|  |  |  |  |
| --- | --- | --- | --- |
| Title | A YOLOv11-Based Deep Learning Framework for Multi-Class Human Action Recognition | | |
| Author(s) Name | Nayeemul Islam Nayeem, Shirin Mahbuba, Sanjida Islam Disha, Md Rifat Hossain Buiyan, Shakila Rahman, M. Abdullah-Al-Wadud, Jia Uddin | | |
| Contact Email(s) | Shakila.rahman@aiub.edu | | |
| Published Journal Name | Computers, Materials & Continua | | |
| Type of Publication | Article | | |
| Volume | 85 | Issue | 1 |
| Publisher | Tech Science Press | | |
| Publication Date | 29 August 2025 | | |
| ISSN | ISSN:1546-2226 | | |
| DOI | https://doi.org/10.32604/cmc.2025.065061 | | |
| URL | https://www.techscience.com/cmc/v85n1/63515 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Human activity recognition is a significant area of research in artificial intelligence for surveillance, healthcare, sports, and human-computer interaction applications. The article benchmarks the performance of You Only Look Once version 11-based (YOLOv11-based) architecture for multi-class human activity recognition. The article benchmarks the performance of You Only Look Once version 11-based (YOLOv11-based) architecture for multi-class human activity recognition. The dataset consists of 14,186 images across 19 activity classes, from dynamic activities such as running and swimming to static activities such as sitting and sleeping. Preprocessing included resizing all images to 512 512 pixels, annotating them in YOLO’s bounding box format, and applying data augmentation methods such as flipping, rotation, and cropping to enhance model generalization. The proposed model was trained for 100 epochs with adaptive learning rate methods and hyperparameter optimization for performance improvement, with a mAP@0.5 of 74.93% and a mAP@0.5-0.95 of 64.11%, outperforming previous versions of YOLO (v10, v9, and v8) and general-purpose architectures like ResNet50 and EfficientNet. It exhibited improved precision and recall for all activity classes with high precision values of 0.76 for running, 0.79 for swimming, 0.80 for sitting, and 0.81 for sleeping, and was tested for real-time deployment with an inference time of 8.9 ms per image, being computationally light. Proposed YOLOv11’s improvements are attributed to architectural advancements like a more complex feature extraction process, better attention modules, and an anchor-free detection mechanism. While YOLOv10 was extremely stable in static activity recognition, YOLOv9 performed well in dynamic environments but suffered from overfitting, and YOLOv8, while being a decent baseline, failed to differentiate between overlapping static activities. The experimental results determine proposed YOLOv11 to be the most appropriate model, providing an ideal balance between accuracy, computational efficiency, and robustness for real-world deployment. Nevertheless, there exist certain issues to be addressed, particularly in discriminating against visually similar activities and the use of publicly available datasets. Future research will entail the inclusion of 3D data and multimodal sensor inputs, such as depth and motion information, for enhancing recognition accuracy and generalizability to challenging real-world environments. | |