CTrPile: A Computer Vision and Transformer Approach for Pile Capacity Estimation from Dynamic Pile Load Test

Sompote Youwai
Department of Civil Engineering
King Mongkut's University of technology Thonburi
Bangkok, Thailand
sompote.you@kmutt.ac.th

Parchya Makam
Department of Civil Engineering
King Mongkut's University of technology Thonburi
Bangkok, Thailand

***Abstract*— Dynamic pile load tests are essential for verifying the ultimate limit state for pile design in geotechnical engineering. However, conventional methods for monitoring these tests, such as strain gauges and accelerometers, are expensive and labor-intensive. This paper proposes a novel method that uses computer vision and artificial markers to measure pile head movement during dynamic pile load tests, and a transformer-based deep learning model to predict pile capacity from the movement data. The proposed method is low-cost, easy-to-use, and accurate, with a mean absolute error of 2.4% for pile capacity prediction using K-fold cross-validation. The paper also presents a sensitivity analysis of the transformer model with respect to the number of heads and layers, which indicated the optimal settings to avoid overfitting of the training data. The paper discusses the limitations of the proposed method, such as the dependency on the camera position and suggests future directions of the research, such as incorporating other features and improving the data quality. The proposed method can be applied in real cases of dynamic pile load tests to increase the number of tests on site and to ensure the safety and reliability of pile design.**

Keywords: Computer vison, Transformer, Pile Load Test

I. Introduction

Pile foundations are a type of foundation that transfer loads from the building to the firm layer of soil. The pile capacity needs to be verified by pile load tests to ensure the safety and correctness of the calculation from the geotechnical engineer. Dynamic pile load testing is a popular method for testing pile capacity. Its results can be used directly as the pile capacity for certain projects with a factor of safety. It involves applying a load to the pile by a drop hammer. The load applied to the pile is measured directly by the strain gauge and accelerometer installed on the pile itself. However, the cost for installing such instruments as well as the data acquisition unit to obtain data is high when testing a large number of piles. Therefore, an alternative method that is cheaper and faster for testing is strongly needed. This study will use the application of computer vision and machine learning to predict the pile capacity by dynamic pile load test.

Artificial markers are effective methods for measuring the deformation of a target. By using a calibrated camera, the vector from the camera to the target and the pose of the target relative to the camera can be estimated. One type of artificial marker is ArUco[1], which is a library for Augmented Reality applications that uses OpenCV [2] to detect and estimate the pose of binary square fiducial markers. ArUcoO markers have several advantages, such as being easy to create, robust, fast and simple to use. They can be applied to various computer vision tasks, such as camera calibration and pose estimation. ArUco markers have been widely used in applications that involve locating the position of robots [4] and flying vehicles [5]. This study hypothesizes that ArUco markers can be used to detect the pile movement in dynamic pile loading tests.

The load capacity of a pile from dynamic pile load test can be determined by several methods: 1) using the Case method [6],which is a direct calculation, or 2) using the Case Pile Wave Analysis Program [7] (CAPWAP). The CAPWAP method is widely used around the world [8]. The force developed during dynamic pile load test is back-calculated by the CAPWAP method to determine the pile capacity. Several researchers have attempted to predict the pile capacity by using machine learning [9]–[12]. They applied a simple fully connected neural network (FNN) to predict the capacity, end bearing and vibration of the pile. However, there was still a high error from the test results. Therefore, it is necessary to apply the current architectures of machine learning to extract features from the results of dynamic pile load test. The sequential type neural network architectures such as convolutional neural network (CNN), long short-term memory (LSTM) and transformer should be evaluated for their effectiveness in predicting the ultimate pile capacity.

The main objective of this study is to propose a method for monitoring dynamic pile load tests with the application of artificial markers, called ARUCO. The displacement of the pile during testing was utilized to predict the pile capacity and compared with the result of the CAPWAP method. The prediction method used different types of neural network architectures, such as CNN, LSTM and transformer with self-attention. The comparison of the results of these architectures was discussed in the paper. The proposed workflow can be applied in real cases of dynamic pile load tests. The amount of dynamic pile load tests on site can be increased or possibly

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00080

tested with every pile on site with only installing the ARUCO target without any expensive instrument and monitoring system

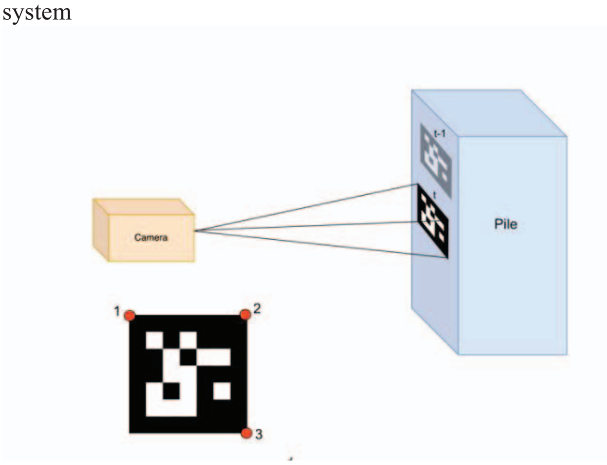

Fig. 1 The concept for calculate of the pile deformation

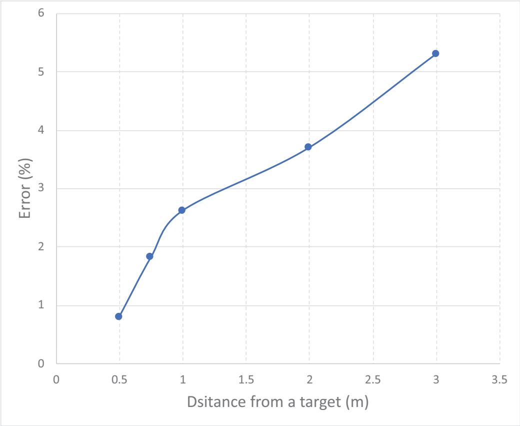

Figure 2 The error for deformation measurement from ARUCO target

II. Computer Vision

This study applied the artificial marker call ArUco attached to pile head to monitor the movement of pile head. ArUco [2] detection is the process of finding and identifying ArUco markers in an image or a video stream. ArUco markers are square-shaped patterns that have a black border and a binary code inside. The binary code determines the marker's unique identifier, which can be used for various applications such as camera pose estimation, augmented reality, and robot navigation. ArUco detection is performed by using the OpenCV library[13], which provides a module called cv2.aruco that contains functions and classes for creating, detecting, and refining ArUco markers. The detection from the liberally was provide the coordinate of the corner for the target of ArUco. The movement of the coordinate coordinate of target can be converted to pile head movement during pile driving.

A GoPro 10 camera was employed in this study to capture high-resolution images (2048 × 1080 pixels) at a high frame rate (240 frames per second). This camera was selected for its commercial-grade performance, low cost ($250 USD), and easy applicability in the construction site. A checkerboard pattern (9x7 grid) was used to calibrate the camera and correct the lens distortion[14]. Camera calibration is a technique to estimate the camera parameters that relate a 3D point in the real world to its 2D projection in the image plane. These parameters include the intrinsic matrix, which contains the focal length, optical center, and skew of the camera, and the distortion coefficients, which account for the radial and tangential distortion of the lens. The OpenCV library was used to perform the calibration by capturing several images of the calibration chessboard from different angles, finding the 2D coordinates of the chessboard corners in the images, and computing the camera parameters [15]. Approximately 50 images were taken with varying camera poses, and the translation and rotation vectors were calculated from the calibration. The camera matrix and distortion coefficient were calculating form the camera calibration.

The displacement of the pile head and the hammer during dynamic pile testing was measured using a 5x5 ArUco target with a size of 150 mm. The target size was chosen to match the pile diameter and to fit into an A4 paper. For larger piles, larger targets could be used. The target was attached to the pile head and the hammer with IDs 0 and 1, respectively. The displacement was computed by tracking the movement of the target relative to the previous frame. The coordinates of the upper left and lower right corners of the target were used as reference points for detection, as shown in Fig. 1 and Equation 1. The displacement in pixels was converted to the real distance in millimeters by multiplying with a distance-to-pixel ratio factor, F (Equation 2). The F value was obtained by dividing the real width of the marker (150 mm) by the measured distance (in pixels) between point 1 and point 2 of the target. The camera accuracy was tested by moving the target up and down by 100 mm. The calibration results are shown in Fig. 2. The error increased with the distance to the target due to the reduction of the target size in pixels. The error could be caused by the pixel detection error with the small distance. The smaller size of the target seemed to be difficult to detect and estimate the pose of the marker [16]. The distance measurement error also increased with the obliqueness of the camera to the marker. This is because the translation vector that represents the position of the marker relative to the camera is affected by the rotation vector that represents the orientation of the marker relative to the camera[17].

$$d = F\sqrt{\left(\left(\frac{x_1+x_3}{2}\right)_t - \left(\frac{x_1+x_3}{2}\right)_{t-1}\right)^2 + \left(\left(\frac{y_1+y_3}{2}\right)_t - \left(\frac{y_1+y_3}{2}\right)_{t-1}\right)^2} \quad (1)$$

$$F = \frac{d_{marker}}{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}} \quad (2)$$

III. Field Experiment

A 23 pile was installed with an ArUco target and a dynamic pile driving equipment (Fig. 3). The dynamic pile driving equipment consisted of 2 strain sensor and accelerometer for monitoring the pile testing. The force and velocity data measured during the pile driving were analyzed by the Case Pile Wave Analysis Program (CAPWAP) [18]. The soil layer parameters were varied until the simulated load matched the measured data at each time step. This approach is the conventional method to determine the pile capacity in civil engineering projects. The pile capacity from CAPWAP approach will be used as a label for trained the machine leaning for the next part.

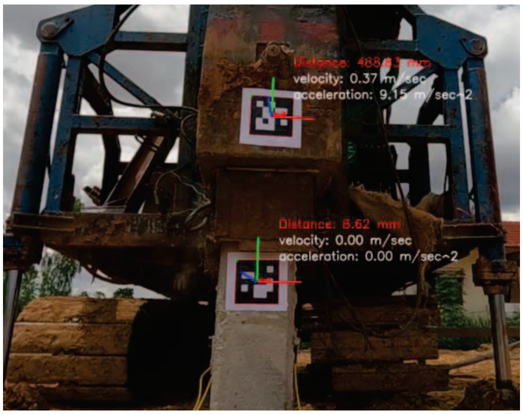

Fig. 3 The ARUCO target installation during testing

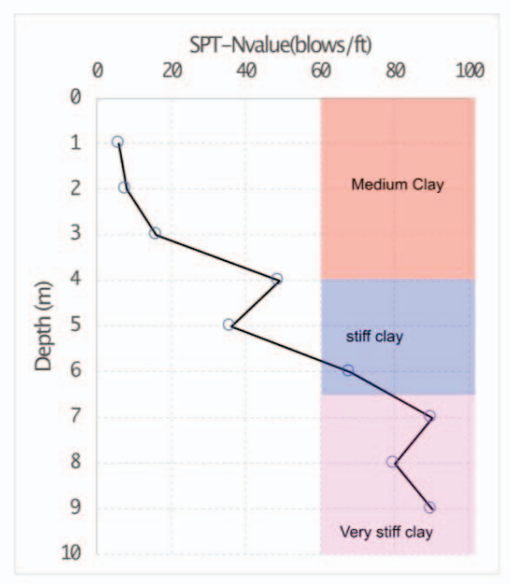

Fig. 4 The value of SPT-N value VS Depth

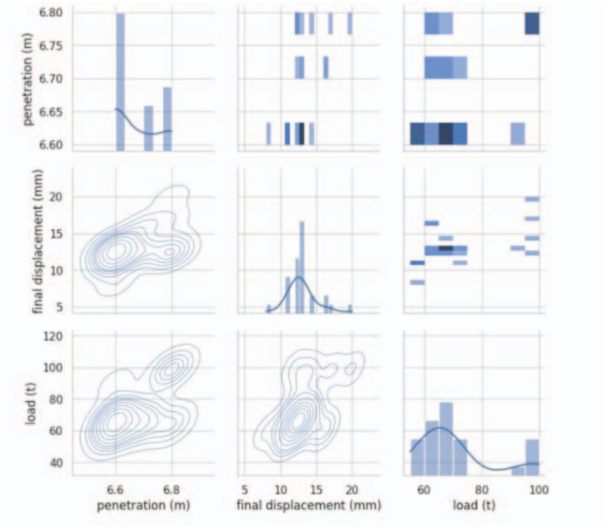

Fig. 5 Feature and label characteristics

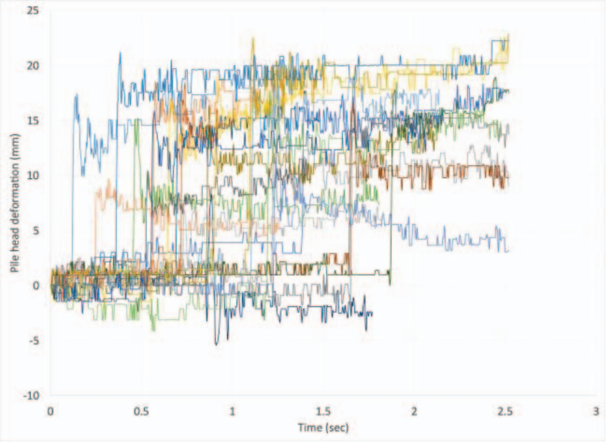

Fig. 6 Pile head deformation from ArUco target

The soil profile from Borehole in this area consists of medium clay from 0 to 4 m (SPT from 5 to 20) and very stiff clay from 5 to 10 m (SPT from 40 to 90) (Fig. 4). The pile penetration depth from pile driving was estimated to be 6.5 m from the ground surface based on the soil profile. The pile could not penetrate further than 6.5 m into the very stiff clay layer. A pile with a diameter of 0.3 m was driven into the soil profile using a hammer with a weight of 5 tons and a height of 2 m from the head of the driven pile. The results from the dynamic pile load test are shown in Fig. 5. The penetration depth of the pile was almost constant, ranging from 6.8 to 6.65 m. The applied load during the dynamic pile load test varied from 60 to 100 tons. The final displacement also varied from 5 to 20 mm. The lower final displacement indicated a higher pile resistance than the higher final displacement.

Figure 6 illustrates the pile head deformation measured by the ArUco target during the pile driving process. The hammer impact induced a rapid variation in the pile head deformation, which ranged from 5 to 20 mm for the first strike. A noticeable rebound of the pile head occurred after each hammer impact. The subsequent oscillation of the deformation value resulted from the wave reflection and vibration in the pile. The camera location, which was close to the pile, might have also introduced some noise in the data due to the camera vibration. The length of the data sequence for measurement varied, so it was padded with zero values to ensure the same sequence length for training a deep learning model.

IV. Machine Learning

This study applied a deep learning sequential model based on the transformer architecture to estimate the load capacity of piles. Unlike the original transformer model, which was designed for classification tasks, this study formulated the problem as a regression task. The final activation function of the final neuron was Linear instead of the activation type for classification. The model architecture consisted of a transformer layer with a multi-head attention mechanism, a dropout layer, and a residual connection, followed by a multilayer perceptron (MLP) (Fig. 7). The input vector was transformed into a one-dimensional vector by flattening, and then passed through the MLP. The MLP gradually reduced the dimension of the vector until it produced a single output value, which represented the load capacity of the pile.

The attention mechanism [19], [20] is a technique that enables a neural network to selectively focus on the most relevant parts of the input or output sequence, depending on the task. It is widely used in natural language processing (NLP) to improve the performance of encoder-decoder models, such as machine translation, text summarization, and speech recognition. The attention mechanism works by projecting the input vector onto a trainable vector at the attention head, which consists of query, key and value vectors (Q, K, and V) (Fig. 8). These vectors can be learned during the neural network training. The attention mechanism can be viewed as a function that takes a query vector, which represents the current state of the decoder, and a set of key-value pairs, which represent the encoded input sequence, as inputs, and returns a weighted sum of the value vectors as output. The weights are computed by measuring the similarity between the query vector and each key vector, using a scoring function such as dot product, cosine similarity, or a neural network. The weights are then scaled by the inverse square root of the key dimension to reduce the effects of high sequence dimension (Eq.3). The weights are then normalized by a Softmax function to form a probability distribution. The result from the Softmax function is then multiplied with the value vector to obtain the output. The output from the attention mechanism is then added with the initial input vector, known as residual connection, to mitigate the effects of gradient descent.

$$Attention(q, k, v) = softmax\left(\frac{qk^T}{\sqrt{d_k}}\right) v \qquad (3)$$

where q *is the* query, k is the key, v is the value and d_k is the scaling factor

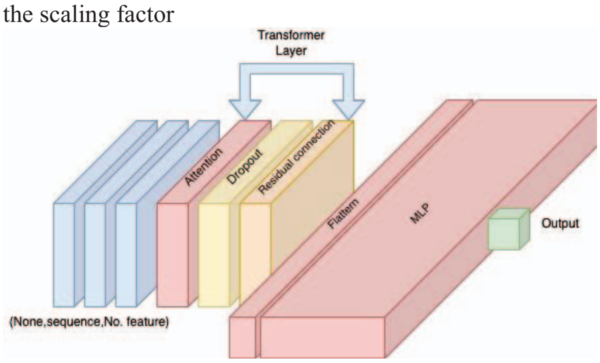

Fig. 7 The architecture of transformer-based model

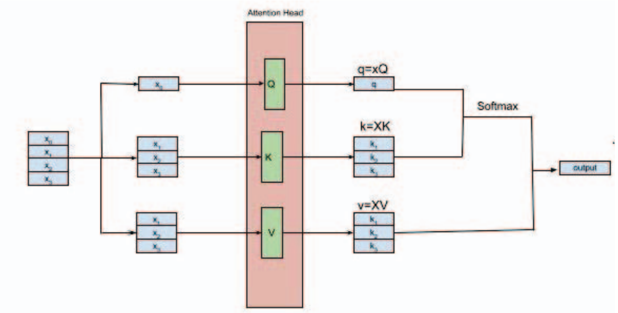

Fig. 8 The diagram of attention mechanism

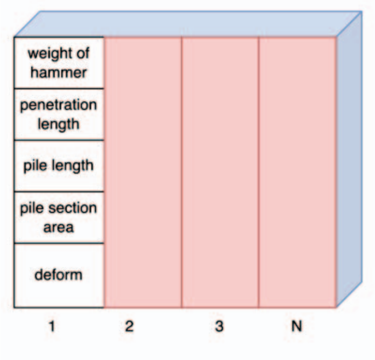

Fig. 9 Input vector of the model

The deformation at each time step was used as a token input for the transformer, as shown in Fig. 9. The first dimension was the deformation at each time step recorded from the ArUco target. The value was normalized by dividing by 20 to set it close to one and prevent gradient explosion during the training of the neural network. The other dimensions of the vector were embedded with features of the pile, such as pile section area, pile length, penetration length, and weight of hammer. For the same pile test case, the feature-embedded vector was similar throughout the sequence of deformation data. It changed with different pile specimens for testing under different conditions, such as different lengths and penetration depths. The features of each pile were normalized by a standard scaler [21], which computed the difference between the value and its mean and divided by its standard deviation.

TABLE 1 MODEL ARCHITECTURE FOR TRANSFORMER

Transformer	
Multi-head attention head_size=5, num_heads=1, dropout =0.2 Residual connection	2 layers
Multi-layer perceptron	

Flatten
Dense(200)
Dense(50)
Dense(10)[4], [5]
Outout(1)	

TABLE 2 MODEL ARCHITECTURE FOR CNN AND LSTM

CNN	LSTM
CNN1D -Filter=256	LSTM (200)
Max pooling-pool size =2	
CNN1D -Filter=256	
Max pooling-pool size =2	
CNN1D -Filter=256	
Max pooling-pool size =5	

The hyperparameter settings of the transformer architecture are shown in Table 1. The first layer consists of a multi-head attention layer with a head size of 5, which matches the dimension of the input token vector. The number of heads is set to 1 to avoid overfitting the model. The transformer architecture relies heavily on self-attention, which can capture long-range dependencies but also be sensitive to noise and outliers. A regularization layer with a dropout rate of 0.2 is applied to reduce the overfitting of the transformer model.

To evaluate the model architecture, this study applied the K-fold cross-validation method [22] , which is suitable for small data sets. The data set was divided into eight equal folds, and one fold was used as the test set while the rest were used as the training set. This procedure was repeated eight times, each time using a different fold as the test set, to assess the generalization performance of the model on unseen and varied data. The mean absolute percentage error (MAPE) was the metric used to measure the prediction accuracy of the model. The hyperparameters were set as follows: batch size = 8, epochs = 400. The model with the lowest mean squared error (MSE) loss on the training set was selected as the representative model. MSE was chosen as the loss function because of the regression nature of the problem. The proposed mode was also evaluate with the other type of sequential model, which are convolution neural network (CNN) and log-short time memory (LSTM). The architectures of the models are shown in Table 2.

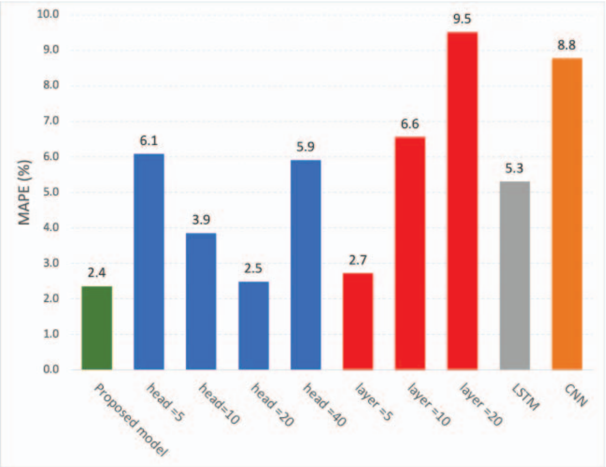

Fig. 10 The mean absolute error (MAPE) from different model architecture

V. Results and Discussion

Fig. 10 shows the comparison of mean absolute percentage error (MAPE) between different types of models. The proposed model varied the hyperparameters of the number of heads and the number of transformer layers. The MAPE value initially increased with the number of heads, but then decreased as the number of heads increased further. Then, the MAPE trend to reduce again when increasing number of head (head-= 40). A higher number of heads also increased the training resource consumption and the tendency to overfit the data, resulting in a higher MAPE value compared to the proposed model. The number of transformer layers also increased the MAPE value of the model, indicating overfitting of the training data. The model could not generate accurate predictions for the unseen data. The proposed transformer-based model outperformed the other types of sequential models, such as convolutional neural network (CNN) and long short-term memory (LSTM). The MAPE of CNN was the highest among the other models. It seemed that it could not capture the overall sequential data, especially the data that had abrupt changes during pie driving. CNNs use convolutional filters that operate on small regions of the input data, which can limit their ability to capture global patterns and long-range dependencies. Transformers, on the other hand, use self-attention mechanisms that can access the whole input data and weigh the importance of different parts. Transformers can process the entire input sequence in parallel, while LSTMs have to process it token by token. This makes transformers faster and more efficient than LSTMs, especially for long sequences.

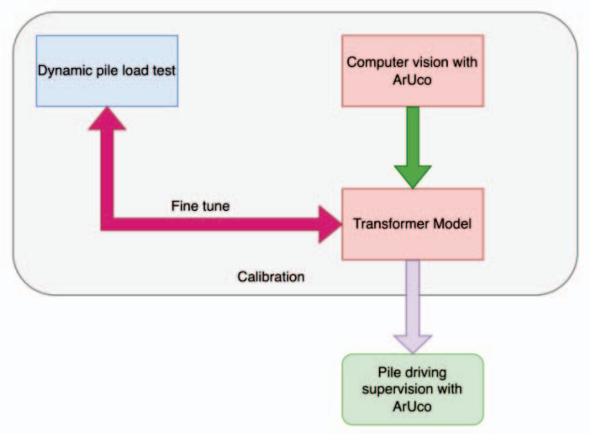
Dynamic pile load test
Computer vision with ArUco
Fine tune
Transformer Model
Calibration
Pile driving supervision with ArUco

Fig. 11 The workflow of the proposed method

This study proposed a transformer-based model that integrated computer vision to predict pile capacity from ArUco target images. The model outperformed other sequential models, but it was limited by the data from one site and a few pile variables. To improve the model's generalizability, more data from diverse sites and features are needed. The model can be applied to site-specific conditions, but it requires calibration with conventional dynamic pile load tests with full instrumentation (Fig. 11). The calibration process involves two steps: (1) calibrating the hybrid computer vision and machine learning model with the real dynamic pile load test results, which are obtained by the CAPWAP method using the velocity and load on the pile measured by the instrument installed on the pile dynamic test; and (2) training the transformer model with the ArUco target images and pile capacity labels from the dynamic pile load tests. The ArUco target is a low-cost tool that can be used for quality control and pile capacity estimation for every pile on site, ensuring the ultimate limit state for pile design. Future work should use higher resolution and frame rate cameras to obtain more accurate and sequential data for the transformer model.

VI. Conclusion

This paper presents a novel method for monitoring dynamic pile load tests using computer vision and artificial markers attached to piles12. The pile head movement detected by the markers is then used as an input for a deep learning model based on the transformer architecture to predict the pile capacity. The main findings of this study are:

• The proposed computer vision method for measuring pile head movement with ArUco targets achieved a satisfactory performance with an error of 2% when the camera was set 2 m from the target.

• The transformer model outperformed other sequential models such as CNN and LSTM in terms of accuracy and robustness for predicting pile capacity from dynamic pile load test data. The mean absolute error for prediction using K-fold cross-validation was as low as 2.4%.

• The mean absolute error of the prediction of pile capacity from dynamic pile load test by using transformer model increased with the number of heads and layers of the transformer, indicating overfitting of the training data.

• We proposed a work flow that uses a hybrid computer vision and transformer model for pile driving supervision. The transformer model requires training before it can be applied to the supervision task.

Acknowlegement

The authors would like to express their gratitude to the Khonkaen Soil Engineering Co. Ltd for assisting them with conducting dynamic pile load tests in fieldwork.

References

[1] European Committee for Standardization, Eurocode 7: Geotechnical design - Part 1: General rules. Brussels: CEN, 2004. [Online]. Available: 1

[2] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," Pattern Recognition, vol. 47, no. 6, pp. 2280–2292, Jun. 2014, doi: 10.1016/j.patcog.2014.01.005.

[3] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[4] S. Roos-Hoefgeest, I. A. Garcia, and R. C. Gonzalez, "Mobile robot localization in industrial environments using a ring of cameras and ArUco markers," in IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/IECON48115.2021.9589442.

[5] A. Khazetdinov, A. Zakiev, T. Tsoy, M. Svinin, and E. Magid, "Embedded ArUco: a novel approach for high precision UAV landing," in 2021 International Siberian Conference on Control and Communications (SIBCON), Kazan, Russia: IEEE, May 2021, pp. 1–6. doi: 10.1109/SIBCON50419.2021.9438855.

[6] F. Rausche, G. G. Goble, and Jr. Likins Garland E., "Dynamic Determination of Pile Capacity," Journal of the Geotechnical Engineering Division, vol. 111, no. 3, pp. 367–383, 1985.

[7] R. Likert, "A method for analyzing pile driving data using wave equations," Journal of Geotechnical Engineering, vol. 100, no. 9, pp. 981–1001, 1974.

[8] M. R. Svinkin, "Sensible determination of pile capacity by dynamic methods," Geotechnical Research, vol. 6, no. 1, pp. 52–67, Mar. 2019, doi: 10.1680/jgere.18.00032.

[9] Y. Robert, "A new approach to the analysis of high-strain dynamic pile test data," Can. Geotech. J., vol. 31, no. 2, pp. 246–253, Apr. 1994, doi: 10.1139/t94-029.

[10] Y. K. Chow, W. T. Chan, L. F. Liu, and S. L. Lee, "Prediction of pile capacity from stress ‐ wave measurements: A neural network approach," Num Anal Meth Geomechanics, vol. 19, no. 2, pp. 107‐126, Feb. 1995, doi: 10.1002/nag.1610190204.

[11] S. K. Das, B. Manna, and D. K. Baidya, "An Artificial Neural Network Approach for Prediction of Dynamic Pile-Soil-Pile Interaction under Vertical Motion," in GeoFlorida 2010, Orlando, Florida, United States: American Society of Civil Engineers, Feb. 2010, pp. 1422–1431. doi: 10.1061/41095(365)143.

[12] R. Gohil and C. R. Parthasarathy, "Intelligent Assessment of Axial Capacity of Pipe Piles Using High Strain Dynamic Pile Load Tests in Offshore Environment," in Soil Behavior and Characterization of Geomaterials, vol. 296, K. Muthukkumaran, R. S. Jakka, C. R. Parthasarathy, and B. Soundara, Eds., in Lecture Notes in Civil Engineering, vol. 296. , Singapore: Springer Nature Singapore, 2023, pp. 271–287. doi: 10.1007/978-981-19-6513-5_24.

[13] "OpenCV: Detection of ArUco Markers." Accessed: May 30, 2023. [Online]. Available: https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html

[14] "OpenCV: Camera Calibration." Accessed: Nov. 16, 2023. [Online]. Available: https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html

[15] M. Sajjad et al., "An Efficient and Scalable Simulation Model for Autonomous Vehicles With Economical Hardware," IEEE Trans. Intell. Transport. Syst., vol. 22, no. 3, pp. 1718–1732, Mar. 2021, doi: 10.1109/TITS.2020.2980855.

[16] Y. Wang, Z. Zheng, Z. Su, G. Yang, Z. Wang, and Y. Luo, "An Improved ArUco Marker for Monocular Vision Ranging," in 2020 Chinese Control And Decision Conference (CCDC), Hefei, China: IEEE, Aug. 2020, pp. 2915–2919. doi: 10.1109/CCDC49329.2020.9164176.

[17] P. Oščádal et al., "Improved Pose Estimation of Aruco Tags Using a Novel 3D Placement Strategy," Sensors, vol. 20, no. 17, p. 4825, Aug. 2020, doi: 10.3390/s20174825.

[18] G. G. Goble, F. Rausche, and G. E. Likins, "The Analysis of Pile Driving - A State of the Art," in Proceedings of the 6th Annual Offshore Technology Conference, Offshore Technology Conference, 1975, pp. 2687–2696.

[19] [D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[20] Vaswani *et al.*, "Attention Is All You Need," 2017, doi: 10.48550/ARXIV.1706.03762.

[21] "sklearn.preprocessing.StandardScaler," scikit-learn. Accessed: Nov. 27, 2023. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[22] Y. Jung and J. Hu, "A K -fold averaging cross-validation procedure," Journal of Nonparametric Statistics, vol. 27, no. 2, pp. 167–179, Apr. 2015, doi: 10.1080/10485252.2015.1010532.