

Article

Research on Visual Pose Detection Method for Bridge Prestressed Corrugated Pipes Using SC-YOLOv11

Dong-Po Chen ¹, Hai-Bin Huang ^{2,*}, Si-Hao Zhang ², Yuan Cheng ¹ and Dong Liang ²

¹ ShiAn Branch, Hebei Expressway Group Co., Ltd., Shijiazhuang 051430, China

² School of Civil and Transportation Engineering, Hebei University of Technology, Tianjin 300401, China

* Correspondence: hbhuang@hebut.edu.cn

Abstract

During the fabrication of prestressed concrete beams, the quality and positional accuracy of the laid corrugated ducts (or prestressing ducts) directly influence the load-bearing capacity and durability of the beams. However, traditional manual inspection is inefficient, highly subjective, and difficult to achieve full coverage. To address this problem, this paper proposes an automated detection method that integrates improved YOLOv11-based pose estimation, robust curve fitting, and image stitching techniques. The method automatically identifies duct positions and evaluates laying quality. By incorporating the SE channel attention mechanism and the SPPFCSPC multi-scale pooling module, the SC-YOLOv11 model is developed, which significantly enhances the detection accuracy of slender corrugated pipe key points in environments with dense rebar occlusion. The RANSAC algorithm is employed to fit curves to the predicted key points, effectively suppressing the influence of outliers. Furthermore, the SIFT algorithm is used for precise stitching of drone-captured segmented images, which are then transformed into a unified front orthographic coordinate system of the entire box girder via perspective transformation, enabling accurate reconstruction of the corrected 2D layout of corrugated ducts across the full beam. Ablation experiments using 5-fold cross-validation demonstrate that SC-YOLOv11 improves mAP_{50} and mAP_{50-95} by 2.6% and 1.2%, respectively, with statistical significance (paired t -test, $p < 0.01$). The model achieves a per-image inference time of 6.37 ms, with 4.34 M parameters and 8.1 GFLOPs, meeting real-time requirements. In a 30 m prefabricated box girder field application, the measured section trajectory fitting curves of the corrugated ducts were compared with the design alignment, successfully identifying two abnormal locations where the laying deviation exceeded the allowable threshold. Cross-validation with on-site inspector records shows that over 92% of the measurement points agree within ± 10 mm. This method achieves a fully automated analysis chain from key point detection and curve fitting to deviation quantification, providing an efficient, non-contact, and traceable intelligent tool for quality control of bridge prestressed systems.

Keywords: corrugated pipe; SC-YOLOv11; RANSAC fitting; image stitching; perspective transformation

Academic Editor: Antonio Caggiano

Received: 1 July 2026

Revised: 1 August 2026

Accepted: 5 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In modern society, bridges serve as critical transportation infrastructure, playing an essential role in spanning complex terrains, reducing road intersections, shortening spatial and temporal distances, and improving traffic capacity [1–3]. With the rapid advancement of urbanization, the scale and complexity of bridge engineering projects have continued to increase. Prestressed reinforcement plays a vital role in bridge structures. By introducing prestressing forces, it effectively enhances the crack resistance of concrete bridges, reduces crack widths under bending moments, and thereby improves the durability and safety of the structure [4,5]. Moreover, prestressed reinforcement helps control long-term deflections, ensures service stability, increases load-bearing capacity to accommodate greater dynamic and static loads, and extends the overall service life of bridges.

In bridge construction, corrugated ducts act as protective conduits for prestressed tendons. Their positions in the rectified image plane, pose, and deformation significantly affect the stress distribution in the tendons and the efficiency of prestress transfer. The installation quality of these ducts directly influences the effectiveness of prestress transmission; therefore, ensuring accurate placement is a critical step in bridge construction [6,7]. However, traditional corrugated duct laying methods rely heavily on manual measurement and empirical judgment. During construction, factors such as environmental variations, formwork displacement, and vibrations from concrete pouring can cause the ducts to bend, twist, or deviate from their designed positions. These geometric deviations not only compromise positioning accuracy but may also lead to prestress eccentricity, stress concentrations, and incomplete grouting, ultimately reducing the bridge's durability and safety. Since such issues are often difficult to detect and rectify after concrete placement, real-time monitoring of corrugated duct pose during construction is of great significance for early risk identification and timely quality control [8,9].

Currently, common on-site detection methods for corrugated ducts include manual measurement, total station surveying, and structural probing [10,11]. Although these approaches can assess laying quality to some extent, they generally suffer from low efficiency and limited capability for continuous monitoring. As bridge projects grow in scale and complexity, traditional methods can no longer satisfy the demands of real-time and intelligent monitoring during construction. Consequently, developing a non-contact, rapid, and intelligent technology for corrugated duct pose detection has become an important direction in the informatization and quality control of bridge construction.

In recent years, computer vision and deep learning technologies have advanced rapidly in civil engineering, offering new tools for intelligent detection during construction [12,13]. Kim et al. [14] proposed an automated method based on an improved Faster R-CNN for detecting unsafe worker behaviors in high-altitude operations and validated its effectiveness and limitations. Yong Y.P. et al. [15] developed a collision avoidance system combining deep learning-based object detection with distance sensors to prevent potential collisions during crane lifting. Fang et al. [16] introduced a Mask R-CNN-based computer vision approach to detect workers crossing structural supports, enabling timely identification of unsafe behaviors. Similar studies by Fang et al. [17], Wei et al. [18], Son et al. [19], Fang et al. [20,21], and Li et al. [22] have demonstrated the effectiveness of CNN-based models (including Faster R-CNN and ResNet) in recognizing construction workers, personal protective equipment (PPE) compliance, and unsafe actions from site images or videos.

These camera-based object detection methods offer notable advantages, including low cost, ease of deployment, and minimal interference with ongoing construction activities, thereby enabling continuous monitoring [23]. Therefore, camera systems and image processing techniques can be effectively applied to detect the placement of corrugated ducts and identify potential quality issues in prestressed tendon arrangements. In bridge

construction environments, corrugated ducts typically exhibit regular curved geometries, and their spatial pose can be reconstructed by detecting and fitting key feature points. Combining deep learning with geometric fitting techniques allows high-precision reconstruction of duct morphology even in complex construction settings.

While CNN-based object detection has seen wide adoption in construction safety, recent advances in key point estimation and curve reconstruction have been largely driven by transformer architectures and lightweight high-resolution networks—yet their application to industrial slender-object scenarios remains unexplored. ViTPose [24] demonstrated that a plain vision transformer can achieve strong human pose estimation performance; TokenPose [25] further introduced learnable keypoint tokens to model spatial relationships among joints. Lightweight models such as HRNet [26] maintain high-resolution representations for precise key point localization, and RTMPose [27] pushes real-time multi-person pose estimation to edge devices. In the domain of curve and line detection, CLRNet [28] has shown promise for structural feature extraction, while real-time detection transformers like RT-DETR [29] combine DETR's paradigm with efficient hybrid encoders for faster inference.

These advances motivate the exploration of a specialized solution for corrugated duct inspection, where the task is uniquely characterized by extreme aspect ratios, dense rebar occlusion, and the need to output ordered key points for subsequent curve fitting. Rather than directly adopting existing models designed primarily for human pose or generic object detection, we build upon the YOLOv11 framework for its proven anchor-free key point head, real-time inference capability, and modular architecture that facilitates targeted modifications. By integrating the SE channel attention mechanism and the SPPFCSPC multi-scale pooling module into this backbone, we construct SC-YOLOv11 to address the specific challenges of slender duct detection under occlusion while maintaining the efficiency required for UAV-based field deployment.

Based on the above background, this paper proposes a bridge prestressed tendon pose detection and quality analysis method that integrates the SC-YOLOv11 network with curve fitting techniques. The method is applied to deformation monitoring and installation quality control of corrugated ducts during the construction phase. Taking site-captured images as input, the approach first employs SC-YOLOv11 for automatic detection and localization of corrugated duct key points. RANSAC [30] is then used to perform robust curve fitting on the detected points. Subsequently, the SIFT [31] algorithm stitches segmented box-girder images, connecting the fitted curves to obtain the complete spatial layout of corrugated ducts across the entire girder. Finally, by comparing the reconstructed 2D trajectories with the design drawings, the deviations and pose changes measured in the rectified image plane are quantified, enabling inference of prestressed tendon positions and comprehensive construction quality assessment. This system facilitates rapid, non-contact detection on construction sites and provides reliable data support for dynamic quality control.

2. Methodology

This paper systematically elaborates on the execution flow of the algorithm, provides an in-depth analysis of the corrugated pipe curve fitting algorithm and image stitching algorithm, and finally introduces the model evaluation methodology.

2.1. Overview

This paper proposes a corrugated pipe quality detection method based on machine vision and curve fitting, optimizing the YOLOv11 (Ultralytics YOLOv11 v11.0.0) [32] model to extract features of corrugated pipes and box girder frameworks. The proposed method is illustrated in the framework diagram shown in Figure 1. First, data processing

is conducted: images of laid corrugated pipes are captured by unmanned aerial vehicles (UAVs) as raw data and annotated, including key information such as corrugated pipe feature points and box girder framework vertices. Since precast box girders are generally long, the UAV captures images at intervals determined by the shooting distance. To ensure that images assigned to different data subsets do not contain the same scene, the image acquisition positions were carefully spaced, and subsequent dataset splitting was performed at the girder-segment level. Second, the SE channel attention mechanism and SPPFCSPC multi-scale pooling module are embedded into the YOLOv11 model to substantially improve key point localization accuracy under dense occlusion—a prerequisite for the subsequent geometric algorithms to function correctly. Training is performed based on image annotations, and the pre-trained optimized YOLOv11 model is employed to analyze the images of laid corrugated pipes, obtaining the vertex coordinate information of the corresponding box girder framework and the feature point coordinate information of the corrugated pipes. Next, the coordinate information of corrugated pipe feature points obtained from different shooting directions is used for curve fitting with the RANSAC algorithm. Subsequently, perspective correction is performed on the box girder images based on the vertex coordinates of the box girder framework obtained by the optimized YOLOv11, yielding the perspective transformation matrix. Box girder images with sequential relationships are stitched using the SIFT algorithm to obtain the image transformation matrix. This method performs perspective transformation after stitching all images, thereby reducing errors caused by multiple perspective transformations. Finally, by aggregating all transformed fitted curves, the overall morphology and distribution of corrugated pipes throughout the entire box girder can be obtained, and the relative position of corrugated pipes with respect to the entire box girder framework can be accurately calculated.

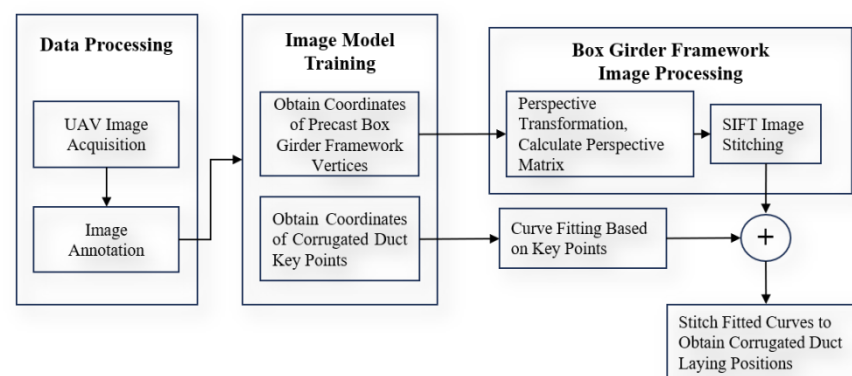


Figure 1. Experimental method framework diagram.

Although YOLOv11, RANSAC, and SIFT are individually mature, simply chaining them together fails in the heavily occluded scenes encountered during corrugated duct inspection. A generic YOLOv11 detector, confronted with ultra-narrow ducts woven through dense rebar, inevitably outputs a high fraction of outlier key points along with substantial localization errors. Under these conditions, RANSAC's ability to recover the correct duct trajectory degrades sharply: with such a low inlier ratio, the consensus set is easily hijacked by noise, producing a fitted curve that bears no physical relation to the actual pipe. At the same time, SIFT-based stitching and perspective correction depend critically on the precision of the detected box-girder corner points. When these key points are poorly localized, the errors propagate into the estimated homographies, accumulate along the image sequence, and ultimately render millimetre-level deviation measurement unreliable. The proposed SC-YOLOv11 is designed to break this chain by integrating SE

channel attention, which suppresses background clutter and strengthens duct-specific features, and SPPFCSPC, which supplies multi-scale context to recover fine spatial detail under occlusion. These modules jointly reduce the key point error (NME from 0.184 to 0.138) and lift the inlier rate high enough for RANSAC to reliably identify the true consensus set. In this pipeline, the improved detection precision is not an incremental gain but the enabling condition for the subsequent geometric stage to produce physically meaningful measurements—something that simply cascading YOLOv11, RANSAC, and SIFT cannot deliver.

2.2. Network Model Structure Optimization

To enhance the robustness of the network model in recognizing partially occluded corrugated pipe targets, it is necessary to optimize the YOLOv11 network framework. To this end, this paper introduces the SE module (Squeeze and Excitation) [33] and replaces the SPPF [34] module in the backbone network with the SPPFCSPC module. The improved network is referred to as SC-YOLOv11, and its structure is shown in Figure 2.

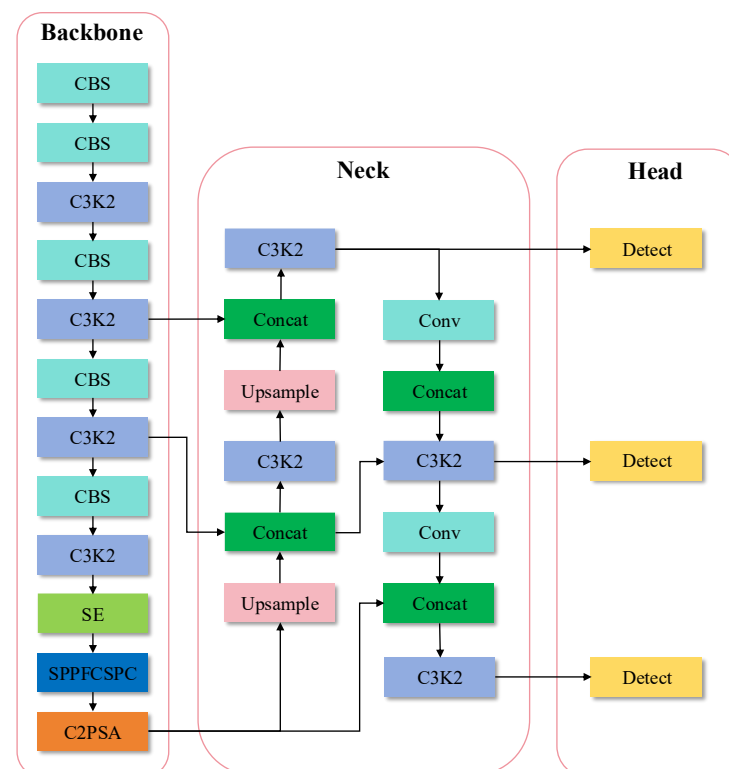


Figure 2. Improved YOLOv11-pose Network Architecture.

In corrugated duct pose estimation, two inherent challenges arise: the ducts' extreme length-to-width ratio makes them highly sensitive to occlusion by the dense rebar framework, and the low contrast against concrete leads to ambiguous feature boundaries. To address both, we propose a synergistic combination of SE and SPPFCSPC. The SE module re-weights channels to enhance duct-specific structural patterns while suppressing rebar and shadow noise; the SPPFCSPC module expands the receptive field and preserves fine local details critical for localising narrow, elongated ducts. Together, SE provides global channel attention to maintain object coherence, while SPPFCSPC supplies multi-scale spatial context to recover from local occlusions. This joint design, rather than a trivial addition of existing modules, forms a dedicated backbone for slender and heavily occluded targets in industrial scenes.

2.2.1. SE Attention Module

When detecting the corrugated ducts used for embedding prestressed tendons in precast box girders, the dense reinforcement framework inside the girder causes severe occlusion of the ducts. This significantly reduces their saliency in images and adversely affects detection accuracy. To mitigate this issue, the SE (Squeeze-and-Excitation) channel attention module is embedded at specific stages of the YOLOv11 backbone network. The module adaptively models inter-channel dependencies through a “squeeze-and-excitation” mechanism, thereby enhancing the response to structural features associated with the corrugated ducts while suppressing background noise and irrelevant information. This effectively improves both the recognition accuracy and robustness of the model for corrugated duct targets in complex internal environments of box girders. The structure of the SE attention module is illustrated in Figure 3.

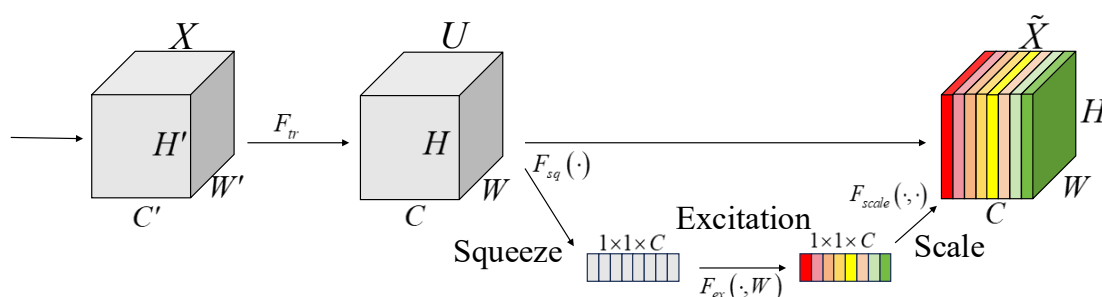


Figure 3. SE Structure Diagram.

As shown in Figure 3, X represents the input feature map with C channels, and height and width of H and W , respectively. The SE module applies squeeze, excitation, and recalibration operations to X sequentially, generating a one-dimensional vector corresponding to the number of channels C , which is used to characterize the importance scores of each channel. These weights are then used to rescale the original feature map, achieving channel-wise feature recalibration and improving the utilization efficiency of informative features. The detailed operational steps of the SE module are as follows:

Squeeze (F_{sq}): The module first performs a global average pooling operation on the input feature map, compressing the two-dimensional spatial information of each channel into a single scalar value that captures the global activation intensity of that channel. The mathematical expression is:

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

where u_c represents the input feature map.

Excitation (F_{ex}): In this stage, the squeezed feature descriptor first passes through a fully connected layer that reduces its dimension from $1 \times 1 \times C$ to $1 \times 1 \times C/r$ using a reduction ratio r . This dimensionality reduction effectively lowers computational complexity. The compressed features are then fed into a ReLU activation function, followed by another fully connected layer that restores the dimension back to $1 \times 1 \times C$. During this process, the inter-channel dependencies are learned to generate weight coefficients for each channel. Finally, a sigmoid activation function is applied to normalize the weights to the range $[0, 1]$.

Scale (F_{scale}): The weight coefficients learned from the excitation operation are then multiplied element-wise with the original feature map to recalibrate the channel-wise features. The mathematical expression is:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = u_c \cdot s_c \quad (2)$$

where s_c represents the normalized weights.

In summary, the SE module obtains a global statistical description of each channel through the squeeze operation and adaptively weights the features according to the excitation operation. By explicitly modeling the interdependencies between channels, it strengthens the response to critical features, suppresses redundant or interfering ones, and thereby enhances the model's capability for feature reconstruction and representation in complex scenes.

2.2.2. SPPFCSPC Pooling Module

In conventional convolutional neural networks, the input image size is typically fixed. Consequently, various cropping and resizing operations are often required during actual detection tasks. These preprocessing steps can lead to a loss of feature representation capability, particularly degrading the detection performance for small and slender targets. To address this limitation, spatial pyramid pooling techniques are commonly inserted into neural networks. In this paper, the original SPPF pooling module in the YOLOv11 backbone is replaced with the more advanced SPPFCSPC pooling module. This substitution not only provides richer multi-scale feature representations but also effectively mitigates the over-smoothing problem caused by pooling operations, resulting in improved detection accuracy and robustness for small-target detection tasks. The structure of the SPPFCSPC pooling module is illustrated in Figure 4.

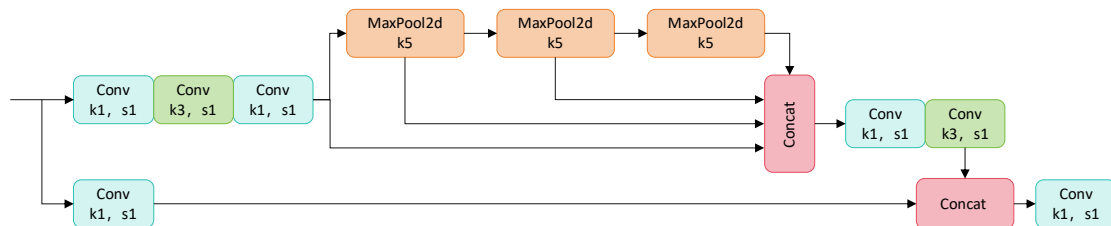


Figure 4. SPPFCSPC Structure Diagram.

The SPPFCSPC module consists of two main sub-modules: the SPP (Spatial Pyramid Pooling) sub-module and the FCSPC sub-module. The SPP sub-module applies spatial pyramid pooling to map the input features into multiple scale spaces, aggregating contextual information across different receptive fields and thereby generating feature representations that capture multi-scale details. In contrast, the FCSPC sub-module employs standard convolutions with multiple kernel sizes to extract features from targets of varying scales, further enhancing the overall feature representation capability. The outputs of these two branches are subsequently fused along the channel dimension. This design improves detection accuracy while maintaining high computational efficiency.

By performing multiple max-pooling operations on the input features, the SPPFCSPC module strengthens the network's ability to model semantic information of targets. Furthermore, the improved structure promotes deep fusion between local fine-grained details and global contextual features. This effectively expands the receptive field, enhances the network's adaptability, and improves detection accuracy in complex scenes.

2.3. Perspective Transformation

In Section 2.1, to accurately obtain the relative positions of the corrugated ducts, it is necessary to convert the captured box girder images into a front orthographic view. This allows for more intuitive comparison and analysis with the design drawings. To eliminate the effects of perspective distortion on the images, this study employs a perspective transformation algorithm to process the original images. The perspective relationships in the images are adjusted to a frontal orthographic view, thereby achieving more accurate

geometric restoration and facilitating subsequent image analysis and comparison. The coordinates of the four vertices of the box girder are first detected using the network model. Perspective transformation is then applied according to the following formula:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

The actual coordinates (x', y') are subsequently obtained through homogeneous coordinate normalization:

$$x' = \frac{x'}{w'}, y' = \frac{y'}{w'} \quad (4)$$

where (x, y) denotes a point in the original image, (x', y') is the corresponding point after perspective transformation, and H is the 3×3 perspective transformation matrix.

Since the perspective transformation matrix is defined up to a scale factor, it possesses eight degrees of freedom. It can represent a general planar projective transformation that includes translation, rotation, scaling, shearing, and perspective distortion. By establishing four pairs of corresponding points between the original image and the desired orthographic view, the perspective transformation matrix H is computed. The curves fitted in the original image can then be corrected into the front orthographic view using this transformation matrix.

To further convert corrected image coordinates to actual physical dimensions, two coordinate systems are defined: the image coordinate system (u, v) with its origin at the upper-left corner of the rectified orthographic image, where u points to the right and v points downwards, measured in pixels; and the physical coordinate system (x', y') with its origin at the lower-left corner of the box-girder framework, where x' is along the beam span direction and y' is vertically upward, in millimeters. Using the four framework vertices detected by SC-YOLOv11 and the known design dimensions—for instance, the width W_{mm} and height H_{mm} —together with their pixel spans W_{pix} and H_{pix} , a mm-per-pixel scale factor is obtained: where $\lambda_x = W_{mm}/W_{pix}$, $\lambda_y = H_{mm}/H_{pix}$. This perspective transformation simultaneously corrects perspective distortion caused by the camera view and residual lens distortion, eliminating the need for a separate intrinsic calibration. All subsequent physical deviations are converted from pixel errors using λ_x and λ_y .

2.4. RANSAC Algorithm

In the experiments of this paper, due to factors such as occlusion during the corrugated pipe recognition process, feature points may not be captured completely and clearly, potentially introducing noise. Therefore, this paper employs the RANSAC algorithm to fit the corrugated pipe laying curve. RANSAC is a random sample consensus estimation method that can, in datasets containing outliers, repeatedly randomly extract minimum subsets from samples for model fitting, evaluate model consistency through the entire dataset, and ultimately obtain model parameter estimates for the maximum consensus set.

Since fitting a cubic polynomial requires estimating four independent parameters, at least four sample points are needed to determine a unique set of parameter solutions under the premise that the design matrix formed by the selected sample points is full-rank. Therefore, the minimum sample set size for each random sampling in this RANSAC fitting is 4. To ensure that there is at least one sampling obtaining an all-inlier sample with a given confidence level p , the following formula is used:

$$K \geq \frac{\log(1-p)}{\log(1-w^s)} \quad (5)$$

where p is the desired confidence level, ε is the prior estimate of the inlier ratio, and s is the minimum number of samples per iteration, with a value of 4 in this case.

In the implementation, the residual threshold was set to 10 pixels to match the typical key point localization error in the images. The maximum number of iterations was fixed at 500. Given the densely occluded construction environment, a conservative inlier ratio of 0.5 was assumed. With a minimum sample size of 4 and 500 iterations, the confidence level $p > 0.999$, i.e., the probability that at least one all-inlier subset is sampled is greater than 99.9%. The optimal model is selected as the candidate with the largest consensus set. Final parameters are then re-estimated from all inliers using least squares and SVD.

In each iteration, 4 coordinate points are randomly sampled without replacement from the dataset, and the coefficient vector $\mathbf{a} = [a_0, a_1, a_2, a_3]^T$ is solved based on these 4 coordinate points. The residuals are calculated point by point according to the coefficient vector, as shown in Equation (6), and samples are classified as inliers or outliers based on the relationship between residuals and the threshold:

$$r_i = |y_i - \hat{y}_i| \quad (6)$$

where r_i is the absolute residual of the i sample point, y_i is the observed value, and $\hat{y}_i$ is the predicted value under the candidate model parameters.

Through multiple iterations, the optimal candidate model is continuously updated, and the best consensus set is ultimately obtained. At this point, a final least squares fitting is performed on this best consensus set to obtain the final function formula. The least squares method first constructs matrices based on the required x_i and y_i for fitting, as shown below:

$$\mathbf{V} = \begin{bmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_m^3 & x_m^2 & x_m & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad (7)$$

where m represents the number of observed data points, and $\mathbf{V}$ is the design matrix constructed from the input data.

After constructing matrices $\mathbf{V}$ and observation vector $\mathbf{y}$, the coefficient vector $\mathbf{a} = [a_0, a_1, a_2, a_3]^T$ is determined by solving Equation (4). In actual computation, the inverse matrix is not directly calculated; instead, SVD [35] is employed to avoid problems caused by numerical instability. SVD decomposition decomposes matrix $\mathbf{V}$ into the product of three matrices: $\mathbf{V} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, thereby obtaining the coefficient $\mathbf{a}$ by solving a stable linear equation, avoiding the accumulation of numerical errors. This method not only improves computational accuracy but also enhances the stability of the model when processing larger datasets.

$$\min_a \|\mathbf{V}\mathbf{a} - \mathbf{y}\|_2^2 \quad (8)$$

2.5. SIFT Algorithm

The Scale-Invariant Feature Transform (SIFT) algorithm is a robust local feature extraction method for images. It can stably detect and describe feature points even under conditions of scaling, rotation, illumination changes, and partial occlusion. The algorithm mainly consists of four steps: scale-space extrema detection, key point localization, orientation assignment, and key point description.

Scale-space extrema detection begins by applying Gaussian blur to the image using convolution kernels of multiple scales, constructing a Gaussian pyramid composed of blurred images at different scales. Next, the difference between adjacent levels of the

Gaussian pyramid is computed to generate the Difference of Gaussian (DoG) pyramid. Then, each pixel in the middle layers of the DoG pyramid is examined. For a given pixel, its value is compared with the 26 neighboring pixels in a 3×3 window across the current scale and the scales immediately above and below. If the pixel value is the maximum or minimum among these 26 neighbors, it is identified as a candidate extremum point.

These candidate key points correspond to local extrema in the DoG space, but their coordinates are discrete. Key point localization refines these candidates by fitting a quadratic function to the DoG space to determine the precise location and scale of each key point.

After accurate localization, the dominant orientation of each key point is determined. In the orientation assignment step, the gradient magnitude and direction are calculated for all pixels in a neighborhood around the key point. A gradient orientation histogram is then constructed within this region, and the direction with the highest accumulated value is selected as the dominant orientation. To achieve rotation invariance, the coordinate axes in the local neighborhood are rotated to align with this dominant orientation.

Finally, key point description is performed. The rotated region around the key point is divided into smaller sub-windows. For each sub-window, the gradient magnitudes and orientations are computed. A Gaussian weighting function is applied to give higher weight to pixels closer to the key point. Gradient orientation histograms are then built for each sub-window, and the accumulated values for each orientation bin are calculated. In this way, the descriptor of each key point is formed by concatenating the histograms from all sub-windows, resulting in a vector that encodes both the magnitude and directional information of the local gradients.

2.6. Model Performance Evaluation

In this study, the performance of the proposed pose estimation algorithm is comprehensively evaluated using the metrics of Precision, Recall, mAP_{50} , and mAP_{50-95} .

The formulas for *Precision* and *Recall* are defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

where TP , FP , and FN denote the number of true positives, false positives, and false negatives, respectively.

Mean Average Precision (mAP) is a key comprehensive metric for evaluating model performance. It considers not only the localization accuracy but also the classification correctness across object categories. The mAP is calculated as the average of the Average Precision (AP) values over all categories. The formula is given by:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (11)$$

where N is the total number of categories, and AP_i is the average precision for the i category. AP_i is obtained as the area under the Precision-Recall (P-R) curve for each category. In this paper, mAP_{50} refers to the mAP value when the Intersection over Union (IoU) threshold is 0.5, while mAP_{50-95} represents the average mAP across IoU thresholds ranging from 0.50 to 0.95, serving as a stricter evaluation criterion.

To assess the quality of the curve fitting, the coefficient of determination R^2 is adopted. It reflects the goodness of fit of the fitting algorithm to the data. The formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted (fitted) value, $\bar{y}$ is the mean of the actual values, and n is the number of samples.

The value of R^2 ranges is $(-\infty, 1]$, representing the proportion of the variance in the dependent variable that is predictable from the independent variable. An R^2 value closer to 1 indicates a better fit. This metric is dimensionless and is not affected by image resolution or data scale, allowing for objective and intuitive comparison of regression performance across different models or datasets. In this study, R^2 is presented in decimal form. We also report the Percentage of Correct Key Points (PCK) and Normalized Mean Error (NME) to evaluate key point localization. $PCK_{0.05}$ and $PCK_{0.1}$ denote the fraction of key points with error below 5% and 10% of the object scale, while NME is the average normalized Euclidean distance between predicted and ground-truth keypoints. In addition, to reflect the actual physical accuracy of key point localization, the physical root-mean-square error (RMSE, in cm) and mean absolute error (MAE, in cm) are computed from key point coordinates using the scale factor λ defined in Section 2.3.

3. Data Preprocessing and Experimental Environment

The dataset used in this study was collected using a UAV (unmanned aerial vehicle) to capture both top-down and near-frontal views of precast box girders. All images have a resolution of 3840×2160 pixels, as shown in Figure 5. Specifically, Figure 5a presents the top-down view captured with the UAV positioned vertically at a height of 15 m above the ground. Figure 5b shows the near-frontal view, where the UAV was positioned approximately 25 m horizontally from the girder at a height of 15 m above the ground, with a pitch angle of -30° . These data were collected at an expressway construction site in Hebei Province, China, involving 22 precast box girders (each with a span of 30 m). For each girder, two independent UAV flights were conducted, one for the overhead view and one for the near-frontal view, totaling 44 flight sorties. All images were captured under clear weather and natural lighting conditions, as the UAV operated only in rain-free and fog-free conditions. The corrugated ducts are plastic pipes with an outer diameter of approximately 50 mm, and partial occlusion exists in some areas. In the top-down view images, two corrugated ducts are typically visible. In the side-view cross-section, five corrugated ducts are physically installed; however, the lowest duct is completely hidden by the dense reinforcement cage and is not visible in the captured images. Therefore, only the upper four ducts are annotated and used in the subsequent detection, curve fitting, and deviation analysis.

During the annotation process, the Labelme tool (v5.8.3) was employed to label the corrugated ducts and the precast box girder framework in the images. In addition, key point annotation rules were defined as follows to enable subsequent pose estimation. For top-down images, where the corrugated ducts run horizontally across the box girder, two boundary points were first placed at a short distance inward from each visible end of the duct. The remaining seven key points were then equally spaced between these two boundary points, resulting in a total of nine key points per duct and corresponding to eight equal intervals. If a duct was partially outside the frame, key points were marked only for the portion within the field of view. For near-frontal (side-view) images, the segment of each duct between two adjacent vertical reinforcement bars was used. Two boundary points were first positioned at a small offset inward from the two reinforcement bars. The remaining three key points were then equally distributed between those two boundary

points, yielding a total of five key points and corresponding to four equal intervals. When a key point location was fully occluded (e.g., by other structural components or heavy shadows), that point was skipped and left unannotated. A pose estimation approach was adopted to facilitate the fitting of the corrugated duct trajectories.

The processed image dataset was subsequently divided into training, validation, and test sets. To avoid data leakage, the split was performed based on capture time: images taken at different times and covering different viewing angles of the box girder were assigned to separate subsets. Consequently, all images from the same or adjacent time periods, which may share overlapping views, are kept entirely within one subset and never appear across different data splits. Since each box girder was photographed in its own separate sessions, splitting the data by capture time is equivalent to splitting by girder. As a result, all images belonging to the same girder are guaranteed to be placed entirely within one subset, and no images of the same girder appear in more than one subset. The split ratio was kept at 8:1:1, and the specific numbers remain as presented in Table 1.



Figure 5. Example of the dataset: (a) Overhead. (b) Front view.

Table 1. Corrugated Duct Pose Estimation Dataset.

View	Training Set	Validation Set	Test Set	Ratio	Total
Overhead	263	33	33	8:1:1	329
Front view	246	30	30	8:1:1	308

In the experiments, the system was configured with PyTorch 2.6.0+cu118, an NVIDIA GeForce RTX 5070Ti GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 16 GB of video memory, an Intel Core i5-14600KF processor (Intel Corporation, Santa Clara, CA, USA; 14 cores and 20 threads), and 32 GB of system memory. The Python environment was version 3.9, and CUDA 11.8 was used to ensure compatibility with PyTorch 2.6.0+cu118. During model training, a total of 200 epochs were performed with a batch size of 4, a default learning rate of 0.01, and an input image size of 1280×1280 pixels.

4. Experimental Results

This paper conducted ablation experiments on the optimized model, demonstrated its prediction performance, and compared the results with evaluation results of other models. Additionally, the corrugated pipe path curve fitting process was presented, along with the stitching and correction of box girders and predicted curves based on the relationships between consecutive images.

4.1. Ablation Experiments

On the dataset constructed in this paper, based on the YOLOv11 model, the SE attention mechanism was embedded into the backbone network, and the SPPF module was replaced with the SPPFCSPC module. The performance data of ablation experiments on *precision*, *recall*, *mAP₅₀*, and *mAP_{50–95}* metrics are shown in Table 2.

Table 2. Ablation Experiment Results.

Model	Precision	Recall	mAP ₅₀	mAP _{50–95}	PCK _{0.05}	PCK _{0.1}	NME
YOLOv11	66.5% ± 0.8%	60.2% ± 0.9%	55.5% ± 0.7%	17.1% ± 0.5%	51.2% ± 1.0%	59.8% ± 0.9%	0.194 ± 0.008
YOLOv11+SE	70.5% ± 0.7%	65.6% ± 0.8%	58.4% ± 0.6%	21.5% ± 0.6%	55.6% ± 0.9%	64.5% ± 0.8%	0.161 ± 0.007
YOLOv11+CBAM	67.6% ± 0.7%	63.1% ± 0.7%	56.1% ± 0.6%	18.9% ± 0.6%	52.8% ± 0.8%	60.5% ± 0.9%	0.183 ± 0.008
YOLOv11+CA	68.4% ± 0.8%	64.6% ± 0.9%	56.8% ± 0.8%	20.5% ± 0.5%	53.8% ± 0.9%	62.8% ± 0.8%	0.191 ± 0.007
YOLOv11+ECA	70.9% ± 0.6%	65.2% ± 0.8%	57.5% ± 0.7%	21.2% ± 0.4%	55.2% ± 0.9%	63.9% ± 0.7%	0.168 ± 0.006
YOLOv11+SPPFCSPC	69.9% ± 0.8%	66.1% ± 0.8%	57.6% ± 0.7%	21.4% ± 0.5%	54.3% ± 1.0%	63.9% ± 0.9%	0.167 ± 0.008
YOLOv11+SE+SPPFCSPC	71.2% ± 0.6%	66.5% ± 0.7%	59.1% ± 0.6%	22.3% ± 0.4%	57.3% ± 0.8%	67.6% ± 0.7%	0.138 ± 0.006

The experimental results clearly demonstrate that introducing the SE attention mechanism and SPPFCSPC module into YOLOv11 can consistently and significantly improve detection performance. Compared with the original YOLOv11, after embedding the SE attention mechanism alone, *precision* improved by 1.1% and *mAP₅₀* improved by 1.9%, indicating that channel attention effectively suppressed background noise interference. Replacing SPPF with the SPPFCSPC module alone improved recall by 1.1%, showing that more efficient multi-scale feature fusion significantly enhanced detection capability for small and occluded targets. When both mechanisms acted simultaneously, optimal performance was achieved, with *precision* increasing by 1.7%, *recall* by 1.3%, *mAP₅₀* by 2.6%, and *mAP_{50–95}* by 1.2%, demonstrating obvious synergistic gains. The novelty of SC-YOLOv11 lies not in the isolated introduction of SE or SPPFCSPC, but in their task-specific integration that jointly addresses the challenges of slender geometry and dense occlusion, a scenario rarely tackled by prior pose estimation networks.

Overall, the combination of SE and SPPFCSPC achieved dual improvements in *precision* and *recall*, enabling the improved YOLOv11 to possess stronger robustness and practical value in complex scenarios. The key point metrics in Table 2 further support these improvements. Baseline YOLOv11 reaches a *PCK_{0.05}* of 0.522 and an *NME* of 0.184. Adding the SE module alone raises *PCK_{0.05}* to 0.556 and reduces *NME* to 0.161. Replacing SPPF with SPPFCSPC yields a *PCK_{0.05}* of 0.543 and an *NME* of 0.167. The complete SC-YOLOv11 achieves the highest *PCK_{0.05}* (0.573) and the lowest *NME* (0.138), confirming substantially

better key point accuracy. To assess statistical significance, 5-fold cross-validation was employed: each configuration was trained five times, each time using a different fold for validation while the remaining four folds were used for training. The reported metrics are averaged across the five folds, and a paired t -test was applied to compare the proposed SC-YOLOv11 against the baseline YOLOv11. The resulting p -values for the key metrics mAP_{50} and $PCK_{0.05}$ were both below 0.01, confirming that the observed improvements are statistically significant.

The prediction effects are shown in Figure 6, where Figure 6a presents the top-down view and Figure 6b presents the frontal view. From the visualization results, it can be observed that the original YOLOv11 predicted numerous outlier key points for corrugated pipes, with some key points significantly deviating from the corrugated pipes, seriously affecting the accuracy of subsequent curve fitting and laying path reconstruction. In contrast, the key points detected by the improved SC-YOLOv11 almost all precisely fell on the corrugated pipes, with outliers substantially reduced, significantly improving the robustness and accuracy of corrugated pipe pose estimation and providing high-quality feature point inputs for subsequent RANSAC curve fitting.

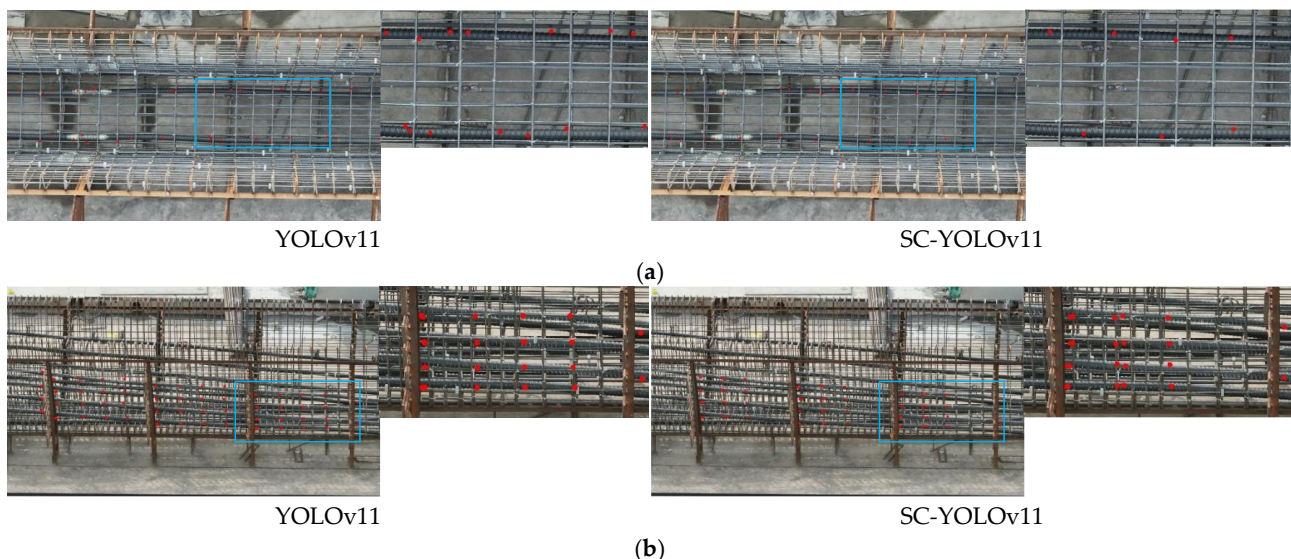


Figure 6. Comparison of Prediction Performance Between the Original and Improved Models: (a) Overhead. (b) Front view.

4.2. Performance Comparison with Different Detection Models

To further validate the detection performance of SC-YOLOv11, comparative experiments were conducted under identical experimental conditions and hyperparameter settings. All comparison models were retrained from scratch using exactly the same dataset and the same training, validation, and test split described in Section 3. The only model-specific modifications involved adjusting the number of output key points to match the corrugated duct key point configuration. Various representative models were selected for comparison, and the results are summarized in Table 3. In addition to classification-related metrics (*Precision*, *Recall*, mAP_{50} , mAP_{50-95}), we also report efficiency indicators (*FPS*, *Params*, *GFLOPs*) to provide a comprehensive assessment of practical deployability.

Table 3. Comparison of detection and efficiency metrics across different models.

Model	<i>Precision</i>	<i>Recall</i>	mAP_{50}	mAP_{50-95}	<i>FPS</i>	<i>Params</i>	<i>GFLOPs</i>
OpenPose [36]	70.5%	63.4%	57.4%	20.9%	16.2	96.4	68.5
PoseNet [37]	62.2%	58.3%	50.6%	16.7%	36.5	34.8	23.6

Lightweight OpenPose [38]	67.5%	60.6%	57.3%	19.5%	136.7	3.27	5.2
DarkPose [39]	69.3%	62.7%	58.4%	19.8%	34.3	28.1	43.4
ViTPose [24]	72.5%	67.2%	59.7%	23.1%	27.7	76.3	113.8
SC-YOLOv11	71.2%	66.5%	59.1%	22.3%	101.6	4.34	8.1

As shown in Table 3, the proposed SC-YOLOv11 outperforms traditional OpenPose-series models and PoseNet across all evaluation metrics. Compared with OpenPose, PoseNet, and Lightweight OpenPose, SC-YOLOv11 raises mAP_{50-95} from 20.9%, 16.7%, and 19.5% to 22.3%, corresponding to improvements of 1.4, 5.6, and 2.8 percentage points, respectively. These gains can be primarily attributed to YOLOv11's single-stage anchor-free instance detection framework, which fundamentally avoids the limb partitioning and erroneous association issues commonly encountered by OpenPose-series methods when using Part Affinity Fields (PAF) in densely occluded rebar scenes. It also overcomes the inherent instability of PoseNet's direct coordinate regression when dealing with targets exhibiting extreme aspect ratios. Furthermore, the incorporation of the SE channel attention mechanism and the SPPFCSPC multi-scale pooling module significantly enhances the model's ability to capture multi-scale features of slender corrugated ducts, thereby achieving superior instance integrity and robustness in complex industrial environments.

SC-YOLOv11 also demonstrates strong competitiveness against DarkPose, achieving slightly higher mAP_{50} (59.1%, compared with 58.4%) and mAP_{50-95} (22.3%, compared with 19.8%), while operating at nearly three times the inference speed (101.6 FPS for SC-YOLOv11 versus 34.3 FPS for DarkPose) with only one-sixth of the parameters (4.34 M against 28.1 M). This indicates that the proposed architectural improvements yield both accuracy gains and substantial efficiency benefits.

When compared with the ViTPose model, SC-YOLOv11 achieves very competitive accuracy: its mAP_{50} (59.1%) and mAP_{50-95} (22.3%) are only 0.6 and 0.8 percentage points lower than ViTPose's 59.7% and 23.1%, respectively. This marginal accuracy gap reflects ViTPose's advantage in global context modeling through its Vision Transformer backbone with multi-head self-attention, which slightly improves fine-grained key point localization for slender ducts. However, this higher precision comes at a considerable cost: ViTPose requires 76.3 M parameters and 113.8 GFLOPs, achieving only 27.7 FPS on the same hardware. By contrast, SC-YOLOv11 contains merely 4.34 M parameters and 8.1 GFLOPs, and runs at 101.6 FPS—over 3.7 times faster than ViTPose while maintaining nearly identical detection quality.

In summary, for the pose estimation task of slender objects such as corrugated duct key points, the proposed SC-YOLOv11 not only surpasses most existing models but also achieves a well-balanced trade-off between detection accuracy and inference speed. It exhibits clear advantages in instance integrity and real-time performance, particularly in complex industrial scenarios.

4.3. Evaluation of Corrugated Duct Fitting Performance

After performing pose estimation on the corrugated ducts in the images using the SC-YOLOv11 model and extracting their key points, the RANSAC algorithm was applied to fit the detected feature points. The resulting fitting curves are shown in Figure 7, where Figure 7a corresponds to the top-down view and Figure 7b to the front orthographic view. The results demonstrate that the fitted curves closely follow the actual 2D trajectories of the corrugated ducts in the rectified images.

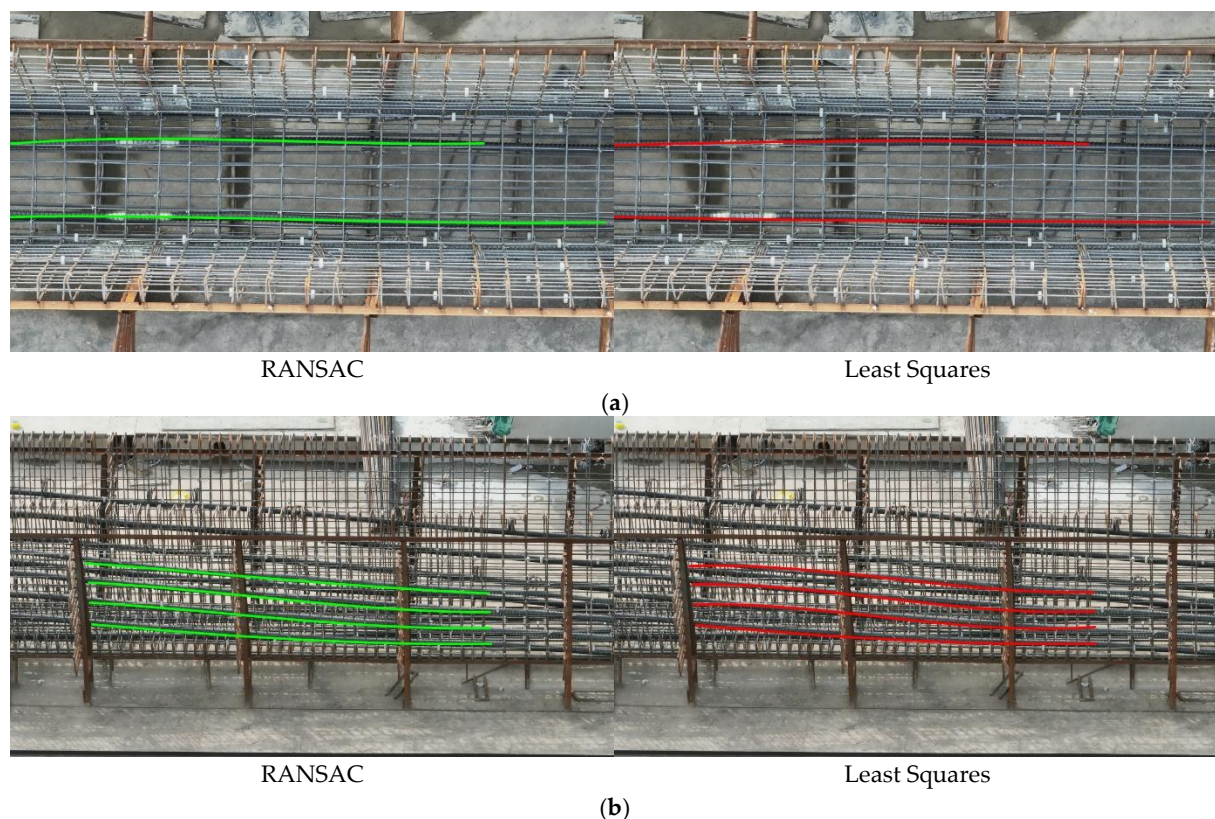


Figure 7. Visual comparison of curve fitting results: (a) Overhead. (b) Front view.

To further evaluate the fitting performance, the least squares method was also applied to the same set of corrugated duct key points for comparison with the RANSAC results. The coefficient of determination R^2 was used as the evaluation metric, and the comparison results are presented in Table 4. Since the RANSAC algorithm relies on random sampling, it may produce slightly different results in each run. Therefore, multiple fitting trials were conducted, and the average R^2 value was calculated to ensure stability and reliability of the evaluation.

Table 4. Comprehensive comparison of fitting performance across different methods.

		R^2	$RMSE$ (mm)	MAE (mm)
Overhead	LSM	94.21%	7.3	5.7
	RANSAC	96.83%	4.8	4.0
	Huber	95.29%	6.5	5.8
	Theil-Sen	89.46%	9.6	8.4
Front view	LSM	92.34%	6.9	5.6
	RANSAC	94.03%	5.3	4.2
	Huber	93.34%	7.8	6.1
	Theil-Sen	73.08%	17.8	14.8

As shown in Table 4, RANSAC yields the highest R^2 and lowest $RMSE/MAE$ among all methods. Huber still incorporates all points and remains biased by the gross outliers common under dense occlusion. Theil-Sen, designed for linear models, cannot suit the cubic polynomial required for duct trajectories (R^2 only 0.8946 and 0.7308). Ordinary least squares is, as expected, the most severely affected. RANSAC explicitly rejects such erroneous key points through its consensus mechanism, producing a clean inlier set. Given that outlier rates can exceed 30% in our construction scene, this hard rejection is critical for physically reliable trajectory reconstruction. This advantage is further amplified by the

complex construction environment: low feature contrast and severe occlusion make precisely identifying the duct axis difficult in both automated detection and manual annotation, which inevitably introduces outliers that other methods fail to fully suppress.

In scenarios containing noticeable outliers and annotation noise, the RANSAC algorithm effectively eliminates abnormal points through its random sampling consensus mechanism and estimates the model using only high-quality inlier sets. This approach significantly improves the accuracy and robustness of the fitted curves. In contrast, the traditional least squares method assigns equal weight to all data points and is highly susceptible to being skewed by outliers, causing the fitted curve to deviate from the true axis of the corrugated duct. Therefore, the strategy combining RANSAC with least squares demonstrates stronger adaptability and stability in real complex engineering environments, making it more suitable for corrugated duct pose estimation tasks.

4.4. Analysis of Prestressed Box Girder Stitching Results

In this study, to acquire high-quality image datasets of corrugated ducts, a UAV was employed to perform segmented aerial photography of the precast box girders. This approach effectively overcomes the limited field of view of a single shot, reduces lens distortion caused by long-distance imaging, and facilitates independent curve fitting of corrugated ducts on each segmented image by subsequent models. After completing curve fitting on all segmented images, the multiple local images are precisely stitched together in the order of capture to reconstruct the complete corrugated duct layout across the entire span of the box girder.

To eliminate perspective differences introduced by segmented shooting, this paper proposes a global perspective correction strategy based on a single reference image. First, the transformation matrices for stitching all segmented images are computed. Then, one segmented image is selected as the reference image. The four corner points of the rebar framework are detected using the model, and the perspective transformation matrix for this reference image is calculated. Subsequently, all segmented images are corrected using the same perspective transformation matrix, achieving batch perspective correction of all corrugated duct regions.

To provide objective and repeatable metrics beyond visual inspection, we further quantified the stitching and rectification quality using the following indicators: re-projection error, stitching error (SSIM), accumulated drift and measurement uncertainty. Table 5 summarizes the results for the overhead and front-view configurations.

Table 5. Quantitative evaluation of image stitching and perspective rectification.

	Re-Projection Error (px)	Stitching Error	Accumulated Drift (px)
Overhead	0.93	0.986	1.21
Front view	0.87	0.992	0.92

The re-projection error was computed as the root-mean-square (RMS) distance between the matched SIFT inliers after homography estimation. Values below 1 px confirm that the estimated geometric transformations are highly accurate. SSIM was evaluated inside the overlapping regions of stitched image pairs, and the results (all above 0.98) indicate that the structure and texture of the corrugated ducts are preserved without perceptible distortion. Accumulated drift was assessed by re-projecting the first frame of a multi-image sequence through the concatenated homographies and measuring the loop-closure error; the sub-2 px drift demonstrates the stability of the SIFT-based stitching chain. Finally, measurement uncertainty was derived by comparing the deviation values obtained from our rectified orthographic images with the independent on-site inspector records. The mean absolute error (MAE) between the two sets of measurements is less than 6 mm, well within the ± 10 mm tolerance specified in JTG/T 3650-2020 [40]. These quantitative

metrics collectively demonstrate that the proposed stitching and rectification pipeline not only produces visually coherent panoramas but also satisfies the geometric accuracy required for reliable quality assessment of prestressed duct layout.

The experimental results are shown in Figures 8 and 9. Figure 8 presents the direct stitching result without perspective correction. The segmented images are accurately aligned in sequence, and the fitted curves closely follow the actual laying paths. Figure 9 shows the corrected result after applying perspective transformation, where the fitted curves are transformed into a front orthographic view. The corrected complete centerlines can be further compared with the theoretical laying paths in the design drawings to quantify spatial deviations. This provides a reliable basis for evaluating the positional accuracy of prestressed corrugated ducts. The proposed method is simple to implement, highly robust, and demonstrates considerable engineering application value.



Figure 8. Image Stitching Result: (a) Overhead. (b) Front view.

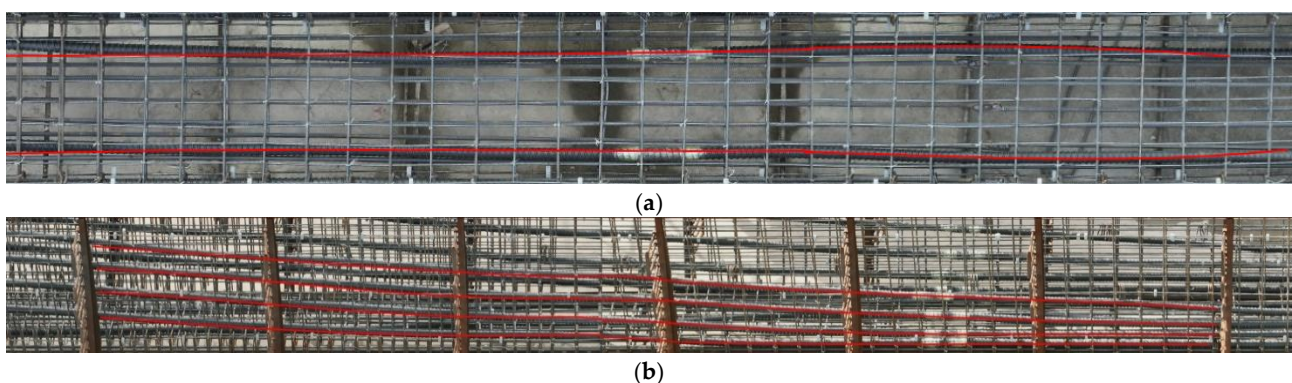


Figure 9. Perspective Correction Result: (a) Overhead. (b) Front view.

4.5. Comparison of Fitting Results with Design Drawings

After fitting the corrugated duct curves, the results were compared with the data from the design drawings. According to the “Technical Specification for Construction of Highway Bridges and Culverts” (JTG/T 3650-2020), the allowable positional deviation for prestressed ducts is ± 10 mm. Therefore, if the vertical deviation between the fitted curve and the design curve exceeds this threshold, the location is flagged and reported to the supervisor. In this study, the position of each captured image relative to the box girder was determined according to the UAV’s shooting distance, allowing the corresponding corrugated ducts to be compared with the design drawings. As noted in Section 3, the lowest of the five physical ducts in the side-view cross-section is totally blocked by reinforcement and is not visible in the images; thus, only the four annotated ducts are analysed. All four fitted curves are shown in Figure 10b.

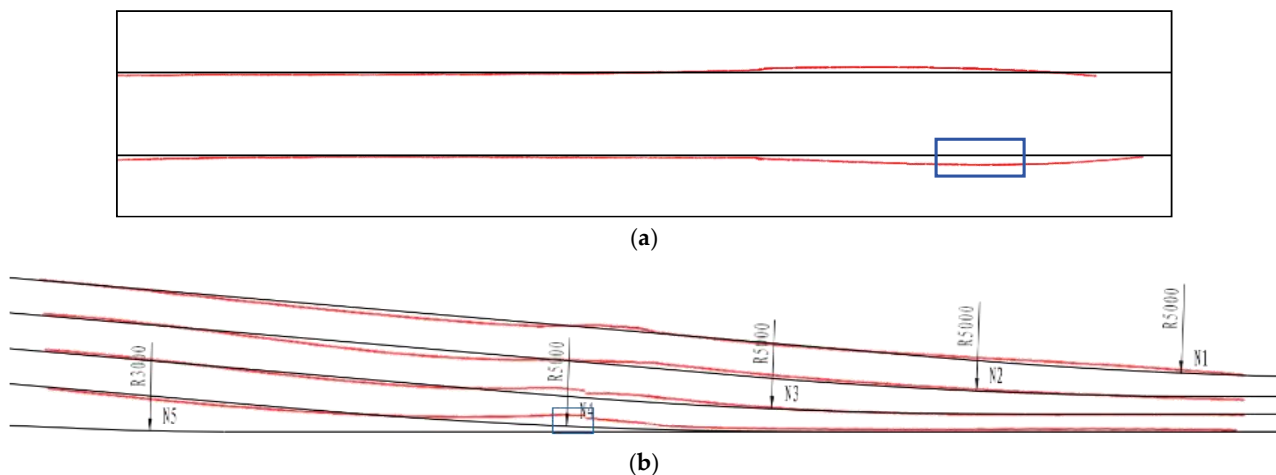


Figure 10. Drawing Comparison Effect Diagram: (a) Overhead. (b) Front view.

As shown in Figure 10, significant deviations between the fitted curves and the design curves are observed within the blue rectangular regions in both figures. The red curves represent the fitted laying paths of the corrugated ducts, while the black curves indicate the designed paths. In Figure 10a, a notable deviation exists within the blue-boxed region, with a maximum vertical deviation of 11.53 cm. Similarly, in Figure 10b, a significant deviation is also present in the blue-boxed region, with a maximum vertical height difference of 7.62 cm. These results demonstrate that the proposed method can effectively detect the laying quality of corrugated ducts. It provides clear and intuitive feedback on potential laying quality issues during actual construction and enables supervisors to identify problems more directly. Cross-validation with on-site inspector records confirmed that the measured deviations agree with actual laying errors. It should be noted that the reported deviations are two-dimensional distances measured in the rectified image plane, which reasonably represent the actual laying errors given that the design drawings also adopt planar projections.

To quantitatively evaluate the method, independent on-site measurements were performed by an inspector using a calibrated steel tape. Given the confined workspace and dense reinforcement inside the box girder, the steel tape provided a practical and efficient means of direct measurement, with sufficient accuracy for centimeter-level deviation assessment. The reference coordinate system was defined with the origin at the lower-left corner of the box girder, the X-axis along the span direction, and the Y-axis vertically upward. Across all matched measurement locations, over 92% of the points agree with the inspector’s records within ± 10 mm.

Due to the large volume of data, the duct curves in Figure 9 are taken as a detailed illustrative example. All 42 sampling points on the corrected duct curves obtained from the two views in Figure 9 were individually compared with the inspector's steel-tape readings. The results are presented in Table 6. For each sampling point, the listed vertical distance was measured from the bottom reinforcement bar visible in the image upward to the duct centerline, in accordance with the on-site steel-tape measurement procedure. The corrugated ducts are numbered from bottom to top: Duct 1 is the lowest visible duct, Duct 2 the next, and so on. Notably, the two locations with the largest inspector-recorded deviations—11.3 cm and 7.4 cm—did not fall exactly on our sampling points. By linearly interpolating the deviations of the adjacent sampling points on either side, we obtained image-based estimates at these locations that agree closely with the inspector's records, thereby completing the cross-validation. Meanwhile, the maximum deviations directly measured on the sampling points by our method were 11.53 cm and 7.62 cm (see Table 6), highlighting the method's capability to capture substantial deviations.

Table 6. Comparison of deviation measurements (cm).

Point ID	View	Steel Tape	Proposed	Difference	Point ID	View	Steel Tape	Proposed	Difference
O1	Overhead	24.4	24.0	0.4	O2	Overhead	70.2	68.6	1.6
O1	Overhead	24.1	24.2	−0.1	O2	Overhead	67.5	68.0	−0.5
O1	Overhead	22.2	23.0	−0.9	O2	Overhead	67.0	67.8	−0.8
O1	Overhead	21.7	22.8	−1.0	O2	Overhead	69.5	70.2	−0.7
O1	Overhead	23.0	22.8	0.2	O2	Overhead	70.8	71.4	−0.6
O1	Overhead	19.1	18.9	0.2	O2	Overhead	71.5	72.2	−0.7
O1	Overhead	19.5	19.4	0.1	O2	Overhead	69.5	69.9	−0.4
F1	Front	43.4	41.8	1.6	F2	Front	69.1	68.4	0.7
F1	Front	31.3	30.6	0.7	F2	Front	56.6	55.9	0.7
F1	Front	19.9	19.5	0.4	F2	Front	43.7	42.9	0.8
F1	Front	15.0	14.2	0.8	F2	Front	33.6	34.3	−0.7
F1	Front	10.9	10.2	0.7	F2	Front	23.8	23.1	0.7
F1	Front	10.7	10.2	0.5	F2	Front	22.9	21.8	1.1
F1	Front	10.5	10.2	0.3	F2	Front	22.5	21.6	0.9
F3	Front	94.3	93.5	0.8	F4	Front	119.3	118.2	1.1
F3	Front	81.4	80.9	0.5	F4	Front	105.7	105.4	0.3
F3	Front	68.2	67.1	1.1	F4	Front	92.8	93.1	−0.3
F3	Front	57.8	58.3	−0.5	F4	Front	80.1	80.8	−0.7
F3	Front	42.1	42.8	−0.7	F4	Front	65.7	65.6	0.1
F3	Front	36.3	35.9	0.4	F4	Front	56.5	56.9	−0.4
F3	Front	33.4	32.9	0.5	F4	Front	48.8	49.0	−0.2

5. Conclusions and Future Work

This paper presents an automated detection and quality assessment method for pre-stressed corrugated ducts in precast box girders, based on the SC-YOLOv11 model integrated with robust curve fitting and multi-image stitching. The proposed approach effectively improves both the efficiency and accuracy of corrugated duct laying quality inspection during the prefabrication process. The method first employs the SC-YOLOv11 model to estimate the pose of corrugated ducts and extract their key points. RANSAC algorithm is then applied to fit curves using the detected key points, thereby determining the precise 2D positions of the ducts. Finally, all fitted curves from segmented images are stitched together based on the image stitching matrix, and perspective transformation is performed to correct the stitched results into a front orthographic view of the entire precast box girder.

- (1) To address the challenges posed by dense reinforcement and severe occlusion of slender corrugated ducts inside precast box girders, this study incorporates the SE channel attention mechanism into the YOLOv11 backbone and upgrades the original SPPF module to the more effective SPPFCSPC module, resulting in the proposed SC-YOLOv11 detection model. Ablation experiments demonstrate that the synergistic effect of these two improvements increases *Precision* by 1.7%, *Recall* by 1.3%, *mAP₅₀* by 2.6%, and *mAP_{50–95}* by 1.2%. These enhancements significantly improve the model's robustness and localization accuracy for detecting slender targets in complex and heavily occluded environments.
- (2) The proposed method accurately locates the vertices of the box girder framework and the key points of corrugated ducts. By combining SC-YOLOv11 predictions with RANSAC-based curve fitting, it enables precise reconstruction of the local 2D geometric morphology of the ducts. Through perspective transformation and the SIFT algorithm, the actual laying positions of the corrugated ducts across the entire girder can be rapidly determined, achieving fully automated quality assessment during the construction phase. Compared with traditional manual measurement, the method substantially enhances detection efficiency and traceability. It can automatically identify potential quality issues, reduce waiting time between construction procedures, and lower the risk of later reinforcement or structural scrapping, demonstrating considerable engineering application value and broad promotion potential.
- (3) Due to the complex internal environment of precast box girders, where corrugated ducts often exhibit low saliency and are frequently occluded by dense reinforcement, the SC-YOLOv11 model still shows certain limitations in pose recognition, particularly in key point localization accuracy. Future work may focus on further optimizing the model architecture or adopting more advanced detection frameworks to improve recognition performance in highly challenging scenes.

In summary, the method proposed in this paper realizes an end-to-end automated analysis pipeline for corrugated duct laying quality—from key point detection and robust curve fitting to image stitching and deviation quantification. It provides an efficient, practical, and intelligent technical solution for real-time monitoring, early warning, and dynamic quality control of prestressed corrugated ducts during the construction stage, significantly improving the efficiency and reliability compared with conventional manual inspection methods.

Author Contributions: Conceptualization, D.-P.C. and D.L.; data curation, Y.C. and H.-B.H.; software, S.-H.Z.; validation, S.-H.Z.; visualization, S.-H.Z.; writing—original Draft, Y.C. and D.L.; writing—review and editing, H.-B.H. and D.L.; supervision, H.-B.H.; funding acquisition, D.-P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 52478310).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Authors Dongpo Chen and Yuan Cheng were employed by the ShiAn Branch of Hebei Expressway Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zucca, M.; Stochino, F.; Vecchi, E.; Crespi, P. On the seismic performance of existing RC motorway viaducts subject to corrosion effects. In *Capacity Assessment of Corroded Reinforced Concrete Structures: From Research to Daily Engineering Evaluation*; International Federation for Structural Concrete: Lausanne, Switzerland, 2025; pp. 282–289.
2. Brighenti, F.; Caspani, V.F.; Costa, G.; Giordano, P.F.; Limongelli, M.P.; Zonta, D. Bridge management systems: A review on current practice in a digitizing world. *Eng. Struct.* **2024**, *321*, 21. <https://doi.org/10.1016/j.engstruct.2024.118971>.
3. Mendoza-Lugo, M.A.; Nogal, M.; Morales-Nápoles, O. Network-level optimisation approach for bridge interventions scheduling. *Struct. Infrastruct. Eng.* **2024**, *early access*. <https://doi.org/10.1080/15732479.2024.2403568>.
4. Rafieizonooz, M.; Jang, H.; Kim, J.; Kim, C.S.; Kim, T.; Wi, S.; Banihashemi, S.; Khankhaje, E. Performances and properties of steel and composite prestressed tendons—A review. *Heliyon* **2024**, *10*, 26. <https://doi.org/10.1016/j.heliyon.2024.e31720>.
5. Xu, Y.; Mao, M.; Huang, X.-H.; Guo, Y.-Q.; Yang, Y.; Dong, Y.-R.; Shah, S.A.A.; Gao, X.; Usman, M.; Jeong, S.-H. Comparative investigation of dynamic damping and fracture in viscoelastic dampers with two different interfacial bonding types: Molecular dynamics simulations and device tests. *J. Build. Eng.* **2025**, *116*, 114582. <https://doi.org/10.1016/j.jobbe.2025.114582>.
6. Muldoon, R.; Chalker, A.; Forde, M.C.; Ohtsu, M.; Kunisue, F. Identifying voids in plastic ducts in post-tensioning prestressed concrete members by resonant frequency of impact-echo, SIBIE and tomography. *Constr. Build. Mater.* **2007**, *21*, 527–537. <https://doi.org/10.1016/j.conbuildmat.2006.04.009>.
7. Gong, S.L.; Sun, F.T.; Chen, K.; Feng, X. Finite-Element Performance Degradation Behavior of a Suspension Prestressed Concrete Arch Bridge with Grouting Defects. *Buildings* **2024**, *14*, 19. <https://doi.org/10.3390/buildings14020399>.
8. Oh, B.D.; Choi, H.; Chin, W.J.; Park, C.Y.; Kim, Y.S. Void Detection inside Duct of Prestressed Concrete Bridges Based on Deep Support Vector Data Description. *Appl. Sci.* **2023**, *13*, 13. <https://doi.org/10.3390/app13105981>.
9. Owerko, P.; Honkisz, M. Innovative technique for identification of prestressing tendons layout in post-tensioned bridges using a probe with MEMS accelerometer. *Struct. Infrastruct. Eng.* **2017**, *13*, 869–881. <https://doi.org/10.1080/15732479.2016.1212905>.
10. Sun, X.T.; Li, D.; He, W.Y.; Wang, Z.C.; Ren, W.X. Grouting Quality Evaluation in Post-Tensioning Tendon Ducts Using Wavelet Packet Transform and Bayes Classifier. *Sensors* **2019**, *19*, 17. <https://doi.org/10.3390/s19245372>.
11. Hurlebaus, S.; Hueste, M.; Karthik, M.; Terzioğlu, T. *Condition Assessment of Bridge Post-Tensioning and Stay Cable Systems Using NDE Methods*; Texas A&M Transportation Institute: College Station, TX, USA, 2016.
12. Fang, W.L.; Love, P.E.D.; Luo, H.B.; Ding, L.Y. Computer vision for behaviour-based safety in construction: A review and future directions. *Adv. Eng. Inform.* **2020**, *43*, 13. <https://doi.org/10.1016/j.aei.2019.100980>.
13. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 26. <https://doi.org/10.1016/j.autcon.2022.104440>.
14. Kim, H.S.; Seong, J.; Jung, H.J. Real-Time Struck-By Hazards Detection System for Small- and Medium-Sized Construction Sites Based on Computer Vision Using Far-Field Surveillance Videos. *J. Comput. Civ. Eng.* **2023**, *37*, 14. <https://doi.org/10.1061/jccee5.Cpeng-5238>.
15. Yong, Y.P.; Lee, S.J.; Chang, Y.H.; Lee, K.H.; Kwon, S.W.; Cho, C.S.; Chung, S.W. Object Detection and Distance Measurement Algorithm for Collision Avoidance of Precast Concrete Installation during Crane Lifting Process. *Buildings* **2023**, *13*, 24. <https://doi.org/10.3390/buildings13102551>.
16. Fang, W.L.; Zhong, B.T.; Zhao, N.; Love, P.E.; Luo, H.B.; Xue, J.Y.; Xu, S.J. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. <https://doi.org/10.1016/j.aei.2018.12.005>.
17. Fang, W.L.; Ding, L.Y.; Luo, H.B.; Love, P.E.D. Falls from heights: A computer vision-based approach for safety harness detection. *Autom. Constr.* **2018**, *91*, 53–61. <https://doi.org/10.1016/j.autcon.2018.02.018>.
18. Wei, R.; Love, P.E.D.; Fang, W.L.; Luo, H.B.; Xu, S.J. Recognizing people's identity in construction sites with computer vision: A spatial and temporal attention pooling network. *Adv. Eng. Inform.* **2019**, *42*, 9. <https://doi.org/10.1016/j.aei.2019.100981>.
19. Son, H.; Choi, H.; Seong, H.; Kim, C. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Autom. Constr.* **2019**, *99*, 27–38. <https://doi.org/10.1016/j.autcon.2018.11.033>.
20. Fang, Q.; Li, H.; Luo, X.C.; Ding, L.Y.; Luo, H.B.; Rose, T.M.; An, W.P. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom. Constr.* **2018**, *85*, 1–9. <https://doi.org/10.1016/j.autcon.2017.09.018>.
21. Fang, Q.; Li, H.; Luo, X.C.; Ding, L.Y.; Luo, H.B.; Li, C.Q. Computer vision aided inspection on falling prevention measures for steepjacks in an aerial environment. *Autom. Constr.* **2018**, *93*, 148–164. <https://doi.org/10.1016/j.autcon.2018.05.022>.

22. Li, J.Q.; Zhou, G.Y.; Li, D.F.; Zhang, M.Y.; Zhao, X.F. Recognizing workers' construction activities on a reinforcement processing area through the position relationship of objects detected by faster R-CNN. *Eng. Constr. Archit. Manag.* **2023**, *30*, 1657–1678. <https://doi.org/10.1108/ecam-04-2021-0312>.
23. Feng, D.M.; Feng, M.Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review. *Eng. Struct.* **2018**, *156*, 105–117. <https://doi.org/10.1016/j.engstruct.2017.11.018>.
24. Xu, Y.; Zhang, J.; Zhang, Q.; Tao, D. ViTPose: Simple vision transformer baselines for human pose estimation. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*; NeurIPS: San Diego, CA, USA, 2022; Volume 35, pp. 38571–38584.
25. Li, Y.; Zhang, S.; Wang, Z.; Yang, S.; Yang, W.; Xia, S.-T.; Zhou, E. TokenPose: Learning keypoint tokens for human pose estimation. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021; Volume 19, pp. 1–15. <https://doi.org/10.48550/arXiv.2104.03516>.
26. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2019; pp. 1–10. <https://doi.org/10.48550/arXiv.1902.09212>.
27. Jiang, T.; Lu, P.; Zhang, L.; Ma, N.; Han, R.; Lyu, C.; Li, Y.; Chen, K. RTMPose: Real-time multi-person pose estimation based on MMPose. *arXiv* **2023**, arXiv:2303.07399. <https://doi.org/10.48550/arXiv.2303.07399>.
28. Zheng, T.; Huang, Y.; Liu, Y.; Tang, W.; Yang, Z.; Cai, D.; He, X. CLNet: Cross layer refinement network for lane detection. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 1–10. <https://doi.org/10.48550/arXiv.2203.10350>.
29. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs beat YOLOs on real-time object detection. *arXiv* **2023**, arXiv:2304.08069. <https://doi.org/10.48550/arXiv.2304.08069>.
30. Sevgen, S.C.; Karsli, F. An Improved RANSAC Algorithm for Extracting Roof Planes from Airborne Lidar Data. *Photogramm. Rec.* **2020**, *35*, 40–57. <https://doi.org/10.1111/phor.12296>.
31. Yang, H.C.; Zhang, S.B.; Wang, Y.B. Robust and Precise Registration of Oblique Images Based on Scale-Invariant Feature Transformation Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 783–787. <https://doi.org/10.1109/lgrs.2011.2181485>.
32. Zou, C.; Yu, S.Q.; Yu, Y.K.; Gu, H.T.; Xu, X.L. Side-Scan Sonar Small Objects Detection Based on Improved YOLOv11. *J. Mar. Sci. Eng.* **2025**, *13*, 25. <https://doi.org/10.3390/jmse13010162>.
33. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. <https://doi.org/10.1109/tpami.2019.2913372>.
34. Gao, X.; Cao, C.; Yi, X. Using the improved YOLOv11 model to enhance computer vision applications for building crack detection algorithms. *Sci. Rep.* **2025**, *15*, 38843. <https://doi.org/10.1038/s41598-025-22160-6>.
35. Dembele, D. A Power Method for Computing Singular Value Decomposition. *arXiv* **2024**, arXiv:2410.23999.
36. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. <https://doi.org/10.1109/tpami.2019.2929257>.
37. Papandreou, G.; Zhu, T.; Kanazawa, N.; Toshev, A.; Tompson, J.; Bregler, C.; Murphy, K. Towards Accurate Multi-person Pose Estimation in the Wild. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2017; pp. 3711–3719. <https://doi.org/10.1109/CVPR.2017.395>.
38. Zhang, X.R.; Xie, Q.J.; Sun, W.; Ren, Y.J.; Mukherjee, M. Dense Spatial-Temporal Graph Convolutional Network Based on Lightweight OpenPose for Detecting Falls. *Comput. Mater. Contin.* **2023**, *77*, 47–61. <https://doi.org/10.32604/cmc.2023.042561>.
39. Zhang, F.; Zhu, X.; Dai, H.; Ye, M.; Zhu, C. Distribution-Aware Coordinate Representation for Human Pose Estimation. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2020; pp. 7093–7102. <https://doi.org/10.1109/CVPR42600.2020.00712>.
40. JT/T 3650-2020; Technical Specification for Construction of Highway Bridges and Culverts. China Communications Press: Beijing, China, 2020.

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