



UAV-assisted bridge alignment measurement using enhanced small target detection and adaptive ellipse fitting

Lu Deng ^{a,b,c}, Cheng Zhang ^a, Weiqi Mao ^d, Feng Zhang ^a, Lizhi Long ^a, Hao Dai ^d, Jingjing Guo ^{a,*}

^a College of Civil Engineering, Hunan University, Changsha, China

^b Key Laboratory of Damage Diagnosis for Engineering Structures of Hunan Province, Hunan University, Changsha, China

^c State Key Laboratory of Bridge Safety and Resilience, Hunan University, Changsha, China

^d China Railway Major Bridge Engineering Group CO.,LTD, China

ARTICLE INFO

Keywords:

Prefabricated bridge
Alignment measurement
Unmanned aerial vehicle (UAV)
Small target detection
Ellipse fitting

ABSTRACT

Prefabricated bridges are preferred in modern construction for their rapid assembly, cost-efficiency, and minimal environmental impact. However, traditional alignment methods, such as total stations and levels, are time-consuming and labor-intensive. This paper proposes a UAV-based alignment measurement system using artificial markers for vertical alignment in prefabricated bridges. The key contributions include: (1) a high-precision UAV system framework based on overlapping marker centers and point lattice stitching; (2) a YOLOv8-based detection network, YOLO-USMD, for precise marker identification in aerial images; and (3) a Dynamic Adaptive Multi-Scale Ellipse Detection (DAMSED) algorithm to improve marker detection in complex images. Field experiments on a prefabricated steel-concrete bridge demonstrated that the proposed method achieved a root mean square error (RMSE) of 2.84 mm for a 30 m range, proving its effectiveness for accurate and efficient alignment assessment in bridge construction.

1. Introduction

Prefabricated bridges, as an innovative construction method, have gained widespread adoption due to their advantages, including rapid assembly, consistent quality, environmental sustainability, and energy efficiency [1]. These bridges are being increasingly used in urban expressways, highways, and large-scale cross-river and cross-sea projects, particularly in scenarios such as emergency response and urban redevelopment [2]. The quality of prefabricated bridges is critical to their safety and service life. Key quality assessment factors include material strength, structural stability, and alignment precision [3]. Among these, bridge alignment is a particularly important indicator as it directly influences the geometric accuracy and structural load distribution of the bridge. Bridge alignment is generally categorized into horizontal and vertical alignment. Horizontal alignment pertains to the geometric layout within the plane, encompassing straight lines, curves, and turning points. This type of alignment ensures that the bridge's layout adheres to design standards and meets traffic safety requirements. While horizontal alignment is essential for ensuring proper traffic flow and adherence to design plans, it is typically determined during the design phase and is less prone to significant deviations during construction due to its

relatively static nature. In contrast, vertical alignment deals with the undulations of the bridge deck and is more dynamic during the construction phase. It ensures that the constructed bridge's elevation and positioning adhere to design specifications. Unlike horizontal alignment, vertical alignment directly reflects the bridge's structural condition and is highly susceptible to deviations caused by construction inaccuracies, material deformation, or environmental factors [4]. These deviations can lead to localized load imbalances, stress concentrations, and potential structural damage. Such issues not only compromise the bridge's durability and operational safety but also increase long-term maintenance costs. Therefore, vertical alignment is selected as the focus for measurement and control during construction, as it plays a pivotal role in ensuring the bridge's structural integrity and safety.

According to the *Engineering Surveying Standards* (GB50026–2020) [5], the maximum allowable error for vertical alignment measurement is generally within 10 mm. Timely quality control also requires rapid detection and immediate results. However, traditional tools like total stations and level instruments, while highly accurate, are labor-intensive and cumbersome [6]. When using a total station to detect the vertical alignment of an assembled bridge, it is essential to first ensure the proper operation of the equipment and establish reference points. The total

* Corresponding author.

E-mail address: guojingjing@hnu.edu.cn (J. Guo).

station is then set up at the survey station, ensuring it is leveled correctly. Subsequently, retro-reflectors are positioned at key measurement points, such as supports and bridge deck joints, and the coordinates of these points relative to the station are measured and recorded. Through coordinate system transformation, the measured data is converted into world coordinates within the control network, which is then compared with the design standards to promptly identify vertical installation errors and assess whether the bridge installation precision meets the required specifications. Effective on-site construction guidance requires prompt feedback of alignment deviations to enable immediate adjustment during installation. However, these traditional tools are not only expensive to purchase and maintain, but they also require skilled personnel for operation and upkeep. Additionally, in challenging construction environments, such as high altitudes or steep cliffs, operators face significant safety risks [7]. Consequently, there is a pressing demand for an alignment measurement technology that is not only efficient, safe, and cost-effective but also capable of addressing the limitations of traditional methods, particularly for prefabricated bridge construction.

In recent years, digital and intelligent technologies for structural geometric information extraction have gained significant attention for bridge alignment measurement. These methods primarily rely on two types of data: point clouds and images. Point cloud-based methods, such as those using LiDAR or structured light cameras, offer high precision and perform well in complex environments. However, they are costly and demand substantial computational resources [8–10]. Image-based methods, on the other hand, utilize RGB cameras and offer a more cost-effective alternative, though their accuracy and stability are often limited [11–13]. For data collection, ground-fixed systems offer stable monitoring but lack the flexibility required for large-scale alignment tasks [14]. Unmanned Aerial Vehicles (UAVs) offer significant advantages in mobility, cost-efficiency, and safety. They enable rapid and continuous alignment measurement by covering large areas quickly, reducing manual labor, and minimizing the risks associated with ground-based inspections. This flexibility allows for frequent data collection, supporting fast monitoring and timely adjustments [15–19]. The typical process of UAV-based alignment measurement includes deploying artificial markers (AM) to enhance positioning accuracy, identifying and locating AM, reconstructing the 3D coordinates of the marker centers, and stitching the coordinates of the marker centers [11,16,17]. Unlike traditional ground-based fixed systems, artificial markers are passive reference points rather than active measurement devices. They can be flexibly deployed and quickly adjusted according to the bridge structure and measurement requirements. Lightweight and portable, artificial markers can be strategically placed at key locations, reducing setup constraints. Additionally, relying on UAV image acquisition, artificial markers provide extensive coverage without the need for repeated installation, making them particularly suitable for measuring large-span bridges and significantly improving measurement efficiency and coverage range. Despite their potential, challenges remain in achieving millimeter-level accuracy and operational efficiency. For example, while Shi et al. [20] demonstrated 2 mm accuracy over 28.5 m range using UAV, this level of precision required densely placed targets and low flight altitudes, which increased operational complexity and costs. In construction site environments, detecting small elliptical targets presents a dual challenge of both marker and ellipse detection. First, due to the small size of markers in high-resolution images, accurate localization becomes more difficult. Additionally, the complexity of the construction site environment, including background interference, obstructions, and lighting variations, further complicates marker recognition, especially when the boundary between the marker and background becomes blurred, leading to potential false negatives and false positives. For ellipse detection, the YOLO framework generates rectangular bounding boxes that cannot accurately enclose tilted markers, resulting in incomplete coverage and affecting detection performance. Moreover, due to lighting changes or viewing angles, the

surface of the marker often produces arc-like shadows, which resemble the shape of the marker itself, increasing the risk of false positives. In particular, under complex lighting conditions, shadows and reflections may be mistakenly identified as additional targets, exacerbating the problem of false detection.

To optimize the balance between accuracy and efficiency in UAV-based bridge alignment measurement, this study introduces a framework to achieve automatic bridge alignment assessment that is designed specifically for complex construction environments. The research focuses on solving key problems, including efficient detection and localization of small markers, precise 3D coordinate reconstruction, and accurate alignment stitching across multi-span bridge segments. To tackle these issues, the framework incorporates the following innovative algorithms: 1) a network named You Only Look Once - UAV Small Marker Detection (YOLO-USMD), designed to rapidly and accurately detect these small markers; 2) a Dynamic Adaptive Multi-Scale Ellipse Detection (DAMSED) algorithm is proposed innovatively, which enhances marker center localization in challenging environments; 3) the five-point algorithm combined with the triangulation principle for camera pose estimation and 3D coordinates reconstruction of the marker centers, 4) a sequential point lattice stitching method based on overlapping markers to achieve high-precision alignment of different bridge segments. Detailed descriptions of the framework and its validation through experiments are provided in the following sections.

2. Related work

This section provides an overview of key technologies relevant to the key challenges in bridge alignment measurement, including small object detection, elliptical detection algorithms for positioning, and methods employing RGB cameras to calculate the 3D coordinates of target points.

2.1. Small object detection algorithms

Small object detection is a particularly challenging task in computer vision, with difficulties arising not only from the inherent characteristics of the targets but also from the balance between data distribution, model design, and computational efficiency. In recent years, with the rapid development of deep learning, object detection algorithms have been widely applied in fields such as autonomous driving, medical image analysis, and video surveillance [21]. However, even mainstream methods like Faster R-CNN [22] and YOLO [23] still exhibit significant limitations when it comes to detecting small objects [24] prompting researchers to continually explore more targeted solutions.

The primary reason for the difficulty in detecting small objects is the scarcity of their feature information. Due to the small size of the target in images, the features are often overshadowed by complex backgrounds, making it challenging for models to effectively extract distinctive features. To address this issue, techniques such as dilated convolutions, multi-scale feature extraction, and attention mechanisms have been introduced to enhance the resolution and contextual information of small objects. Among them, dilated convolutions help to capture more contextual information by increasing the receptive field, thereby improving the detection accuracy of small objects. For example, Cui et al. [25] proposed the Context-Aware Block Net, which improves detection performance through a pyramid dilated convolution module but faces challenges related to sparsity and computational complexity. Fu et al. [26] introduced dilated convolutions into the SSD network (DFSSD), enhancing information flow between low-level and high-level feature maps, achieving outstanding performance on the PASCAL VOC2007 and KITTI datasets. Mahaur et al. [27] optimized YOLOv5 to improve the accuracy of traffic sign and signal light detection; however, computational overhead and the hollow effect remain challenges. Multi-scale feature extraction enhances a model's ability to recognize objects of different sizes. Zhang et al. [28] proposed EPSANet, which utilizes a pyramid structure and channel squeeze operations to significantly

improve the detection sensitivity of small objects. However, as the scale increases, information redundancy and computational overhead gradually emerge, affecting real-time performance and complexity, particularly in autonomous driving and security surveillance. Attention mechanisms have been widely applied to strengthen the model's focus on key regions. Yang et al. [29] improved Faster-RCNN by introducing multiple attention modules, enhancing small object detection accuracy, especially in low-resolution and complex backgrounds. Lim et al. [30] developed a residual attention mechanism that reduces background interference, further improving detection performance under low-resolution conditions, though computational overhead may still impact real-time detection. To improve small object detection accuracy, researchers have proposed feature fusion strategies. Lin et al. [31]'s FPN integrates multi-scale features effectively through a top-down pathway and lateral connections, significantly enhancing small object detection. PANet [32] strengthens bottom-up pathways and adaptive feature pooling, achieving excellent performance in small object detection, though with higher computational complexity. MLFPN [33] further enhances small object detection by alternating slim U-shaped modules and feature fusion modules, though this also increases computational complexity. BiFPN [34] addresses information loss in low-resolution feature processing encountered by FPN through bidirectional feature fusion, delivering superior performance across multiple datasets. However, it struggles to capture the details of small objects during feature fusion.

Another challenge in small object detection is the insufficient number of small object samples. In standard datasets, the proportion of small objects typically does not exceed 40 %, leading to a lack of diverse samples for model training [35]. To address this issue, data augmentation techniques have become a key breakthrough. Data augmentation effectively expands the training dataset, and both supervised and unsupervised methods have shown significant potential in this field. Supervised data augmentation methods expand small object samples using target label information, improving detection accuracy and robustness. Zeng et al. [36] proposed a multi-scale feature fusion method that leverages techniques such as AutoAugment and Mixup to automatically select augmentation strategies, enhancing small object detection performance, reducing overfitting, and improving model adaptability to different scenarios. However, excessive augmentation may lead to computational overhead and performance issues with extremely small or heavily occluded objects. Kisantal et al. [37] improved small object diversity through oversampling and copy-pasting techniques, though this approach may negatively impact large object detection. Zhang et al. [38] introduced the CDCI strategy, which enhances detection accuracy by expanding contextual information around small objects, although false positives remain a challenge in dense and occluded scenes. Wang et al. [39] increased training data diversity with image flipping, color jittering, and noise addition, achieving outstanding results in dead tree detection, though false positives may still occur in complex environments. Unsupervised data augmentation improves model performance by using synthetic data, reducing dependence on annotated datasets. For example, Bosquet et al. [40] proposed the DS-GAN method, which synthesizes small objects and inserts them into images, enhancing small object detection accuracy, particularly in autonomous driving and satellite image analysis. Li et al. [41]'s Perceptual GAN method employs super-resolution to enhance small object features, improving detection accuracy, though differences between generated features and real objects in complex backgrounds limit its application. Akyon et al. [42] proposed a tiling augmentation technique that divides images into small patches, improving small object detection accuracy in high-resolution images, but the associated computational and memory overhead may negatively impact large object detection.

Accurate anchor box matching is another major challenge in small object detection. The Intersection over Union (IoU) values of small object regions can drastically decrease due to slight offsets, resulting in poor performance of traditional IoU loss in these scenarios. The

Generalized Intersection over Union (GIoU) [43] addresses the zero-gradient problem of IoU when bounding boxes do not overlap by introducing an additional term, effectively improving localization accuracy and convergence speed for small object detection, especially in complex backgrounds and occlusion scenarios. However, Giou entails considerable computational overhead, particularly in high-dimensional spaces or for targets with complex shapes. The Distance Intersection over Union (DIoU) [44] optimizes the distance between the center points of boxes, mitigating gradient vanishing issues, significantly accelerating convergence, and enhancing accuracy for small object detection, though it still requires integration with other techniques for extremely small objects and complex backgrounds. SIoU [45] optimizes bounding box regression by considering angle, distance, and shape, improving localization accuracy in multi-scale object detection, but its effectiveness remains limited for complex shapes or large-scale objects. Soft-CIoU [46] improves upon CIOU loss to enhance the learning capability for small objects, achieving outstanding performance in complex backgrounds and occlusion scenarios, albeit with increased computational overhead. The Normalized Wasserstein Distance (NWD) [47] models bounding boxes as Gaussian distributions, addressing IoU's sensitivity issues in small object detection, significantly improving accuracy, particularly on the AI-TOD and VisDrone2019 datasets. However, its computational complexity remains high, especially in scenarios with complex targets and substantial background noise.

Although substantial advancements have been made in small object detection, detecting AM in complex environments such as bridge construction sites remains challenging. Small objects in natural scenes, such as vehicles and pedestrians, often exhibit distinct shape and color features, allowing existing algorithms to perform well. However, AM often lack pronounced differences in shape, color, or size from their surroundings, especially in construction environments where they tend to blend into the background. This results in a higher likelihood of missed detections or false positives. Furthermore, most existing datasets primarily are designed for small objects in natural scenes, offering limited applicability to AM. Consequently, directly applying these algorithms and datasets to construction scenarios yields suboptimal results. Addressing this issue necessitates the development of dedicated algorithms and tailored datasets that are specifically designed for efficient and accurate detection of AM in such environments.

2.2. Elliptical detection algorithms for localization

After detecting small markers in aerial images, the next critical step is to precisely locate their centers for accurate spatial reconstruction. Since artificial markers are designed as circular targets, their center points serve as reference coordinates for alignment measurement. However, due to factors like shooting angles and geometric distortion, circles in images often appear as ellipses. This deformation makes direct circular detection unreliable, necessitating the use of robust elliptical detection algorithms to accurately determine marker centers. In response to this challenge, various classic and influential algorithms have been proposed: McLaughlin et al. [48] proposed a two-stage decomposition-based random Hough transform (RHT) algorithm. By dividing the parameter space into two stages (first determining the ellipse center, then solving for the remaining parameters), this method effectively reduces computational time and complexity. However, it lacks a robust verification strategy, which can still lead to false detections. Wu et al. [49] proposed a real-time robust ellipse detection method based on direct least-square fitting, which can accurately detect ellipses in images and achieves high processing speed. However, the method may be affected by noise and non-elliptical edge points when handling complex scenes. Prasad et al. [50] introduced a method that utilizes the curvature and convexity characteristics of edges to generate smooth edge contours. They applied a 2D Hough transform to rank candidate edges and generate better ellipse hypotheses. However, the computational efficiency of this approach remains relatively low.

Fornaciari et al. [51] proposed an algorithm to improve detection speed by detecting and removing line segments. This method accelerates ellipse detection by comparing the area of regions to assess the convexity of arc segments. However, it leads to a decrease in localization accuracy. Jia et al. [52] improved the detection strategy of [51] by calculating the feature number of connected components to distinguish between lines and conic sections, thereby more efficiently filtering line segments. Although the speed has improved, this method is prone to repeated detections. Pătrăucean et al. [53] introduced the ELSDc algorithm, which extends line segment detection to a multi-functional detector. This method uses an improved line segment detection (LSD) algorithm to detect line segments and directly generates candidate lines and ellipses from grayscale images. By employing local grouping and verification strategies, it enhances detection performance. However, it may overlook global information, leading to an increased false detection rate. Lu et al. [54] proposed an efficient ellipse detection algorithm based on arc segment-supported line segments. They group arc segment-supported line segments based on geometric features and generate initial ellipse candidates using a combination of local and global methods. Detection accuracy is then improved through multi-level clustering and strict verification, significantly enhancing detection efficiency and robustness in complex scenarios. However, this method relies on preset parameters, which may not be suitable under varying image conditions, thus limiting detection accuracy and stability. Cheng et al. [11] proposed a deep learning-based ellipse detection method in markers for evaluating the installation quality of close-range pre-fabricated components. This method relies on close-range frontal imaging and achieves precise installation guidance through marked point detection and 3D reconstruction. However, it does not consider small object detection in oblique photography, which limits its accuracy in UAV-based bridge alignment measurement.

2.3. 3D coordinate estimation for targets using RGB camera systems

Techniques for estimating 3D coordinates of targets using RGB cameras primarily include stereoscopic camera systems, binocular systems, and multi-view geometry-based triangulation. Stereoscopic cameras typically employ a fixed short baseline design, which is suitable for close-range (0.1–10 m) scenarios, offering high accuracy at short distances. However, their accuracy diminishes for long-range measurements [11,55–58].

Binocular systems, which consist of two independent cameras, offer greater flexibility by allowing adjustable baseline length. This adaptability enables them to maintain high accuracy and wide-range measurements in scenarios ranging from 0.2 to 25 m, making it particularly suitable for long-distance measurements [59–61]. In contrast, multi-view geometry-based triangulation captures images from different viewpoints. It uses feature point matching and camera pose estimation to reconstruct the 3D structure of the target. Finally, triangulation determines the spatial coordinates of the feature points. This method provides high accuracy and robustness in large-scale, long-distance, and complex scenarios. It is well-suited for large-scale measurement tasks that require flexible shooting positions. However, it has a high computational cost. When handling many viewpoints, feature matching and pose estimation require significant processing power. A high-performance computer is needed for efficient execution [62–64].

Given the large-scale and long-distance characteristics of bridge construction in this study, the method of using UAVs for multi-view image acquisition combined with triangulation to solve the 3D coordinates of marker centers is selected. This method balances precision and efficiency, meets the flexibility demands of complex environments, and provides a reliable foundation for accurate alignment measurements.

3. Methodology

Building on the insights and advancements outlined in the previous sections, this study proposes a comprehensive approach to bridge alignment measurement that combines deep learning-based target recognition and localization with efficient camera pose estimation and 3D reconstruction strategies.

As illustrated in Fig. 1, the proposed method establishes a systematic framework for accurate and efficient bridge alignment measurement. The process begins with the rapid identification of small markers in aerial images using a designed YOLO-USMD network, ensuring high detection accuracy and efficiency. The precise localization of marker centers is achieved through the DAMSED algorithm. Subsequently, the five-point algorithm is applied to estimate camera poses, while the triangulation principle is used to reconstruct the 3D coordinates of the markers. To further enhance accuracy, global optimization is performed using bundle adjustment. To address the challenges of long-span bridges, a sequence point lattice stitching method based on overlapping markers is introduced, enabling the alignment of adjacent measurement segments to generate a unified 3D coordinate system for accurate detection of large-scale bridge alignment. By seamlessly integrating marker detection, 3D coordinate reconstruction, and alignment stitching, the proposed approach offers a robust solution for UAV-based bridge alignment measurement in complex construction environments.

3.1. Deep learning-based detection of small marker centers in UAV tilted photographic images

AM are specially designed visual patterns that are installed on the bridge deck to assist in detecting and localizing target points during image-based measurements. These markers typically consist of distinguishable geometric shapes with unique features to facilitate identification. The markers used in this study are shown in Fig. 2 and are classified into six types, labeled (a) to (f). Each type consists of a central large circle for precise localization of its center coordinates. The rectangular structures surrounding the circle are used to uniquely number markers. The high contrast and structured design of these markers ensure reliable detection even under varying lighting and environmental conditions, making them suitable for complex construction sites.

In UAV oblique photogrammetry, circular markers often appear as ellipses due to shooting angles and geometric distortions, making fast detection challenging for traditional algorithms. This study introduces a deep learning-based approach to detect marker centers in aerial images, streamlining alignment measurement through three key steps:

- (1) Dataset Establishment: Target patterns are designed, and a comprehensive dataset is built using data augmentation techniques such as Mosaic augmentation, horizontal flipping, noise addition, and brightness adjustment.
- (2) Marker Identification: A YOLO-USMD network is trained to detect and classify markers, removing irrelevant background and isolating regions of interest for further processing. Details of this step are presented in Section 3.1.1.
- (3) Ellipse detection and localization: The DAMSED refines localization with multi-level verification and inlier refinement strategies, ensuring accurate detection even in complex environment. Details of this step are presented in Section 3.1.2.

3.1.1. UAV small marker detection

This study employs a target detection algorithm for classifying and localizing AM, selecting YOLO for its speed, accuracy, and ease of use. Specifically, YOLOv8 network is chosen for detection, because of its improved network design and optimization techniques. It offers enhanced performance in speed and accuracy while remaining well-documented, robust, and reliable for practical applications [65]. The

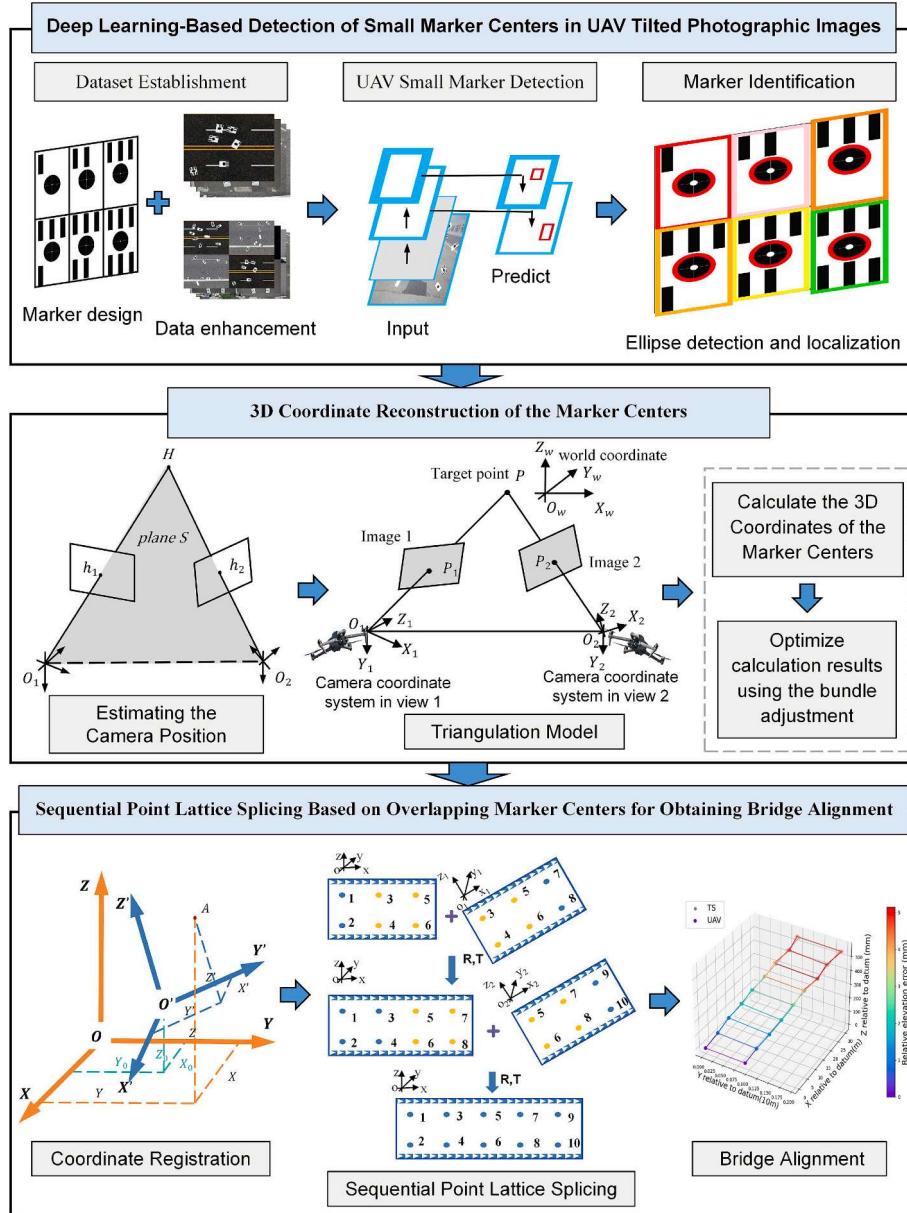


Fig. 1. Overview of the UAV-based photogrammetry system for bridge alignment measurement using AM.

architecture of YOLOv8 consists of three main components: the backbone, the neck, and the detection head. The backbone, built on Cross Stage Partial Network (CSPNet), hierarchically processes the input images to extract multi-scale features while optimizing model efficiency. The neck employs a standard Feature Pyramid Network (FPN) to fuse multi-scale features, enabling better adaptation to both large and small object detection tasks. However, the one-way feature fusion method in FPN can struggle with effectively representing small-scale features. Finally, the detection head takes a decoupled design to predict object classes and locations, efficiently handling object detection tasks. In high-resolution UAV images, AM occupy very few pixels, leading to challenges for the original YOLOv8 model. The standard feature pyramid struggles to capture these small targets, resulting in missed detections. Additionally, the original Intersection over Union (IoU) loss function is overly sensitive to slight positional changes, causing false detections and multiple detections of the same marker.

To effectively address the challenges of small target detection in high-resolution UAV images, this paper improves the YOLOv8 network and proposes the YOLO-USMD target detection network, as shown in

Fig. 3. Overall, the target detection network addresses the challenges of small marker detection in high-resolution UAV images by improving feature extraction, multi-scale fusion, and anchor box localization precision, progressively enhancing the model's detection performance for small markers after training. Small markers, due to their small pixel proportion and vulnerability to background interference, pose significant challenges for traditional methods in terms of feature extraction, multi-scale feature representation, and bounding box localization. To address these challenges, a deformable attention mechanism [66] is integrated after the SPPF module in the feature extraction stage, enabling the model to focus on critical regions where small markers are located while reducing information redundancy and computational overhead. This enhancement provides higher-quality feature inputs for subsequent processing. Building on this foundation, the head of the network is restructured. Compared to the original design, three additional convolutional layers (Conv) were added before feature fusion for preliminary feature extraction and feature channel adjustment, laying the foundation for subsequent multi-scale feature fusion. These Conv not only strengthen the feature extraction process but also provide more

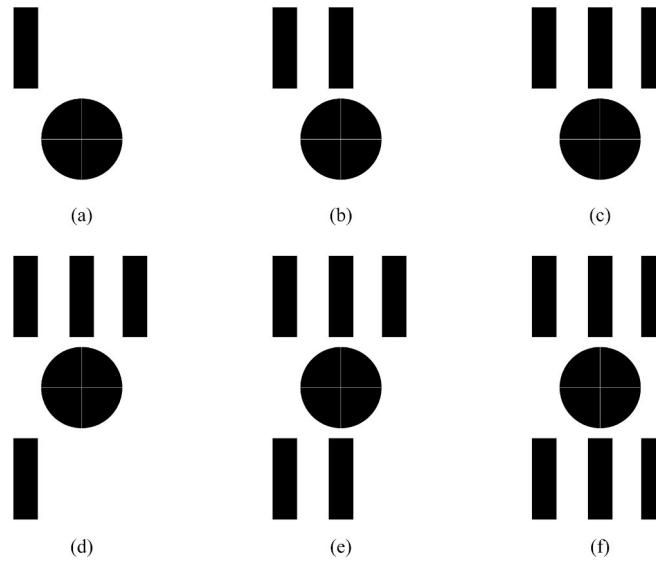


Fig. 2. AM: (a) marker type 1, (b) marker type 2, (c) marker type 3, (d) marker type 4, (e) marker type 5, and (f) marker type 6.

representative multi-scale inputs for the BiFPN module, optimizing the integration of features across different scales. Additionally, by unifying the channel numbers across feature maps at different levels, the efficiency of feature fusion is improved, information loss is reduced, and the representation capability of multi-scale features is enhanced. As a result, the network performs more effectively in handling marker size variations caused by changes in object distance and shooting angles. During the feature fusion phase, feature maps are no longer directly concatenated after upsampling or downsampling. Instead, bidirectional fusion of multi-scale feature maps is achieved through the BiFPN module [34], enhancing the expressiveness of multi-scale features. As shown in Fig. 3, different colors of circular modules represent different feature fusion nodes. The red circles denote the fusion of low-level features (P3), preserving more detailed information. The yellow circles indicate the fusion of high-level features (P5), which enhance semantic representation by incorporating fine-grained details from lower levels. The purple circles represent intermediate fusion nodes, which integrate top-down and bottom-up information flows and optimize the transmission of features (P4). After each fusion step, the feature maps are further refined using designated node processing modules. Finally, these processed feature maps (P3, P4, P5) are passed to the detection head to perform target detection across different scales. Fig. 4 provides a detailed illustration of the modifications made to YOLOv8, resulting in the YOLO-USMD structure. Additionally, the NWD [47] is introduced as a new metric to replace the traditional IoU loss function. The NWD models bounding boxes as 2D Gaussian distributions, offering greater flexibility in handling geometric deformations, such as transformations from

circles to ellipses. Unlike IoU, NWD does not require precise alignment, making it robust to angular and geometric variations. Even when the bounding box and ground truth box do not overlap ($|P \cap G| = 0$), NWD can still compute a meaningful distance, ensuring continuous optimization. Furthermore, NWD is less sensitive to scale and positional shifts, making it well-suited for small object detection by reducing the reliance on precise alignment. It also provides higher-quality positive samples for small objects, mitigating the issue of excessive negative samples seen with IoU, thereby improving training effectiveness. In complex deformation and misalignment scenarios, NWD minimizes errors, enhancing the accuracy and robustness of UAV-based detection tasks.

3.1.2. Ellipse detection and localization

To overcome the limitations of existing methods in handling noise, this paper proposes DAMSED based on Lu's method [54]. The dynamic multi-level angular coverage validation strategy, as shown in Fig. 5, initializes the counter to 0 and the remaining angle θ to 360° , representing a complete coverage range. The process begins by sequentially evaluating whether each candidate ellipse satisfies the angular coverage condition $\alpha < \theta$. For ellipses that do not meet this condition, the candidate ellipse is preliminarily validated to be an ellipse target, completing the angular coverage validation and proceeding to subsequent calculations. For ellipses that satisfy the condition $\alpha < \theta$, the counter is incremented by 1, and the remaining angle θ is dynamically updated using the formula $\theta = 360 - \text{counter} \times \gamma$. The iteration continues until θ exceeds the angular coverage threshold δ , at which point the dynamic multi-level angular coverage validation process is

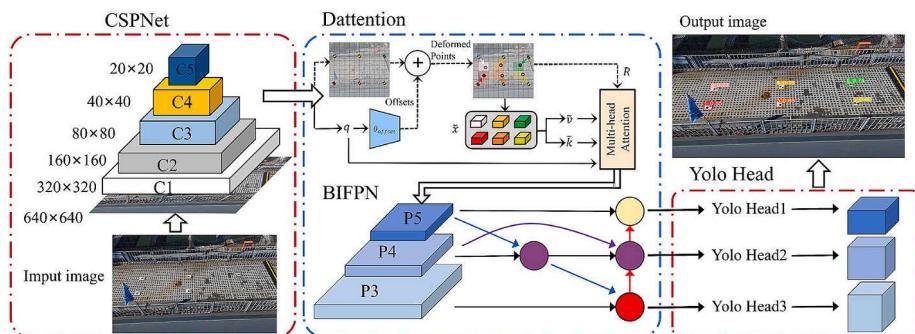


Fig. 3. YOLO-USMD network.

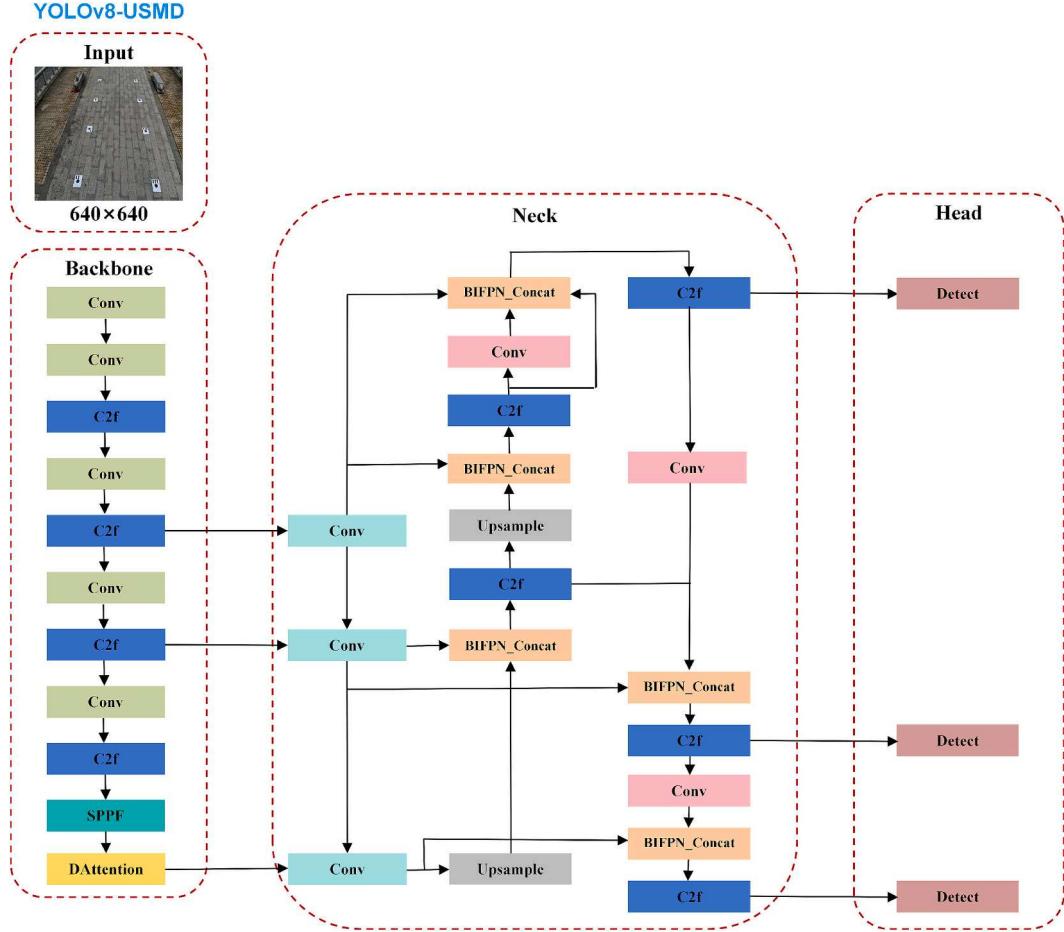


Fig. 4. YOLO-USMD structure.

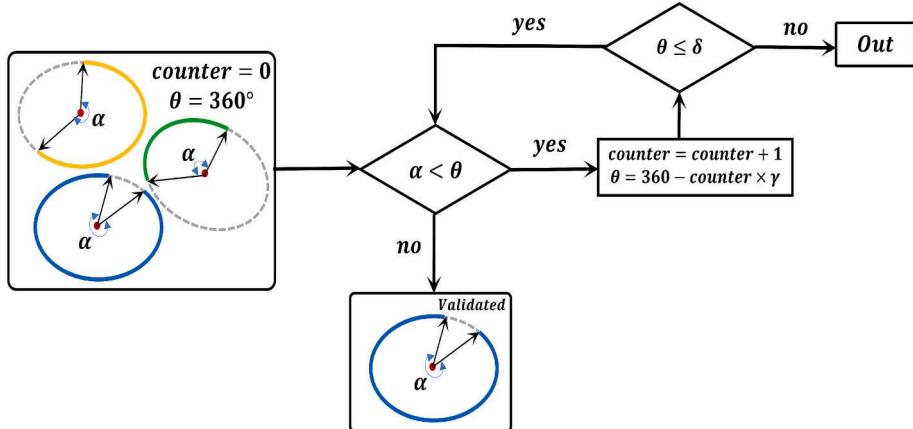


Fig. 5. Dynamic multi-level angle coverage validation strategy.

terminated. The parameter γ represents the step increment, a fixed value used to adjust the remaining angle during each iteration. A smaller γ is more suitable for fine-grained validation, while a larger γ is better suited for rapid coverage adjustments. By dynamically adjusting the coverage range through iterative processes, this strategy improves adaptability and accuracy compared to rigid single-threshold methods. Furthermore, successive rounds of validation gradually reduce the number of candidate objects, keeping the computational workload of subsequent algorithms relatively manageable. This makes the strategy particularly efficient in complex scenarios. To finely distinguish the true ellipse

edges, an enhanced in-point refinement mechanism is employed, which integrates normal vector analysis and connectivity information. Initially, candidate edge points are extracted using an edge detection algorithm, and noise points are filtered based on normal vectors and neighborhood consistency. For the retained edge points, their connectivity on the ellipse boundary is further analyzed to remove isolated noise points. The normal direction of in-points is then dynamically adjusted using error feedback to enhance the accuracy of the ellipse boundary points. Finally, ellipse fitting is performed on the optimized edge points using the least squares method. This process dynamically

adjusts the normal directions of in-points to adapt to different image features, optimizing the selection and filtering of in-points. Through iterative fitting and validation strategies, the algorithm continually optimizes the ellipse parameters in each iteration, gradually approaching the optimal solution.

By applying the deep learning-based method for detecting the centers of small markers in UAV oblique photogrammetry images, this approach enables high-precision localization of AM. The pixel coordinates of each marker center, obtained through this method, serve as the raw data for computing their spatial coordinates.

3.2. 3D coordinate reconstruction of the marker centers

The 3D coordinate reconstruction is a critical step in ensuring the accuracy of bridge alignment measurement. It converts the 2D pixel coordinates of marker centers into 3D spatial coordinates, enabling precise representation of the bridge alignment in a global coordinate system. This process begins by using the Five-Point Algorithm [67] to compute the essential matrix, which determines the relative positions and orientations of the cameras capturing the images and establishes the geometric foundation for spatial reconstruction. Subsequently, the triangulation principle is applied to calculate the 3D coordinates of marker centers by integrating pixel coordinates with camera geometry. Finally, bundle adjustment is employed to globally optimize the 3D coordinates of all marker centers, minimizing computational errors.

3.2.1. Estimating the camera position

Consider two images, where the line connecting the camera centers O_1 and O_2 serves as the baseline. This baseline defines a pencil of planes, which intersects the two image planes, forming a plane S . The camera centers O_1 and O_2 , the corresponding image points h_1 and h_2 , and the 3D target point H all lie within this plane, as shown in Fig. 6. Therefore, the corresponding points on the images can be related through the fundamental matrix F , which describes the geometric constraint between the two camera images, satisfying the following relationship:

$$h_2^T F h_1 = 0 \quad (1)$$

By using the Zhang's calibration method [68], the intrinsic parameter matrix K can be obtained. In this case, the fundamental matrix F degenerates into the essential matrix ε . The essential matrix ε describes the rotational and translational relationship between the two camera perspectives, and the relationship between them is as follows:

$$K^{-T} \varepsilon K^{-1} = F \quad (2)$$

For two images that satisfy the overlap condition, the Five-Point Algorithm estimates the essential matrix ε by solving a system of equations formed by five baseline equations derived from five pairs of corresponding points. The essential matrix ε determines the pose and position of the camera during the capture process. In the five-point algorithm, the essential matrix generally yields multiple possible solutions. Each computed essential matrix corresponds to four different camera matrix solutions, representing four possible relative poses. To

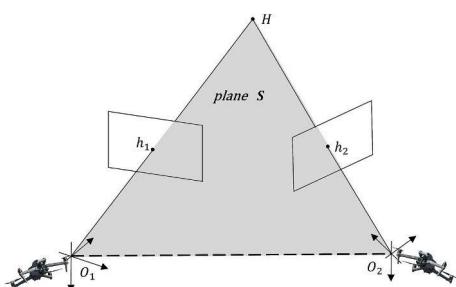


Fig. 6. Geometry relationship of corresponding points.

determine the correct solution, the cheirality constraint is applied by triangulating the 3D coordinates of the five points and verifying whether they lie in front of the camera. If any reconstructed points appear behind the camera, the corresponding solution is discarded. By evaluating all candidate solutions, the one where all triangulated points are in front of the camera is selected as the correct solution, ensuring consistency with the real-world scene. By solving the overlapping images taken from different camera positions, the relative camera positions between the images can be computed, thus allowing the determination of the camera's position in three-dimensional space.

The five-point algorithm [67] and eight-point algorithm [69] are commonly used for camera pose estimation. The eight-point algorithm has a simpler computation process, solving the essential matrix directly through linear equations and singular value decomposition, which makes it computationally efficient. It is suitable for cases where feature point matching is accurate, and image quality is high. However, this method is more sensitive to noise and requires at least eight pairs of matching points. If the input points are not properly normalized, numerical stability may decrease. In contrast, the five-point algorithm [67] solves a nonlinear polynomial system to compute the essential matrix, providing better numerical stability, especially in noisy environments. It requires only five pairs of matching points, reducing the need for artificial markers. Moreover, the algorithm leverages known camera intrinsic parameters to effectively mitigate ambiguity caused by planar structures and ensures the uniqueness of camera pose estimation through the properties of the essential matrix. Therefore, it is more suitable for planar or near-planar scenarios, but this does not imply that the five-point algorithm can independently resolve the depth ambiguity caused by coplanarity. Therefore, in this study, we chose the five-point algorithm and a multi-view measurement strategy to minimize marker usage and improve adaptability in UAV-based imaging scenarios.

3.2.2. Triangulation model

Once the spatial positions of the cameras are obtained, the 3D coordinates of a target point can be reconstructed using the principle of triangulation. This can be done by calculating the pixel positions of corresponding points in two images captured from different viewpoints, along with the camera positions in each image. Assume that a point P in space is captured from two different viewpoints. As shown in Fig. 7, $p_1(u_1, v_1)$ and $p_2(u_2, v_2)$ represent the pixel coordinates of the same point in images 1 and 2, respectively. The coordinate system of viewpoint 1 is represented by $O_1 - X_1 Y_1 Z_1$, which coincides exactly with the world coordinate system $O_w - X_w Y_w Z_w$. This is because, when reconstructing the 3D coordinates of the marker center, it is necessary to choose a world

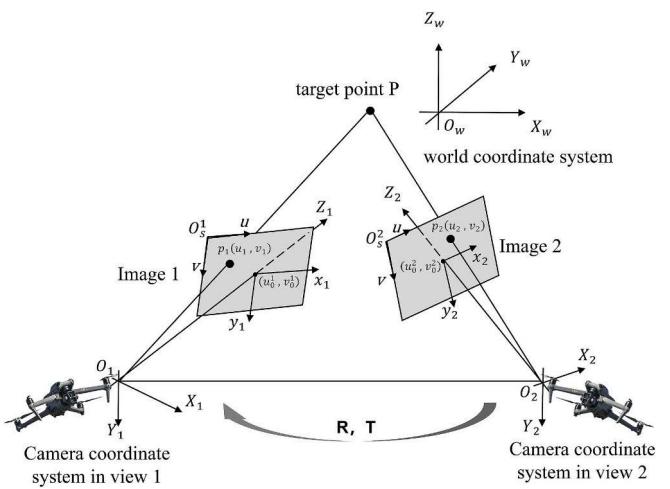


Fig. 7. Calculation of the coordinates of a point in space based on the principle of triangulation.

coordinate system, and the coordinate system of camera 1 is selected as the world coordinate system (WCS) to simplify the computation. The coordinate system of viewpoint 2 is represented by $O_2 - X_2 Y_2 Z_2$, where O_1 and O_2 are the optical centers of the two cameras.

The transformation relationship between the two viewpoint coordinate systems can be expressed as:

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} = M_{12} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = [R \ T] \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{pmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (3)$$

where M_{12} is the transformation matrix from the world coordinate system (WCS) to the viewpoint 2 coordinate system, and $R = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix}$ and $T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$ are the rotation matrix and translation vector, respectively, from the WCS to the viewpoint 2 coordinate system.

$[X_w, Y_w, Z_w]^T$ is the spatial coordinate of the target center in the WCS; $[X_2, Y_2, Z_2]^T$ is the spatial coordinate of the target center in the viewpoint 2 coordinate system. According to the camera imaging model, the relationship is given by:

$$\begin{aligned} Z_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} &= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, Z_2 \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} \end{aligned} \quad (4)$$

$$x = ud_x - u_0, y = vd_y - v_0 \quad (5)$$

where d_x is the length of a single pixel in the u -axis direction, d_y is the length of a single pixel in the v -axis direction, (x, y) represents the image coordinates of the target center, (u_0, v_0) represents the pixel coordinates of the image center, and f is the camera focal length, which can be obtained through camera calibration. Therefore, the spatial coordinates of

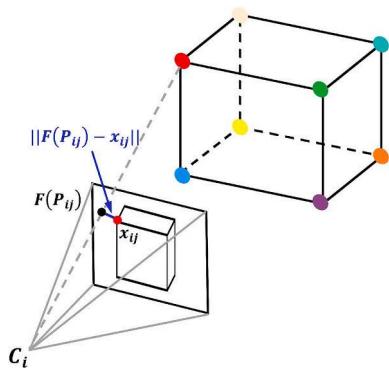


Fig. 8. Bundle adjustment.

the target center in the WCS can be expressed as:

$$X_W = \frac{Z_1 x_1}{f} = \frac{Z_w}{f} (u_1 d_x - u_0) \quad (6)$$

$$Y_W = \frac{Z_1 y_1}{f} = \frac{Z_w}{f} (v_1 d_y - v_0) \quad (7)$$

$$Z_W = \frac{f(t_x - x_2 t_2)}{x_2(r_7 x_1 + r_8 y_1 + fr_9) - f(r_1 x_1 + r_2 y_1 + fr_3)} \quad (8)$$

3.2.3. Optimize calculation results using the bundle adjustment

For each marker point, triangulation and iterative calculations are performed to obtain its three-dimensional coordinates in the world coordinate system, which represents a local optimization solution. Bundle adjustment [70], on the other hand, works in the global sense by minimizing the error between the real pixel locations on each image and the reprojected pixel locations, as shown in Fig. 8.

The minimization of the re-projection error in bundle adjustment is formulated as follows, requiring multiple iterations for solving:

$$\min_{P_{ij}} \sum_{i=1}^n \sum_{j=1}^m \|F(P_{ij}) - x_{ij}\|^2 \quad (9)$$

where n is the number of spatial points, m is the number of camera positions, $F(P_{ij})$ is the reprojected point of the i -th feature point on the j -th image, x_{ij} is the feature point on the j -th image at the i -th location, and $\|a - b\|$ represents the Euclidean distance between points on the image, defined by different vectors a and b .

Using the method described in Section 3.2, the three-dimensional coordinates of the marker centers are estimated, which is critical for ensuring the precision of bridge alignment measurements.

3.3. Sequential point lattice splicing based on overlapping marker centers for obtaining bridge alignment

For bridge alignment measurements, the measurement length often exceeds the camera's field of view. To complete such large-scale photogrammetry tasks, it is necessary to obtain alignment information for different segments and then to be merged into a common coordinate system. This process can be decomposed into the following two parts:

3.3.1. Coordinate registration

The main challenge in coordinate registration is converting from the measurement coordinate system to the world coordinate system. A simplified three-dimensional reference transformation model, suitable for large angular deviations, is employed for this conversion [71]. This

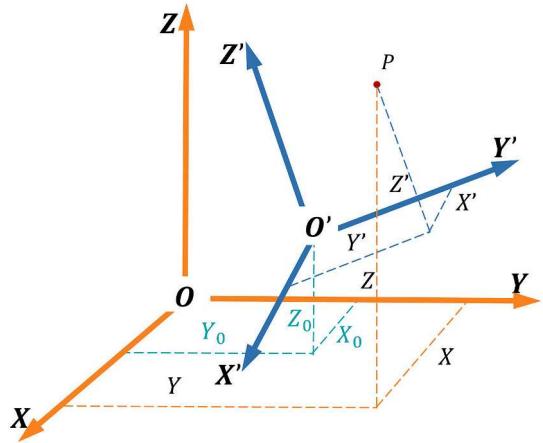


Fig. 9. Relationship between spatial coordinate systems.

model uses linear approximation to manage small coordinate changes, while addressing nonlinear factors separately, improving the computational efficiency and robustness of the dynamic system. Let two spatial Cartesian coordinate systems be denoted as $O - XYZ$ and $O' - X'Y'Z'$, and let a point P in space have coordinates (X, Y, Z) in system $O - XYZ$ and (X', Y', Z') in system $O' - X'Y'Z'$.

Different spatial coordinate systems and their relationships are illustrated in Fig. 9. The coordinate transformation of point P between the two coordinate systems is expressed as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = sR_0 \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} + T_0 \quad (10)$$

where s is the scale factor between the two coordinate systems, $R_0 =$

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \text{ and } T_0 = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}$$

are the rotation matrix and translation vector between the two coordinate systems, respectively, $a_1, a_2, a_3; b_1, b_2, b_3; c_1, c_2, c_3$ are the direction cosines of the X , Y , and Z axes in $O' - X'Y'Z'$, and X_0, Y_0, Z_0 are the coordinates of the point O' in $O - XYZ$. Expanding Eq. (10) using a Taylor series yields:

$$\begin{aligned} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= \begin{bmatrix} X_0^0 \\ Y_0^0 \\ Z_0^0 \end{bmatrix} + s^0 \begin{pmatrix} a_1^0 & a_2^0 & a_3^0 \\ b_1^0 & b_2^0 & b_3^0 \\ c_1^0 & c_2^0 & c_3^0 \end{pmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} + \begin{bmatrix} dX_0 \\ dY_0 \\ dZ_0 \end{bmatrix} \\ &\quad + \begin{bmatrix} a_1^0 x + a_2^0 y + a_3^0 z \\ b_1^0 x + b_2^0 y + b_3^0 z \\ c_1^0 x + c_2^0 y + c_3^0 z \end{bmatrix} \\ &+ \begin{bmatrix} s^0 x & s^0 y & s^0 z & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s^0 x & s^0 y & s^0 z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s^0 x & s^0 y & s^0 z \end{bmatrix} \\ &\bullet [da_1 \ da_2 \ da_3 \ db_1 \ db_2 \ db_3 \ dc_1 \ dc_2 \ dc_3]^T \quad (11) \end{aligned}$$

where the superscript 0 denotes the approximate values, and $dX_0, dY_0, dZ_0, ds, da_1, da_2, da_3, db_1, db_2, db_3, dc_1, dc_2, dc_3$ are the corresponding correction values. The Eq. (11) can be written in the form of an error equation as follows:

$$V = BX + L \quad (12)$$

In the equation, $V = [V_X \ V_Y \ V_Z]^T$ is the error vector of the observed values; $B = [E \ A \ D]$, where $A = \begin{bmatrix} a_1^0 x + a_2^0 y + a_3^0 z \\ b_1^0 x + b_2^0 y + b_3^0 z \\ c_1^0 x + c_2^0 y + c_3^0 z \end{bmatrix}$, $D = \begin{bmatrix} s^0 x & s^0 y & s^0 z & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s^0 x & s^0 y & s^0 z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s^0 x & s^0 y & s^0 z \end{bmatrix}$, $E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and $X = [dX_0 \ dY_0 \ dZ_0 \ ds \ da_1 \ da_2 \ da_3 \ db_1 \ db_2 \ db_3 \ dc_1 \ dc_2 \ dc_3]^T$ are the unknown parameter vectors; $L = \begin{bmatrix} X_0^0 \\ Y_0^0 \\ Z_0^0 \end{bmatrix} + s^0 \begin{pmatrix} a_1^0 & a_2^0 & a_3^0 \\ b_1^0 & b_2^0 & b_3^0 \\ c_1^0 & c_2^0 & c_3^0 \end{pmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$, and the rotation matrix contains only

three independent parameters, and its elements satisfy the following constraint:

$$\left\{ \begin{array}{l} a_1^2 + a_2^2 + a_3^2 = 1 \\ b_1^2 + b_2^2 + b_3^2 = 1 \\ c_1^2 + c_2^2 + c_3^2 = 1 \\ a_1 a_2 + b_1 b_2 + c_1 c_2 = 0 \\ a_1 a_3 + b_1 b_3 + c_1 c_3 = 0 \\ a_2 a_3 + b_2 b_3 + c_2 c_3 = 0 \end{array} \right. \quad (13)$$

Therefore, the condition equations for the adjustment calculation can be formulated as:

$$CX + W = 0 \quad (14)$$

In the equation,

$$C = \begin{bmatrix} 0 & 0 & 0 & 2a_1^0 & 2a_2^0 & 2a_3^0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2b_1^0 & 2b_2^0 & 2b_3^0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2c_1^0 & 2c_2^0 & 2c_3^0 \\ 0 & 0 & 0 & a_1^0 & a_2^0 & 0 & b_2^0 & b_1^0 & 0 & c_2^0 & a_1^0 & 0 \\ 0 & 0 & 0 & a_3^0 & 0 & a_1^0 & b_3^0 & 0 & b_1^0 & c_3^0 & 0 & c_1^0 \\ 0 & 0 & 0 & 0 & a_3^0 & a_2^0 & 0 & b_3^0 & b_2^0 & 0 & c_3^0 & c_2^0 \end{bmatrix}, W = \begin{bmatrix} a_1^0 + a_2^0 + a_3^0 - 1 \\ b_1^0 + b_2^0 + b_3^0 - 1 \\ c_1^0 + c_2^0 + c_3^0 - 1 \\ a_1^0 a_2^0 + b_1^0 b_2^0 + c_1^0 c_2^0 \\ a_1^0 a_3^0 + b_1^0 b_3^0 + c_1^0 c_3^0 \\ a_2^0 a_3^0 + b_2^0 b_3^0 + c_2^0 c_3^0 \end{bmatrix}.$$

By solving the system of Eqs. (12) and (14) and substituting the coordinates of the three known points in coordinate systems $O - XYZ$ and $O' - X'Y'Z'$, the unknown parameter vector X can be determined using the indirect adjustment method with the imposed constraints. The use of three points for coordinate transformation is chosen because they are sufficient to determine the relationship between coordinate systems, require less computational effort, and facilitate the explanation of the fundamental methodology. The coordinates of these three points are obtained through total station measurements or a previous coordinate transformation. Once the values of X are obtained, the coordinate transformation for all points can be completed according to Eq. (10).

3.3.2. Sequence point lattice stitching

When using a total station (TS) to measure the bridge alignment, control points are typically selected at the positions of the corresponding crossbeams, and the coordinates of these control points are measured to compute the bridge deck alignment. The objective of this study is to obtain bridge deck alignment by solving the coordinates of various AM on the bridge surface. To achieve this, this study stitches together marker centers from different images, using the overlapping marker centers from each image as reference points during the stitching process. The stitching process is illustrated in Fig. 10, where the yellow points represent the overlapping marker centers from each pair of measurements, serving as reference points in the stitching process.

Firstly, the initial coordinates of four marker centers are measured using a TS, and these are used as reference points to establish the world coordinate system. Reconstruction and calculations are performed for the first set of images, and the world coordinates of the initial four marker centers are input to obtain the 3D coordinates of the marker centers in the world coordinate system. Then, each subsequent set of images is reconstructed and solved in turn, yielding the marker center coordinates in the respective and independent camera coordinate systems. The transformation between the independent camera coordinate systems and the world coordinate system can be achieved through

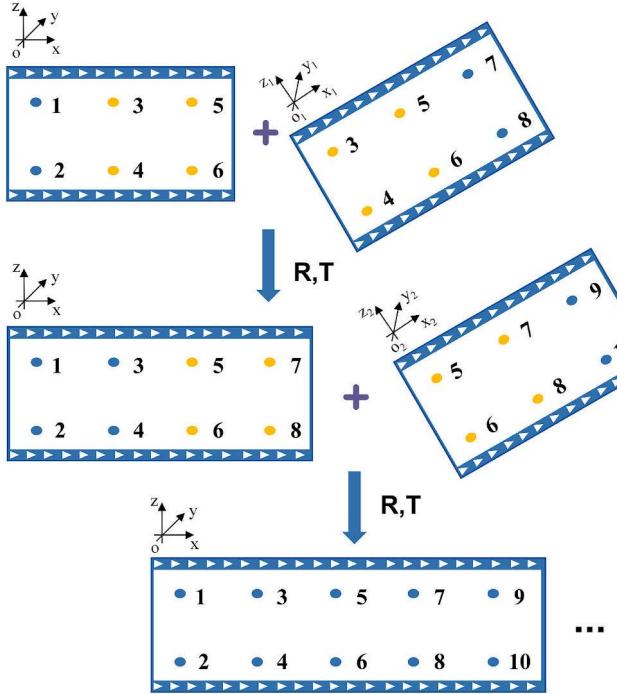


Fig. 10. Stitching process of sequential point lattice coordinates based on overlapping marker centers.

multiple coordinate registrations. It is worth noting that, for simplification, Section 3.3.1 selected the three most fundamental points as an example to illustrate the basic concept of coordinate transformation. However, in the actual experiments, four points were used to enhance transformation accuracy by providing additional constraints while keeping computational costs manageable.

3.3.3. Bridge alignment

Through the registration and stitching of sequential point lattices, the 3D coordinates of all marker centers in a unified coordinate system are obtained. The planar distance of marker centers relative to the reference point is used as the horizontal axis values, and the elevation of marker centers relative to the reference point is used as the vertical axis values for linear fitting. This process yields complete bridge alignment data, clearly illustrating the variations in the bridge alignment along the longitudinal axis, thereby facilitating subsequent structural analysis and adjustments.

4. Experimental validation

This section validates the proposed UAV-based bridge alignment measurement system through two experiments: one for parameter optimization and the other for real-world application. It also evaluates detection accuracy and compares 3D measurements with total station results.

4.1. Experimental setup

To evaluate the feasibility of the proposed bridge alignment measurement technology, two experiments (i.e., Experiment 1 and Experiment 2) were conducted. Experiment 1 primarily explored the influence of different factors on detection accuracy and optimized the system configuration based on the experimental results. Experiment 2 applied the optimized configuration to a real bridge structure, validating its performance in an actual environment. The UAV used in these tests was the DJI Air 3, and the measurement results were compared with those from the Topcon GM-52 TS to validate the accuracy of the UAV-based alignment measurement system. A flowchart of the testing process is shown in Fig. 11.

The Topcon GM-52 is a high-performance TS known for its excellent accuracy and reliability in measurement. It offers an angular measurement accuracy of 2 s and a distance measurement accuracy of 1.5 mm + 2 ppm. It is capable of measuring distances from 0.3 m to 500 m without a prism, and with a single prism, the measurement range extends from 1.3 m to 4000 m.

The DJI Air 3 is a high-performance, lightweight consumer-grade UAV. Its dual-camera system includes a 48MP wide-angle camera and a 70 mm equivalent focal length telephoto camera. The focus range spans from 1 m (wide-angle) or 3 m (telephoto) to infinity, with an effective pixel count of up to 48 million. The data can be promptly transmitted to remote devices for timely processing through the DJI Fly app.

For Experiment 1, based on the principle of UAV oblique photogrammetry, multiple experimental configurations were designed to investigate the impact of four factors on detection accuracy: marker spacing, flight altitude, gimbal tilt angle, and the number of viewpoints. It is worth noting that using the principle of triangulation, the 3D coordinates of marker points can be computed with images from two viewpoints. However, a single measurement may be affected by random errors, which can impact accuracy. To improve precision, we employed a multi-view measurement strategy, where multiple measurements of the same set of marker points were taken from different viewpoints. Outliers were then removed, and the average of the remaining data was computed. This strategy does not rely solely on a single pair of images. Instead, it leverages multiple observations of the same markers, allowing us to overcome potential ambiguities introduced by limited depth or coplanarity in any one image pair. In this sense, this method integrates multi-view geometry as a crucial complement to the five-point pose estimation. The specific parameter settings are shown in Table 1. By analyzing the effects of these factors, the optimal configuration and flight parameters were identified to improve measurement accuracy while maintaining high operational efficiency in practical applications.

For Experiment 2, the optimal configuration and flight parameters from Experiment 1 were applied to a segment of a stone arch bridge with 4 piers and 3 spans (Fig. 12(a)), as shown in Fig. 12(b), to evaluate the performance improvements of key modules in the UAV-based alignment measurement system under real-world bridge conditions. The experiment consisted of two parts: the first part validated the capability of the YOLO-USMD-based small marker detection algorithm in aerial images to accurately identify and localize AM. Additionally, the elevation values of the marker centers measured by the UAV-based alignment measurement system were compared with those obtained by a TS to verify the system's high-precision measurement capability in practical bridge



Fig. 11. Test flow chart.

Table 1
Experimental configuration parameters.

Parameter	Set value
marker spacing	2 m, 3.5 m, 5 m, 6.5 m, 8 m
flight altitude	4 m, 5.5 m, 7 m, 8.5 m, 10 m
gimbal tilt angle	35°, 45°, 55°, 65°, 75°
the number of viewpoints	2, 3, 4, 6, 8

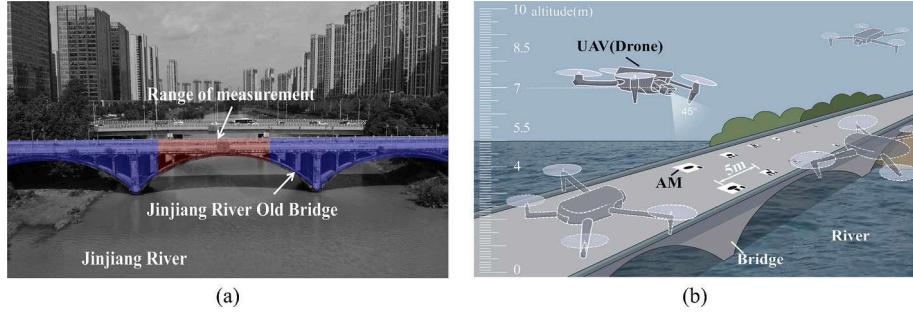


Fig. 12. Alignment measurement setup: (a) selected range of measurement and (b) arrangement of AM on bridge deck.

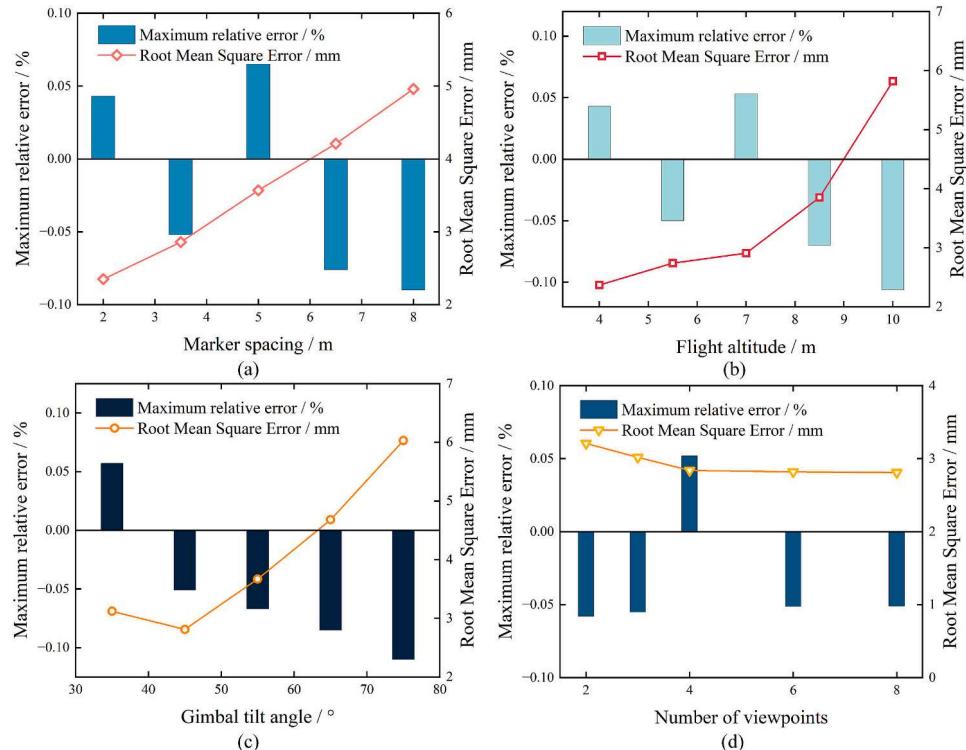


Fig. 13. Influencing factors for multi-view 3D coordinate reconstruction of UAVs: (a) marker spacing, (b) flight altitudes, (c) gimbal tilt angle, and (d) number of viewpoints.

inspection tasks.

The specific experimental details of the proposed YOLO-USMD network are as follows. The hardware environment included a high-performance computer. The main system was equipped with an Intel (R) Core (TM) i5-10,500 processor, 64 GB of RAM, and an NVIDIA GeForce RTX 4060 GPU. The software environment was based on a 64-bit Windows 10 operating system. Python 3.10 was used as the programming language, and the deep learning framework was PyTorch 2.1.0. All computations were performed using the CUDA kernel. For the experimental parameters, the batch size was set to 16, and the image size was 640×640 pixels. The initial learning rate (lr_0) was 0.01, and the final learning rate (lr_f) is set to 0.01 times the initial learning rate. The number of training epochs was 250, and the weight decay coefficient (weight_decay) was 0.0005.

4.2. Experiment 1 results and analysis

Experiment 1 evaluated the impact of flight parameters and marker arrangements on bridge alignment measurement accuracy. The results, shown in Fig. 13, guided the selection of optimal conditions. First,

regarding marker spacing, the experiment demonstrated that as spacing increased, the recognition accuracy declined. However, within a certain range (within 5 m), both high accuracy and measurement efficiency could be maintained. Therefore, 5 m was selected as the final marker spacing. For flight altitude, the experiment found that a height range of 5.5 to 7 m struck a good balance between image clarity and coverage, which further improved detection efficiency and accuracy. The gimbal tilt angle adjustment experiment showed that a tilt angle of 45° provided an optimal balance between the visibility of markers and coordinate reconstruction performance, thus this angle was selected as the final setting. Increasing the number of viewpoints effectively improved the precision of triangulation, but when the number of viewpoints exceeded four, accuracy improvements became marginal. Therefore, four viewpoints were chosen to balance efficiency and precision. These tests identified the optimal experimental conditions, providing a solid foundation for bridge measurement in Experiment 2.

4.3. Experiment 2 results and analysis

This section presents the experimental results and analysis of the

proposed UAV-based bridge alignment measurement method. It evaluates the performance of the YOLO-USMD network and the DAMSED algorithm in accurately detecting and localizing markers on the bridge deck. The results are further validated by comparison with total station (TS) measurements to assess the accuracy of the 3D coordinate reconstruction.

4.3.1. Measurement of marker centers on the bridge deck

This section describes the performance of the YOLO-USMD network and the DAMSED algorithm in accurately detecting and localizing markers on the bridge deck.

4.3.1.1. Target recognition and localization based on YOLO-USMD. The objective of this section is to validate the performance of the YOLO-USMD-based small marker detection algorithm for identifying and localizing AM on bridge decks in aerial images. The proposed method consists of two main steps: the YOLO-USMD network is used to detect markers and obtain the bounding box positions in the images, and ellipse

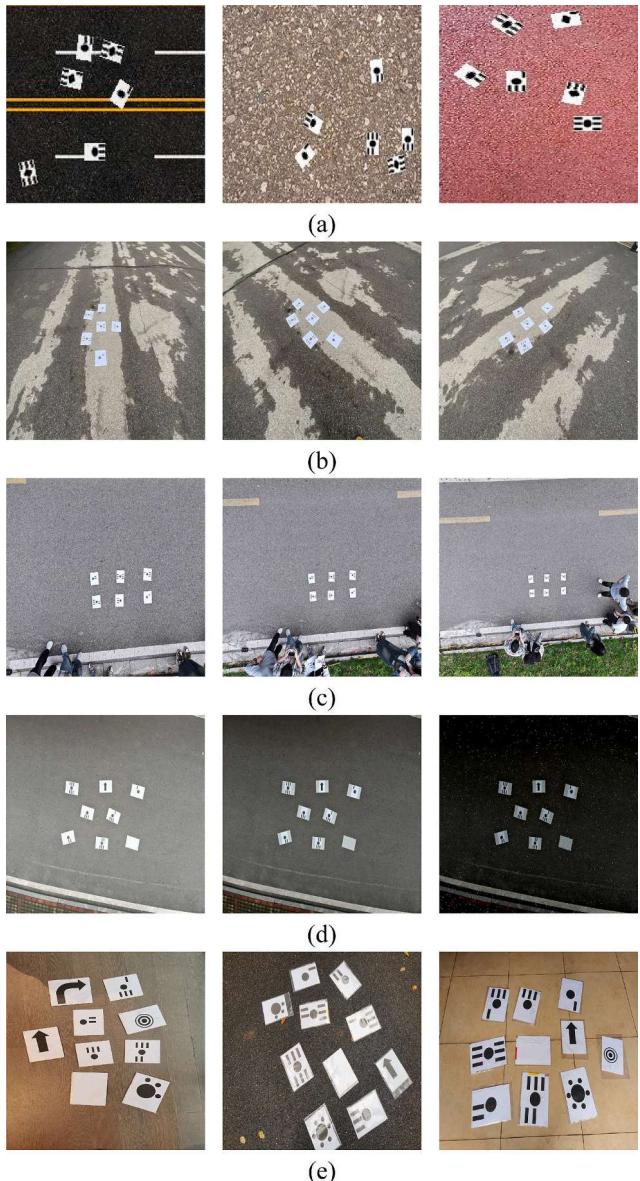


Fig. 14. Artificial target point dataset: (a) diverse surfaces, (b) diverse shooting angles, (c) diverse flight heights, (d) lighting variations, and (e) interference from other markings.

detection is subsequently applied only to the regions of interest in each image to determine the pixel coordinates of the marker centers. To achieve this, two types of experiments were designed: the first type involved constructing a diverse dataset to train and validate the YOLO-USMD network, as well as conducting ablation studies to compare the contribution of each module to the improvement in detection accuracy and efficiency; the second type involved comparing the performance of DAMSED with state-of-the-art methods to validate its comprehensive advantages in speed and accuracy. The experimental results demonstrate that the deep learning-based small marker detection algorithm exhibits significant advantages in balancing precision and efficiency for detecting AM on bridge decks in aerial images. A detailed analysis of the experimental results is provided below:

The YOLO-USMD network was trained to classify and detect AM on bridge decks under varying conditions, including diverse surfaces, shooting angles, flight heights, lighting variations, and interference from other markings. A dataset of 1729 images was collected for this purpose, as shown in Fig. 14, with 1389 images used for training and 340 images for validation. The dataset consists of two parts: images captured on-site using different UAV (DJI M300RTK, DJI Air3, DJI Mavic3), and images with target markers randomly and non-overlappingly placed on various road surfaces.

To validate the effectiveness of each module in the YOLO-USMD network, this study first conducted individual ablation experiments on the DAttention, BiFPN, and NWD modules to analyze their contributions to overall performance. Then, the YOLOv10n, YOLOv11n, and the improved YOLO-USMD models were compared for overall performance. Since this study aims to develop a lightweight model that can provide fast detection results at a low cost, the n-series models were selected for comparison. YOLOv8n is a lightweight model that offers fast computation but lower detection accuracy. YOLOv10n and YOLOv11n are newer versions in the YOLO model series, with their performance slightly surpassing that of YOLOv8n. Among them, YOLOv10n delivers more competitive results but still falls short of surpassing YOLO-USMD. YOLO-USMD builds upon YOLOv8n and further enhances detection performance by incorporating several optimized modules. As shown in Table 2, the baseline model YOLOv8n achieved a precision of 77.1 %, a recall of 80.8 %, a mAP@50 of 86.5 %, and a mAP@50–95 of 52.9 %. However, its missed detection rate and false detection rate were relatively high, at 9.52 % and 16.67 %, respectively. With the addition of the DAttention module, the model's precision and recall were improved to 84.4 % and 84.5 %, respectively, and the mAP@50 increased to 90.9 %. The missed and false detection rates were reduced to 4.75 % and 8.34 %, demonstrating that the DAttention module effectively enhanced feature representation capabilities. After incorporating the BiFPN module, the recall was further improved to 86.0 %, and the missed detection rate was reduced to 5.64 %. Additionally, the parameter count decreased to 2.11 M, indicating that the module improved detection performance while optimizing computational efficiency. With the inclusion of the NWD module, the precision reached 82.3 %, and the mAP@50 was 90.4 %, highlighting the module's positive contribution to small object detection and localization. The final YOLO-USMD network, integrating all modules, demonstrated the best detection performance, achieving a precision of 87.1 % and a recall of 86.1 %, with a mAP@50 of 93.5 % and a mAP@50–95 of 59.9 %. Meanwhile, the missed and false detection rates were reduced to 1.19 % and 1.65 %, respectively. These results validate the effectiveness of the YOLO-USMD network and establish its reliability as a robust solution for small object detection tasks. The newly released YOLOv10n and YOLOv11n models have introduced improvements over YOLOv8n; however, their performance still does not surpass that of YOLO-USMD. As shown in Table 2, YOLOv10n achieved a precision of 82.6 %, a recall of 83.7 %, and a mAP@50 of 89.5 %, while YOLOv11n attained a precision of 80.4 %, a recall of 82.9 %, and a mAP@50 of 88.4 %. As shown in Fig. 15, to clearly and intuitively compare the detection performance of different models, we provide the Precision-Recall (PR) curves for each model. Compared to YOLO-USMD, both models

Table 2

Target detection and recognition results of different models.

Model	P (%)	R (%)	mAP@50	mAP@50–90 (%)	Parameters (M)	FLOPs (G)	Miss (%)	False (%)
YOLOv8n	0.771	0.808	0.865	0.529	3.01	8.1	9.52	16.67
YOLOv8n (with DAttention)	0.844	0.845	0.909	0.565	3.27	8.3	4.75	8.34
YOLOv8n (with BiFPN)	0.819	0.86	0.909	0.568	1.99	7.1	5.64	10.84
YOLOv8n (with NWD)	0.823	0.834	0.904	0.565	3.01	8.1	5.47	9.73
YOLO-USMD	0.871	0.861	0.935	0.599	2.26	7.3	1.19	1.65
YOLOv10n	0.826	0.837	0.895	0.556	2.27	6.5	7.16	12.55
YOLOv11n	0.804	0.829	0.884	0.559	2.58	6.3	8.25	14.36

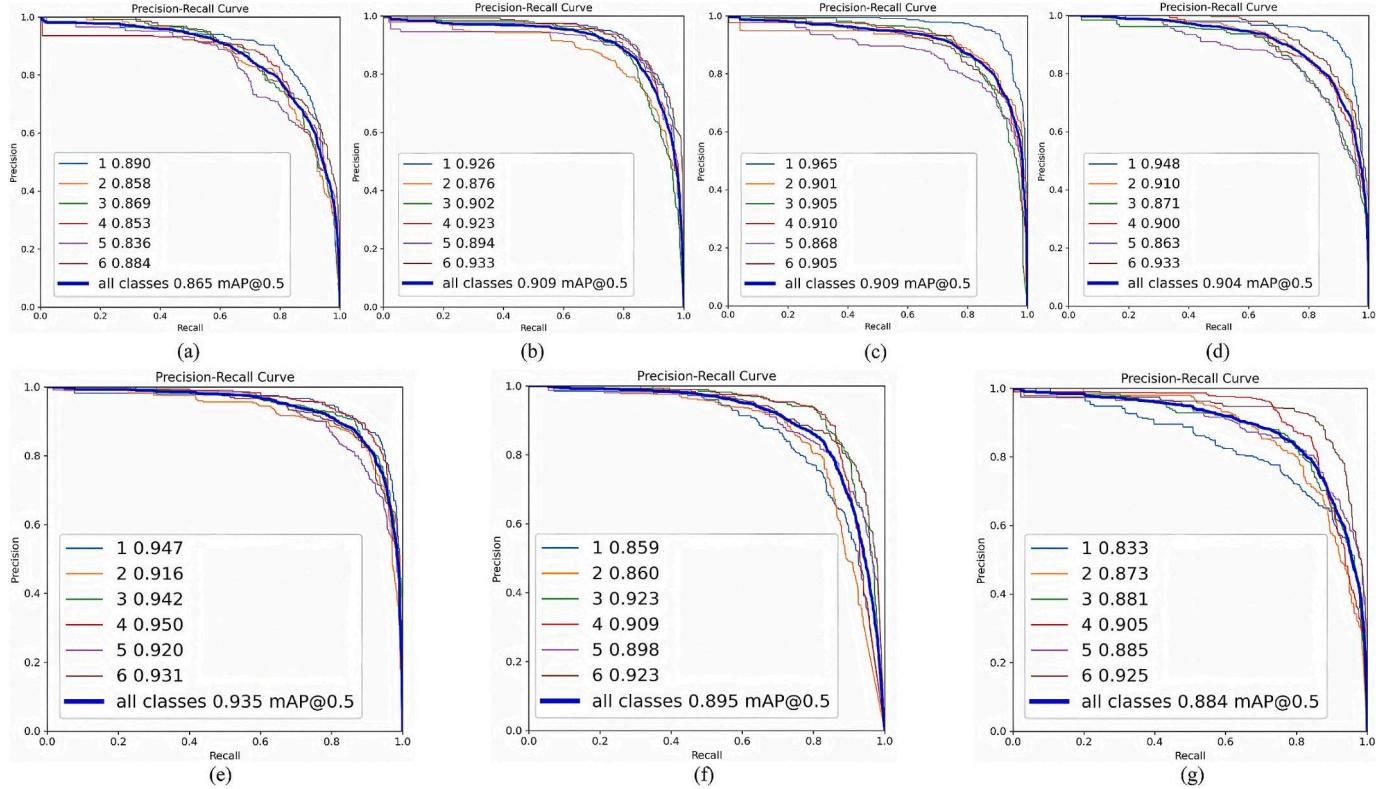
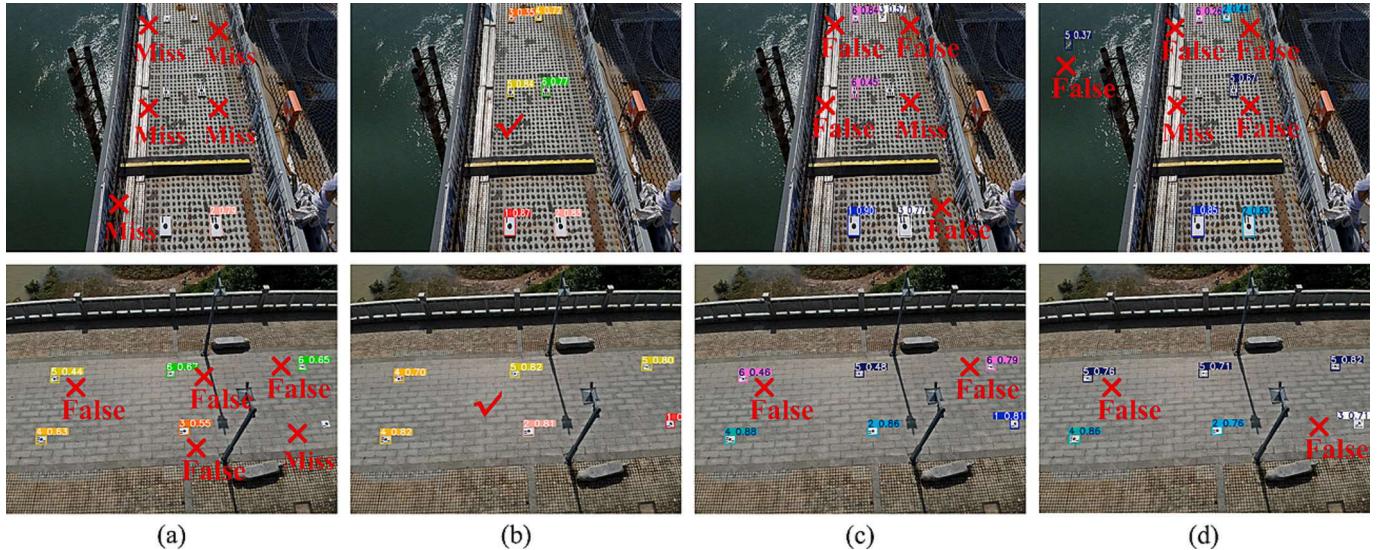
**Fig. 15.** Precision-recall curve for different models: (a) YOLOv8n, (b) YOLOv8n (with DAttention), (c) YOLOv8n (with BiFPN), (d) YOLOv8n (with NWD), (e) YOLO-USMD, (f) YOLOv10n, and (g) YOLOv11n.**Fig. 16.** Comparison of target detection effect: (a) Yolov8n, (b) YOLO-USMD, (c) Yolov10n, and (d) Yolov11n.

Table 3

Detection efficiency results of different models.

Processing stage	YOLOv8n	YOLOv8n (with DAttention)	YOLOv8n (with BiFPN)	YOLOv8n (with NWD)	YOLO-USMD	YOLOv10n	YOLOv11n
Preprocess	1.1 ms	1.5 ms	1.2 ms	0.9 ms	0.9 ms	1.3 ms	1.2 ms
Inference	11.0 ms	20.7 ms	4.9 ms	4.5 ms	4.3 ms	2.3 ms	8.9 ms
Loss	0.0 ms	0.0 ms	0.0 ms	0.0 ms	0.0 ms	0.0 ms	0.0 ms
Postprocess	2.8 ms	10.4 ms	1.9 ms	1.0 ms	1.0 ms	0.9 ms	1.9 ms
Total Time	14.9 ms	32.6 ms	8.0 ms	6.4 ms	6.2 ms	4.5 ms	12.0 ms

exhibited higher missed and false detection rates, further reinforcing the superiority of the YOLO-USMD model. Fig. 16 visually demonstrated the improvement in detection accuracy achieved by the YOLO-USMD model. The YOLOv8n model showed numerous false detections and a certain degree of missed detections, while YOLOv10n and YOLOv11n showed improved performance but still fell short of practical requirements. In contrast, the YOLO-USMD model accurately detected almost all targets, demonstrating superior detection capability and reliability.

To comprehensively evaluate the performance of different models, detection efficiency is also a critical metric in addition to detection accuracy. Table 3 presents a comparison of detection efficiency across various models. Although the introduction of attention mechanisms may have increased computational overhead, YOLO-USMD still performed excellently in terms of overall detection time, with a total time of 6.2 ms, slightly higher than YOLOv10n's 4.5 ms. This advantage can be attributed to the following optimizations: First, BiFPN replaced the PAN structure in YOLOv8, utilizing weighted feature fusion to reduce the number of Concat operations, effectively lowering computational costs. Second, by fixing the head_channel to 256, the feature channel configuration was optimized to avoid unnecessary computational overhead caused by excessive channel numbers. These optimizations allowed YOLO-USMD to maintain high detection accuracy while significantly reducing computational load, thus achieving a balance between detection accuracy and speed.

4.3.1.2. Ellipse fitting and center localization based on DAMSED. The YOLO-USMD model was applied to identify 106 different images containing markers, extracting pixel coordinates of the bounding box vertices. Ellipse detection was then performed within these region-of-interests, which not only reduced false detections but also significantly lowered computational costs. To evaluate the performance of the proposed DAMSED method, it was compared with three well-established and open-source ellipse detection algorithms: RHT [48], Wu's method [49], and Lu's method [54]. The comparison focused on missed detection rate and processing time. As presented in Table 4, the DAMSED method achieved the lowest missed detection rate of 1.84 %, significantly outperforming the other methods, where RHT, Wu's method, and Lu's method exhibited missed detection rates of 39.55 %, 24.05 %, and 7.41 %, respectively. Additionally, the processing time for DAMSED was 0.25 s per image, which is comparable to the 0.23 s of Lu's method, and slightly slower than Wu's method (0.18 s), but faster than RHT (0.36 s). Visual results in Fig. 17 further highlight the advantages of the DAMSED method. The RHT [48] method failed to accurately distinguish the markers from shadows. This caused poor fitting and made the detection

of the marker's central ellipse difficult, leading to many false detections. Although Wu's method [49] effectively detected the target ellipse, the blurred edges of the rectangular shapes in the markers, caused by the distant shooting, led to inaccurate recognition. As a result, the method mistakenly identified the surrounding rectangles as ellipses, resulting in a large number of false detections. Lu's method [54] accurately detected most of the target ellipses, but due to the blurred edges of the markers or the influence of the shooting angle, a small number of detections were missed. Whereas DAMSED accurately identified all target centers with precision. The pixel coordinates of the detected ellipse centers were output as the final results, demonstrating the robustness and reliability of the proposed method. These findings validate that the DAMSED method not only surpasses other methods in detection accuracy but also retains competitive processing speed, making it highly effective for real-time small marker detection in UAV-based bridge alignment tasks.

4.3.2. Accuracy evaluation through comparison with TS

Experiment 2 adopted the optimal experimental configuration with a marker spacing of 5 m (lateral spacing of 2 m), a flight altitude of 7 m, a gimbal tilt angle of 45°, and 4 viewpoints. Following the method described in Section 3.3, the 3D physical coordinates of all marker centers can be calculated. Table 5 presents the computational efficiency of 3D coordinate reconstruction for a set of marker centers (six markers). The elevation values of the target centers obtained by the UAV-based alignment measurement system were compared with the elevations measured by the TS. The same set of markers was measured three times to eliminate experimental variability, and the elevation errors at each measurement point are shown in Fig. 18(a). To minimize the impact of randomness on the results, we repositioned the markers at different sections of the bridge and conducted three additional repeated measurements. The elevation errors at each measurement point are shown in Fig. 18(b). As the stitching process progresses, cumulative errors result in an increasing in elevation errors along the longitudinal axis of the bridge. By measurement point 10, the maximum errors reached 4.015 mm, -4.079 mm, and -3.996 mm, as well as 4.122 mm, 4.058 mm, and -4.103 mm, respectively.

5. Field experiment

This section describes a field experiment conducted on the Xiashesi bridge to evaluate the performance of the proposed UAV-based alignment measurement system. It includes the bridge background, experimental setup and results, as well as a discussion on practical application challenges and future optimization directions.

5.1. Bridge description

To evaluate the effectiveness of the proposed detection system, a field test was conducted during the construction of a steel-concrete composite modular highway bridge across the Xiang River in Xiangtan, China. The Xiashesi bridge, located in Xiangtan City, Hunan Province, China, is an under-construction urban steel-concrete composite modular highway bridge crossing the Xiang River. The bridge spans are arranged as: 50 + 120 + 200 + 120 + 50 m, with a total length of 540 m, as shown in Fig. 19. The main bridge has a straight alignment in plan and a longitudinal slope of 2.1 %. The superstructure of the main bridge

Table 4

Effectiveness of different methods for ellipse detection.

Data	Index	RHT [48]	Wu's method [49]	Lu's method [54]	DAMSED
Artificial targets on bridges taken from different viewpoints	Missed detection rate (%) Time/s	39.55 0.36	24.05 0.18	7.41 0.23	1.84 0.25

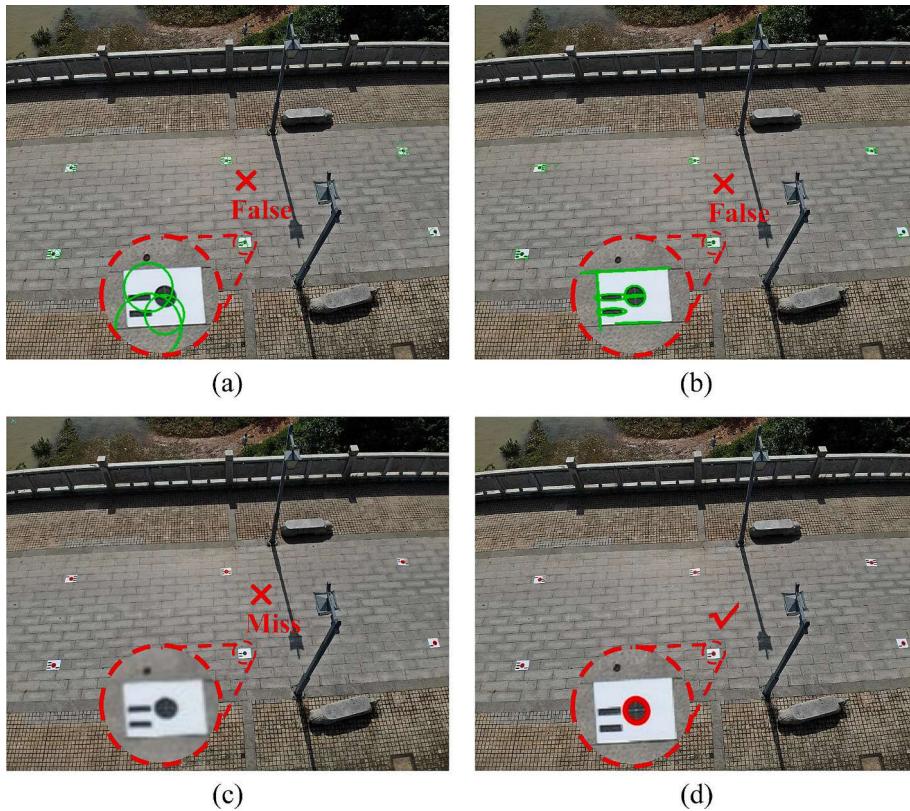


Fig. 17. Comparison of elliptical test results: (a) RHT [48], (b) Wu's method [49], (c) Lu's method [54] and (d) DAMSED.

Table 5
Computational efficiency of marker center 3D coordinate reconstruction.

Computational module	Time per batch of markers
Marker Identification	20-25 ms
Marker Center Localization	900-1000 ms
Camera Position Estimation	400-500 ms
Triangulation	350-400 ms
Total Computation	1700-2000 ms

consists of a steel-concrete composite double-box girder, with a total bridge width of 37.5 m and a single width of 18 m. The girder height varies from 2.45 to 2.85 m, and construction starts from both ends of the bridge. All of the main bridge piers are rectangular, with pier heights ranging from 22.607 to 25.607 m.

During measurement, the downstream arch rib closure segment (ZH segment) of the bridge was lifted and installed. At that time, only the

small pile-number side of the closure segment was welded at the ring joint, while the large side with the smaller pile number was positioned using a steel plate and had not yet been welded and secured. In order to ensure that the installed segment's alignment met design requirements and to provide a reference for subsequent welding of the closure segment, an alignment measurement and control process was required for the installed segment. The superstructure of the bridge is made of steel-concrete composite box girders, while the substructure consists of piers and cross beams, all of which were prefabricated in nearby factories and transported to the construction site for assembly. The advantage of this approach is that it can shorten the construction cycle by more than 30 %.

5.2. Field experiment and results

To meet on-site measurement requirements and overcome environmental constraints, AM were installed on the cross beams of the Xiashesi

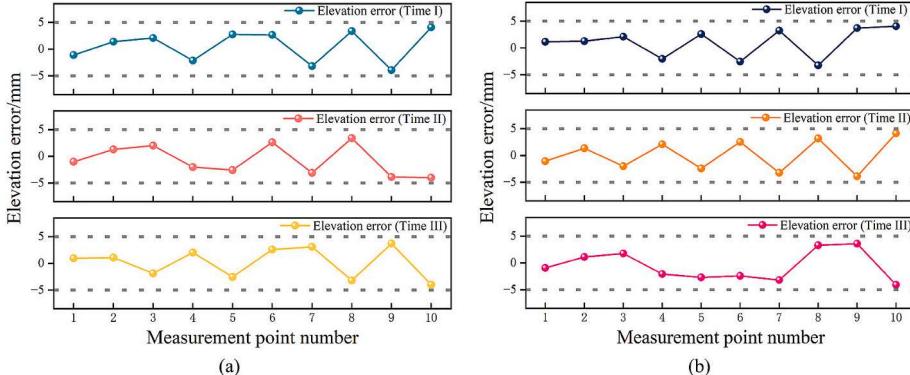


Fig. 18. Elevation error of measurement points: (a) elevation errors of repeated measurements and (b) elevation errors after marker repositioning.

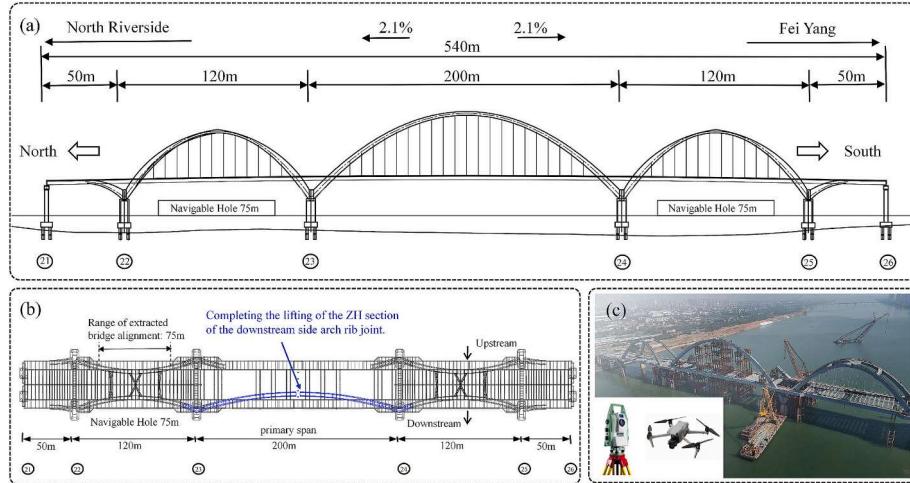


Fig. 19. Xiashesi bridge: (a) span layout of the entire bridge, (b) current construction progress, and (c) construction site and installation.

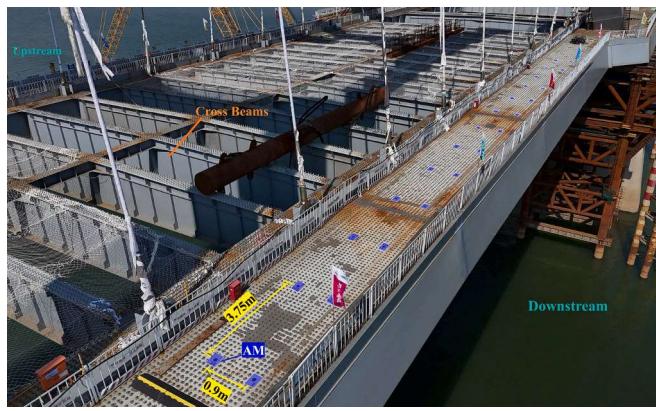


Fig. 20. Placement of markers on the downstream side of the north span.

Table 6
Elevations of marker centers and elevation errors.

Point number	TS (m)	UAV (m)	Elevation errors (mm)
1	54.9869	54.986368	-0.532
2	55.0283	55.029591	1.291
3	55.0641	55.065282	1.182
4	55.0953	55.093270	-2.030
5	55.1409	55.139403	-1.1497
6	55.1668	55.167919	1.119
7	55.2147	55.212279	-2.421
8	55.2369	55.239792	2.892
9	55.2793	55.276705	-2.595
10	55.3117	55.309082	-2.618
11	55.3478	55.350855	3.055
12	55.3801	55.383161	3.061
13	55.4071	55.410635	3.535
14	55.4362	55.432535	-3.665
15	55.4767	55.481212	4.512
16	55.5042	55.508587	4.387
17	55.5381	55.542721	4.621
18	55.5711	55.566464	-4.636

bridge. Markers were placed within a 30 m range on the downstream side of the north span, with a longitudinal spacing of 3.75 m and a lateral spacing of 0.9 m. The exact arrangement of the markers was shown in Fig. 20. The UAV flight height was controlled at 7 m, and the gimbal pitch angle was set to 45°, capturing images four perspectives. The results obtained using the UAV-based method were compared with those from the Leica TM50 TS, known for its angular measurement accuracy of

1 s and distance measurement accuracy of 1 mm + 1.5 ppm. The TS method required 15 min for data collection and an additional 6 min for manual registration within the 30 m range, totaling 21 min. In comparison, the UAV-based method required a similar total time of 22 min, which included 10 min for marker placement, 8 min for multi-view photo acquisition, and 4 min for calculating marker center coordinates.

After completing the sequential point lattice stitching process based on overlapping markers, the elevation values of all measured points in the world coordinate system were obtained, as shown in Table 6. The relative errors between the elevation of the artificial marker centers and the TS measured elevation were shown in Fig. 21, with all errors remaining within 5 mm. The comparative analysis of bridge alignment extraction using UAV and TS methods showed that the Root Mean Square Errors (RMSE) of the two columns of markers were 2.84 mm and 2.96 mm, respectively. These results demonstrated that the UAV-based measurement method achieved high accuracy.

By performing line fitting on the measurement points, the elevation variation of this segment was obtained, as shown in Fig. 22. The results indicate that the measured slope closely aligns with the theoretical slope of 1.834 %, which was calculated using finite element simulations under current conditions. In comparison, the TS results were slightly closer to the design slope. The discrepancy between the slope measured by the TS and the theoretical slope is a common phenomenon, influenced by several factors. First, the environmental temperature and the temperature of the steel structure during measurement deviate from the temperature assumed in the theoretical calculations. Second, there were steel pipe pile supports in certain positions of the main girder, which caused a discrepancy between the support boundary conditions for the main girder in the field and the boundary conditions used in the theoretical simulation. Third, inherent measurement errors within the TS itself. Finally, the influence of wind loads also contributed to the discrepancy. Despite these factors, the obtained bridge alignment measurements were consistent with the theoretical alignment under the current conditions, demonstrating the reliability and accuracy of the proposed method.

5.3. Discussion of field application

The field application of the bridge alignment measurement framework requires close integration with existing construction processes and relies on the collaboration of various personnel to ensure its effectiveness. During construction, workers must install artificial marker points at predefined locations on the bridge deck or cross beams, coordinating this task with other assembly steps to minimize disruption and improve efficiency. UAV operators are responsible for planning and executing

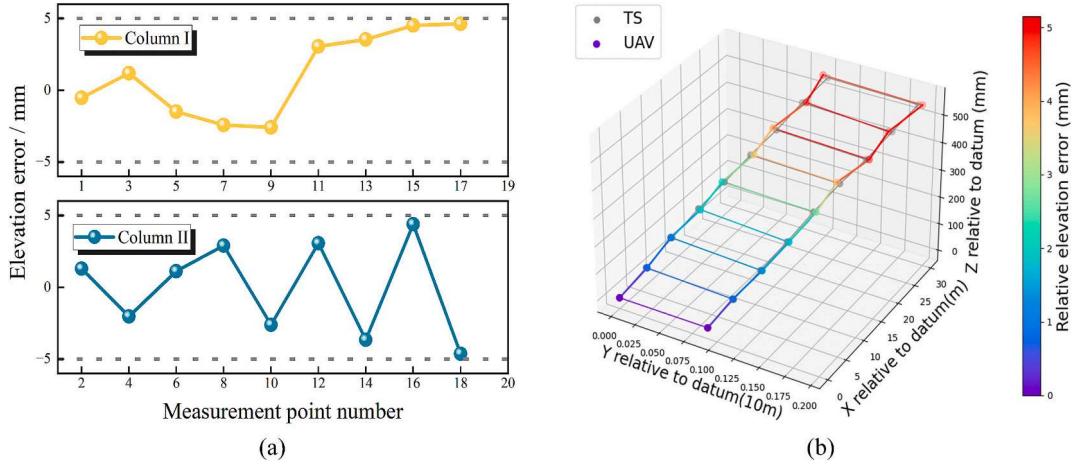


Fig. 21. Elevation error of measurement points on the Xiashesi bridge: (a) Elevation error and (b) Elevation error heatmap.

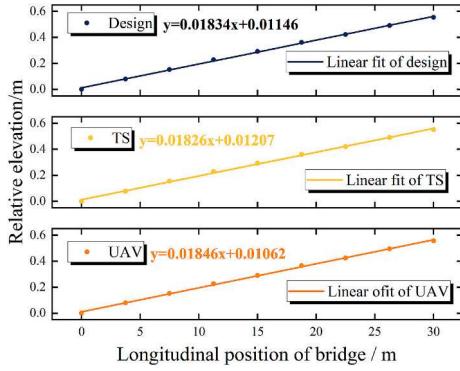


Fig. 22. Comparison of alignment fitting results.

flight missions, controlling flight parameters such as altitude and gimbal angle, to ensure ideal image capture under varying environmental conditions. Meanwhile, remote computational technicians process the collected data using advanced target detection, localization, and stitching algorithms to obtain the alignment data of the bridge.

Although the proposed UAV-based alignment measurement system requires only 1–2 s for data processing per image, a significant amount of time is needed for the setup of AM prior to measurement. Further exploration will focus on the following potential solutions: On the one hand, AM could be integrated into prefabricated bridge components during factory production, thereby reducing the need for on-site marker placement. On the other hand, the possibility of automating the placement of markers or replacing them with alternative methods, such as natural feature point detection techniques, will be explored. These improvements are expected to reduce manual interventions, enhance accuracy, and increase the automation level of data collection and analysis. Additionally, when the measurement span exceeds 30 m, the error exceeds 5 mm. This issue primarily arises from the following factors: 1) small cumulative deviations in the rotation matrix and displacement vector during multiple coordinate system transformations; 2) amplification of measurement errors in reference point locations during the stitching process; 3) unavoidable deviations in the installation of manual marker points, which may affect overall accuracy. Further research will focus on effective methods for correcting accumulated errors. Additionally, the generalizability of the optimal parameters was not discussed in detail in this study. Further investigation and validation of this issue will be conducted in future work. Moreover, the current detection system still relies on manual operations during data collection and processing. Future work will concentrate on the

development of a more intelligent real-time processing system to enable autonomous UAV flight, data collection, and analysis automation.

6. Conclusion

This paper presented a UAV-based alignment measurement technology using AM for alignment measurement during the construction phase of prefabricated bridges. The proposed high-precision alignment measurement system, based on overlapping marker points and sequential point lattice stitching, overcomes the inefficiencies associated with frequent station relocation and setup in TS measurements. Additionally, non-contact measurement enhances safety in complex construction environments, providing a fast and reliable solution for on-site alignment measurement. It is worth noting that the prompt alignment measurement feedback provided by this method refers to post-flight processing turnaround, not real-time onboard computation. Furthermore, to address the challenge of accurately identifying and locating small targets in aerial images captured in complex construction scenarios, an improved deep learning-based target recognition and localization method is designed, effectively increasing the recognition and localization accuracy of AM.

To verify the effectiveness of the proposed method, two experiments were conducted. Results of Experiment 2 demonstrated that under the optimal parameters obtained from Experiment 1, the elevation errors at all measurement points were within 5 mm when compared to the TS measurements, which satisfied the practical engineering requirements. Besides, a field experiment was conducted on a prefabricated steel-concrete bridge under construction. Within a 30 m range, a RMSE of 2.84 mm, as compared to the electronic TS measurements, was achieved, meeting the requirements of practical engineering. Therefore, the proposed method can accurately measure the deviation of the bridge alignment from the theoretical alignment using the spatial coordinates of the AM.

In conclusion, the key contributions of this study lie in advancing UAV-based alignment measurement by introducing a framework that improves small target detection in aerial imaging and offers a scalable, practical solution for fast bridge construction monitoring and quality control. However, there are still some limitations that require further investigation. First, the installation of AM remains time-consuming. Besides, for measurement spans exceeding 30 m, due to cumulative deviations in coordinate transformations, small errors in reference point measurements, marker installation deviations, and system errors, the error tends to increase with the shooting distance. Therefore, this method is more suitable for short-distance measurements, and long-distance measurements may be significantly affected by errors. Further optimization of correction methods is required to address these

challenges. Additionally, future research will prioritize the development of more intelligent systems to enable autonomous UAV operation and automated data processing, thereby improving accuracy and efficiency.

CRediT authorship contribution statement

Lu Deng: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Cheng Zhang:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **WeiQi Mao:** Resources, Funding acquisition. **Feng Zhang:** Software, Methodology, Data curation. **Lizhi Long:** Software, Methodology, Investigation. **Hao Dai:** Resources, Funding acquisition. **Jingjing Guo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgment

The authors would like to acknowledge the financial support provided by the National Key Research and Development Program of China (Grant No. 2023YFC3806800), the National Natural Science Foundation of China (Grant No. 52308312 and 52278177), and the Postgraduate Scientific Research Innovation Project of Hunan Province (Grant No. CX20240430).

Data availability

Data will be made available on request.

References

- [1] R. Hällmark, H. White, P. Collin, Prefabricated bridge construction across Europe and America, *Pract. Period. Struct. Des. Constr.* 17 (2012) 82–92, [https://doi.org/10.1061/\(ASCE\)SC.1943-5576.0000116](https://doi.org/10.1061/(ASCE)SC.1943-5576.0000116).
- [2] Zhang Ziyang, Deng Kailai, Xu Tengfei, State-of-the-art review of prefabricated concrete bridge structures in 2019, *J. Civ. Environ. Eng.* 42 (2020) 183–191, <https://doi.org/10.11835/j.issn.2096-6717.2020.104>.
- [3] Z. Ma, Y. Liu, J. Li, Review on automated quality inspection of precast concrete components, *Autom. Constr.* 150 (2023) 104828, <https://doi.org/10.1016/j.autcon.2023.104828>.
- [4] Y. Han, D. Feng, W. Wu, X. Yu, G. Wu, J. Liu, Geometric shape measurement and its application in bridge construction based on UAV and terrestrial laser scanner, *Autom. Constr.* 151 (2023) 104880, <https://doi.org/10.1016/j.autcon.2023.104880>.
- [5] Announcement of the Ministry of Housing and Urban-Rural Development on the Publication of the National Standard 'Engineering Surveying Standard', Access date: 2.9.2024. Available from: https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/202107/20210730_761396.html, 2024.
- [6] R.F. Annan, Y.Y. Ziggah, J. Ayer, C.A. Odutola, Accuracy assessment of heights obtained from total station and level instrument using total least squares and ordinary least squares methods, *Geoplann. J. Geomat. Plan.* 3 (2016) 87, <https://doi.org/10.14710/geoplanning.3.2.87-92>.
- [7] L. Beltramone, A. Rindinella, C. Vanneschi, R. Salvini, Multitemporal monitoring of rocky walls using robotic total station surveying and persistent Scatterer interferometry, *Remote Sens.* 16 (2024) 3848, <https://doi.org/10.3390/rs16203848>.
- [8] A. Khaloo, D. Lattanzi, K. Cunningham, R. Dell'Andrea, M. Riley, Unmanned aerial vehicle inspection of the Placer River trail bridge through image-based 3D modelling, *Struct. Infrastruct. Eng.* 14 (2018) 124–136, <https://doi.org/10.1080/15732479.2017.1330891>.
- [9] D. Moon, S. Chung, S. Kwon, J. Seo, J. Shin, Comparison and utilization of point cloud generated from photogrammetry and laser scanning: 3D world model for smart heavy equipment planning, *Autom. Constr.* 98 (2019) 322–331, <https://doi.org/10.1016/j.autcon.2018.07.020>.
- [10] T.F. Lam, H. Blum, R. Siegwart, A. Gawel, S.L. Sensor, An open-source, real-time and robot operating system-based structured light sensor for high accuracy construction robotic applications, *Autom. Constr.* 142 (2022) 104424, <https://doi.org/10.1016/j.autcon.2022.104424>.
- [11] Y. Cheng, F. Lin, W. Wang, J. Zhang, Vision-based trajectory monitoring for assembly alignment of precast concrete bridge components, *Autom. Constr.* 140 (2022) 104350, <https://doi.org/10.1016/j.autcon.2022.104350>.
- [12] S. Dong, J. Li, J. Ma, R. Pan, Z. Zhang, X. Kang, Illumination variation robust circular target based on digital image correlation method, *Measurement* 226 (2024) 114088, <https://doi.org/10.1016/j.measurement.2023.114088>.
- [13] R. Sun, S. Wang, M. Li, Y. Zhu, An algorithm for large-span flexible bridge pose estimation and multi-keypoint vibration displacement measurement, *Measurement* 240 (2025) 115582, <https://doi.org/10.1016/j.measurement.2024.115582>.
- [14] M. Rashidi, M. Mohammadi, S. Sadeghlu Kivi, M.M. Abdolvand, L. Truong-Hong, B. Samali, A decade of modern bridge monitoring using terrestrial laser scanning: review and future directions, *Remote Sens.* 12 (2020) 3796, <https://doi.org/10.3390/rs12223796>.
- [15] H. Yoon, J. Shin, B.F. Spencer Jr., Structural displacement measurement using an unmanned aerial system, *Comput.-Aided Civ. Infrastruct. Eng.* 33 (2018) 183–192, <https://doi.org/10.1111/mice.12338>.
- [16] Y. Han, G. Wu, D. Feng, Vision-based displacement measurement using an unmanned aerial vehicle, *Struct. Control. Health Monit.* 29 (2022) e3025, <https://doi.org/10.1002/stc.3025>.
- [17] S. Ri, J. Ye, N. Toyama, N. Ogura, Drone-based displacement measurement of infrastructures utilizing phase information, *Nat. Commun.* 15 (2024) 395, <https://doi.org/10.1038/s41467-023-44649-2>.
- [18] J. Cui, Y. Qin, Y. Wu, C. Shao, H. Yang, Skip connection YOLO architecture for noise barrier defect detection using UAV-based images in high-speed railway, *IEEE Trans. Intell. Transp. Syst.* 24 (2023) 12180–12195, <https://doi.org/10.1109/TITS.2023.3292934>.
- [19] T. Xu, Y. Wu, Y. Qin, S. Long, Z. Yang, F. Guo, Automatic steel girder inspection system for high-speed railway bridge using hybrid learning framework, *Comput.-Aided Civ. Infrastruct. Eng.* (2024), <https://doi.org/10.1111/mice.13409>.
- [20] X. Shi, Q. Xu, H. Ma, Parametrical study and optimization using unmanned aerial vehicle measurement in bridge construction alignment, *J. Tongji Univ. Nat. Sci.* 50 (2022) 32–41, <https://doi.org/10.11908/j.issn.0253-374x.21097>.
- [21] Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu, Object detection with deep learning: a review, *IEEE Trans. Neural Netw. Learn. Syst.* 30 (2019) 3212–3232, <https://doi.org/10.1109/TNNLS.2018.2876865>.
- [22] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 1137–1149, <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [23] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 2016, pp. 779–788, <https://doi.org/10.1109/CVPR.2016.91>.
- [24] Y. Liu, P. Sun, N. Wergeles, Y. Shang, A survey and performance evaluation of deep learning methods for small object detection, *Expert Syst. Appl.* 172 (2021) 114602, <https://doi.org/10.1016/j.eswa.2021.114602>.
- [25] L. Cui, P. Lv, X. Jiang, Z. Gao, B. Zhou, L. Zhang, L. Shao, M. Xu, Context-aware block net for small object detection, *IEEE Trans. Cybern.* 52 (2022) 2300–2313, <https://doi.org/10.1109/TCYB.2020.3004636>.
- [26] J. Fu, Y. Nie, W. Gong, Y. Wu, S.S.D. An Improved Network for small object detection based on dilated convolution and feature fusion, in: 2021 IEEE 4th Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. IMCEC, 2021, pp. 1261–1265, <https://doi.org/10.1109/IMCEC51613.2021.9482158>.
- [27] B. Mahaur, K.K. Mishra, Small-object detection based on YOLOv5 in autonomous driving systems, *Pattern Recogn. Lett.* 168 (2023) 115–122, <https://doi.org/10.1016/j.patrec.2023.03.009>.
- [28] H. Zhang, K. Zu, J. Lu, Y. Zou, D. Meng, EPSANet: an efficient pyramid squeeze attention block on convolutional neural network, in: L. Wang, J. Gall, T.-J. Chin, I. Sato, R. Chellappa (Eds.), *Comput. Vis. – ACCV 2022*, 2023, pp. 541–557, https://doi.org/10.1007/978-3-031-26313-2_33.
- [29] L. Yang, J. Zhong, Y. Zhang, S. Bai, G. Li, Y. Yang, J. Zhang, An improving faster-RCNN with multi-attention ResNet for small target detection in intelligent autonomous transport with 6G, *IEEE Trans. Intell. Transp. Syst.* 24 (2023) 7717–7725, <https://doi.org/10.1109/TITS.2022.3193909>.
- [30] J.-S. Lim, M. Astrid, H.-J. Yoon, S.-I. Lee, Small object detection using context and attention, in: 2021 Int. Conf. Artif. Intell. Inf. Commun. ICAIIC, 2021, pp. 181–186, <https://doi.org/10.1109/ICAIC51459.2021.9415217>.
- [31] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 2017, pp. 936–944, <https://doi.org/10.1109/CVPR.2017.106>.
- [32] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path Aggregation Network for Instance Segmentation, 2018, pp. 8759–8768, <https://doi.org/10.48550/arXiv.1803.01534>.
- [33] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, H. Ling, M2Det: a single-shot object detector based on multi-level feature pyramid network, *Proc. AAAI Conf. Artif. Intell.* 33 (2019) 9259–9266, <https://doi.org/10.1609/aaai.v33i1.33019259>.
- [34] M. Tan, R. Pang, Q.V. Le, EfficientDet: scalable and efficient object detection, in: 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR, 2020, pp. 10778–10787, <https://doi.org/10.1109/CVPR42600.2020.01079>.
- [35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Comput. Vis. – ECCV 2014*, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
- [36] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, X. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–14, <https://doi.org/10.1109/TIM.2022.3153997>.

- [37] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, K. Cho, Augmentation for Small Object Detection, arXiv, 2019, <https://doi.org/10.48550/arXiv.1902.07296>. Preprint. Available from:.
- [38] H. Zhang, M. Li, D. Miao, W. Pedrycz, Z. Wang, M. Jiang, Construction of a feature enhancement network for small object detection, *Pattern Recognit.* 143 (2023) 109801, <https://doi.org/10.1016/j.patcog.2023.109801>.
- [39] X. Wang, Q. Zhao, P. Jiang, Y. Zheng, L. Yuan, P. Yuan, LDS-YOLO: a lightweight small object detection method for dead trees from shelter forest, *Comput. Electron. Agric.* 198 (2022) 107035, <https://doi.org/10.1016/j.compag.2022.107035>.
- [40] B. Bosquet, D. Cores, L. Seidenari, V.M. Brea, M. Mucientes, A.D. Bimbo, A full data augmentation pipeline for small object detection based on generative adversarial networks, *Pattern Recognit.* 133 (2023) 108998, <https://doi.org/10.1016/j.patcog.2022.108998>.
- [41] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, S. Yan, Perceptual generative adversarial networks for small object detection, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 2017, pp. 1951–1959, <https://doi.org/10.1109/CVPR.2017.211>.
- [42] F.C. Akyon, S. Onur Altinuc, A. Temizel, Slicing aided hyper inference and fine-tuning for small object detection, in: 2022 IEEE Int. Conf. Image Process. ICIP, 2022, pp. 966–970, <https://doi.org/10.1109/ICIP46576.2022.9897990>.
- [43] H. Rezatofighi, N. Tsai, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: a metric and a loss for bounding box regression, in: 2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR, 2019, pp. 658–666, <https://doi.org/10.1109/CVPR.2019.00075>.
- [44] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: faster and better learning for bounding box regression, *Proc. AAAI Conf. Artif. Intell.* 34 (2020) 12993–13000, <https://doi.org/10.1609/aaai.v34i07.6999>.
- [45] Z. Xu, Z.H.U. Zhengwei, G.U.O. Yuying, L.I.U. Hui, Z. Hui, Multi-scale remote sensing small target detection based on cosSTR-YOLOv7, *Electron. Opt. Control.* 31 (2024) 28–34, <https://doi.org/10.3969/j.issn.1671-637x.2024.04.006>.
- [46] S.-J. Ji, Q.-H. Ling, F. Han, An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information, *Comput. Electr. Eng.* 105 (2023) 108490, <https://doi.org/10.1016/j.compeleceng.2022.108490>.
- [47] J. Wang, C. Xu, W. Yang, L. Yu, A Normalized Gaussian Wasserstein Distance for Tiny Object Detection, arXiv, 2022, <https://doi.org/10.48550/arXiv.2110.13389>. Preprint. Available from:.
- [48] R.A. McLaughlin, Randomized Hough transform: improved ellipse detection with comparison, *Pattern Recognit. Lett.* 19 (1998) 299–305, [https://doi.org/10.1016/S0167-8655\(98\)00010-5](https://doi.org/10.1016/S0167-8655(98)00010-5).
- [49] J. Wu, Robust real-time ellipse detection by direct least-square-fitting, in: 2008 Int. Conf. Comput. Sci. Softw. Eng., 2008, pp. 923–927, <https://doi.org/10.1109/CSSE.2008.789>.
- [50] D.K. Prasad, M.K.H. Leung, S.-Y. Cho, Edge curvature and convexity based ellipse detection method, *Pattern Recognit.* 45 (2012) 3204–3221, <https://doi.org/10.1016/j.patcog.2012.02.014>.
- [51] M. Fornaciari, A. Prati, R. Cucchiara, A fast and effective ellipse detector for embedded vision applications, *Pattern Recognit.* 47 (2014) 3693–3708, <https://doi.org/10.1016/j.patcog.2014.05.012>.
- [52] Q. Jia, X. Fan, Z. Luo, L. Song, T. Qiu, A fast ellipse detector using projective invariant pruning, *IEEE Trans. Image Process.* 26 (2017) 3665–3679, <https://doi.org/10.1109/TIP.2017.2704660>.
- [53] V. Pătrăucean, P. Gurdjos, R. Grompone von Gioi, Joint a contrario ellipse and line detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 788–802, <https://doi.org/10.1109/TPAMI.2016.2558150>.
- [54] C. Lu, S. Xia, M. Shao, Y. Fu, Arc-support line segments revisited: an efficient high-quality ellipse detection, *IEEE Trans. Image Process.* 29 (2020) 768–781, <https://doi.org/10.1109/TIP.2019.2934352>.
- [55] L. Wu, Y. Su, Z. Chen, S. Chen, S. Cheng, P. Lin, Six-degree-of-freedom generalized displacements measurement based on binocular vision, *Struct. Control. Health Monit.* 27 (2020) e2458, <https://doi.org/10.1002/stc.2458>.
- [56] L. Deng, T. Sun, L. Yang, R. Cao, Binocular video-based 3D reconstruction and length quantification of cracks in concrete structures, *Autom. Constr.* 148 (2023) 104743, <https://doi.org/10.1016/j.autcon.2023.104743>.
- [57] C. Zhou, Y. Gao, E.J. Chen, L. Ding, W. Qin, Deep learning technologies for shield tunneling: challenges and opportunities, *Autom. Constr.* 154 (2023) 104982, <https://doi.org/10.1016/j.autcon.2023.104982>.
- [58] W. Wang, J. Peng, W. Hu, J. Wang, X. Xu, Q. Zaheer, S. Qiu, A multi-degree-of-freedom monitoring method for slope displacement based on stereo vision, *Comput.-Aided Civ. Infrastruct. Eng.* 39 (2024) 2010–2027, <https://doi.org/10.1111/mice.13173>.
- [59] R. Zhu, D. Jiang, Z. Huang, L. Xie, D. Zhang, Q. Fei, Full-field modal identification using reliability-guided frequency-domain-based digital image correlation method based on multi-camera system, *Measurement* 211 (2023) 112567, <https://doi.org/10.1016/j.measurement.2023.112567>.
- [60] S. Hu, J. Zhang, An adaptive kernel correlation filter algorithm and synchronization error correction method for 3D motion reconstruction of flexible flapping wings using binocular vision, *Measurement* 228 (2024) 114388, <https://doi.org/10.1016/j.measurement.2024.114388>.
- [61] H. Habeenu, P. McGetrick, S. Taylor, D. Hester, UAS-based bridge displacement measurement using two cameras with non-overlapping fields of view, *Autom. Constr.* 167 (2024) 105687, <https://doi.org/10.1016/j.autcon.2024.105687>.
- [62] Z. Hu, Y. Hou, P. Tao, J. Shan, IMGTR: image-triangle based multi-view 3D reconstruction for urban scenes, *ISPRS J. Photogramm. Remote Sens.* 181 (2021) 191–204, <https://doi.org/10.1016/j.isprsjprs.2021.09.009>.
- [63] T. Schenk, From point-based to feature-based aerial triangulation, *ISPRS J. Photogramm. Remote Sens.* 58 (2004) 315–329, <https://doi.org/10.1016/j.isprsjprs.2004.02.003>.
- [64] D. Wei, H. Guo, Y. Wan, Y. Zhang, C. Li, G. Wang, Clustering, triangulation, and evaluation of 3D lines in multiple images, *ISPRS J. Photogramm. Remote Sens.* 218 (2024) 678–692, <https://doi.org/10.1016/j.isprsjprs.2024.10.001>.
- [65] G. Jocher, J. Qiu, A. Chaurasia, Ultralytics yolov8, 2023.
- [66] Z. Xia, X. Pan, S. Song, L.E. Li, G. Huang, Vision Transformer with Deformable Attention, arXiv, 2022. Preprint. Available from: <http://arxiv.org/abs/2201.00520>.
- [67] D. Nister, An efficient solution to the five-point relative pose problem, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (2004) 756–770, <https://doi.org/10.1109/TPAMI.2004.17>.
- [68] Z. Zhang, A flexible new technique for camera calibration, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 1330–1334, <https://doi.org/10.1109/34.888718>.
- [69] R.I. Hartley, In defense of the eight-point algorithm, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (1997) 580–593, <https://doi.org/10.1109/34.601246>.
- [70] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment — a modern synthesis, in: B. Triggs, A. Zisserman, R. Szeliski (Eds.), *Vis. Algorithms Theory Pract.*, 2000, pp. 298–372, https://doi.org/10.1007/3-540-44480-7_21.
- [71] Y. Chen, Y. Shen, D. Liu, A simplified model of three dimensional-datum transformation adapted to big rotation angle, *J. Wuhan Univ. Geomat. Inf. Sci.* 29 (2004) 1101–1105, <https://doi.org/10.13203/j.whugis2004.12.016>.