ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.0423
The 10th International Conference on Construction Engineering and Project Management
Jul. 29-Aug.1, 2024, Sapporo

Context- and Shape-Aware Safety Monitoring for Construction Workers

Wei-Chih Chern[1]; Kichang Choi[2] Vijayan Asari, Ph.D.[3]; Hongjo Kim, Ph.D.[4]

[1]*Dept. of Electrical and Computer Engineering, Ph.D. Student , University of Dayton, U.S.A. Email: chernw1@udayton.edu*
[2]*Dept. of Civil and Environmental Engineering, Ph.D. Student, Yonsei Univ., Seoul, Korea. Email: amki1027@yonsei.ac.kr*
[3]*Dept. of Electrical and Computer Engineering, Professor, University of Dayton, U.S.A. Email: vasari1@udayton.edu*
[4]*Dept. of Civil and Environmental Engineering, Assistant Professor, Yonsei Univ., Seoul, Korea. Email: hongjo@yonsei.ac.kr(corresponding author)*

Abstract: The task of vision safety monitoring in construction environments presents a formidable challenge, owing to the dynamic and heterogeneous nature of these settings. Despite the advancements in artificial intelligence, the nuanced analysis of small or tiny personal protective equipment (PPE) remains a complex endeavor. In response to this challenge, this paper introduces an innovative safety monitoring system, specifically designed to enhance the safety monitoring of working both at ground level and at elevated heights. This novel system integrates a suite of sophisticated technologies: instance segmentation, shape classification, object tracking, a visualization report, and a real-time notification module. Collectively, these components coalesce to deliver a safety monitoring solution, ensuring a higher standard of protection for construction workers. The experimental results…..

Key words: safety monitoring, construction, worker, detection, segmentation

1. INTRODUCTION

Vision safety monitoring represents a cutting-edge solution designed to mitigate accidents and fatalities on construction sites. Statistical evidence indicates that falls, slips, and trips are the second leading cause of fatalities in the construction industry [1]. Despite this, accurately identifying workers at heights, such as those on scaffolds or elevated areas, poses significant challenges. This difficulty arises primarily because personal protective equipment (PPE) like safety straps and hooks, essential for worker safety, are often hard to detect due to their small size and potential occlusion from view. Previous research has highlighted the promising capabilities of construction safety monitoring technologies [2,3,4,5,6].

For a safety monitoring system to be effective and deployable in real-world settings, it must comprise several components. This study introduces a comprehensive construction worker safety monitoring system that integrates three essential modules: recognition models, analytical algorithms, and notification & reporting mechanisms. The system employs a state-of-the-art speed-accuracy balanced YOLOv8 [7] instance segmentation model to accurately identify and delineate six specific categories: hardhats, straps, harnesses, hooks, height workers, and ground workers. This model facilitates the detailed analysis of each identified object. Additionally, an efficient vision transformer, FastViT [8] classification model, is utilized to ascertain whether safety straps are being used correctly. As detecting the existence of the PPE do not guarantee the proper usage, it is necessary to examine whether hardhats, harnesses, and straps are equipped and utilized. The analytical component of the system features the PPE Assignment & Connectivity Analysis [3] method. This method assigns detected PPE to the correct workers using bounding box data and verifies the proper equipment of PPE through segmentation mask information. Moreover, a specially customized multiple object tracking algorithm

for worker, hashing-supported cascaded buffered intersection over union (HC-BIoU) [9], is employed to improve worker identification and maintain a record of each worker's safety history, which is crucial for ongoing safety assessments. The system's notification and reporting module is designed to alert site supervisors immediately via email notifications and generate PDF reports post-monitoring sessions for detailed reviews.

To validate the accuracy and reliability of the instance segmentation and classification models, the training process incorporated images from the YUD-COSA-V2 [3] dataset and an additional collection of newly acquired images, totaling 9,847 and 1,658 images for training and validation without overlapping in construction scenes, respectively. The safety strap classification model was specifically trained on images categorized into 'used' and 'not used', to differentiate between the proper and improper use of safety straps. Training images and annotations for both the instance segmentation and classification models are shown in Fig. 1. Furthermore, to examine the system's overall effectiveness for safety assessment, two new clips were annotated to identify workers as safe or unsafe, with the first clip sampled at 10 frames per second (FPS) over 1,200 frames, the second clip sampled at 8 frames per second (FPS) over 1,000 frames, as demonstrated in Fig. 2.

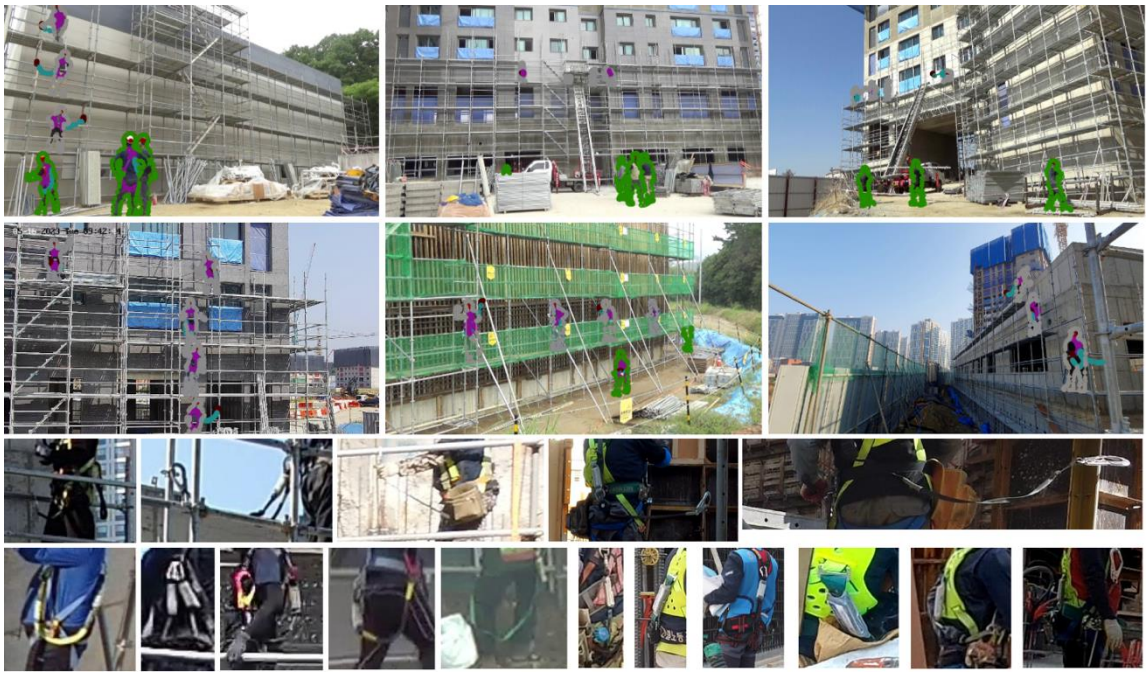

Figure 1. Visualization of annotations used for training instance segmentation and classification models. (1st & 2nd Rows) Polygon annotations of hardhat, strap, harness, hook, height worker (gray polygons), and ground worker (green polygons). (3rd Row) Strap "used" class instances. (4th Row) Strap "not-used" class instances.

Figure 2. Visualization of safety assessment testing images with bounding box annotations. Green boxes represent height-safe, red boxes represents heigh-unsafe, yellow boxes represent ground-unsafe, and blue boxes represents ground-safe. Some workers were annotated as occluded for being covered by other objects without proper vision to the workers.

2. METHODOLOGY

The proposed safety monitoring system is composed of several key components: a YOLOv8 instance segmentation model [7], a FastViT strap classification model [8], a HC-BIoU worker tracker [9], and modules dedicated to alarm logging and notification, as illustrated in Figure 3.

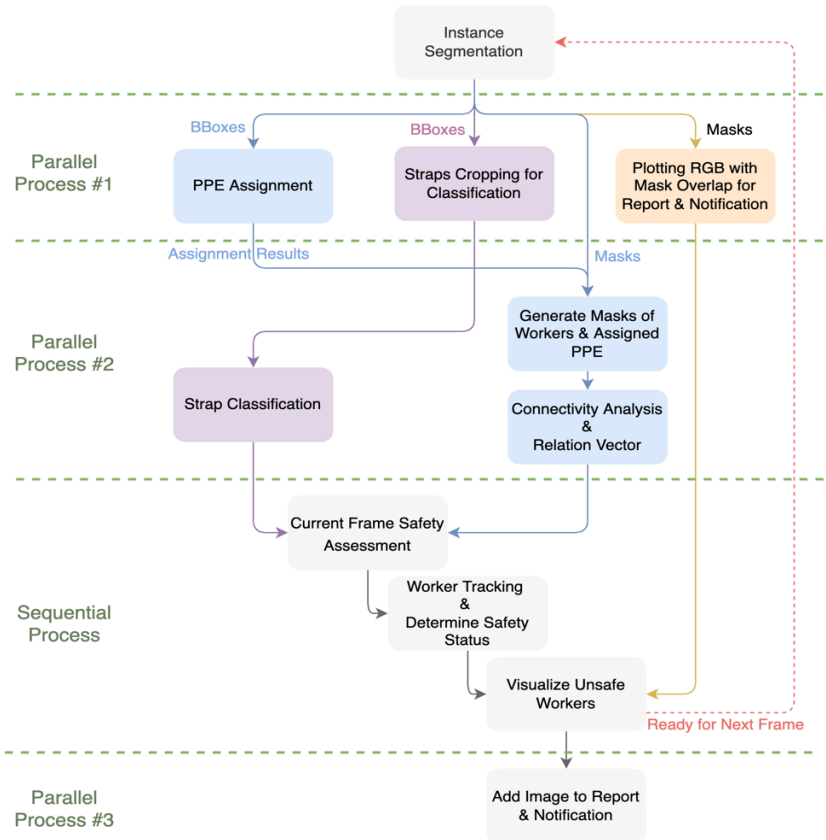

Figure 3. Visualization of the proposed monitoring system flowchart.

To ensure the selection of the most effective recognition models for this system, a meticulous process was followed. For the instance segmentation, the YOLOv8-Medium model was trained using a

learning rate of 0.001, the AdamW optimizer, and an input resolution of 1280x720. This training phase also incorporated various image data augmentation techniques, such as horizontal flipping, color jittering, and Copy-Paste [10], to enhance the model's ability to generalize across different scenarios. Similarly, the FastViT strap classification model underwent training with a learning rate of 0.0001, the AdamW optimizer, and an input resolution of 224x224. The training was further supplemented with data augmentation techniques including horizontal flipping, rotation, color jittering, MixUp [11], and random erase [12] to improve its performance.

Upon obtaining bounding boxes and segmentation masks for PPE and workers, the system employs PPE Assignment and Connectivity Analysis [3]. This analysis assigns PPE to workers based on a predefined formula, ensuring accurate tracking and safety compliance within the monitored environment. PPE Assignment is formulated as follow:

$$Overlap(p_i, w_j) = \frac{I(p_i, w_j)}{A(p_i)}, \quad (1)$$

where the function I represents the intersection between the bounding boxes of each detected PPE (p_i) and worker (w_j), the function A represents the area of the PPE. Each PPE will be assigned to a worker that shares the largest overlap with. Upon completing the initial assignment phase, the Connectivity Analysis process employs segmentation masks to meticulously evaluate each worker and their corresponding PPE, ensuring the correct usage. This involves linking the PPE to workers based on the detailed pixel-level masks provided by using the floodfill algorithm. The analysis then verifies the correct application of PPE for each worker, generating a relation vector for safety assessment in the current frame. This vector, equal in length to the number of target classes, uses boolean values to represent PPE usage status: '1' indicates proper use, while '0' signifies non-compliance.

Prior to determining the safety status of workers in the current frame, it is also imperative to conduct a thorough examination of the workers' safety straps through the strap classification model besides Connectivity Analysis. To achieve a more accurate classification, the bounding boxes around the straps are enlarged to encompass additional contextual information, aiding in the assessment of the strap's shape and its correct usage. This study noticed two primary configurations of safety straps: 'U' shapes and stretched formations. However, as depicted in Figure 4, the presence of a 'U' shape does not unequivocally indicate correct usage. Workers may, for convenience, attach one end of the strap to the safety harness at the back and the other to the chest area, a practice that does not follow the safety guideline. Given these challenges and the dynamic nature of worker movements, a specialized strap classification model has been trained by the diverse dataset to accurately assess the correct usage of safety straps, accounting for the nuances of various attachment methods and strap configurations.

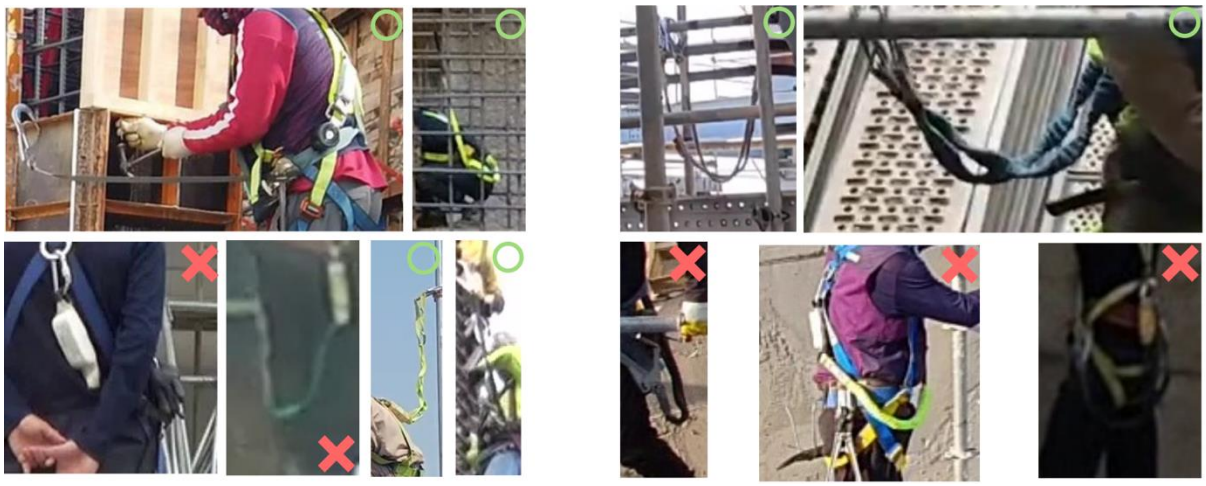

Figure 4. Example of U- and stretched shapes of safety straps. Green circles represents the strap in active use, while red crosses suggest the otherwise.

To enhance the reliability of worker tracking in scenarios where the instance segmentation model may falter, the HC-BIoU [9], a worker-specific tracking algorithm, has been integrated. This algorithm employs a strategy of bounding box matching for multiple object tracking, augmented by a color hashing technique for re-identification purposes. This feature is crucial for determining whether a

new worker instance has been previously identified or is yet to be recognized. Furthermore, the HC-BIoU tracker has been refined to monitor both the historical and current safety statuses of workers, thereby facilitating real-time assessments of their safety conditions. Figure 5 demonstrates the methodology employed by the monitoring system to ascertain a worker's safety status, utilizing a safety history vector with a depth of five as an example. By calculating the mode of this vector, the system can accurately gauge the worker's safety status, which is particularly beneficial in the complex environment of a construction site. This approach aids in minimizing the incidence of false alarms and noise in safety alerts, which can arise from the inherent challenges associated with classifying the correct use of safety straps, as highlighted in Figure 4.

Tracker's Worker Safety Histories						
Worker #1	Safe	Unsafe	Unsafe	Unsafe	Unsafe	Decision: Unsafe
Worker #2	Safe	Unsafe	Safe	Unsafe	Safe	Decision: Safe
Worker #3	Unsafe	Unsafe	Safe	Safe	Safe	Decision: Safe

Figure 5. Visualization of worker safety histories tracking within the worker tracker.

The system's reporting and notification module is designed to visually delineate safe and unsafe workers through the application of bounding boxes and masks. These visualized images are then compiled into a PDF report. Additionally, an email notification, featuring the visualized image, is dispatched to designated recipients as shown in Figure 6. This integrated approach ensures that relevant supervisors are promptly informed about the safety status of workers, enabling timely interventions when necessary.

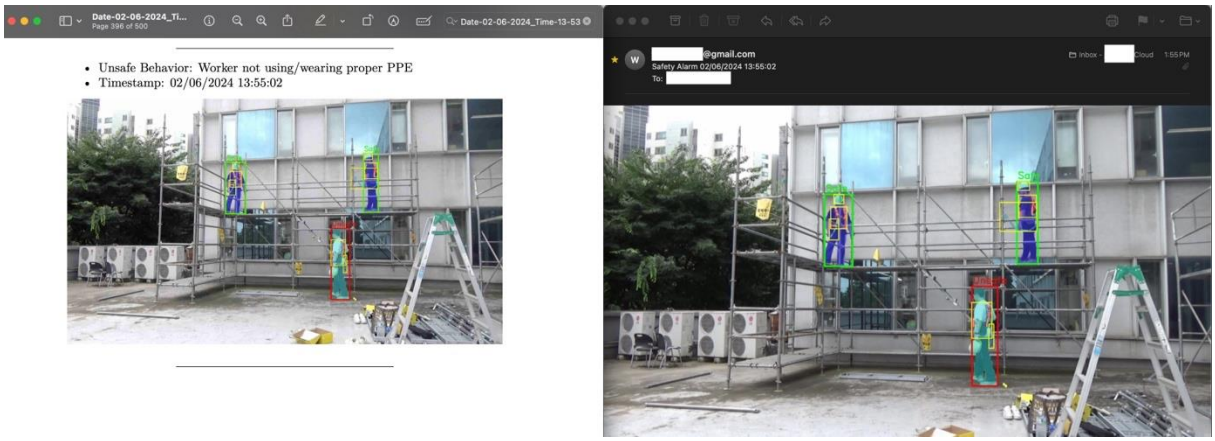

Figure 6. Visualization of PDF report (left) and E-Mail notification (right) for reviews and real-time alarming to site supervisors.

3. EXPERIMENTAL RESULTS

In assessing the performance of the YOLOv8-Medium instance segmentation model, it is noted that the model achieved mean average precision (mAP) scores of 72.0% and 65.5% for detection and segmentation, respectively, on the validation set as shown in Fig. 7. The performance for the FastViT strap classification model on the validation set are 82.32% and 79.20% in accuracy and F1 score, respectively.

The efficacy of the proposed context- and shape-aware safety monitoring system was evaluated through its safety assessment capabilities in the two testing scenes. Specifically, the system demonstrated an accuracy of 78.12% and an F1 score of 72.72% when analyzing the safety conditions of workers recognized in scene #1. In scene #2, the system showed enhanced performance, achieving an accuracy of 80.69% and an F1 score of 79.59%, as presented in Table 2. Collectively, the system's performance across the two testing scenes resulted in an average accuracy of 79.41% and an F1 score of 76.16%. Visualization of the safety monitoring results on the two testing scenes are shown in Fig. 8.

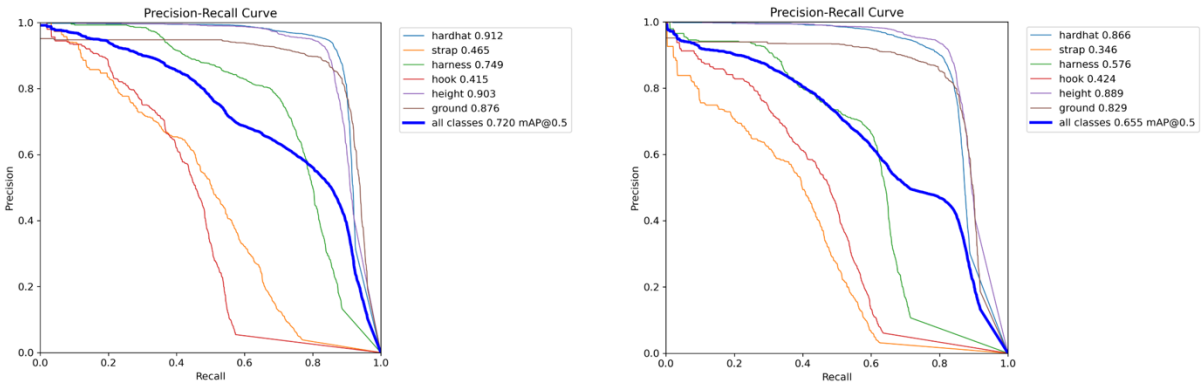

Figure 7. YOLOv8's mean average precision (mAP) curves for detection (left) and segmentation (right).

Table 1. Accuracy, and F1 score of the strap classification.

	Accuracy	F1
Scores	82.32%	79.20%

Table 2. Accuracy, and F1 score of the proposed monitoring system to the testing scenes.

	Scene #1	Scene #2
Accuracy	78.12%	80.69%
F1 Score	72.72%	79.59%
Avg. Accuracy	79.41%	
Avg. F1	76.16%	

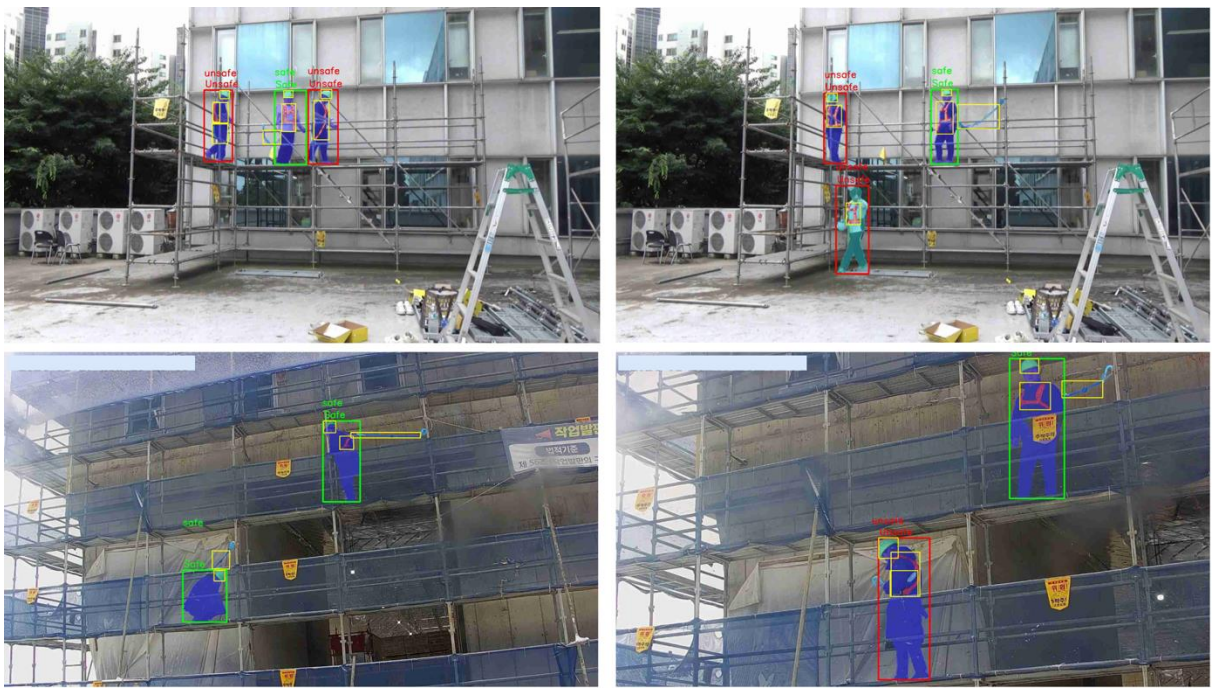

Figure 8. Visualization of safety monitoring results. The proposed system draw red bounding boxes around unsafe workers, green bounding boxes on safe workers, yellow bounding boxes indicates

recognized PPE. First and second column of the texts above the bounding boxes indicates ground truth, and predicted annotation respectively.

4. CONCLUSION

This research focuses on enhancing construction safety monitoring with an emphasis on fall protection by evaluating the proper usage of safety straps. It introduces a sophisticated monitoring system illustrated in Figure 3, which integrates recognition models, analytical algorithms, and a comprehensive notification and reporting framework. The experimental outcomes indicate an overall accuracy of 79.41%, and F1 score of 76.16%. Figures 1 and 4 demonstrate the challenge in the system's ability to consistently recognize the smaller PPE and classify the correct deployment of safety straps. The accuracy of strap classification is impeded by various challenges, such as adverse lighting conditions, the small size of straps, or occlusion caused by occlusion placement and interference from other construction materials.

To improve the recognition models within the monitoring framework, future research will explore the application of stable diffusion models [13,14], to augment the current dataset. This enhancement aims to refine the training performance by augmenting existing data in a realistic matter. The deployment of diffusion models could facilitate a range of applications, from text-to-image to image-to-video transformations, thereby improving the models' ability to generalize or achieve zero-shot inference across new and diverse construction environments. Moreover, the integration of Contrastive Language-Image Pre-Training (CLIP) [15] into the system promises to enhance zero-shot learning capabilities in classification tasks by leveraging textual data within the model architecture. By including the dataset expansion capabilities of stable diffusion models with the advanced classification potential of the CLIP framework, there is a significant opportunity to boost the efficacy and reliability of instance segmentation and classification models. This improvement is pivotal for developing a more robust and effective safety monitoring system for construction workers, ensuring higher safety standards and reducing the risk of accidents on construction sites.

ACKNOWLEDGEMENT

This research was conducted by the support of the “2023 Yonsei University Future-Leading Research Initiative (No. 2023-22-0114)” and the “National R&D Project for Smart Construction Technology (No. RS-2020-KA156488)” funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

REFERENCES

[1] Census of fatal occupational injuries summary. Technical Report USDL-22-2309, U.S. Bureau of Labor Statistics, 2022

[2] Nipun D. Nath, Amir H. Behzadan, and Stephanie G. Paal. Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction, 112:pp. 103085, April 2020. ISSN 09265805. doi: 10.1016/j.autcon.2020.103085.

[3] Wei-Chih Chern, Jeongho Hyeon, Tam V. Nguyen, Vijayan K. Asari, and Hongjo Kim. Context-aware safety assessment system for far-field monitoring. Automation in Construction, 149:pp. 104779, May 2023. ISSN 09265805. doi: 10.1016/j.autcon.2023.104779

[4] JackC.P. Cheng, Peter Kok-Yiu Wong, Han Luo, Mingzhu Wang, and Pak Him Leung. Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. Automation in Construction, 139:pp. 104312, July 2022. ISSN 09265805. doi: 10.1016/j.autcon.2022.104312

[5] Minsoo Park, Dai Quoc Tran, Jinyeong Bak, Seunghee Park. Small and overlapping worker detection at construction sites. Automation in Construction, 151:pp. 104856, July 2023. ISSN 09265805. doi: 10.1016/j.autcon.2023.104856

[6] Siyeon Kim, Seok Hwan Hong, Hyodong Kim, Meesung Lee, Sungjoo Hwang. Small object detection (SOD) system for comprehensive construction site safety monitoring. Automation in Construction, 156:pp. 105103, December 2023, ISSN 09265805. doi: 10.1016/j.autcon.2023.105103

[7] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, January 2024. URL https: //github.com/ultralytics/ultralytics.

[8] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, Anurag Ranjan. FastViT: A Fast Hybrid Vision Transformer Using Structural Reparameterization. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2023, pp. 5785-5795

[9] Wei-Chih Chern, Vijayan Asari, Hongjo Kim. Hashing-Based Object Tracking for Construction Site Safety Monitoring across Different Domains. ASCE International Conference on Computing in Civil Engineering, June 2023. pp. 500-507

[10] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation.

CoRR, abs/2012.07177, 2020. doi: https://doi.org/10.48550/arXiv.2012.07177.

[11] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. CoRR, 2017. http://arxiv.org/abs/1710.09412

[12] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang. Random Erasing Data Augmentation. CoRR, 2017. http://arxiv.org/abs/1708.04896

[13] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 10684-10695

[14] Omer Bar-Tal, Hila Chefer, Omer Tov, Charles Herrmann, Roni Paiss, Shiran Zada, Ariel Ephrat, Junhwa Hur, Guanghui Liu, Amit Raj, Yuanzhen Li, Michael Rubinstein, Tomer Michaeli, Oliver Wang, Deqing Sun, Tali Dekel, Inbar Mosseri. Lumiere: A Space-Time Diffusion Model for Video Generation. arXiv, 2024.

[15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. CoRR, 2021. https://arxiv.org/abs/2103.00020