International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Comparison of Single-Shot and Two-Shot Deep Neural Network Models for Pose Estimation in Assistance Living Application

May Phyo Ko[a]*, Chaw Su[b]

[a,b]*University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin and 05081, Myanmar*
[a]*Email: mayphyoeko@gmail.com*
[b]*Email: mschawsu.it@gmail.com*

Abstract

Estimating human posture from an image or video is an essential task in computer vision. This task has detected body key points from a camera for body posture and gesture recognition technology, which enables the following applications: assisted living in the case of fall detection, yoga pose identification, character animation, and an autonomous drone control system. The rapid development of AI-based posture estimation algorithms for picture recognition has resulted in the availability of quick and dependable solutions for recognizing the human body joint in collected films. One major issue in human posture assessment is the system's capacity to perform with high accuracy in real-time under shifting ambient conditions. The ultimate goal of the proposed transfer learning-based posture estimation assignment is to achieve real-time speed with virtually no drop accuracy. In this research paper, assisted living program (ALP) is implemented by using a single-shot deep estimation network and a pose key points angular feature. Experimental results show that transfer learning-based pose identifies and estimates posture with a frame rate of about 30 frames per second and a detection accuracy rate of 96.81%.

Keywords: Pose Estimation; Real-Time; Transfer Learning; Key Point.

Received: 4/23/2024
Accepted: 6/23/2024
Published: 7/1/2024

---

* Corresponding author.

1. Introduction

An assisted living program is a sort of accommodation for older people who require daily care but not as much care as a skilled nursing home. It provides a greater basis of support for older people than independent living, but less than a nursing home or memory care facility. In the proposed assistance, living pose estimation might be used to monitor and analyze the motions of important body areas, as well as categorize patient movements.

Traditional human position estimates make it difficult to identify the location of human joints like elbows and wrists. The suggested approach focuses on body key point landmarks as well as the primary joint angle between body key points of a person. Transfer learning-based pose estimation is implemented with Deep Neural Network (DNN) approaches. Unlike non-deep neural techniques, transfer learning is the reuse of a previously learned model on a new task. In transfer learning, an algorithm uses knowledge obtained from one activity to enhance generalizations about another. Learning-based deep pose estimation is implemented with wearable sensors or surveillance cameras that are widely used for every monitoring system. Some of the most common sensors used for pose estimation are accelerometers, motion sensors, wearable sensors, and gyroscopes. Figure 1 shows the deep neural network-based pose estimation.

Figure 1: Deep Neural Network Based Pose Estimation

Transfer learning-based deep pose estimation is combined with machine learning (ML) approaches, in which a model that has previously been trained on one task is fine-tuned for a new related task. Training a new ML model is a time-consuming and intense process that necessitates a large quantity of data, computer power, and multiple iterations before it is ready for use. A popular stance estimate made with wearable sensors solves some of the issues that older people face when going about their daily lives, such as the discomfort of having sensors on their body parts for long periods of time. As a result, a proposed video-based motion capture pose estimation strategy is used to recognize the activities of older people in an alternate way. Deploying pose estimation models in real world applications often faces challenges related to computational constraints and real time performance requirements. Furthermore, changing weather conditions and different human appearances can reduce the model's accuracy.

2. Related Work

Alexander Toshev [1] suggested a method for estimating human poses using deep neural networks. Pose estimation is expressed as a DNN-based regression issue for body joints. The author shows a cascade of such

DNN regressors that produce high-precision pose estimates. This approach has the advantage of reasoning about stance holistically, as well as a simple yet effective formulation that takes advantage of current breakthroughs in deep learning.

Shubham Shinde [2] has proposed a real-time model for human action recognition in video based on YOLO. The authors of that paper have demonstrated that YOLO is an effective method and is comparatively fast for recognition and localization in the Liris Human Activities dataset. They found that, even in some cases, a single frame is sufficient for the recognition of action. This proposed system achieved an average end-to-end recognition rate of 89.88% on 367 actions from 167 videos.

Thin Zar Wint Cho [3] has proposed a human activity recognition system for skeleton joint data from the Kinect sensor based on joint distance features. They have compared non-static k-mean and static k-mean for human activity on the testing set of the new dataset. Skeleton joint data can only be extracted from the front view. It sometimes doesn't catch on to body movement.

Win Myat Oo [4] has developed a human activity recognition system. They have proposed a new feature extraction method, a new morphological operation method, 'Vertical Binary Bridge', a pose classification method using Support Vector Machines (SVM) and a new frequency-based feature selection method. The authors of that paper have modified it to generate an alarm when a person comes under tracking, so that system will be useful for security purposes.

Chen [5] proposed a method of human activity recognition for elderly people. Using machine learning and deep learning approaches, their method detected six activities, including sitting, walking, going upstairs, going downstairs, standing, and lying. According to their method, the recognition average accuracy of deep learning-based short-term memory networks is 95.04%, and machine learning-based support vector machine (SVM) is 89.07% for each activity.

3. Proposed System and Methodology

3.1. Transfer Learning based Proposed Model

Human pose estimation is a difficult method for identifying and categorizing the joints in the human body. It is a method of obtaining a set of coordinates for each joint or heatmap intensity, referred to as a key point or landmark, that can outline a person's stance. There are two types of 2D human pose estimation methods: bottom-up estimation and top-down estimation.

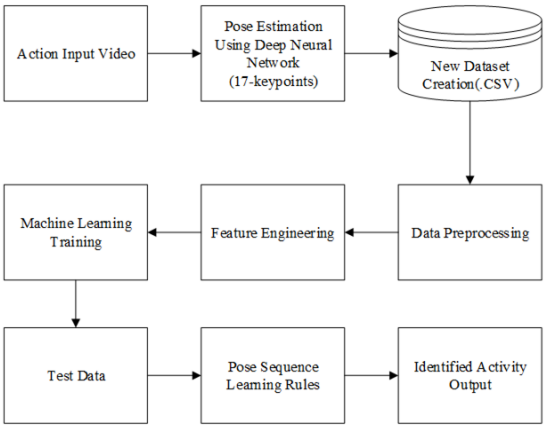

Figure 2: Overview of Proposed Assistance Living Application

In this research, the top-down estimation strategy is used. The top-down strategy is that human body joint is firstly identified and then the key point coordinates for each detected human body. Previous studies on assisted living detection relied primarily on image processing to detect objects in a picture and then classify them using machine learning methods. Many researchers are working to develop a trustworthy pose estimation-based assistive living program. There are three types of pose estimation models: machine learning-based pose estimation models, deep learning-based pose estimation models, and transfer-learning-based pose estimation models. The proposed system is based on a transfer- learning-based posture estimation model. Figure 2 depicts a broad block diagram of the planned assisted living application.

The core pipeline of the proposed transfer learning-based posture estimation model consists of several phases. Firstly, body skeleton joints are extracted from a human image using a reliable pre-trained deep pose estimate method. Next, a new.csv coordinate key point dataset is provided to the system. The next step is pre-processing. After that, the feature extractor is used to get the required features. Next, a machine learning model expects human postural motions using feature engineering. The primary contribution of the recommended system is the suggestion of sequence rules. The proposed system's last phase is to identify the routine actions of humans.

The proposed method employs more than one model to achieve greater accuracy in the complex application of transfer learning-based pose estimation. The output class of the proposed system is categorized into seven normal behaviors and one pathological behavior based on each person's daily activities. The evaluation is carried out on two datasets: the public dataset and the custom personal anomalous behavior dataset.

3.2. Deep Learning based Object Detector Network

In today's environment, computer vision technology has emerged as a critical area of focus for assistive living applications. Object detection has evolved into one of the fundamental challenges of computer vision, serving as the foundation for many vision tasks. In recent years, numerous researchers have used conventional computer vision algorithms to recognize objects. There are numerous challenges with traditional computer vision algorithms. Deep learning techniques are currently being applied to tackle problems faced in standard computer

vision algorithms. Whether researcher need to understand the relationship between images and objects or detect fine categories, it delivers accurate data.

In general, deep neural network-based object identification approaches are typically classified into two types: (1) deep learning algorithms with two-shot object detectors and (2) deep learning algorithms with one-shot object detectors. The industry's three primary pose estimation algorithms are Keypoint R-CNN, YOLO, and SSD. The proposed work presents two variants of the pose estimation model: SSD, a one-shot object detector, and Keypoint RCNN, a two-shot object detector. Although the two networks differ significantly in their construction, they are fundamentally the same. The SSD is a one-shot object detecting network that yields modestly greater performance. Region proposal algorithms often offer slightly higher accuracy but are slower to run, whereas single-shot methods are more efficient and have comparable accuracy.

3.3. Single Shot Multi Box Detector

The Single Shot Multi Box Detector (SSD) uses a single deep neural network to recognize human body landmarks in photos or videos. Its detector is a deep neural network family model. It makes advantage of multi-scale capabilities and default boxes, as well as decreasing photo resolution to improve speed. This allows SSDs to achieve real-time performance with almost little reduction in accuracy. During training, SSD uses a matching step to match the appropriate anchor box to the bounding boxes of each ground truth object in an image. Essentially, the anchor box with the most overlap with an item determines its class and placement. At prediction time, the network calculates scores for the presence of each item type in each default box and modifies the box to better reflect the object shape. Furthermore, the network incorporates predictions from many feature maps with differing resolutions to accommodate objects of different sizes. The SSD model detects several things in a picture in a single shot. The model divides the bounding box output space into default boxes with different aspect ratios and scales based on the feature map location. During prediction, the network distributes scores to each object category that appears in each default box. The network integrates predictions from many feature maps with differing resolutions, making it simple to handle objects of different sizes.

Figure 3 depicts the architecture of SSD. An SSD has two parts: a backbone model and an SSD head. As a feature extractor, the backbone model is often made up of a pre-trained image classification network. The SSD head is just one or more convolutional layers added to the backbone, with outputs read as bounding boxes and object classes corresponding to the spatial location of the final layer's activations. Non-maximum suppression (NMS) is a post-processing technique widely used in object detection to minimize repeated detections and pick the most relevant bounding boxes that match to the identified objects.

The base network is in charge of extracting basic information from the input image, while the detection head network interprets the bounding boxes and class scores of objects in the spatial position of the final layer's activations in the subsequent regression encoder network.

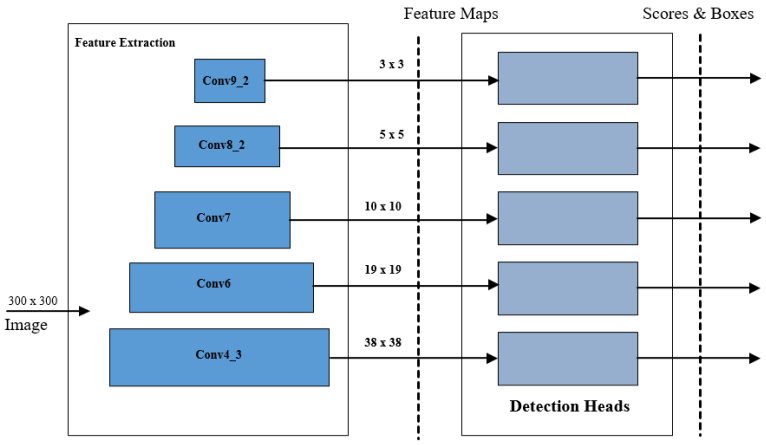

Figure 3: Overview of SSD Architecture

3.3.1 Grid Cell

The Instead of employing a sliding window, SSD divides the image into grid cells, with each cell responsible for identifying objects in that zone. Detecting objects is basically anticipating the class and location of an object within a given region. If no object is present, the network treats it as a background class, and the location is disregarded. Each grid cell can output the position and shape of the object it holds.

3.3.2 Anchor Box

Each grid cell in SSD can be assigned several anchor boxes. These anchor boxes are pre-defined, with each responsible for a specific size and form within a grid cell. The anchor box with the greatest degree of overlap with an object predicts its class and position. This attribute is utilized to train the network as well as to forecast detected items and their positions after training. Each anchor box is defined by an aspect ratio and a zoom level.

3.3.3 Aspect Ratio

Not all objects are square-shaped. Some are longer, while others are wider to varied degrees. To cater for this, the SSD architecture supports pre-trained anchor box aspect ratios. The ratios parameter specifies the varying aspect ratios of the anchor boxes associated with each grid cell at each zoom/scale level.

3.4 Keypoint Region based Convolutional Neural Network

Keypoint region based Convolutional Neural Network (Keypoint RCNN) is a family of Faster RCNN. Keypoint R-CNN is a region-based object detection technique. It is a pre-trained model that can detect 17 critical joints on the human body. Key point detection is commonly utilized in areas such as human pose estimate, object pose estimation, face identification and matching, fashion landmark detection, facial emotion recognition, and so on. Keypoint RCNN is trained on the MS-COCO dataset, which offers different annotation types for object detection, segmentation, and image captioning. The original COCO offered 80 classes for Detection and Segmentation.

However, for proposed Keypoint RCNN model, the annotations are offered only for the person class. ResNet-50 is the backbone network of Keypoint RCNN. It has a 50-layer convolutional neural network that are 48 convolutional layers, one Max-Pool layer and one average pool layer.

The feature pyramid network (FPN) is the main feature extractor utilized in object identification networks to extract multi-scale features. FPN is used to improve the quality of features, which can be accomplished by combining high- and low-resolution features. High-resolution maps contain low-level features, whereas low-resolution maps contain high-level features. The architecture is divided into two parts: a bottom-up pathway and a top-down pathway, as shown in Figure 4. A region-based pre-trained network is utilized to extract the feature vectors of critical points obtained from the backbone network. The main contribution of key point-based Faster RCNN is the detecting head. It has two essential components: the object detection head and the key point detection head. The fully connected layer that forms the object detection head can be separated into two FC blocks.

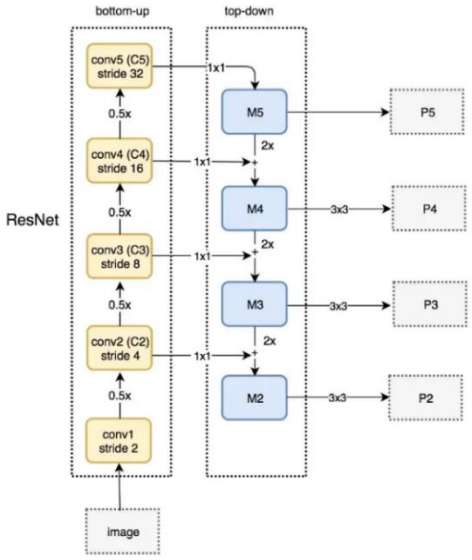

Figure 4: Feature Pyramid Network

One utilizes the output size [N, C] to predict the class-scores for the recommended item, while the other uses the output size [N,4*C] to change the box coordinates for the suggested object. By not encoding a key point of the detected object, the proposed description network of the key point detection head alters the existing Mask RCNN slightly. The key point detection head's output is [N, K=17,56,56]. Two final class scores are obtained: one for the person class and one for the background, for a total of [N, 2].

3.5 Pose Land Marker Model

The pose land marker model scans 17 critical points on a person's body to calculate angles and recreate poses. The specific locations of the detected human key spots are depicted in Figure 5.

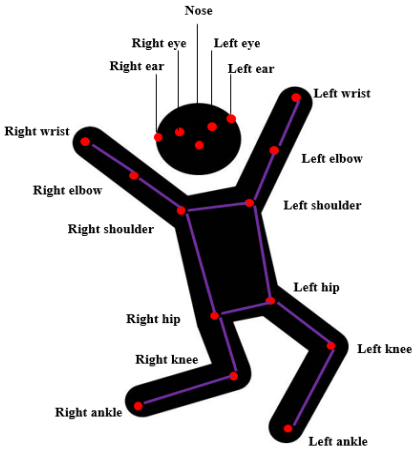

Figure 5: Pose Land Marker Model

The labels 0-16 indicate the main points of the human body in the following order: nose, right eye, left eye, right ear, left ear, right shoulder, left shoulder, right elbow, left elbow, right wrist, left wrist, right hip, left hip, right knee, left knee, right ankle, left ankle, and so on. The prediction can be better interpreted because the 17 landmarks picked are related with the key joints and body components required for daily activities.

4. Modeling and Analysis

Reading and understanding the data: After entering video frames into the pose estimation method, a key points dataset is produced. The overview of the newly obtained deep posture estimation neural network dataset for the proposed system is depicted in Figure 6. The dataset's actual appearance, together with its features and characteristics, are depicted in the picture. Feature engineering is applied to datasets consisting of 65 columns.

	NOSE_X	NOSE_Y	NOSE_DISTANCE	NOSE_SLOPE	NOSE_ANGLE	LEFT_SHOULDER_X
0	110.2554	-14.66667	15.1187	77.31268	-1.28421	10.82798
1	111.2531	-14.8	15.26884	77.93106	-1.25773	10.88877
2	110.2225	-15.7143	15.07615	77.62087	-1.27083	10.84621
3	110.1635	-18.3333	15.01645	77.62087	-1.27083	10.84621
4	111.162	-18.5	15.16693	77.62087	-1.27083	10.84621
5	113.3986	-11.8947	15.70439	75.49338	-1.15152	10.49032
6	114.3951	-12	15.85515	75.49338	-1.15152	10.49032
7	115.3137	-13.5294	15.94787	75.49338	-1.15152	10.49032
8	115.3137	-13.5294	15.94787	75.49338	-1.15152	10.49032
9	103.982	-10.35	14.29907	75.04166	-1.24468	10.49148

10 rows x 65 column

Figure 6: Understanding the data

4.1 Preprocessing

Preprocessing the incoming data is a critical step in posture estimation using machine learning techniques. Skeletal

data is used in this study to classify eight types of motion activity. Normalizing the data becomes important due to variations in body forms and movement variances between people. This work focuses on normalizing skeletal data by specifying body joint angles and converting it to an independent coordinate system. Because the retrieved data has some missing values, this step comprises calculating, displaying, and updating those values for all features.

4.2 Feature Engineering

Feature engineering is the process of developing new features or modifying existing ones in order to improve the performance of a machine-learning model. It entails extracting useful information from raw data and converting it into a model-friendly format. Numerous feature engineers employ the posture estimation system to produce a satisfactory outcome, and numerous features are taken into account. The centroid of an object in an image represents its geometric center or center of mass. In more sophisticated cases involving grayscale or color images, centroid computation may use additional factors such as intensity values or color channels. Addressing computing issues for real-time deployment of centroid-based algorithms in applications such as robotics, autonomous driving, and augmented reality. The proposed framework corresponds to a single-person posture estimation framework, and it employs 17 landmark topologies, which could help in the identification of the model in our research. In Figure 6 shows the pre-trained pose output demonstrating human pose estimation through key point detection. The two points of left shoulder and right hip are calculated by using triangular formula to determine the center points of whole body. To determine this joint angle, the vector of each body part is generated using two key points, and the angle between the two vectors is obtained using the formula presented in equation 2. To identify the relevant features for better outcomes, the suggested system in this work employs angular features such as key point coordinates, distance, and angle. Figure 7 shows the vectors for each body part in the daily stance. The angles of various crucial points can be used to determine the posture during regular tasks. These pose coordinate landmarks are used to calculate the angles at the chosen key point. Once all of the angles necessary to establish the pose have been determined. In the first step, the distance between two body components was calculated using the x and y coordinates of the posture estimation point. The Euclidean distance equation 1 is used to calculate the distance (d).

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (1)$$

where (x1, y1) and (x2, y2) are coordinate of two body key points. The angle from defined key locations was determined in the second step utilizing the vector and magnitude of the key points provided. The angle between two vectors is calculated using equation 2.

$$\theta = cos^{-1} \frac{A \times B}{|A| \times |B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (Ai)^2} \times \sqrt{\sum_{i=1}^{n} (Bi)^2}} \qquad (2)$$

where A and B denote the lines that connect the proper joints. The subsequence diagram depicts the angle relationships between each feature point.

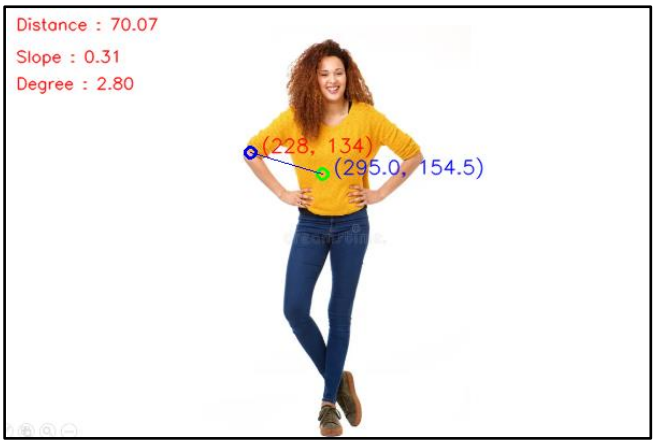

Figure 7: Centroid based Feature using pre-trained pose estimation

The essential characteristics of the proposed system are key point coordinates of human skeletal joints and angles. The determined angles were kept as features, as shown in Table 1. These angles were then given and used to train and test classifiers for daily posture differentiators.

Table 1: Description of Joint Angle based Features

Feature	Description	Feature	Description
LS	Left Shoulder	RS	Right Shoulder
LH	Left Hip	RH	Left Hip
LK	Left Knee	RK	Right Knee
LA	Left Ankle	RA	Right Ankle

4.3 Model Training

As the analysis of data is complete, modeling of data must be done using a classification-based machine learning algorithm. So, the system uses 4 different machine learning algorithms namely- logistic regression, support vector machine classifier, random forest classifier and naïve bayes classifier. Table 2 shows performance pose classification of machine learning models. The support vector machine (SVM) classifier is having the best accuracy score among all the classifiers employed with an accuracy score of 96%. Random forest classifier performs worst among all with an accuracy score of 89%.

5. Experimental Results and Discussion

In this section, the proposed transfer learning-based pose estimation approach has been evaluated on the public UR human pose dataset and a new custom dataset recorded by the surveillance camera. From these datasets, body joint features are extracted based on a pre-trained object detector network, and these features are clustered to create a new CSV dataset that is implemented with a centroid based pose angular feature vector. According to the preceding centroid based pose feature testing results, Tables 4, 5 and 6 were evaluated on three indoor recorded videos, and the accuracy of sitting poses was 96.68%, standing poses were 94.7%, and lying stances were 94.1%.

These accuracy tests were run on three videos that lasted an average of 34 seconds. Accuracy best predicts the successful classification out of a total number of samples and determines the model's ability to correctly predict the target value.

For an image/static video/live video after pose prediction, a threshold value is set up below which the system generates the output as no pose detected. The confidence level required for pose detection is 96%. Figure 8 show the experimental results of pose activities with landmark detection. According to the results, compared with one-shot object detection model, two-shot object detection model reduces the proposal region detection module, the model structure is simplified, and the detection is more suitable to perform real-time detection. Whereas two-shot object detectors base proposed system are very accurate, single shot object detectors based proposed system are computationally faster while sacrificing accuracy. Table 7 show the experimental results of pre-trained SSD and Faster RCNN are presented in details. The model single-shot multi-box detector (SSD) has a frame rate of about 30 frames per second and a detection accuracy rate of 96.81%, which is greater than the frame rate of about 20 frames per second and detection accuracy rate of Faster RCNN, which is 94.44%.

Table 2: The performance pose classification of the single-shot detector and Two-shot detector on Custom Dataset

Machine Learning Model	Single-Shot Pose Detetor			Two-Shot Pose Detector		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Logistic Regression	0.94	0.94	0.94	0.93	0.92	0.93
SVM Classifier	0.96	0.96	0.95	0.94	0.94	0.93
Random Forest Classifier	0.89	0.90	0.89	0.87	0.88	0.90
Naïve Bayes Classifier	0.91	0.93	0.89	0.87	0.88	0.91

Figure 8 show the experimental results of pose activities with landmark detection.

Figure 8: Pose Detection Successful

The activities of human's pose differ from traditional picture classification problems. The foundation of human pose estimation requires a succession of frames to extract information and forecast human activity. Table 3 show the standard sequence of proposed system.

Table 3: Proposed Standard Sequence for Activity

Sequences	Selected Feature	Label	Activity Recognition
11….22222	{1 2}	SQ_1	Sitting Down
222….1111…11	{2 1}	SQ_2	Standing Up
1111…2222…333	{1 2; 2 3}	SQ_3	Lying Down
111….1…333	{1 3}	SQ_4	Falling Down

Table 4: Accuracy of Standing Pose in Three Video

Metrics	Video 1	Video 2	Video 3
Standing Pose	400	420	450
Not Standing Pose	600	580	550
TP	388	410	442
TN	549	537	517
FP	12	10	8
FN	51	43	33
Accuracy	93.7%	94.7%	95.9%
Average Accuracy	94.7%		

Table 5: Accuracy of Sitting Pose in Three Video

Metrics	Video 1	Video 2	Video 3
Sitting Pose	410	430	400
Not Sitting Pose	590	570	600
TP	400	415	390
TN	570	550	580
FP	10	15	10
FN	20	20	20
Accuracy	97.0%	96.5%	97.0%
Average Accuracy	96.8%		

Table 6: Accuracy of Lying Pose in Three Video

Metrics	Video 1	Video 2	Video 3
Lying Pose	490	470	450
Not Lying Pose	510	530	550
TP	467	451	432
TN	472	496	505
FP	23	19	18
FN	38	34	45
Accuracy	93.9%	94.7%	93.7%
Average Accuracy	94.1%		

Table 7: Comparison of the score of Single-Shot Detector-based ALP and Two-Shot Detector-based ALP

SN.	Pre-Trained Deep Pose Estimation Neural Network	Machine Learning Model	Mean Average Precision	Frame Per Second (FPS)
1	SSD	SVM	0.96	30 fps
2	Keypoint R-CNN	SVM	0.93	20 fps

6. Conclusion

Transfer learning based human pose estimation is up-to-date AI-based learning techniques on assistance living program. In this study, an assistance pose classifier was successfully developed which works perfectly on images, static video, and live video of any user. This paper compares two popular pose estimation detectors, SSD and Key point R-CNN, to learn efficient deep pose estimation models. The performance of these models is evaluated by using mean average precision (mAP) and frame per second (FPS) as evaluation metric.

Media pipe pose estimation library is used for human pose estimation which returns body key points, these data points form the basis of a new dataset. Then data preprocessing takes place in which target variables are changed. After this normalization of data occurs for better performance of machine learning algorithms and finally feature engineering of features starts where various joint angles of the body are calculated using the angle formula. As the data is completely pre-processed data is finally fed to machine learning models.

Evaluation of these models is done on test data and is compared based on accuracy score. Support Vector Machine classifier achieves a maximum score of 96% among all classifiers. For classification a threshold value is used which is set at 96% below which no pose detected is given as output to the user. The suggested pose estimation system can recognize single-person and multi-person actions. It processes more quickly, however, there are a lot of missing postures. It focuses on low-computation technology such as desktop applications it is user-friendly and mobile devices such as laptop computers. The suggested transfer learning-based posture estimation system

analyzes an elderly person's daily actions and determines if they are normal or aberrant. The suggested transfer learning-based approach works effectively with both a single camera and a cross-view. Finally, this proposed system can reduce the redundancy of features that act as noise for the classifier during the recognition process. The proposed transfer learning-based solution is intended to be used in any e-personalized assisted living program.

Acknowledgements

The author especially would like to take this opportunity to express my sincere gratitude, respect, and regard for my supervisor Dr. Chaw Su, Professor, Department of Computer Engineering, University of Technology (Yatanarpon Cyber City) for giving me guidance, constant encouragement, patience, and trust to work on this paper.

References

[1] A.Hatsham, Y.Chen "Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches", ICIEM Access ,2022.

[2] A.Latee Haroon, "Effective Human Activity Recognition Approach using Machine Learning", Journal of Robotics Control,2021.

[3] Muhnad, "A Novel Feature Selection Method for Video-Based Huma Activity Recognition Systems", IEEE Access, 2019.

[4] Shubham, "YOLO based Human Action Recognition and Localization", (RoSMa),2018.

[5] Tin Zar Wint Cho, "Performance Analysis of Human Action Recognition System between Static k-Means and Non-Static k-Means", IJSR,2017.

[6] Win Myat Oo, "Feature Based Human Activity Recognition using Neural Network", IEEE,07 February 2021.

[7] May Phyo Ko, "Human Activity Recognition System Using Angle Inclination Method and Keypoints Descriptor Network", IEEE, Conference of Young Researchers in Electrical and Electronic Engineering (ElCon), 2024.

[8] May Phyo Ko, " Keypoints Feature based Activity Recognition Using Deep Neural Network", ISSN:2709-6505, Journal of Research and Innovation, vol.6,2023.

[9] Eisha Akanksha, "A Feature Extraction Approach for Multi-Object Detection Using HoG and LPT", International Journal of Intelligent Engineering and Systems, 2021.

[10] K. Wei, "Multiple Branches Faster RCNN for Human Parts Detection and Pose Estimation", Springer International Publishing, 2020.

[11] Muhammad, "Effective Human Activity Recognition Approach using Machine Learning", Journal of Robotics Control, 2021.

[12] S-Xiang, "RGB+2D skeleton: local hand-crafted and 3D convolution feature coding for action recognition", Springer,2021.

[13] V.Parameswari, "Human Activity Recognition using SVM and Deep Learning", European Journal of Molecular and Clinical Medicine,2020.

[14] Valentin Bazarevsky Tyler Zhu, "On-device Real-time Body Pose Tracking", arXiv:2006.10204v1

[cs.CV] 17 Jun 2020

[15] Utkarsh Bahukhandi, "Yoga Pose Detection and Classification Using Machine Learning Techniques", International Research Journal of Modernization in Engineering Technology and Science, vol -3 /issue:12/ December-2021.

[16] Wei Lui, "Single Shot Multi Box Detector", arXiv:1512.02325v5 [cs.CV] 29 Dec 2016.

[17] Utkarsh Kharb, "Review and Analysis of Various Human Pose Estimation Models", 2nd International Conference on Advancement in Electronic & Communication Engineering (AECE 2022) July 14-15, 2022.

[18] Jyoti Jangade, "Study on Deep Learning Models for Human pose Estimation and its Real Time Application", IEEE,6th International Conference on Information Systems and omputer Networks (ISCON), 2023.