Cite this article
Han SJ, Han SH, Choo YW, Kim J and Yoon S (2026)
Machine learning analyses for dynamic images recorded from physical model tests.
International Journal of Physical Modelling in Geotechnics **26(4)**: 225–238,
https://doi.org/10.1680/jphmg.25.00007

Research Article
Paper 2500007
Received 31/01/2025; Accepted 22/04/2026

International Journal of Physical Modelling in Geotechnics

Machine learning analyses for dynamic images recorded from physical model tests

Seung Jae Han
Department of Civil and Environmental Engineering, Kongju National University, Cheonan-si, Republic of Korea

Se Hee Han
Department of Civil and Environmental Engineering, Kongju National University, Cheonan-si, Republic of Korea

Yun Wook Choo
Department of Civil and Environmental Engineering, Kongju National University, Cheonan-si, Republic of Korea (corresponding author: ywchoo@kongju.ac.kr)

Jungeun Kim
Department of Computer Science and Engineering, Inha University, Incheon, Republic of Korea

Susik Yoon
Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea

This study aims to measure vibrational displacement of structures in physical model tests using high-speed recorded images and an open-source machine learning–based model, YOLOv5. Two dynamic experiments were performed, and their images were recorded: a vibrated block on a shaker and a single-degree-of-freedom structure on dynamic centrifuge tests. For the shaker experiment images, four parameters were examined: different training methods, pre-trained models, partial area patterns, and pattern types. The patterns include black-and-white checks, circles, squares, crosses, and X shapes. For the different training methods, training with one labelled image and its copies showed better performance than training with all images. The pre-trained model analysed other videos small errors although the errors increased with greater camera-to-structure distances. Among the types of patterns, the X pattern performed the best, with ∼2% errors and an average coefficient of determination of 0.9928. The images from the dynamic centrifuge test were analysed by the machine learning code and compared with results from two other popular object-tracking software programs. The tracking results from the machine learning model showed performance comparable to that of the other image-based tracking programs. The results suggest that YOLO-based image tracking can be effectively applied to laboratory vibration tests.

Keywords: centrifugal experiments/dynamic displacement/dynamics/image analysis/machine learning/model tests/shaker/testing and evaluation/vibration

1. Introduction

Traditionally, the displacement of vibrating structures is measured by contact or point sensors, such as accelerometers and displacement transducers, which are attached and point to a single spot on the structure for measurement. These methods involve perturbation caused by sensor attachment to the target object and limitations in the sensing area on the object. Recently, image analysis techniques using cameras have been utilised to address these limitations.

Image analysis techniques have been effectively applied in various cases of geotechnical engineering. Wahyudi *et al*. (2012) studied the effects of initial relative density on the shear band of sand specimens in a hollow cylinder torsional shear apparatus using image analysis, where black dots were pasted on the face of the outer membrane covering the specimen and analysed with software (Move-tr/2D Ver. 7.21) to capture the local deformation of specimens. Kaddhour *et al*. (2013) used X-ray micro-tomography to analyse the water retention behaviour and localised deformation in sand, using a 3SR X-ray scanner applied to a triaxial compression apparatus to measure porosity and the degree of saturation, as well as particle displacements and rotations. Altuhafi *et al*. (2013) introduced a pragmatic approach for quantitative shape analysis, measuring convexity, sphericity, and aspect ratio. The soil sample in a hopper with controlled vibrations generated a relatively steady flow of particles, while a camera recorded a sequence of binary images (frames) of the particles using the QICPIC apparatus, but these were analysed using measures implemented in imaging software like Axiovision, ImageJ, and MATLAB.

On the other hand, particle image velocimetry (PIV) and digital image correlation (DIC) techniques have been widely adopted in laboratory and model tests in geotechnical engineering to observe the deformation and displacement of soil and structures. White (2002) and White *et al*. (2003) introduced PIV in the form of the GeoPIV package to trace the movement of soil patches within each image. Costa and Kodikara (2012) used a double-ring test to evaluate the fracture behaviour of shrinking specimens and the crack front during desiccation using a digital camera and the

GeoPIV to measure deformation and strain within the soil medium. Choo *et al*. (2013) also utilised the GeoPIV package to analyse the deformation and failure zones of clay media created by the penetration of mandrels. Arshad *et al*. (2014) analysed the cone penetration process in silica sand with uniform density using DIC techniques. A half-circular steel chamber with a reinforced poly sheet was used to capture images during cone penetration tests. The images were processed using VIC-2D software to compute soil displacement at shallow and deep penetration. Yuan *et al*. (2017) performed a scaled model test to investigate the deflection of a laterally loaded pile and soil deformation in loose sand, using images captured by two digital cameras, and the displacement fields were calculated using the PIVview2CDemo software. Chavda *et al*. (2020) and Park *et al*. (2024) utilised GeoPIV-RG to evaluate failure zones in sand developed by ring-shaped footings and the penetration of projected piles. While PIV and DIC provide high-precision, sub-pixel measurements of full-field deformation, they generally require controlled lighting, high-quality imaging systems, and are computationally expensive. Moreover, these techniques are more suitable for tracking continuous deformation fields than rigid-body motion. In contrast, machine learning–based marker tracking methods, such as the You Only Look Once (YOLO) model, offer faster processing and less hardware requirements for multiple objects, making them a promising alternative for rigid-body vibration measurement. This study aims to explore a new possibility of machine learning–based image processing techniques for measuring vibration displacement. An open-source model, YOLOv5, was utilised to establish a procedure to analyse the displacement of vibrating objects captured by relatively low-cost portable cell phones and a high-speed camera in two experiments: a vibrating block on a shaker and a dynamic centrifuge test.

2. Methodology

2.1 Machine learning algorithm for image processing

In this study, we analysed dynamically recorded images from the shaker and centrifuge experiments using an object detection model called YOLO. YOLO is designed to enable real-time object detection with a single-stage detection approach, where the detection and classification of objects in an image occur simultaneously (Redmon, 2016). Figure 1 presents the overall architecture of the original YOLOv5 network used in this study (Jocher, 2020). The model consists of three main components: a backbone, a neck, and detection heads. In the backbone, convolutional layers are used to gradually reduce the spatial size of the input image while increasing the number of feature maps. These feature maps contain important visual information such as edges, textures, and object shapes. A Spatial Pyramid Pooling Fast layer is placed at the end of the backbone to enlarge the receptive field and improve the representation of objects at different scales. The neck combines feature maps from different layers through up-sampling and concatenation operations. This process allows information from both high-resolution and low-resolution feature maps to be integrated, enabling the detection of objects at different scales. In the final stage, three detection heads generate predictions at different scales. Each detection head outputs bounding-box co-ordinates, objectness scores, and class probabilities using pre-defined anchor boxes. This multi-scale, anchor-based prediction enables the network to efficiently detect small, medium, and large objects within a single forward pass. In this study, the original YOLOv5 architecture was adopted without structural modifications.

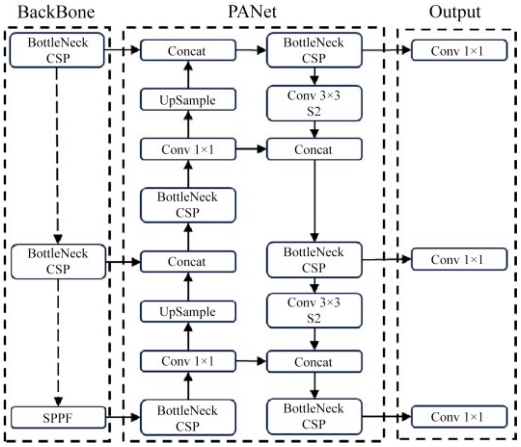

Figure 1. Architecture of YOLOv5 (modified from Jocher, 2020)

In this study, the procedure outlined in Figure 2(a) was followed to utilise YOLO in measuring vibration displacement of structures in physical model tests. The flowchart consists of two major stages: (*a*) the training process and (*b*) the object detection and tracking process. In the training process, the image dataset was prepared initially by collecting experimental images and manually annotating the ground-truth bounding boxes for each marker. The annotated dataset was then divided into the training and validation datasets, after which the YOLO model was trained by iteratively updating the network parameters to minimise the loss function. At this stage, the loss function represents the discrepancy between the predicted bounding boxes and the ground-truth bounding boxes. Once the model demonstrates stable performance on the validation set, it is finalised for use in the inference stage. YOLO predicts multiple bounding boxes for objects in an image, after being trained on hand-labelled ground truth bounding boxes corresponding to those objects. A single convolutional network in YOLO extracts object features from the image, and the fully connected layers predict output probabilities and co-ordinates of bounding boxes. YOLO divides an input image into a fixed number of grids (see S × S grids in Figure 2(b)). Each grid cell is responsible for detecting an object if the centre of the object falls within that cell. YOLO calculates confidence scores for each grid

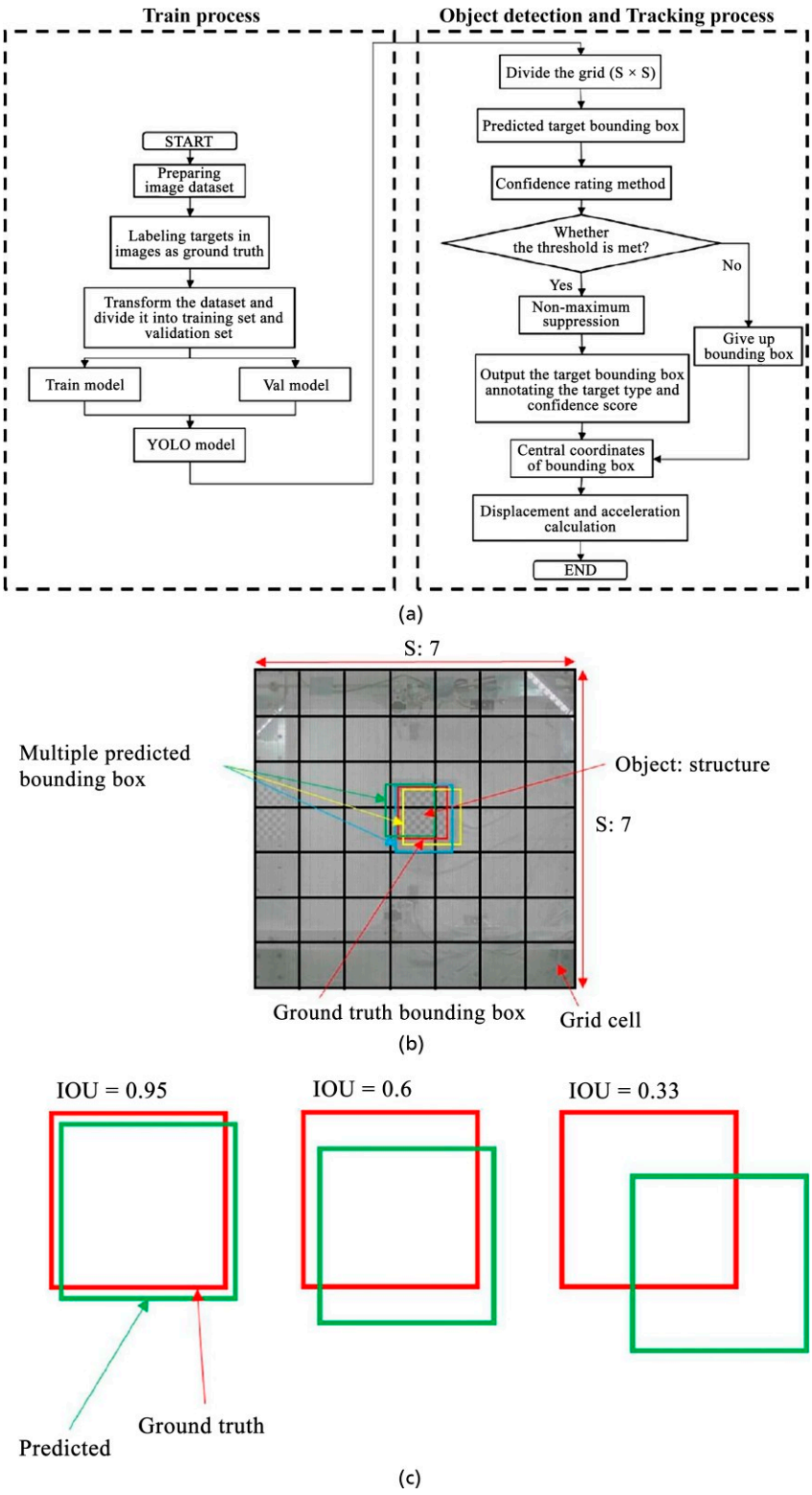

Figure 2. YOLO algorithm: (a) YOLO flow chart; (b) bounding boxes for object detection; and (c) IOU

cell for multiple predicted bounding boxes (see Figure 2(b)). These confidence scores are defined as the probability of whether the corresponding bounding box actually contains the object and how accurate the box is. If no object exists in that cell, the confidence scores should be zero. Otherwise, the confidence score is calculated using the intersection over union (IOU) between the predicted bounding box and the ground truth bounding box. The IOU ranges from 0 to 1, with values closer to 1 showing a higher match with the ground truth bounding box (see Figure 2(c)). Only one set of class probabilities is predicted per grid cell, regardless of the multiple predicted bounding boxes. During training, only one bounding box from a grid cell is assigned the responsibility of predicting an object, based on the highest current IOU, which corresponds to the Non-Maximum Suppression (NMS) logic that is used to eliminate redundant, overlapping detections. After NMS selects a single final bounding box, the centre co-ordinates of the predicted bounding box are tracked on a frame-by-frame basis and are used to compute the structural displacement and acceleration.

2.2 Laboratory experiment with shaker

To assess the feasibility of the YOLO image analysis, a laboratory setup with a shaker system (K2007E01, The Modal Shop) was constructed as presented in Figure 3. A plastic block with a patterned surface was mounted on the shaker. An accelerometer (353B16, PCB Piezotronics) was attached to the top of the block, which has a sensitivity of 9.85 mV/g and a measurement range on the order of ± 500 g. A laser displacement transducer (IL-S025, Keyence) was installed above the block and positioned next to the accelerometer, featuring a measurement range of 20–30 mm, a reference distance of 25 mm, a linearity of ± 0.1% F.S., and a minimum sampling period of 0.33 ms. It was intended for both sensors to trace the motion of the block. A smartphone camera (Galaxy S21+) was used to capture the motion of the pattern at 960 images per second, with a recording resolution of 1280 × 720 pixels.

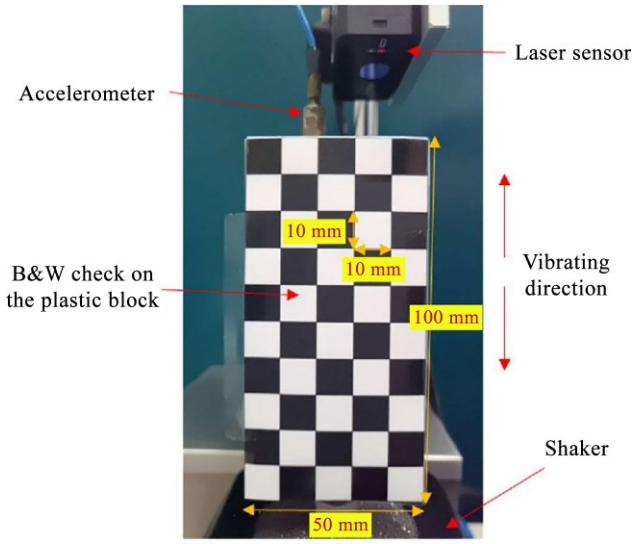

Figure 3. Vibrating block on shaker photographed at a distance of 100 mm (1280 × 720 pixels)

The shaker was firmly installed on the ground, and a sinusoidal signal with an amplitude of 300 mV and a frequency of 30 Hz was fed to the shaker using a signal generator (33120a, Agilent). The acceleration and displacement signals of the block were recorded using an oscilloscope (Infini Vision DSO-X 2004A, Keysight). A single-frequency input at 30 Hz was selected after testing 10–50 Hz to compromise different dynamic characteristics between the laser displacement sensor and accelerometer.

All experiments were recorded three times for each of the five patterns with a constant camera-to-object distance of 100 mm: black-and-white check (B&W, hereafter), circle, square, cross, and X. Their identifications (ID, hereafter) are BW1, O, S, C, and X (see Table 1). The B&W pattern was additionally tested at camera-to-object distances of 150 mm and 200 mm, denoted as BW2 and BW3, to enhance video resolution and observe the effect of different recording setups. The unit size of the five patterns is 10 mm by 10 mm, and the total size of the B&W check is 50 mm by 100 mm (see Table 1 and Figure 4).

In Table 1, the Test ID and Video ID were generally assigned based on the first letter of each pattern name. However, for the circle pattern, the ID ‘O’ was used instead of ‘C’ to avoid confusion with the cross pattern.

Table 2 tabulates the information of video images trained and analysed in this study. To evaluate the effect of labelling methods, four different methods are established: M1–M4, where the prefix ‘M’ represents the initial letter of the word ‘Method’. M1 uses all 540 images recorded from BW1-1 as ground truth, while M2 uses only the first image and its 539 copies as ground truth. Figure 4 presents an example of labelling for Case2 with M2. An additional labelling method, M3, is defined, using a partial area of the patterns to evaluate potential improvements with smaller-sized labelling, as shown in Figure 5. Two sizes are selected in Figure 5. One is a ground truth bounding box of size 10 × 10 mm, whose centre is located at the intersection of the B&W check in Figure 5(a), and its corresponding analysis is Case3. The other is a larger area encompassing one black square and one white square, as shown in Figure 5(b), with its corresponding analysis labelled as Case4. M4 involves labelling and tracking the entire pattern by labelling the first image and creating 999 copies, representing an increase in the number of images compared with M2. To check the applicability of a pre-trained model for tracking other separately recorded videos, BW1-1 to BW1-3 are analysed by the pre-trained model trained with BW1-1 using M4 as Case5 to Case7, respectively

Table 1. Testing conditions

		Testing conditions				
Test ID	Video ID	Pattern	Vibration frequency: Hz	Video recording speed: image/s	Distance between camera and target: mm	Pattern size: mm
BW1	BW1-1	Black-and-white check	30	960	100	50 × 100
BW1	BW1-2	Black-and-white check	30	960	100	50 × 100
BW1	BW1-3	Black-and-white check	30	960	100	50 × 100
BW2	BW2	Black-and-white check	30	960	150	50 × 100
BW3	BW3	Black-and-white check	30	960	200	50 × 100
O	O1	Circle	30	960	100	10 × 10
O	O2	Circle	30	960	100	10 × 10
O	O3	Circle	30	960	100	10 × 10
S	S1	Square	30	960	100	10 × 10
S	S2	Square	30	960	100	10 × 10
S	S3	Square	30	960	100	10 × 10
C	C1	Cross	30	960	100	10 × 10
C	C2	Cross	30	960	100	10 × 10
C	C3	Cross	30	960	100	10 × 10
X	X1	X	30	960	100	10 × 10
X	X2	X	30	960	100	10 × 10
X	X3	X	30	960	100	10 × 10

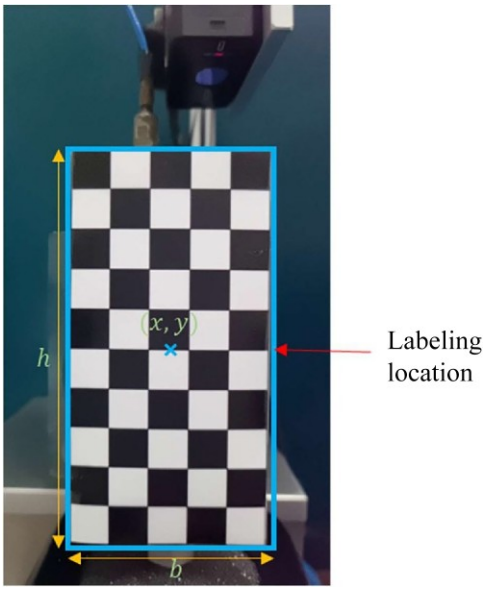

Figure 4. One example image for Case2 with M2 labelling and its position

(see Table 2). The other four patterns (O, S, C, and X) were examined in the same way, resulting in Case12 to Case23 (see Figure 6). BW2 and BW3 were analysed using the pre-trained model with BW1-1 as Case8 and Case9. In addition, Case10 (the result of BW2 analysed using the trained model with BW2 for both training and tracking) is compared to Case8, and Case11 (the result of BW3 analysed using the trained model with BW3 for both training and tracking) is compared to Case9.

2.3 Dynamic centrifuge model tests

The YOLO image analysis was additionally applied to high-speed camera video recorded from dynamic centrifuge tests performed with an inflight earthquake simulator at 60 g centrifugal acceleration. The target object for the image analysis was a single-degree-of-freedom (SDoF) structure supported by an underground structure buried in a soil model made of dense dry silica sand with a relative density of 80% (see Figure 7(a)). The dynamic centrifuge tests were conducted at the Korea Advanced Institute of Science and Technology Geotechnical Centrifuge Testing Center, and an equivalent shear beam container was used to minimise wave reflection at the container boundaries during dynamic loading. The soil model was prepared in the container using an air-pluviating method to achieve a uniform density. The soil model had approximate dimensions of 0.84 × 0.63 × 0.60 m at the model scale. The shear wave velocity of the soil was measured using bender elements installed in the soil layer, and the average shear wave velocity was ∼173 m/s. The dynamic centrifuge tests were conducted sequentially, beginning with the free-field ground condition and followed by tests with structures installed on the soil model. The SDoF system exhibited a resonant frequency of ∼86 Hz in the model scale (0.69 Hz in the prototype scale), as determined from the impact hammer test with the fixed boundary condition. Details of the centrifuge tests are also reported in Park *et al*. (2025, 2026).

The motion of the structure (i.e. the top mass of the SDoF structure) was captured by a high-speed camera (Phantom v5.1, Phantom) at 1024 frames per second (hereafter, fps), with an image resolution of 1024 × 1024 pixels, and a B&W pattern of size 110 × 110 mm was attached to the top mass (see Figure 7(b)).

Table 2. Information on videos analysed in this study

Analysis ID	Train video	Tracking video	Train method	Number of datasets	Training/labelling condition
Case1	BW1-1	BW1-1	M1	540	Fully labelled images
Case2	BW1-1	BW1-1	M2	540	Single-image labelling (copy)
Case3	BW1-1	BW1-1	M3	1000	Partial labelling (B&W intersection)
Case4	BW1-1	BW1-1	M3	1000	Partial labelling (two squares)
Case5	BW1-1	BW1-1	M4	1000	Whole-pattern labelling (expanded set)
Case6	BW1-1	BW1-2	M4	1000	Pre-trained model (same pattern)
Case7	BW1-1	BW1-3	M4	1000	Pre-trained model (same pattern)
Case8	BW1-1	BW2	M4	1000	Pre-trained model (same pattern)
Case9	BW1-1	BW3	M4	1000	Pre-trained model (same pattern)
Case10	BW2	BW2	M4	1000	Self-trained model for comparison with pre-trained cases
Case11	BW3	BW3	M4	1000	Self-trained model for comparison with pre-trained cases
Case12	O1	O1	M4	1000	Pre-trained model applied to different pattern types
Case13	O1	O2	M4	1000	Pre-trained model applied to different pattern types
Case14	O1	O3	M4	1000	Pre-trained model applied to different pattern types
Case15	S1	S1	M4	1000	Pre-trained model applied to different pattern types
Case16	S1	S2	M4	1000	Pre-trained model applied to different pattern types
Case17	S1	S3	M4	1000	Pre-trained model applied to different pattern types
Case18	C1	C1	M4	1000	Pre-trained model applied to different pattern types
Case19	C1	C2	M4	1000	Pre-trained model applied to different pattern types
Case20	C1	C3	M4	1000	Pre-trained model applied to different pattern types
Case21	X1	X1	M4	1000	Pre-trained model applied to different pattern types
Case22	X1	X2	M4	1000	Pre-trained model applied to different pattern types
Case23	X1	X3	M4	1000	Pre-trained model applied to different pattern types

A video record of excitation induced by the Kobe earthquake, with peak ground accelerations of 0.440 g at the surface and 0.227 g in the bedrock, was selected among the sequential excitations in increasing acceleration levels. The video selected is the maximum amplitude excited at the last stage to obtain the clearest image data. The ground accelerometers were installed in vertical arrays at several horizontal locations within the soil model. Accelerometers were installed at each level of the soil and structure, but data from two accelerometers, A068 and A064, attached to the top mass, were selected in this study to compare with the image analysis results.

Since the high-speed camera was installed far from the model within the centrifuge, the quality of the video from the dynamic centrifuge tests had to be improved. The original images were zoomed in, and their contrast was adjusted to minimise errors caused by insufficient lighting and blurry images. A total of 957 images were extracted.

Finally, the YOLO results were compared with the results analysed by two popular software programs: TEMA Motion and Tracker. TEMA Motion is a commercial software program developed by (Image Systems Motion Analysis, 2026) while Tracker is an open-source software program developed by Douglas Brown (2009). TEMA Motion is specialised and widely used for high-speed camera image analysis, offering strengths in ensuring the accuracy and repeatability of displacement measurements. Tracker, on the other hand, allows experimental data to be analysed without significant cost, providing various functions and being highly accessible.

2.4 Analysis method

2.4.1 YOLO setup

In deep learning, an epoch is a unit that refers to the model being trained on the entire dataset once. One epoch consists of training the model on the entire dataset, divided into smaller sets called batches, with each batch having a specified size. These parameters affect memory usage and the time required for training. In this study, the number of epochs was set to 1000, and the batch size was set to 8. The ratio of the training and validation datasets was set to 8:2. The detection confidence threshold was set to 0.5. The input image size for the shaker test was 1280 × 1280 pixels, while that for the centrifuge tests was 1024 × 1024 pixels. Training and detection processes were conducted on Google Colab, a cloud-based platform that utilised up to 32GB of RAM, an Intel(R) Xeon(R) CPU @ 2.30 GHz, and an NVIDIA Tesla T4 GPU.

The coefficient of determination (R^2) and error (%) were used to evaluate the performance of the analysis.

1. $$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

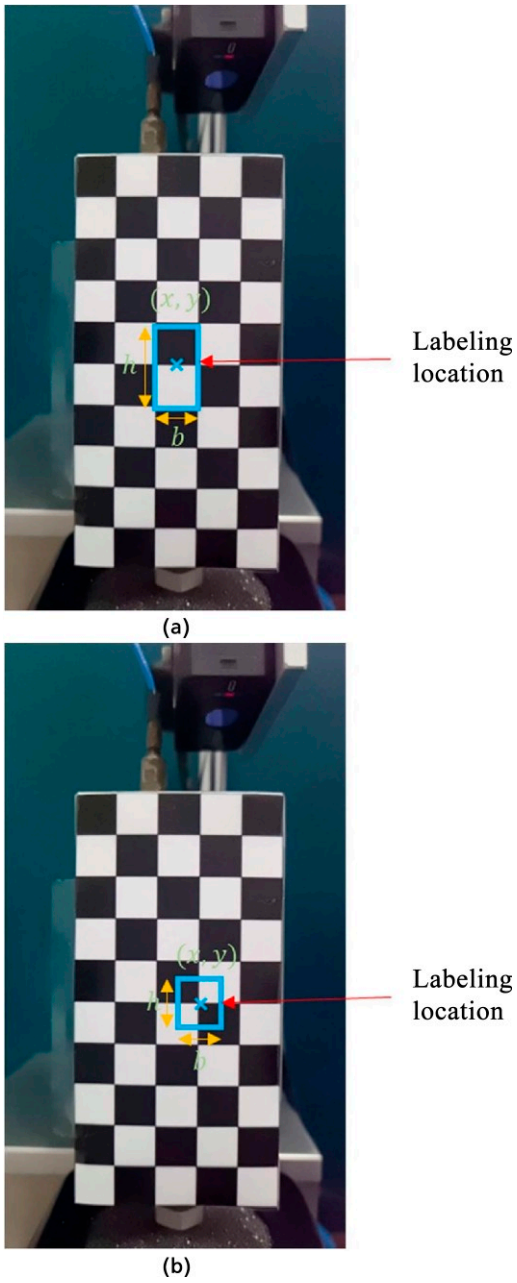

Figure 5. Example of partial area labelling and tracking: (a) partial area tracking unit located at the intersection of black-and-white checks in the centre (10 × 10 mm); and (b) larger partial area tracking unit (20 × 10 mm)

$$2.\quad \text{Error}\ (\%) = \frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\overline{y}_i$ is the mean of the actual values.

2.4.2 Image correction

During the image analysis process, poor resolution and low-quality images can result in unclear and blurred object boundaries, leading to potential errors in the analysis. In particular, the centrifuge experiment images required improvement since they were captured from a long distance. Image enlargement and contrast enhancement techniques were applied to improve resolution and minimise errors. From the original image of the centrifugal model experiment, a partial area of 240 × 240 pixels was selected from the centre of the object. This area was resampled to match the original image size of 1024 × 1024 pixels, as shown in Figure 8. The pixel size of this zoomed image was 0.15714 mm/pixel.

The RSWHE-M method was utilised to enhance contrast (Kim and Chung, 2008). The RSWHE-M method is a histogram equalisation technique that preserves brightness while effectively enhancing image contrast. The input image is recursively divided into two sub-histograms based on the average brightness. An adjusted histogram is then generated by applying a weight that considers the average brightness of each sub-histogram. Smoothing is independently performed on these adjusted histograms, which are then combined to produce the final image. Finally, the brightness of the original image is preserved using the mean-preserving transformation. According to Patel *et al.* (2013), RSWHE-M has demonstrated superior performance in preserving brightness and enhancing contrast compared with other existing techniques.

2.4.3 Signal processing

The procedures established in this study for computing displacement and acceleration are as follows. Firstly, the displacement in pixels from YOLO analysis is computed by measuring the change in the centre co-ordinate relative to its position in the first frame. Secondly, the displacement in pixels is converted to physical unit displacement (i.e. mm in this study) by multiplying the pixel-to-mm sensitivity, which is determined from the ratio of the labelled bounding-box dimensions (in pixels) to the actual size of the pattern. Thirdly, the physical displacement time series is differentiated numerically twice to obtain the acceleration.

The YOLO-based displacement data were processed to correspond to the accelerometer and laser sensor records. Since the sampling rate of the image frames was lower than that of the sensors, the YOLO displacement time series was resampled using linear interpolation to match the time stamps of the sensor data. The unnecessary data at the beginning and end of the recordings were removed to ensure the same signal duration and the same number of samples between the image-based and sensor-based data.

3. Results of experiments with shaker

3.1 Summary of training and analysis status

Table 3 presents the resulting status parameters for training and analysis. After YOLO object training, Train_box/loss and Val_box/loss

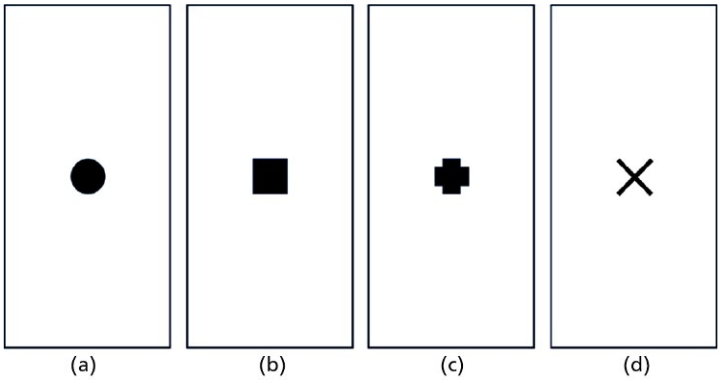

Figure 6. Pattern according to analysis ID: (a) Case12 to Case14; (b) Case15 to Case17; (c) Case18 to Case20; and (d) Case21 to Case23

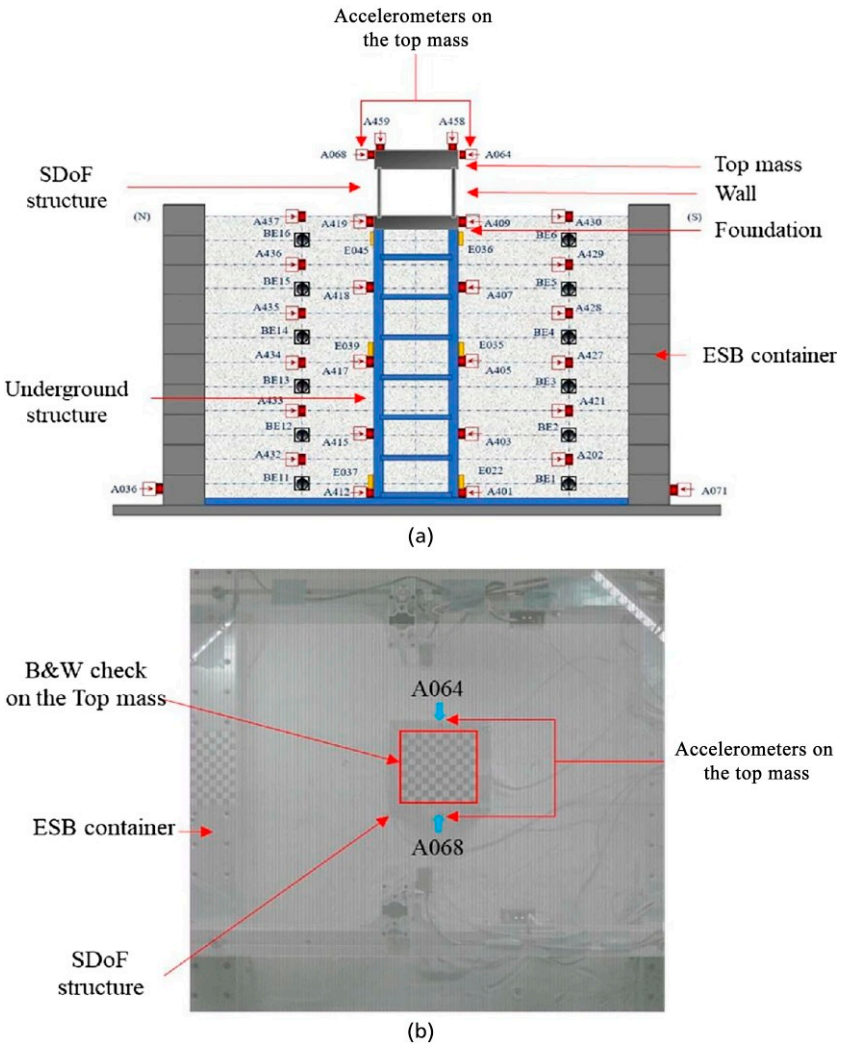

Figure 7. Dynamic centrifuge test setup: (a) cross section of the centrifuge model; and (b) high-speed camera image

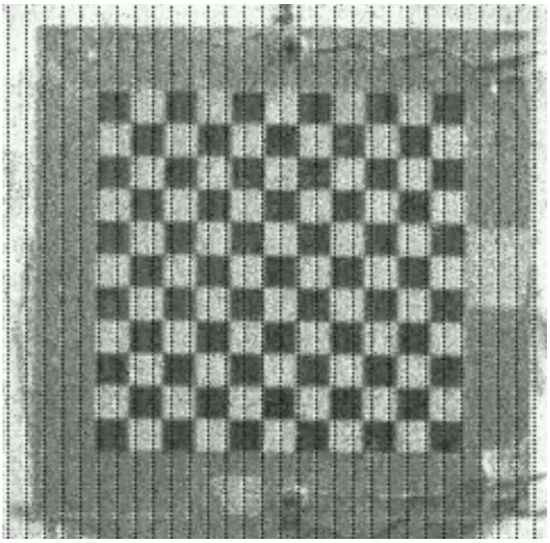

Figure 8. An example image with modified magnification and contrast enhancement

represent the bounding box co-ordinate prediction errors for the training and validation datasets, respectively. These metrics are generally used as indices to evaluate how accurately the trained model predicts the bounding boxes of actual objects. The training status parameters indicate the maturity of YOLO-trained models, which can vary depending on the training methods, the number of images used for training, and the types of patterns. The closer this result is to 0, the better the training performance. As a result, the training status parameters of Case3 to Case23 in Table 3 are very small (i.e. 0.000427–0.000772), and their labelling methods are the same as that of Case2. Therefore, in the next section, only the results of Case1 and Case2 are compared and discussed.

Figure 9(a) compares one example of time histories analysed by YOLO and measured by the laser sensor. Using data points from these histories, the displacement error (against the laser sensor measurement) and acceleration error (against the accelerometer measurement) were calculated using Equation 2. The average errors at all peaks are listed as displacement error and acceleration error in Table 3. The closer these values are to 0, the more closely they match the actual measurements. In addition, to visually compare the analysis cases, a linear regression was performed between the analysis results and their corresponding measured records, as plotted in Figure 9(b). The R^2 value and the slope deviation from the ideal 1:1 line of the regression is also calculated and listed in Table 3. Here, the slope deviation refers to the difference between the slopes of the regression line and the ideal 1:1 line (see Figure 9(c)). The closer the R^2 value is to 1, the more similar the analysis results are to the actual values.

3.2 Effect of training methods

The training results show that for Case1, the train/box_loss was 0.004113, and the val/box_loss was 0.000753. In comparison,

Table 3. Status parameters for training and analysis

	Train status parameter		Analysis status parameter			
Analysis ID	Train_box/loss	Val_box/loss	Displacement error: %	Acceleration error: %	R^2	Slope deviation
Case1	0.004113	0.000753	5.247	9.076	0.9795	0.0966
Case2	0.000248	0.000736	5.026	2.196	0.9725	0.0115
Case3	0.000774	0.001191	20.089	15.071	0.9772	0.1903
Case4	0.000606	0.000537	11.076	5.832	0.9886	0.0958
Case5	0.000682	0.000219	5.0265	2.413	0.9829	0.025
Case6	0.000682	0.000219	3.796	1.560	0.9857	0.0259
Case7	0.000682	0.000219	6.697	0.965	0.9822	0.0455
Case8	0.000682	0.000219	2.676	1.762	0.9895	0.0288
Case9	0.000682	0.000219	8.820	14.059	0.9711	0.098
Case10	0.000427	0.000217	0.856	3.544	0.9805	0.0247
Case11	0.00052	0.000632	5.608	0.094	0.9708	0.0292
Case12	0.000611	0.000387	6.721	0.988	0.9907	0.0604
Case13	0.000611	0.000387	6.046	1.096	0.9884	0.058
Case14	0.000611	0.000387	6.627	1.073	0.9882	0.0599
Case15	0.000656	0.000917	5.279	6.518	0.9906	0.081
Case16	0.000656	0.000917	7.777	6.212	0.9914	0.0765
Case17	0.000656	0.000917	8.530	7.646	0.9874	0.0791
Case18	0.000656	0.000724	6.579	2.912	0.991	0.0763
Case19	0.000656	0.000724	8.046	2.789	0.9927	0.0732
Case20	0.000656	0.000724	8.406	1.837	0.9899	0.0653
Case21	0.000676	0.000847	2.419	2.727	0.9931	0.0166
Case22	0.000676	0.000847	2.630	2.234	0.9924	0.0244
Case23	0.000676	0.000847	2.781	2.266	0.9921	0.0145

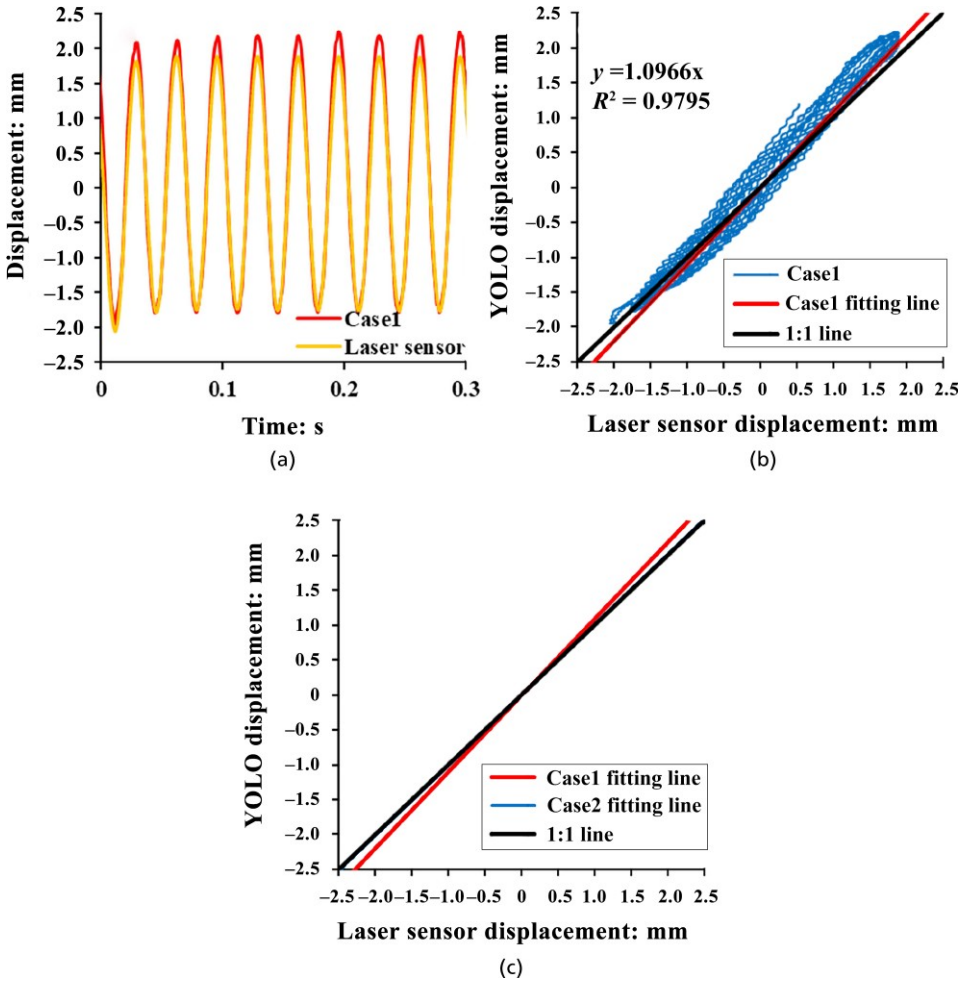

Figure 9. (a) Time histories of Case1 prediction and measured laser sensor displacement signal; (b) fitting result between Case1 prediction and laser measurement signals; and (c) comparison of Case1 and Case2 fitting results

Case2 had a train/box_loss of 0.000248 and a val/box_loss of 0.000736, indicating that Case2 predictions were closer to the actual values than those of Case1. The coefficient of determination was 0.9795 for Case1 and 0.9725 for Case2. The slope deviation was 0.0966 for Case1 and 0.0115 for Case2, with Case1 showing a larger deviation. The average peak-to-peak amplitude of Case1 was 3.987 mm, while that of Case2 was 3.839 mm. The deviation between these two values was 0.148 mm, resulting in an error of 3.714%. The displacement error for the laser sensor measurements was 5.247% for Case1 and 5.026% for Case2, showing a difference of less than 1%. However, the acceleration error for the accelerometer measurements was 9.076% for Case1 and 2.196% for Case2, demonstrating that Case2 performed 6.88% better.

Although Case2 yielded lower errors than Case1 across most evaluation metrics, the applicability of Case2 is limited, as the copied-frame labelling strategy is valid only under controlled laboratory conditions. These conditions require consistent illumination, rigid-body motion without pattern deformation, and a camera–object distance that remains nearly constant throughout the experiment, causing more severe human errors possibly involved in manual labelling of the ground-truth. Therefore, while Case2 demonstrates improved performance, its use is restricted to environments where these assumptions are satisfied.

In terms of training time, Case1 took 5.924 h, while Case2 required 6.058 h. Although Case1 took 0.134 h less to train, the labelling task for Case1 required more than an hour, making the process more time-consuming overall. The higher error associated with Case1 is likely due to the manual labelling process used to define the ground truth for the object. As a result, the remainder of the analysis in this study was conducted using Case2.

3.3 Effect of pre-trained model

Training is quite time-consuming; thus, the performance of the pre-trained model was examined by comparing Case5 to Case11. The training of Case5, Case10, and Case11 took 11.63 h, 11.212 h, and 11.31 h, respectively, and Case5 to Case9 required the same amount of time as Case5 because their training models were identical. The displacement errors for Case5, Case6, and Case7 were 5.0265%, 3.796%, and 6.697%, respectively, and the acceleration errors were 2.413%, 1.560%, and 0.965% (see Table 3). The displacement error for Case8 was recorded as 2.676%, and the acceleration error was 1.762%, showing a difference of about 2% when compared to the displacement error of 0.856% and the acceleration error of 3.544% for Case10. The displacement error for Case9 was recorded as 8.820%, and the acceleration error was 14.059%, showing a higher error of about 13% in acceleration compared with the displacement error of 5.608% and the acceleration error of 0.094% for Case11.

Case5 to Case7 were compared to the laser displacement measurements. The R^2 values of the fitting lines for Case5, Case6, and Case7 were 0.9829, 0.9857, and 0.9822, respectively, all being close to the 1:1 line. Their regression lines showed small slope deviations, ranging from 0.025 to 0.0455. The R^2 value for Case8 was 0.9895, slightly higher than the value of 0.9805 for Case10. Similarly, the R^2 value for Case9 was 0.9711, slightly higher than the value of 0.9708 for Case11. The slope deviation for Case8 was low at 0.0288, for Case10 at 0.0247, and for Case11 at 0.0292. However, Case9, which was tracked using a pre-trained model on an image taken at a distance of 200 mm, had a relatively high slope deviation of 0.098.

These results demonstrate that the pre-trained model can be applied to separately recorded videos, but a higher error is observed when the images are taken from a greater distance.

3.4 Partial area tracking of black-and-white check pattern

For training ground truth, the whole pattern can be labelled as demonstrated in Figure 4, or only a partial area of the pattern can be labelled, as shown in Figure 5. In this study, Case3 and Case4 were trained with two small partial areas of the pattern and compared to Case5, which was trained with the whole pattern. For Case3, the error of the displacement measured by YOLO relative to that by the laser sensor was 20.089%, and the error of the acceleration by YOLO relative to the accelerometer was 15.071%. For Case4, the error of the YOLO displacement relative to the one by laser sensor was 11.076%, and the error of YOLO acceleration relative to the one by the accelerometer was 5.832%. These are much larger than the displacement error of 5.027% and the acceleration error of 2.413% measured for Case5, where the whole pattern was used for training (Table 3). The R^2 value of Case3 is 0.9772, and the slope deviation is 0.1903. The R^2 value of Case4 is 0.9886, and the slope deviation is 0.0958. The values of Case4 were closer to the actual values than those of Case3, but both methods measured relatively large errors. The reasons for the larger error could be as follows: Firstly, the inaccurate detection of boundary lines, resulting in a change in the size of the predicted bounding box. Secondly, YOLO may have interpreted the B&W check differently or missed some of them.

3.5 Performance of various patterns

Four additional patterns were examined, as shown in Figure 6. The analysis IDs for B&W check, circle, square, cross, and X are Case5 to Case7, Case12 to Case14, Case15 to Case17, Case18 to Case20, and Case21 to Case23, respectively (Table 2). The training time was 13.97 h for the circle, 19.04 h for the square, 11.31 h for the cross, and 16.59 h for the X. Of the five patterns (B&W check, circle, square, cross, and X), the cross required the shortest training time, and the square required the longest.

The displacement errors were 6.050%–6.721%, 5.279%–8.530%, 6.579%–8.406%, and 2.419%–2.781% for the circle, square, cross, and X, respectively, and the acceleration errors were 0.988%–1.096%, 6.212%–7.646%, 1.837%–2.912%, and 2.234%–2.727%, respectively (see Table 3). The X showed the lowest error for displacement, and the circle showed the lowest error for acceleration. For both the displacement error and acceleration error, the square showed the highest error. The R^2 values were 0.9882–0.9907, 0.9874–0.9914, 0.9899–0.9927, and 0.9921–0.9931, with X recording the highest value and the slope being closest to the 1:1 line. The X, with consistent displacement and acceleration errors around 2% and the highest R^2 value, was evaluated as the best pattern (see Table 3).

The superior tracking performance of the X pattern can be attributed to the hierarchical way the Convolutional Neural Network architecture in YOLO processes visual information, where the model recognises objects by first identifying simple edges and then combining them into complex features such as junctions. Unlike other patterns, the X pattern's diagonal intersections provide a rich set of unique visual anchors that stand out against typical horizontal and vertical background noise. These distinct geometric features allow YOLO to maintain precise localisation of the object's co-ordinates, ensuring the detection remains highly sensitive and responsive to rapid displacements.

4. Application to dynamic centrifuge tests

4.1 Yolo results compared to other programs

Figure 10 presents a comparison of the acceleration time histories and Fast Fourier Transform (FFT) of YOLO to the acceleration signal of the accelerometers attached to the structure. All signals were processed using a band-pass filter with a frequency range from 0.2 to 300 Hz before comparison. Additional smoothing methods, including Savitzky–Golay smoothing and moving-average smoothing, were examined prior to numerical differentiation of the YOLO-derived displacement signal. However, they did not produce a significant difference in the resulting acceleration response, and therefore no additional smoothing was applied. The amplitudes

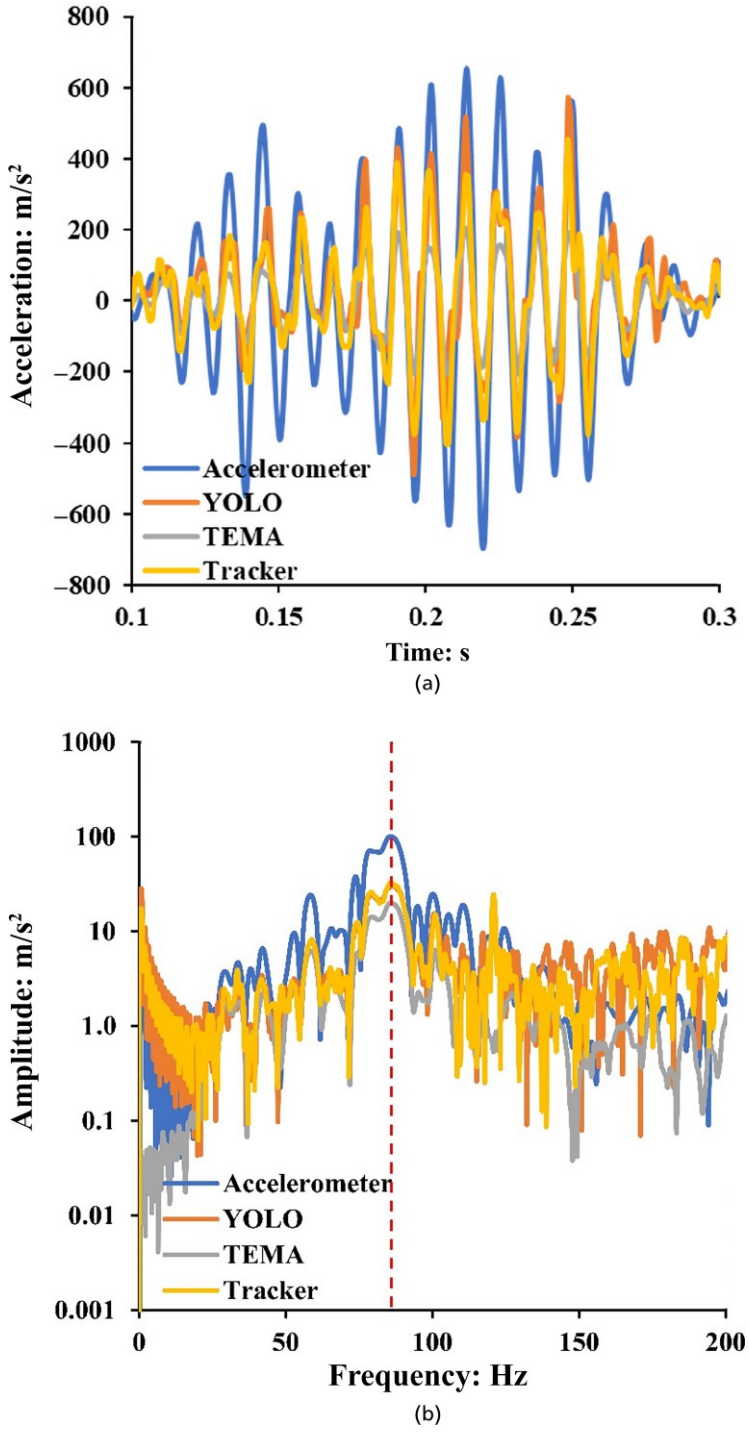

Figure 10. Comparison of accelerometer measurement and three analysed results: (a) time histories, and (b) FFT

near dominant peaks derived from the YOLO-based displacement differentiation were relatively lower. Such discrepancies are expected because the video data were sampled at 1024 fps (i.e. ∼14 frames a wave between peaks), whereas the accelerometer signals were recorded at a higher sampling frequency, 4096 Hz. This mismatch would contribute higher error in image-based methods to accurately capture the true extrema of rapidly varying acceleration signals in the case of this study. This confirms that the higher errors

in image-based method is originate from the limitation of the recording device rather than the performance of YOLO.

The maximum amplitudes in the FFT results showed a large difference (see Figure 10(b)). However, the peak frequency ranges from 85.799 to 86.105 Hz, showing that the accelerometer and the three program results have consistent peak frequencies (see Table 4). This large difference in amplitude is due to the poor quality of the images used, caused by the distance between the object structure and the camera. It is concluded that the frequency response of the object is well captured by the image analysis including YOLO and others.

Figure 11 compares the seismic displacement traces analysed by YOLO, TEMA, and the Tracker using the same high-speed camera video recorded from the centrifuge test. Since any displacement sensor was not installed, the YOLO image analysis result was compared with the other program results. At the maximum peak in time history, errors of 15.266% with TEMA and 8.251% with the Tracker were observed; while at the minimum peak in time history, the errors were 17.399% with TEMA and 4.440% with the Tracker. The peak frequencies from the three tools were consistently observed within the range of 85.917–86.105 Hz, confirming that the overall results are in good agreement (see Table 5). Because all three methods analysed the same video data with identical sampling characteristics, these results reflect a relative comparison of image-based tracking performance rather than an absolute validation against ground-truth displacement, confirming the comparable performance of YOLO analysis to the other software packages in frequency domain.

Table 4. Comparison of analysis results for acceleration signals

	FFT	
Measuring equipment	Peak frequency: Hz	Amplitude at peak frequency: m/s^2
Accelerometer	85.799	101.027
YOLO	86.105	30.871
TEMA Motion	85.917	20.088
Tracker	85.853	33.233

Table 5. Comparison of analysis results for displacement signals

	FFT	
Measuring equipment	Peak frequency: Hz	Amplitude at peak frequency: mm
YOLO	85.979	0.1146
TEMA motion	86.105	0.1002
Tracker	85.917	0.1255

5. Conclusions

In this study, one of the machine learning–based image analysis techniques, YOLO, was examined regarding its applicability to dynamic measurement of vibrating objects. As a result, the following findings were obtained:

1. To optimise training time and reduce errors, a comparison was made between labelling all images from the training video and labelling and copying only the first image from the training video. The results showed that labelling and copying only the first image was overall more effective for both the training state parameters and the analysis state parameters. However, this copied-frame labelling strategy is valid only under strict laboratory conditions, where the illumination remains constant, the pattern exhibits rigid-body motion without deformation, and the camera–object distance remains unchanged.
2. To verify whether a pre-trained model can be applied to track different videos separately, B&W checks were recorded three times at a distance of 100 mm, and videos recorded at camera-to-object distances of 150 mm and 200 mm were analysed to assess the effect of camera distance. The errors at 100 mm and 150 mm showed low values, whereas the error at 200 mm showed a relatively large value.
3. The applicability of training with only a partial portion of the pattern was examined. When the ground-truth region was limited

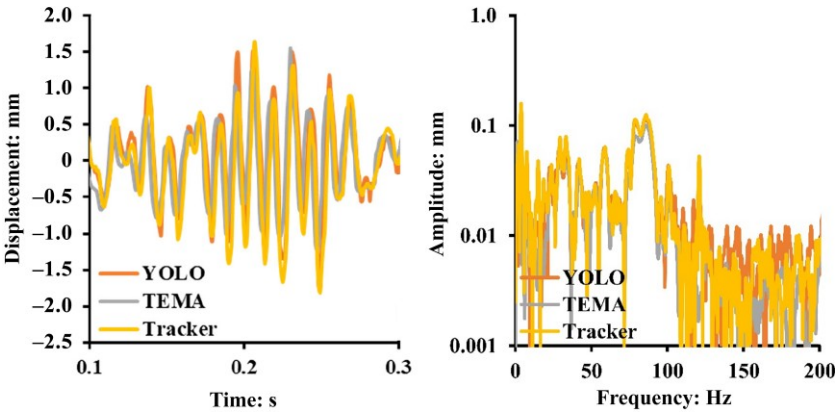

Figure 11. Comparison of seismic displacement signals analysed by YOLO, Tracker, and TEMA Motion: (a) time histories, and (b) FFT

to a partial area of the pattern, the resulting displacement and acceleration errors increased compared to training with the entire pattern. This demonstrates that labelling the entire pattern provides greater performance in dynamic tracking. Separately, to analyse the results of different pattern images, five patterns (B&W check, circle, square, cross, and X) were recorded three times each from a distance of 100 mm and analysed using a pre-trained model. The results of the analysis showed that the X pattern had the highest R^2, with the displacement error and acceleration error mostly constant at 2%. This superior performance likely stems from the clear intersection point and distinct linear features of the X pattern, which provide stable visual cues for localisation and contribute to more consistent bounding-box predictions compared with the other patterns.

4. YOLO was applied to analyse the images recorded from dynamic centrifuge tests. The amplitudes near the dominant peaks derived from the YOLO-based displacement differentiation appeared lower than those obtained from the accelerometer measurements. However, the FFT results showed that the peak frequency was found consistently at 85.799–86.105 Hz, indicating consistent frequency components. Therefore, part of the discrepancy can be attributed to the limited sampling rate of the high-speed camera, which limits the ability of image-based methods to accurately capture the true peak values of rapidly varying acceleration signals. The results indicate that the performance of YOLO analysis is comparable to that of the other image-based software packages in the frequency domain.

The results suggest that dynamic displacement measurement using the YOLO algorithm is feasible with low-cost equipment under the tested laboratory condition. The proposed approach is primarily suited for rigid-body vibration tracking, where pattern deformation is negligible. Further research, including sinusoidal tests over a broader frequency range and enhancements in image resolution and noise reduction, will be necessary to improve the accuracy and broader applicability of the YOLO-based analysis.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00406320 and RS-2026-25468576).

REFFRENCES

Altuhafi F, O'Sullivan C and Cavarretta I (2013) Analysis of an image-based method to quantify the size and shape of sand particles. *Journal of Geotechnical and Geoenvironmental Engineering* **139(8)**: 1290–1307.

Arshad MI, Tehrani FS, Prezzi M and Salgado R (2014) Experimental study of cone penetration in silica sand using digital image correlation. *Géotechnique* **64(7)**: 551–569.

Brown D (2009) Tracker, See. https://physlets.org/tracker/ (accessed 07/05/2026).

Chavda JT, Mishra S and Dodagoudar GR (2020) Experimental evaluation of ultimate bearing capacity of the cutting edge of an open caisson. *International Journal of Physical Modelling in Geotechnics* **20(5)**: 281–294.

Choo YW, Kim JH, Park HI and Kim DS (2013) Development of a new asymmetric anchor plate for prefabricated vertical drain installation via centrifuge model tests. *Journal of Geotechnical and Geoenvironmental Engineering* **139(6)**: 987–992.

Costa S and Kodikara J (2012) Evaluation of J integral for clay soils using a new ring test. *Geotechnical Testing Journal* **35(6)**: 1–9.

Image Systems Motion Analysis (2026) TEMA Motion. See https://temaplatform.com/ (accessed 22/05/2026).

Jocher G (2020) YOLOv5, See https://github.com/ultralytics/yolov5 (accessed 07/05/2026).

Kaddhour G, Andò E, Salager S *et al.* (2013) Application of X-ray tomography to the characterisation of grain-scale mechanisms in sand. In *Multiphysical Testing of Soils and Shales*. Springer Berlin Heidelberg, Berlin, Germany, pp. 195–200.

Kim M and Chung MG (2008) Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement. *IEEE Transactions on Consumer Electronics* **54(3)**: 1389–1397.

Park S, Kim GY and Chang I (2024) Experimental study on the effect of surface-projected conditions on the mechanical behavior of pile embedded in sand. *International Journal of Geo-Engineering* **15(1)**: 22.

Park SJ, Falcon SSD, Van Nguyen D, Choo YW and Kim D (2026) Centrifuge modeling on seismic response for soil-foundation-building systems supported by shallow foundations and deep basements. *International Journal of Geo-Engineering* **17(1)**: 5.

Park SJ, Van Nguyen D, Kim D and Choo YW (2025) Period identification of multi degree of freedom structure-shallow foundation-ground system: dynamic centrifuge test. *Geomechanics and Engineering* **43(3)**: 167–195.

Patel O, Maravi YP and Sharma S (2013) A comparative study of histogram equalization based image enhancement techniques for brightness preservation and contrast enhancement. *Signal & Image Processing: An International Journal* **4(5)**: 11–25.

Redmon J (2016) You only look once: unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 1–12.

Wahyudi S, Miyashita Y and Koseki J (2012) Shear banding characteristics of sand in torsional shear test evaluated by means of image analysis technique. *Bulletin of ERS* **(45)**: 1–8.

White DJ (2002) *An investigation into the behaviour of pressed-in piles*. PhD thesis, University of Cambridge, Cambridge, UK.

White DJ, Take WA and Bolton MD (2003) Soil deformation measurement using particle image velocimetry (PIV) and photogrammetry. *Géotechnique* **53(7)**: 619–631.

Yuan B, Xu K, Wang Y, Chen R and Luo Q (2017) Investigation of deflection of a laterally loaded pile and soil deformation using the PIV technique. *International Journal of Geomechanics* **17(6)**: 04016138.