Option 2

ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.1307

The 10th International Conference on Construction Engineering and Project Management

Jul. 29- Aug. 1, 2024, Sapporo

Optimization of 3D ResNet Depth for Domain Adaptation in Excavator Activity Recognition

Seungwon SEO[1] and Choongwan KOO[2]*

[1] *Department of Architecture, Incheon National University, Republic of Korea,* E-mail address: ssw9541@inu.ac.kr

[2] *Division of Architecture & Urban Design, Incheon National University, Republic of Korea,* E-mail address: cwkoo@inu.ac.kr

Abstract

Recent research on heavy equipment has been conducted for the purposes of enhanced safety, productivity improvement, and carbon neutrality at construction sites. A sensor-based approach is being explored to monitor the location and movements of heavy equipment in real time. However, it poses significant challenges in terms of time and cost as multiple sensors should be installed on numerous heavy equipment at construction sites. In addition, there is a limitation in identifying the collaboration or interference between two or more heavy equipment. In light of this, a vision-based deep learning approach is being actively conducted to effectively respond to various working conditions and dynamic environments. To enhance the performance of a vision-based activity recognition model, it is essential to secure a sufficient amount of training datasets (i.e., video datasets collected from actual construction sites). However, due to safety and security issues at construction sites, there are limitations in adequately collecting training dataset under various situations and environmental conditions. In addition, the videos feature a sequence of multiple activities of heavy equipment, making it challenging to clearly distinguish the boundaries between preceding and subsequent activities.

To address these challenges, this study proposed a domain adaptation in vision-based transfer learning for automated excavator activity recognition utilizing 3D ResNet (residual deep neural network). Particularly, this study aimed to identify the optimal depth of 3D ResNet (i.e., the number of layers of the feature extractor) suitable for domain adaptation via fine-tuning process. To achieve this, this study sought to evaluate the activity recognition performance of five 3D ResNet models with 18, 34, 50, 101, and 152 layers, which used two consecutive videos with multiple activities (5 mins, 33 secs and 10 mins, 6 secs) collected from actual construction sites. First, pretrained weights from large-scale datasets (i.e., *Kinetic-700* and *Moment in Time (MiT)*) in other domains (e.g., humans, animals, natural phenomena) were utilized. Second, five 3D ResNet models were fine-tuned using a customized dataset (14,185 clips, 60,606 secs). As an evaluation index for activity recognition model, the F1 score showed 0.881, 0.689, 0.74, 0.684, and 0.569 for the five 3D ResNet models, with the 18-layer model performing the best. This result indicated that the activity recognition models with fewer layers could be advantageous in deriving the optimal weights for the target domain (i.e., excavator activities) when fine-tuning with a limited dataset. Consequently, this study identified the optimal depth of 3D ResNet that can maintain a reliable performance in dynamic and complex construction sites, even with a limited dataset. The proposed approach is expected to contribute to the development of decision-support systems capable of systematically managing enhanced safety, productivity improvement, and carbon neutrality in the construction industry.

Key words: excavator, activity recognition, domain adaptation, 3D ResNet, transfer learning, fine-tuning