions of pedestrians in danger zones. Finally, the detection of seepage is still based on empirical judgment and lacks a theoretical baseline and standardized discernment criteria based on temperature anomaly recognition [68].

In future research, detection algorithms for various types of damages should be developed, and they should be mapped to real-time digital models [69]. For network optimization, significant computational resources should be invested to fine-tune the network architecture and hyperparameters while also increasing the dataset size to further enhance the pedestrian detection accuracy. Additionally, methods that integrate image and LiDAR data should be considered to quickly detect targets while simultaneously determining the corresponding point cloud locations, thereby enabling the detection of multiple targets on complex geometric surfaces. In the application of edge devices, ensuring

smooth data transmission is a crucial issue that needs to be addressed. Additionally, there is a need to consider and mitigate the problem of data transmission delays caused by network instability. Moreover, inspection results for different time periods should be compared, and the progression of damages should be analyzed [70] to identify the rate of change without requiring initial conditions [71], enabling long-term dam health monitoring. With sufficient data, it should be possible to predict the dam response and the evolution of damages [72].

Institutional Review Board Statement: No ethical approvals are required.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets presented in this article are not readily available because authors do not have the relevant permissions.

Conflicts of Interest: Author Yiqing Si was employed by the company Chongqing Surveying and Design Institute Co. Ltd. of Water Resources. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Habets, F.; Molénat, J.; Carluer, N.; Douez, O.; Leenhardt, D. The cumulative impacts of small reservoirs on hydrology: A review. *Sci. Total Environ.* **2018**, *643*, 850–867. [CrossRef] [PubMed]
2. Kang, F.; Li, J.; Zhao, S.; Wang, Y. Structural health monitoring of concrete dams using long-term air temperature for thermal effect simulation. *Eng. Struct.* **2019**, *180*, 642–653. [CrossRef]
3. Gong, J.; Zou, D.; Kong, X.; Liu, J.; Qu, Y. An approach for simulating the interaction between soil and discontinuous structure with mixed interpolation interface. *Eng. Struct.* **2021**, *237*, 112035. [CrossRef]
4. Chen, X.; Li, D.; Tang, X.; Liu, Y. A three-dimensional large-deformation random finite-element study of landslide runout considering spatially varying soil. *Landslides* **2021**, *18*, 3149–3162. [CrossRef]
5. Arvor, D.; Daher, F.R.; Briand, D.; Dufour, S.; Rollet, A.J.; Simoes, M.; Ferraz, R.P. Monitoring thirty years of small water reservoirs proliferation in the southern Brazilian Amazon with Landsat time series. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 225–237. [CrossRef]
6. Mignot, E.; Dewals, B. Hydraulic modelling of inland urban flooding: Recent advances. *J. Hydrol.* **2022**, *609*, 127763. [CrossRef]
7. Zhou, R.; Su, H.; Wen, Z. Experimental study on leakage detection of grassed earth dam by passive infrared thermography. *NDT E Int.* **2022**, *126*, 102583. [CrossRef]
8. Sedláček, J.; Bábek, O.; Grygar, T.M.; Lenďáková, Z.; Pacina, J.; Štojdl, J.; Hošek, M.; Elznicová, J. A closer look at sedimentation processes in two dam reservoirs. *J. Hydrol.* **2022**, *605*, 127397. [CrossRef]
9. Guo, Y.; Cui, Y.a.; Xie, J.; Luo, Y.; Zhang, P.; Liu, H.; Liu, J. Seepage detection in earth-filled dam from self-potential and electrical resistivity tomography. *Eng. Geol.* **2022**, *306*, 106750. [CrossRef]
10. Kim, S.Y.; Kwon, D.Y.; Jang, A.; Ju, Y.K.; Lee, J.S.; Hong, S. A review of UAV integration in forensic civil engineering: From sensor technologies to geotechnical, structural and water infrastructure applications. *Measurement* **2023**, *224*, 113886. [CrossRef]
11. Spencer, B.F., Jr.; Hoskere, V.; Narazaki, Y. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* **2019**, *5*, 199–222. [CrossRef]
12. Zhu, Y.; Tang, H. Automatic damage detection and diagnosis for hydraulic structures using drones and artificial intelligence techniques. *Remote Sens.* **2023**, *15*, 615. [CrossRef]
13. Guzman-Acevedo, G.M.; Quintana-Rodriguez, J.A.; Vazquez-Becerra, G.E.; Garcia-Armenta, J. A reliable methodology to estimate cable tension force in cable-stayed bridges using Unmanned Aerial Vehicle (UAV). *Measurement* **2024**, *229*, 114498. [CrossRef]
14. Jiang, S.; Wang, Y.; Zhang, J.; Zheng, J. Full-field deformation measurement of structural nodes based on panoramic camera and deep learning-based tracking method. *Comput. Ind.* **2023**, *146*, 103840. [CrossRef]

15. Xu, Z.; Wu, Y.; Lu, X.; Jin, X. Photo-realistic visualization of seismic dynamic responses of urban building clusters based on oblique aerial photography. *Adv. Eng. Inform.* **2020**, *43*, 101025. [CrossRef]

16. Wu, Y.; Meng, F.; Qin, Y.; Qian, Y.; Xu, F.; Jia, L. UAV imagery based potential safety hazard evaluation for high-speed railroad using Real-time instance segmentation. *Adv. Eng. Inform.* **2023**, *55*, 101819. [CrossRef]

17. Feng, S.; Gao, M.; Jin, X.; Zhao, T.; Yang, F. Fine-grained damage detection of cement concrete pavement based on UAV remote sensing image segmentation and stitching. *Measurement* **2024**, *226*, 113844. [CrossRef]

18. Narazaki, Y.; Hoskere, V.; Chowdhary, G.; Spencer, B.F., Jr. Vision-based navigation planning for autonomous post-earthquake inspection of reinforced concrete railway viaducts using unmanned aerial vehicles. *Autom. Constr.* **2022**, *137*, 104214. [CrossRef]

19. Xu, Y.; Zhang, J. UAV-based bridge geometric shape measurement using automatic bridge component detection and distributed multi-view reconstruction. *Autom. Constr.* **2022**, *140*, 104376. [CrossRef]

20. Chen, G.; Liang, Q.; Zhong, W.; Gao, X.; Cui, F. Homography-based measurement of bridge vibration using UAV and DIC method. *Measurement* **2021**, *170*, 108683. [CrossRef]

21. Zhou, R.; Wen, Z.; Su, H. Automatic recognition of earth rock embankment leakage based on UAV passive infrared thermography and deep learning. *ISPRS J. Photogramm. Remote Sens.* **2022**, *191*, 85–104. [CrossRef]

22. Qin, L.; Yan, C.; Yu, L.; Chai, M.; Wang, B.; Hayat, M.; Shi, Z.; Gao, H.; Jiang, X.; Xiong, B.; et al. High-resolution spatio-temporal characteristics of urban evapotranspiration measured by unmanned aerial vehicle and infrared remote sensing. *Build. Environ.* **2022**, *222*, 109389. [CrossRef]

23. Zhong, X.; Zhao, L.; Wang, J.; Zhang, X.; Nie, Z.; Li, Y.; Ren, P. A retrieval method for land surface temperatures based on UAV broadband thermal infrared images via the three-dimensional look-up table. *Build. Environ.* **2022**, *226*, 109793. [CrossRef]

24. Zhou, X.; Zhu, Q.; Zhang, Q.; Du, Y. The full-field displacement intelligent measurement of retaining structures using UAV and 3D reconstruction. *Measurement* **2024**, *227*, 114311. [CrossRef]

25. Zhao, S.; Kang, F.; Li, J.; Ma, C. Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction. *Autom. Constr.* **2021**, *130*, 103832. [CrossRef]

26. Chen, S.; Fan, G.; Li, J. Improving completeness and accuracy of 3D point clouds by using deep learning for applications of digital twins to civil structures. *Adv. Eng. Inform.* **2023**, *58*, 102196. [CrossRef]

27. Szostak, B.; Specht, M.; Burdziakowski, P.; Stateczny, A.; Specht, C.; Lewicka, O. Methodology for performing bathymetric measurements of shallow waterbodies using an UAV, and their processing based on the SVR algorithm. *Measurement* **2023**, *223*, 113720. [CrossRef]

28. Lee, G.; Moon, S.; Hwang, J.; Chi, S. Development of a real-time noise estimation model for construction sites. *Adv. Eng. Inform.* **2023**, *58*, 102133. [CrossRef]

29. Jiang, S.; Wu, Y.; Zhang, J. Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system. *Autom. Constr.* **2023**, *146*, 104685. [CrossRef]

30. Chen, J.; Li, S.; Lu, W. Align to locate: Registering photogrammetric point clouds to BIM for robust indoor localization. *Build. Environ.* **2022**, *209*, 108675. [CrossRef]

31. Tan, Y.; Li, G.; Cai, R.; Ma, J.; Wang, M. Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Autom. Constr.* **2022**, *139*, 104284. [CrossRef]

32. Ren, C.; Jiao, Y.; Liu, Y.; Shang, H. Optimal camera focal length detection method for GPS-supported bundle adjustment in UAV photogrammetry. *Measurement* **2024**, *228*, 114329. [CrossRef]

33. Chen, J.; Lu, W.; Lou, J. Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 1079–1098. [CrossRef]

34. Wu, C.; Chen, B. An automatic measurement system for the wall thickness of corrugated plate based on laser triangulation method. *Adv. Eng. Inform.* **2023**, *55*, 101814. [CrossRef]

35. Lenda, G.; Marmol, U. Integration of high-precision UAV laser scanning and terrestrial scanning measurements for determining the shape of a water tower. *Measurement* **2023**, *218*, 113178. [CrossRef]

36. Zhuge, S.; Xu, X.; Zhong, L.; Gan, S.; Lin, B.; Yang, X.; Zhang, X. Noncontact deflection measurement for bridge through a multi-UAVs system. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 746–761. [CrossRef]

37. Xu, Z.; Kang, R.; Lu, R. 3D reconstruction and measurement of surface defects in prefabricated elements using point clouds. *J. Comput. Civ. Eng.* **2020**, *34*, 04020033. [CrossRef]

38. Tan, Y.; Li, S.; Liu, H.; Chen, P.; Zhou, Z. Automatic inspection data collection of building surface based on BIM and UAV. *Autom. Constr.* **2021**, *131*, 103881. [CrossRef]

39. Liu, D.; Chen, J.; Hu, D.; Zhang, Z. Dynamic BIM-augmented UAV safety inspection for water diversion project. *Comput. Ind.* **2019**, *108*, 163–177. [CrossRef]

40. Chang, G.; Xu, T.; Wang, Q. Error analysis of the 3D similarity coordinate transformation. *GPS Solut.* **2017**, *21*, 963–971. [CrossRef]

41. Lee, E.; Park, S.; Jang, H.; Choi, W.; Sohn, H.G. Enhancement of low-cost UAV-based photogrammetric point cloud using MMS point cloud and oblique images for 3D urban reconstruction. *Measurement* **2024**, *226*, 114158. [CrossRef]

42. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In Proceedings of the Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999; Springer: Berlin/Heidelberg, Germany, 2000; pp. 298–372. [CrossRef]

43. Claessens, S.J. Efficient transformation from Cartesian to geodetic coordinates. *Comput. Geosci.* **2019**, *133*, 104307. [CrossRef]

44. Kern, A.; Bobbe, M.; Khedar, Y.; Bestmann, U. OpenREALM: Real-time mapping for unmanned aerial vehicles. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020; pp. 902–911. [CrossRef]

45. Han, Y.; Wu, G.; Feng, D. Vision-based displacement measurement using an unmanned aerial vehicle. *Struct. Control Health Monit.* **2022**, *29*, e3025. [CrossRef]

46. Jiang, S.; Zhang, J.; Gao, C. Bridge Deformation Measurement Using Unmanned Aerial Dual Camera and Learning-Based Tracking Method. *Struct. Control Health Monit.* **2023**, *2023*, 4752072. [CrossRef]

47. Hui, Y.; Wang, J.; Li, B. STF-YOLO: A small target detection algorithm for UAV remote sensing images based on improved SwinTransformer and class weighted classification decoupling head. *Measurement* **2024**, *224*, 113936. [CrossRef]

48. Li, Y.; Bao, T.; Li, T.; Wang, R. A robust real-time method for identifying hydraulic tunnel structural defects using deep learning and computer vision. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 1381–1399. [CrossRef]

49. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. Available online: https://github.com/ultralytics/ultralytics (accessed on 11 March 2024).

50. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]

51. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986. [CrossRef]

52. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 16133–16142. [CrossRef]

53. Li, C.; Zhou, A.; Yao, A. Omni-dimensional dynamic convolution. *arXiv* **2022**, arXiv:2209.07947. [CrossRef]

54. Zhang, Z.; Liu, Y.; Xue, W. MS-IRTNet: Multistage information interaction network for RGB-T semantic segmentation. *Inf. Sci.* **2023**, *647*, 119442. [CrossRef]

55. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 2778–2788. [CrossRef]

56. Zhao, S.; Kang, F.; Li, J. Concrete dam damage detection and localisation based on YOLOv5s-HSC and photogrammetric 3D reconstruction. *Autom. Constr.* **2022**, *143*, 104555. [CrossRef]

57. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef] [PubMed]

58. Li, X.; Chen, J.; He, Y.; Yang, G.; Li, Z.; Tao, Y.; Li, Y.; Li, Y.; Huang, L.; Feng, X. High-through counting of Chinese cabbage trichomes based on deep learning and trinocular stereo microscope. *Comput. Electron. Agric.* **2023**, *212*, 108134. [CrossRef]

59. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.

60. Agisoft Metashape. Available online: www.agisoft.com (accessed on 11 March 2023).

61. Zhang, S.; Xie, Y.; Wan, J.; Xia, H.; Li, S.Z.; Guo, G. Widerperson: A diverse dataset for dense pedestrian detection in the wild. *IEEE Trans. Multimed.* **2019**, *22*, 380–393. [CrossRef]

62. Du, D.; Zhu, P.; Wen, L.; Bian, X.; Lin, H.; Hu, Q.; Peng, T.; Zheng, J.; Wang, X.; Zhang, Y.; et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019. [CrossRef]

63. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014. [CrossRef]

64. Luo, H.; Liu, J.; Fang, W.; Love, P.E.; Yu, Q.; Lu, Z. Real-time smart video surveillance to manage safety: A case study of a transport mega-project. *Adv. Eng. Inform.* **2020**, *45*, 101100. [CrossRef]

65. Liu, G.; Zhong, Z.; Ma, K.; Bo, W.; Zhao, P.; Li, Y.; Zhang, Z.; Zhang, P. Field experimental verifications of 3D DDA and its applications to kinematic evolutions of rockfalls. *Int. J. Rock Mech. Min. Sci.* **2024**, *175*, 105687. [CrossRef]

66. Chen, D.; Huo, L.; Song, G. High resolution bolt pre-load looseness monitoring using coda wave interferometry. *Struct. Health Monit.* **2022**, *21*, 1959–1972. [CrossRef]

67. He, L.; He, X.; Huang, Y.; Yang, C. A new approach for estimating geocenter motion based on BDS-3 plane-specific orbit error correction model. *GPS Solut.* **2023**, *27*, 204. [CrossRef]

68. Cheng, C.; Shang, Z.; Shen, Z. Automatic delamination segmentation for bridge deck based on encoder-decoder deep learning through UAV-based thermography. *NDT E Int.* **2020**, *116*, 102341. [CrossRef]

69. He, Z.; Wang, Y.H.; Zhang, J. Generative Structural Design Integrating BIM and Diffusion Model. *arXiv* **2023**, arXiv:2311.04052.

70. Liu, X.; Li, Z.; Sun, L.; Khailah, E.Y.; Wang, J.; Lu, W. A critical review of statistical model of dam monitoring data. *J. Build. Eng.* **2023**, *80*, 108106. [CrossRef]

71. Ni, F.; Zhang, J.; Taciroglu, E. Development of a moving vehicle identification framework using structural vibration response and deep learning algorithms. *Mech. Syst. Signal Process.* **2023**, *201*, 110667. [CrossRef]

72. Yuan, D.; Gu, C.; Wei, B.; Qin, X.; Gu, H. Displacement behavior interpretation and prediction model of concrete gravity dams located in cold area. *Struct. Health Monit.* **2023**, *22*, 2384–2401. [CrossRef]