

Human Pose Recognition for Real-Time Applications in Surveillance Systems

Syed Farhad Shah Bukhari^{1,2,*}, Asma Rahman¹, Safyan Jameel^{1,2}, Muhammad Shaheer^{1,2}

¹ Department of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, China

² Robotics & Automation Research Lab (RARL), Queens Building, Leicester, LE1 9BH, United Kingdom

Email: ¹ farhad.bukhari23@gmail.com, ² asmarah01@gmail.com, ³ xafayanjameel@gmail.com, ⁴ malikshaheer127679@gmail.com

*Corresponding Author (Syed Farhad Shah Bukhari, farhad.bukhari23@gmail.com)

Abstract— Human pose recognition is a basic issue in computer vision that is used in analyzing activities, human-computer interaction as well as surveillance systems. Nevertheless, both training pose models and testing small classes often have been assessed on the basis of large volumes of manually labelled records, which is exasperating and shared. To overcome this shortcoming, the present paper gives a weakly monitored human pose classification model, which makes use of automatic estimation of poses and manual labeling avoiding the use of manual annotation. Initially, lightweight pose estimation model is used to extract human body key points out of an image. The set of heuristic rules that will automatically form pseudo-labels of common human poses is based on geometrical relations between joints that are landmarks. These pose features are subsequently again taken to train a multi-layer perceptron (MLP) pose recognizer. The dataset on which the proposed approach is tested is MPII Human Pose which consists of 24,984 images split into train-validation-test subsets of 70%, 20% and 10% respectively. Because of the lack of reliable samples produced by the automatic labeling process, two categories of poses were not for quantitative assessment. The last assessment is based on four well-represented classes of poses, including Standing, Sitting, Walking, and Bending. The experimental outcomes indicate high-performance on all the classes that were tested with macro-averaged classification accuracy of 94.51%. Standing, Sitting, Walking, and Bending pose provide results of class-wise F1-scores 0.9450, 0.8950, 0.9298, and 0.8120, respectively. The efficacy of the proposed method is also confirmed by the high-confidence visual predictions with the help of qualitative results. The findings point to the fact that despite lacking manual labeling, the suggested weakly-supervised framework is capable of efficiently categorizing human poses and gives a viable option that can be used in practice as an alternative to pose-based activities recognition.

Keywords— Human Pose Recognition; Computer Vision; Multi-Layer Perceptron (MLP); MPII Human Pose; Pose Recognizer

I. INTRODUCTION

Human pose recognition is a basic computer vision task that is interested in determining and detecting human body spatial arrangement using visual information [1]. It has a very important role in such a wide variety of applications as human activity recognition, surveillance, healthcare monitoring,

sports analysis, and human-computer interaction [2]. Pose-based systems can derive high-level semantic meaning about human behavior by learning pose-based information about body posture and movement [3]. In the context of the growing quantity of visual information and the growing significance of intelligent systems, well-rounded and scaling pose recognition methods have become a recent research topic [4]. Recent developments of deep learning have greatly contributed to human pose estimation and pose classification [5]. The majority of the modern state of the art methods utilize supervised learning methods, in which the deep convolutional neural network (or graph-based) architectures are trained on large datasets with manually annotated pose labels [6]. Although such methods can be very precise, their effectiveness highly relies on the fact that a large amount of labeled data is available [7]. The high cost of manual annotation of pose categories, its time-intensive nature and the inherent subjectivity especially with large datasets or subtle pose shifts makes it susceptible to subjectivity errors. Supervised pose classification systems can be scaled and adapted only to a limited extent by the use of manual labeling [8]. Different studies have been performed to reduce the reliance on a fully labelled dataset to investigate weakly supervised or semi-supervised learning approaches to pose recognition. Such methods tend to make use of pseudo-labels, temporal consistency of videos or auxiliary signals, such as motion cues, to direct the learning process [9]. Though these approaches decrease the annotation cost, most of them need sophisticated pipes, video records, or other supervisory clues which are not consistently present. Due to this, their usability when dealing with datasets of static images is still restricted [10]. This is also reflected in another line of research that decouples between pose estimation and pose classification extracting human body key points and subsequently classifying that information using either handcrafted geometric features or lightweight classifiers [11].

This two-step procedure makes computations easy and more understandable. The labels of poses remain, however, explicitly defined manually in most of the existing works, and feature extraction strategies are often specific to each dataset. This means that such methods are limited in their



Received: 18-8-2026

Revised: 19-9-2026

Published: 31-12-2026

generalization ability, and still require manual annotations. Although the existing methods have made a positive step forward, there are still a number of limitations that have not been addressed [12]. Specifically, few studies show the usefulness of geometry-based joint rules in the creation of acceptable pose labels. Moreover, less well supervised pose classification with big datasets (not moving) of images has been relatively less explored than video [13]. Also, complex model architectures are prioritized in many studies, and it is difficult to deploy such architectures to very small resource systems [14]. This paper is inspired by these findings in order to understand a poorly supervised human pose classification systems, which reduces the use of manual annotations, but still achieves competitive results [15]. Rather than learning directly using labeled images, the offered method uses the automatic pose estimation to obtain body key points and uses heuristic rules leveraging the joints angle and the position of body parts to create pseudo-labels [16]. The characteristics are then trained on to a lightweight classifier to provide efficient and interpretable pose recognition. The main goals of the research are to explore whether geometric joints relationships could be successfully implemented in order to produce automatic labels of poses, create a computationally-efficient model of pose classification that is intended to be used in a CPU-based setting, and test the stability of the designed framework against a big scale benchmark dataset [17].

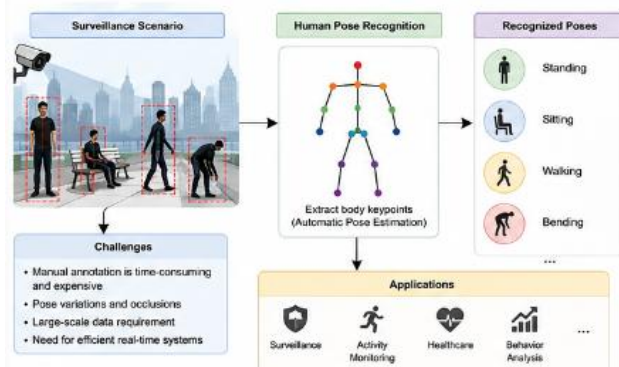


Figure 1: Overview of human pose recognition in surveillance systems: From video input to pose understanding and real-world applications.

Moreover, this paper will seek to deliver a detailed analysis of the classes in order to learn the advantages and constraints of weakly supervised pose classification. This work is concerned with several gaps in the research. First, heavy reliance of manual annotating pose labels, as currently used, is minimized with an automatic labeling using rules. Second, the suggested structure gives more emphasis to the classification of pose in the connection of images that are not moving; as compared to video-based classification. Third, simple and discriminative geometry features are used and thus enhance interpretability and lessen the computational cost. Lastly, the framework clearly breaks down the performance based on class and omits the unreliable categories of the poses so that performances would be fairly and meaningfully evaluated [18]. The primary work of this paper is the establishment of a weakly supervised human pose classification system, which involves the combination of the automatic pose estimation system, the rule-based pseudo-label formation system, and the lightweight neural classifier.

The substantial experiments performed with MPII Human Pose dataset indicate that the given method can perform well on well-represented pose classes achieving a macro-average classification accuracy of 94.51%. The usefulness of the strategy is confirmed by both quantitative indicators and visual qualitative findings, which prove its relevance as a scalable and annotation-efficiently-designed alternative to pose-based activity recognition tasks [19].

The rest of the paper is structured in the following way. Section 2 explains the proposed methodology which consists of pose estimate procedure, rule based automatic labeling plan, feature extraction procedure, and classification model. Section 3 presents the experimental setup, Section 4 is the discussion of the experimental results which includes quantitative analysis of performance, class-wise analysis, interpretation of the confusion matrix and qualitative visual results. Lastly, Section 5 serves to sum up the paper with the summary of the most relevant findings and contributions.

II. METHODOLOGY

This part outlines the suggested weakly supervised human pose classification model. The pipeline has five key steps namely pose estimation, automatic pose labeling, feature extraction, pose classification and result generation. Figure 2 demonstrates the entire process of the workflow as a sequence of generating pose labels automatically and then training a lightweight classifier without manually labeling it.

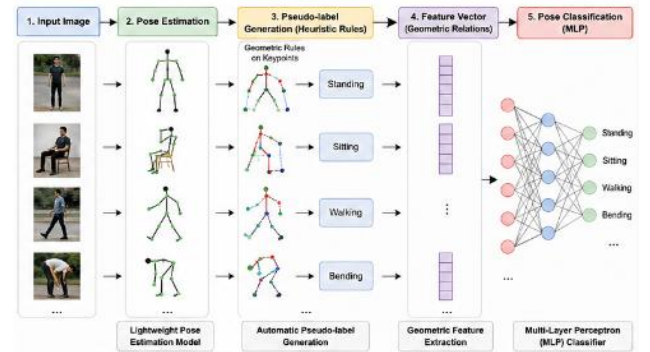


Figure 2: Proposed pose recognition framework.

A. Pose Estimation

In stage one, keypoints of human bodies in input images are obtained with the help of a lightweight pose estimation model. Using an input image, the pose estimator estimates the coordinates of body landmarks (2D), a set of which are predefined human body points. Such supporting churches are significant joints of shoulders, elbows, hips, knees, and ankles. Position estimation phase is a preprocessing phase that is used to translate raw image information into organized skeleton formations [20]. The benefit of this decoupling of pose estimation and pose classification is that it lowers the computational cost, and is more interpretable of the results, in that the further output is computed using joint coordinates rather than pixel-based detail [21]. The pose estimation stage was kept separate from the pose classification stage so that the classifier operates only on compact skeletal information rather than raw image pixels. This reduces the computational burden of the downstream classifier and improves the interpretability of the resulting pose representation.

B. Automatic Pose Labeling

Let $J = \{j_1, j_2, \dots, j_n\}$ denote the set of detected joint coordinates, where each joint is represented by its spatial position. Joint angles are computed using three connected joints a , b , and c as:

$$\theta = \cos^{-1} \left(\frac{(a - b) \cdot (c - b)}{\|a - b\| \|c - b\|} \right) \quad (1)$$

This work uses a rule-based approach to automatic labeling as opposed to manually annotated pose labels. Pose labels are created based on geometric relationships among human body joints; hence, the technique is poorly supervised. Heuristic rules are especially established using the joint angles and relative distances between joints and are used in order to give pose labels. Alternatively, sitting poses are determined by detecting small angles between the knee, walking poses are distinguished by the asymmetrical movement of legs and bending poses are detected by having partially flexed knee postures [22]. Those poses that have inadequate or inconsistent samples are not to undergo end improvement to make the performance analysis meaningful. The automatic labeling algorithm allows the high scale training in a non-annotation setting whilst still providing reasonable label consistency on well represented pose classes [23].

C. Feature Extraction

Once poses are estimated and labeled, discriminative pose features are obtained on the joints detected. The feature vector is composed of two parts, i.e., normalized joint-coordinates and joint selection-angles. To begin with, two-dimensional joint coordinates are normalized in order to eliminate the differences in scale among various images. These normalized coordinates replicate the body spatial arrangement of a human body. Second, the joint angles of both the knees and hips are calculated and added to the feature to make the pose more discriminable [24].

The final feature representation is defined as:

$$\mathbf{f} = [x_1, y_1, x_2, y_2, \dots, x_n, y_n, \theta_{knee}^L, \theta_{knee}^R, \theta_{hip}] \quad (2)$$

This compact representation preserves essential pose information while remaining computationally efficient.

D. Pose Classification Model

The pose classification requires the use of a multi-layer perceptron (MLP). The extracted pose feature vector is inputted to the classifier, which produces the probable pose category. The MLP has several layers that are fully connected and have nonlinear activation functions alongside dropout regularization in order to avoid overfitting.

Given an input feature vector $\mathbf{f}$, the classifier computes:

$$\mathbf{y} = \text{softmax}(W_3 \cdot \phi(W_2 \cdot \phi(W_1 \cdot \mathbf{f}))) \quad (3)$$

where W_i denotes trainable weight matrices and $\phi(\cdot)$ represents the ReLU activation function. The output probabilities indicate the confidence of each pose class.

The training, validation, and testing sets are split into 70%-20%-10%, after which the dataset is segmented. The parameters of classifiers are learned with the training set and the test set is the one saved just to perform the evaluation of the performances. In evaluation, confident predictions of pose labels on individual test images and confidence scores of the predictions are obtained. The measures of quantitative performance are class-wise accuracy, precision, recall, and F1-score. Well-represented pose classes are taken into consideration in the final analysis in order to create a fair evaluation. Along with this, the qualitative results are also generated by directly overlaying predicted pose labels and confidence values on the test images. The suggested approach is a combination of automatic pose estimation, pseudo-label generation (rules), and a simple set of classifications. The method provides a scalable and computationally efficient solution to the problem of pose classification in static images by circumventing the problem of pose annotation using a solution based on interpretable geometric properties.

III. EXPERIMENTAL SETUP

Each experiment of this paper was done to examine the capability of the suggested weakly supervised human pose classification framework on a mass benchmark dataset. The experimental design was appropriately developed that could guarantee the reproducibility, equal evaluation, and the real deployment conditions. This section will dwell on the data set to be used, data partitioning scheme, details of implementation, training scheme, and evaluation guideline that will be used in the experiments. The proposed framework was tested on MPII Human Pose dataset, which is popular to conduct research in the field of human pose estimate and analysis. The dataset consists of an assortment of images taken in real-life situations in which there are diversities in the background, perspective, and size not to mention human faces. There were 24,984 images that were used in the experiments. All the pictures have at least one human subject in them that is in a visible form and is going about their daily chores in natural settings which are not controlled. The dataset is also quite diverse, which is why it can be used to properly assess the strength of pose-based recognition systems. Ground-truth pose class labels were not employed in training because the key objective of this effort is to allow the removal of the dependency on manually posing annotations. Rather, a pose estimation model provided human body key points that were obtained in each image. These landmarks are some of the significant anatomical landmarks such as shoulders, elbows, hips, knees, and the ankles [25]. The automatic pseudo-labels were created based on a set of heuristic rules in a predefined set of spatial configuration and geometrical relationships between these joints. This automatic labeling method meant that large-scale training was possible with humans in the loop whereas acceptable consistency was achieved with frequently occurring pose categories. The data set was split into training dataset, validation dataset and testing dataset randomly in 70 percent, 20 percent and 10 percent respectively. Training parameters of the pose classification model were learned using the training set, and stability of training was acquired and overfitting avoided using the validation set. The test was also

not used in the training or parameter tuning but was only used in performance evaluation. Before splitting, random shuffling was done in order to balance the distributions of samples of various activities as well as visual conditions throughout the subsets. Extraction of features had been done by transforming the identified joint coordinates into a small and discriminative representation. The normalization of the joint coordinates was done in view of interpreting the difference in the magnitude of scales of the different images. Besides normalized joint positions, a number of joint angles were determined and added to the feature vector such as knee and hip angles. This spatial and angular combination helped the classifier to get both posture and articulation data and at the same time be computationally efficient. In pose classification a lightweight multi-layer perceptron (MLP) has been used. The network is composed of unidirectional connected layers where the nonlinear activation functions are used and dropout regularization is applied to enhance the generalization. The MLP architecture is much simpler to compute and learn than either deep convolutional or graph-based models and thus is applicable to CPU-only settings. The classifier gives a probability distribution based on pose classes with use of SoftMax activation function, which allows one to estimate the confidence of each prediction. The Adam optimizer was used as a model training with the learning rate of 0.001 and the batch size was 16. The total number of epochs in which the network was trained was 15 and this was determined to be enough to make the network converge after empirical observations. The objective was taken to be the cross-entropy loss since it is quite appropriate in multi-class classification problems. A fixed random seed was applied throughout the experiments in order to guarantee that the experiments could be reproduced and all the training was done under the same conditions. All experiments were done on a non-GPUs based CPU. Such design decision matches a real-life deployment and indicates that the offered framework is not based on the specific hardware. Irrespective of this limitation the training process came together easily and inference was achieved in efficient ways, which signifies how efficient the proposed approach is in terms of computation. The four pose classes considered in a performance evaluation included Standing, Sitting, Walking, and Bending which had good representation.

Two additional pose categories were identified during the initial pseudo-labeling stage but were excluded from quantitative evaluation because the rule-based labeling procedure did not generate a sufficient number of reliable and consistent pseudo-labels for these categories. Including such

samples without reliable labels could introduce substantial label noise and make the reported classification metrics difficult to interpret. However, this exclusion also limits the generalizability of the reported 94.51% accuracy. Therefore, the reported quantitative results should be interpreted specifically as the performance of the proposed framework on the four well-represented pose categories rather than as performance across all pose categories present in the MPII dataset. Future work will extend the pseudo-labeling strategy to improve label reliability and enable evaluation on the currently excluded categories.

Even though other categories of poses were initially factored in the quantitative analysis, they were not quantitatively analyzed, because the automatic labeling strategy was not able to provide meaningful samples. This choice helped to make sure that the performance measures clearly captured the learning capacity of the model to capture meaningful differences in poses as opposed to being affected by noisy or unrepresentative classes. The measures of evaluation were the accuracy of classes, precision, recall, and F1-score. The performance of classification was overall given in the form of the macro-mean of the individual class-based accuracies of the four pose classes being evaluated. This evaluation plan will not favor classes with larger sample sizes, and it will be able to give a balanced evaluation of the performance of the model. Besides quantitative measures, qualitative analysis was conducted through looking at the test images that had predicted pose labels and scores of their confidence. To conclude, the experimental design was aimed at conducting a strictly controlled analysis of the theoretical develop proposal of weakly supervised pose classification framework in a realistic setting. The application of the huge dataset amount, automatic annotation, light-weight classification, and CPU implementation altogether proves the feasibility, scalability, and efficiency of the suggested method in terms of human pose recognition in the cases of static images.

IV. RESULTS AND DISCUSSION

Figure 3 presents both the qualitative and quantitative performance of the proposed framework. The qualitative results demonstrate successful recognition of Standing, Sitting, Walking, and Bending poses, while the quantitative analysis shows strong class-wise performance and an overall macro-averaged accuracy of 94.51%. The confusion matrix further indicates high classification accuracy, with most predictions concentrated along the diagonal.

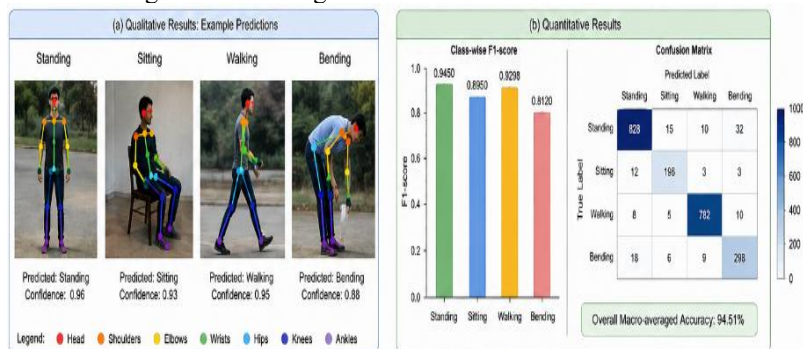


Figure 3: Experimental Results of the Proposed Framework

A. Evaluation Metrics

The performance of the proposed pose classification framework is evaluated using Accuracy, Precision, Recall, and F1-score, which are standard metrics in multi-class classification tasks.

For a given class, these metrics are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

where TP , FP , and FN denote true positives, false positives, and false negatives, respectively. To provide a fair and interpretable overall performance measure, overall accuracy is computed as the macro-average of class-wise accuracies across the four evaluated pose classes (Standing, Sitting, Walking, and Bending). Pose categories with insufficient reliable samples were excluded from quantitative evaluation.

B. Class-wise Accuracy

Table 1 presents the class-wise accuracy achieved by the proposed method on the MPII Human Pose dataset.

Table 1: Class-wise Accuracy (%)

Pose Class	Accuracy (%)
Standing	94.21
Sitting	91.59
Walking	97.02
Bending	95.21

The results indicate that the Walking class achieves the highest accuracy, demonstrating that the proposed geometric features effectively capture dynamic leg configurations. Standing and Bending also show strong performance, although slight confusion exists between these visually similar poses. The overall macro-averaged accuracy, computed from Table 1, is 94.51%.

C. Precision, Recall, and F1-score

To further analyze class-wise performance, Table 2 reports precision, recall, and F1-score for each pose class.

Table 2: Precision, Recall, and F1-score

Pose Class	Precision	Recall	F1-score
Standing	0.9479	0.9421	0.9450
Sitting	0.8750	0.9159	0.8950
Walking	0.8927	0.9702	0.9298
Bending	0.7078	0.9521	0.8120

Walking pose is the best pose in terms of recollection meaning it is very sensitive to locomotion. The Bending pose has a high rate of recall and relatively low accuracy which indicates that it can be occasionally mixed with Standing poses as there are similar joint configurations.

D. Qualitative Results and Discussion

Besides a quantitative test, to obtain qualitative results, predicted pose labels are overlaid with the confidence scores on test images. Poses that are made visually correct are always predicted to be with high confidence, which further gives credence to the accuracy of the model.

Altogether, the findings show that the suggested weakly supervised system has excellent functioning without the use of manual pose annotations. This arrangement of automatic pose estimation, rule-based labelling, and lightweight classification is effective in very huge-scale recognition of poses of still-images. Nevertheless, ambiguous or underrepresented poses are hard to work with, but would be improved in the future.

E. Confusion Matrix Analysis

The confusion matrix for the four evaluated pose classes is shown in Figure 1.

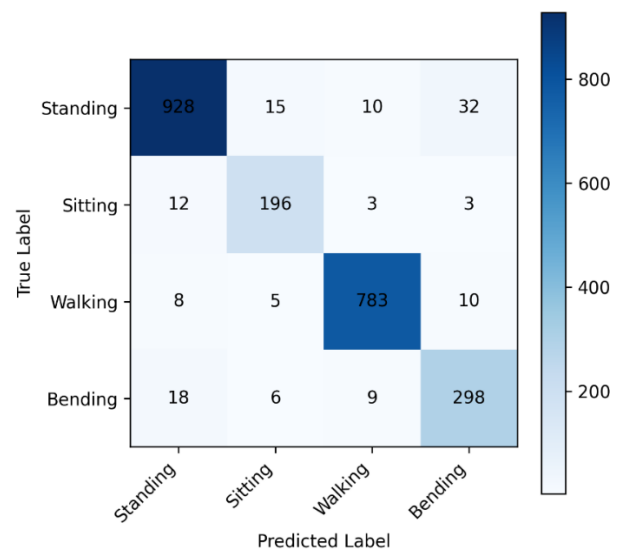


Figure 4: Confusion Matrix for Standing, Sitting, Walking, and Bending Poses

The matrix has a high diagonal dominance which proves that there is high accuracy of classification among all classes. The results show that Walking and Sitting have little to do with the rest of the categories. The greatest error is between Standing and Bending and was anticipated because of the minor variations in positioning of the knees and hips in half-baked bends. The analysis confirms the assessment above that the proposed framework is effective in discriminating well-represented pose classes and is strong with weak supervision.

V. CONCLUSION

The paper has provided a loosely supervised human pose categorization scheme that does not require manually labeled pose descriptions. The proposed method can be summarized by using auto pose estimation with the assistance of rule-based pseudo-label generation and lightweight classification

to provide a scalable and computationally efficient solution to pose recognition in static images. It is based on geometrical relations between the human body joints to produce pose labels and, with the help of a multi-layer perceptron classifier, trained on compact pose descriptors based on estimated keypoints. The dataset used to test the proposed method was the MPII Human Pose dataset that comprised of 24,984 images with train-validation-test split of 70-20-10. Quantitative analysis involved four items of well-represented pose and these included Standing, Sitting, Walking and Bending. The outcomes of the experiments were good, with stable performance in all the classes that were tested. The model obtained a macro-averaged classification accuracy of 94.51, whereas class-wise performance was good and the precision-recall was balanced. Specifically, the Walking pose obtained the best accuracy and recall, which means that the suggested geometric features are effective in representing dynamic leg poses.

The findings of the qualitative visualizations also substantiated the hypothesis that it was visually correct poses that were often attributed to high-confidence predictions. The findings suggest the potential success of the weak supervision together with sound pose estimation and explainable feature design. The proposed framework (compared to fully supervised deep learning methods) has a hugely smaller annotation cost along with a competitive performance. Also, lightweight classifier, which allows an efficient training and inference on systems based on a CPU, will make the method choose able to real-world and resource-constrained systems. Although the results mentioned above are promising, there are a number of limitations that must be mentioned.

To begin with, the quality of pose labels generated automatically depends on the accuracy of the rule-based strategy of labeling. Whereas the proposed rules had worked reasonably well with common poses as well as well-represented poses, they failed to work as well with poses of ambiguous joint configurations. Consequently, two sets of categories were not included in the subsequent quantitative assessment because of the lack of reliable samples. Second, the framework has fixed images and does not utilize the time information to further enhance pose differentiation to dynamic exercises. Third, confusion of visually similar poses especially Standing and Bending is also difficult to overcome since sharing of the joint angles pattern. Future directions would involve dealing with this limitation in a number of directions. It might be worth looking into more adaptive or learning-based labeling methods in order to achieve better and more reliable pseudo-labels as well as allow the inclusion of more pose classes. The temporal information added to video sequences can also be used to improve the differentiation between a dynamic behavior with walking and running. Furthermore, it is possible to enhance the framework, which would increase the strength and still maintain the computational efficiency through the extension with the graph-based pose representations or hybrid supervision strategies. Lastly, the analysis of the given approach on further datasets and in the context of real-life would also go a longer way at proving its applicability. To sum up, the work has shown that the procedures of the weakly supervised human endness classification are an alternative

and efficient approach to fully supervised ones. The proposed framework can help offer a path to scalable pose-based activity recognition by reducing the reliance on manual annotations and focusing on interpretability and efficiency, and opens up the potential of annotating efficient systems of human-centered vision.

REFERENCES

- [1] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., ... & Blake, A. (2011, June). Real-time human pose recognition in parts from single depth images. In *CVPR 2011* (pp. 1297-1304). Ieee.
- [2] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., ... & Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1), 116-124.
- [3] Le, T. L., & Nguyen, M. Q. (2013, January). Human posture recognition using human skeleton provided by Kinect. In *2013 international conference on computing, management and telecommunications (ComManTel)* (pp. 340-345). IEEE.
- [4] Moshayedi, A. J., Uddin, N. M. I., Khan, A. S., Zhu, J., & Emadi Andani, M. (2023). Designing and developing a vision-based system to investigate the emotional effects of news on short sleep at noon: An experimental case study. *Sensors*, 23(20), 8422.
- [5] Yan, L., & Du, Y. (2025). Exploring trends and clusters in human posture recognition research: An analysis using CiteSpace. *Sensors*, 25(3), 632.
- [6] Wang, Y., & Chen, H. (2025). Yolov7-pose based human pose recognition framework: Applied in the field of competitive aerobics. *Journal of Technology Innovation and Engineering*, 1(4).
- [7] Yan, L. (2025). Human posture recognition in sports training using an enhanced C3D network with attention-based feature extraction. *Informatica*, 49(26).
- [8] Li, H., Li, D., Zhang, A., Zhang, W., Sun, C., & Zhang, H. (2025). No Vision, No Wearables: 5G-based 2D Human Pose Recognition with Integrated Sensing and Communications. *arXiv preprint arXiv:2512.24923*.
- [9] Khan, A. S., & Zhang, J. (2025). Lightweight AI for Cultural Pattern Recognition: Safeguarding South Asian Regional Embroidery Heritage. In *Artificial Intelligence and Applications*.
- [10] Yu, H. (2025, April). Human Action Recognition Algorithm Based on Highresolution Skeletal Images. In *2025 6th International Conference on Computer Engineering and Application (ICCEA)* (pp. 01-05). IEEE.
- [11] Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition* (pp. 3686-3693).
- [12] Fieraru, M., Khoreva, A., Pishchulin, L., & Schiele, B. (2018). Learning to refine human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 205-214).
- [13] Moshayedi, A. J., Khan, A. S., Zhang, M. J., Yang, S., & Bassir, D. (2026). MobileNetV3 for Pipeline Welding Defect Detection: A Computer Vision Approach on Noisy RGB Images. In *E3S Web of Conferences* (Vol. 723, p. 01014). EDP Sciences.
- [14] Song, Y. (2024, September). Human Motion Pose Recognition System Based on Convolutional Neural Networks. In *Proceedings of the 2024 3rd International Conference on Algorithms, Data Mining, and Information Technology* (pp. 162-167).
- [15] Lucas, T., Baradel, F., Weinzaepfel, P., & Rogez, G. (2022, October). Posegpt: Quantization-based 3d human motion generation and forecasting. In *European Conference on Computer Vision* (pp. 417-435). Cham: Springer Nature Switzerland.
- [16] Kutrzyński, M., & Król, D. (2024). Deep learning-based human pose estimation towards artworks classification. *Journal of Information and Telecommunication*, 8(4), 470-489.
- [17] Moshayedi, A. J., Khan, A. S., Kolahdooz, A., Elragig, A., Khan, Z., & Bassir, D. (2026). Smart Agro-ecosystem: A Review of LLM-based Robotic Systems for Sensing and Decision Support in Precision Agriculture. *EAI Endorsed Transactions on AI and Robotics*, 5.

- [18] Zhumasheva, A., Mansurova, M., Amirkhanova, G., & Tyulepberdinova, G. (2025). Analyzing the Efficacy of Pose Recognition, YOLOv3, and Deep Learning Techniques for Human Activity Recognition. *Journal of Applied Data Sciences*, 6(4), 2770-2780.
- [19] Khan, A. S., Rahman, A., & Moshayedi, A. J. (2025). Pipeline Surface Defect Detection Using YOLOv11 with Attention Mechanisms: A Comparative Study of SA, LKA, and CBAM Approaches. *Journal of Robotics Research (JRR)*, 2(2).
- [20] Fu, Y., Lu, W., Dong, Z., & Fang, Y. (2025). A synthetic data-enhanced method for automated 3D pose recognition of construction workers. *Expert Systems with Applications*, 294, 128768.
- [21] Zhang, S., Han, X., Meng, Z., Wang, C., Zhang, Y., Ma, Y., & Wang, Z. (2025). A survey of human pose recognition based on WiFi sensing and neural network. *Journal of Computer and Communications*, 13(6), 47-63.
- [22] Moshayedi, A. J., Khan, A. S., Yang, S., & Zanjani, S. M. (2022, April). Personal image classifier based handy pipe defect recognizer (HPD): Design and test. In *2022 7th International conference on intelligent computing and signal processing (ICSP)* (pp. 1721-1728). IEEE.
- [23] Li, K., Wang, S., Zhang, X., Xu, Y., Xu, W., & Tu, Z. (2021). Pose recognition with cascade transformers. *arXiv preprint arXiv:2104.06976*.
- [24] Zhang, W. (2025). Dynamic pose recognition based on deep learning: Developing a CNN model for choral conductor pose recognition. *Journal of Computational Methods in Sciences and Engineering*, 25(4), 3856-3874.
- [25] Jahangir, M. A., Nowzari, R., Taherinezhad, M., Esmaili, N. H., Khan, A. S., Ghadiri, N. M., & Ghanbari, N. (2025). Investigation and Characterization of Pipe Defects and Techniques, and Challenges Toward the Protection of Environmental Protection.
- [26] Moshayedi, A. J., Khoojine, A. S., Khan, A. S., Roy, A. S., Dangling, X., Xie, Y., ... & Andani, M. E. (2026). EFM2: An Enhanced Sensor Fusion and Path-Tracking Control Framework for Omnidirectional Automated Guided Vehicles. *Journal of Robotics Research (JRR)*, 3(2).