Cureus Journal of Computer Science

Open Access Review Article

Artificial Intelligence in Fitness: Pose Estimation and Movement Correction

Dhanashree A. Phalke [1] , Varsha Kotipalli [1] , Priya Ranjan [1] , Yash Pawar [1] , Pranjal Bharat [1]

1. Computer Engineering, D. Y. Patil College of Engineering, Pune, IND

Corresponding author: Dhanashree A. Phalke, daphalke@dypcoeakurdi.ac.in

Review began 01/06/2025
Review ended 03/25/2025
Published 03/28/2025

DOI:
https://doi.org/10.7759/s44389-024-00747-w

Abstract

In this study, we focus on recent developments in the use of artificial intelligence (AI) and deep learning for fitness movement assessment and human pose estimation. The incorporation of AI techniques has become important in fitness tracking; therefore, methods such as convolutional neural networks and pose estimation models, including PoseNet and ConvNeXt, are increasingly utilized. Identified article topics are based on 2D and 3D pose estimation methods, real-time feedback technology, and augmented reality implementation plans. We also state that these technologies can be used to correct exercise postures, prevent injuries, and, generally, augment all fitness-related exercises because of accurate real-time feedback. The survey is rounded off with a focus on mobility and issues concerning it, while also extending the reach towards mobile-based applications.

Categories: AI applications, Expert Systems, Health Informatics
Keywords: deep learning, human pose estimation, convolutional neural networks (cnn), 2d/3d pose estimation, pose correction

Introduction And Background

Introduction

The average blending of artificial intelligence (AI) with fitness has led to the adoption of technologies that seek to optimize workout experience. For instance, real-time body posture correction, human pose estimation or movement appraisal, and real-time pose correction have been the main areas for implementing deep learning techniques. Since physical fitness can be easily assessed and controlled with the help of portable gadgets and augmented reality (AR) technologies, increased accuracy in the assessment and timely correction [1] of movement in the course of training is possible, which is quite promising for users and trainers.

Human pose estimation [2], where the goal is to localize certain joints in images or videos, has turned out to be an essential component of fitness apps. For movement analysis, both 2D and 3D pose estimation models have been used with postural and technique feedback provided to the user. Recent developments of convolutional neural networks (CNNs), including PoseNet [3] and ConvNeXt [4], among others, make it possible to identify positions of the human body and draw body maps on the scene even on low-memory devices such as the mobile platforms.

Here, in this paper, we discuss and review the advanced deep learning methodologies for fitness movement assessment and pose estimation. The techniques in human pose recognition [5] provide real-time feedback, along with the roles of fitness, with special emphasis on ways to prevent exercise adherence issues and enhance form. This survey also explores the utilization of AR and mobile technologies in advancing fitness and extending its interactivity. Finally, we discuss the possible limitations and future research directions in this rapidly growing field.

Background and related work

The applications of deep learning in fitness applications have improved the ability to identify movements and estimate poses enormously. Monitoring of physical activity, assessment of movements, and postural adjustments are some of the roles previously played by fitness trainers and instructors [6], which have now partly shifted to AI-based solutions. This section offers a review of the fundamental of deep learning approaches encompassing human pose estimation and movement assessment in fitness applications.

Deep Learning in Fitness

Some sub-branches of deep learning such as CNNs have been found useful in the identification and categorization of human motions in fitness facilities [7]. It is worth mentioning that CNNs are intended to learn features of the images or the frames of the videos on their own, which makes them suitable for pose estimation, action recognition, etc. It is capable of learning large inputs about human motion and then recognizing specific patterns [8] in the input data.

Human Pose Estimation

Human pose estimation is a computer vision where one is able to point out the landmarks on the human body in the form of joints or movement in images or videos. Therefore, based on a certain type, the different pose estimation models can be classified under 2D and 3D. Pose estimation models can be categorized into 2D and 3D approaches:

2D pose estimation: In 2D pose estimation, models detect keypoints in the 2D plane of an image, identifying body parts such as elbows, knees, and shoulders. These models, like PoseNet, are lightweight and widely used in mobile and web-based applications due to their efficiency. PoseNet [9] has been successfully applied in fitness applications where real-time correction of posture is essential.

3D pose estimation: More advanced 3D pose estimation models predict the spatial (3D) arrangement of key body points, allowing for more precise analysis of movement. For example, ConvNeXtPose, a model described in the study by Newell et al. [10], offers higher accuracy by mapping joints in 3D space. This technique is more computationally intensive but opens new possibilities for fitness applications, such as tracking complex movements and offering in-depth feedback through AR.

Real-Time Feedback in Fitness

The ability to provide real-time feedback is a major advantage of deep learning in fitness. Real-time pose estimation systems allow users to receive instant corrections on their posture during exercises. For instance, the Fitness Tutor system [11] uses PoseNet to match users' exercise postures with reference images of trainers. This system enables users to receive corrections on body positioning without assistance from another person, enhancing performance during home exercises. Such real-time feedback options make it easy to participate in fitness training without exercise trainers, ensuring that users are attended to in real-time, even when exercising remotely or at home.

AR and Mobile-Based Applications

AR is among the most viable areas of fitness where pose estimation will be effective. AR fitness applications, such as those designed with the help of ConvNeXtPose [12], allow users to overlay guides or avatars on top of their workout videos in real-time. These applications provide an excellent way to offer feedback to users regarding their form, enabling real-time corrections through an interactive medium. The synergy of AR and deep learning pose estimation presents considerable opportunities for enhancing user interaction, exercise adherence, and workout correctness without the need for expensive equipment. Also, some fitness technology applications are now being developed for mobile platforms. Recent deep learning models, including OpenPose [13] and ConvNeXtPose, have been proposed for mobile devices due to their efficient computational requirements. Despite being lightweight, these models provide sufficient accuracy for real-time feedback, making it easier for users to access fitness tools on the go.

Existing Fitness Evaluation Systems

Several deep learning-based systems for the evaluation of fitness movements have been discussed in this paper. He et al. [14] proposed a system using CNNs to classify movements and offer relevant scores to users. This system enables users to track changes over time and provides feedback on their movement profiles [15]. Similarly, Ueta [8] presents an AR application that estimates the user's pose and gives instantaneous feedback during exercises.

These systems demonstrate that, along with continuous advancements in deep learning, fitness technologies allow for accurate movement assessment and the creation of new fitness programs based on users' records [16].

Review

Key techniques for pose estimation and movement evaluation

In fitness applications, human pose estimation and movement evaluation were done in a traditional way, but now deep learning methods, especially CNNs, have changed how it is done. This section addresses some of the most critical approaches that include CNNs, pose estimation techniques both in 2D and 3D, real-time feedback system, and ARs, with the main focus being mathematical applications [17] that provide the basic foundation for these technologies.

Convolutional Neural Networks

CNNs are an essential part of the human pose estimation systems. These networks pass a sequence of convolutional filters that help in extracting relevant features when it comes to modeling the joints of the body in images or even videos [18]. The general operation of a CNN can be expressed as follows:

$$z(l) = \sigma\left(W(l) * z(l-1) + b(l)\right) \quad (1)$$

where the parameters are detailed as follows:

z(l): The feature map at layer *l*. This represents the processed information at the current convolutional layer and serves as the input for the next layer.

W(l): The weight matrix at layer *l* This represents the learned features. These weights are adjusted during training to help the model learn spatial hierarchies in images.

$*$: Convolution operation, where the input $z(l-1)$ is convolved with the filter $W(l)$, producing a feature map for the next layer. The convolution operation is key to detecting patterns in images that correspond to human joints.

z(l−1): The feature map from the previous layer. It serves as input for the current convolutional layer and holds essential spatial information about the image.

b(l): The bias term, added to the result of the convolution to help the model learn more complex features. It ensures that the activation function can adjust more flexibly.

σ: The activation function (e.g., ReLU), applied to the output of the convolution. This introduces non-linearity, which is essential for the model to learn complex relationships in the data.

Comparison of existing 3D pose estimation techniques

Several techniques are used for 3D pose estimation, each with varying accuracy, computational complexity, and robustness. The main categories include traditional optimization-based approaches, deep learning-based models, and hybrid methods.

Traditional Optimization-Based Methods

These methods rely on geometric constraints and handcrafted features to estimate 3D poses [16].

Examples: Perspective-n-Point (Figure *1*) and Bundle Adjustment.

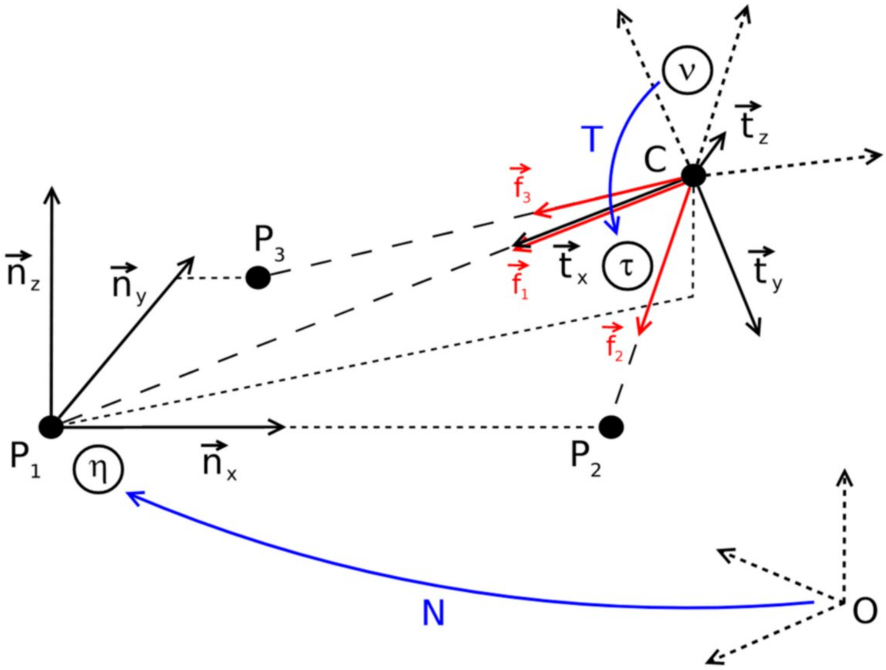

FIGURE 1: Perspective-n-Point Pose Computation

OpenCV: Perspective-n-Point pose computation, www.bharatagritech.com

Advantages:

1) Interpretable and mathematically grounded

2) Works well with limited data

Disadvantages:

1) Prone to errors due to occlusions and ambiguities

2) Less effective in complex real-world scenarios [19]

Deep Learning-Based Methods

These methods leverage CNNs or transformer-based architectures to learn pose representations from large datasets.

Examples: OpenPose (Figure *2*), PoseNet [5], High-Resolution Networks, and Graph Convolutional Networks (Figure *3*)

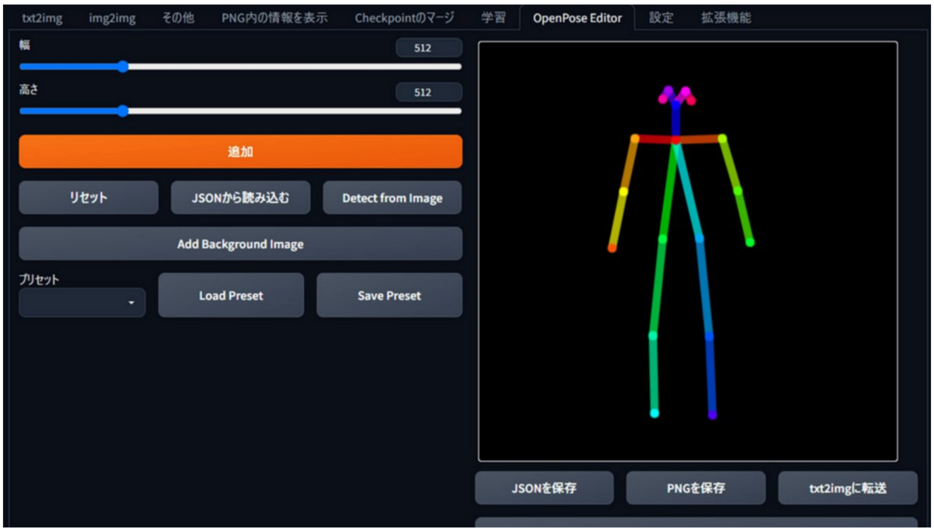

FIGURE 2: OpenPose Estimation

https://aienthusiastic.com/use-openpose-with-controlnet-and-its-editor/

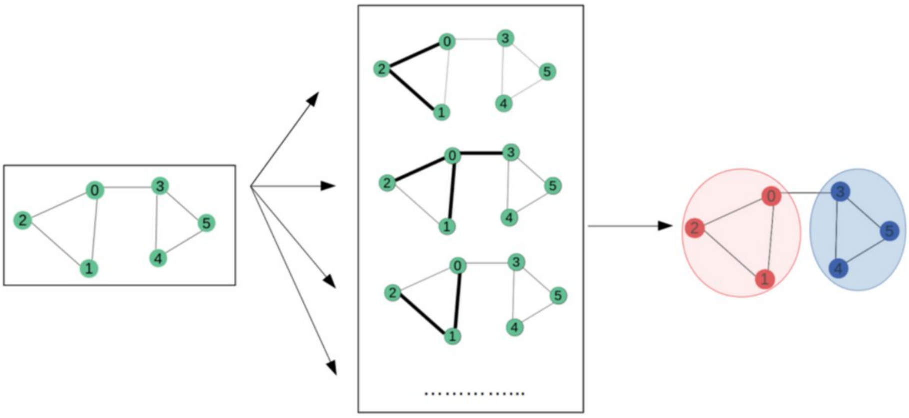

FIGURE 3: Graph Convolutional Network

Layers in a graphical flow, with connections between them. Source: ResearchGate

Advantages:

1) High accuracy due to large-scale learning

2) Robust to occlusions and variations

Disadvantages:

1) Requires large amounts of annotated training data

2) High computational cost

3) Pose estimation models

Hybrid Methods

Combining optimization-based and deep learning approaches can lead to more robust and efficient solutions.

Examples: Deep kinematic priors [12] and physics-informed models

Advantages:

1) Improved generalization across different datasets

2) Incorporates domain knowledge for better accuracy

Disadvantages:

1) Can be challenging to balance computational efficiency and model performance

Pose estimation

In 2D pose estimation, models detect human joints in the image plane (x_i, y_i) for each joint.

Pose estimation can be framed as a regression problem, where a deep learning model is trained to predict the joint locations. The loss function for 2D pose estimation is typically based on the mean squared error between the predicted joint coordinates:

$$L_{2D} = \frac{1}{N} \sum \left((x'_i - x_i)^2 + (y'_i - y_i)^2 \right) \tag{2}$$

$x_i, \quad y_i$: The ground truth coordinates of the i-th joint in the image.

$x'_i, \quad y'_i$: The predicted coordinates of the i-th joint.

N: The total number of joints being predicted.

This loss function quantifies the error between the predicted and actual joint coordinates. The goal is to minimize this error to improve the model's accuracy in joint location prediction.

Real-Time Movement Evaluation

Real-time feedback is essential in fitness applications to help users correct their form immediately. This feedback loop calls for the use of pose estimation models that will be in a position to analyze data and produce a feedback [20] within milliseconds. A study proposed a Fitness Tutor system that tracks body movements using the PoseNet model [21] and offers corrective feedback in real-time. The system minimizes the error in body joint positions between the user and a reference posture through a form of real-time error minimization, which can be expressed as:

$$E(t) = \sum_i |p'_i(t) - p_i(t)| \tag{3}$$

$p_i(t)$: The position of the i-th joint at time t in the user's real-time pose.

$p'_i(t)$: The reference joint position at time t from the ideal posture.

$E(t)$: The total error between the user's and the reference pose, calculated for each joint over time.

This error helps track deviations from the ideal posture, providing immediate feedback for corrections.

Table *1* presents a comparative overview of pose estimation models, PoseNet, OpenPose, HRNet [22], and BlazePose, used in fitness applications. It highlights their key features, advantages, and limitations. This comparison helps in understanding the trade-offs between accuracy, efficiency, and hardware requirements for each model.

Model	Type	Description	Advantages	Limitations
PoseNet	2D	Utilizes convolutional neural networks to estimate human poses from images.	Lightweight; suitable for mobile applications.	Limited accuracy in complex poses.
OpenPose	2D	Detects multiple person poses in real-time using part affinity fields.	High accuracy; real-time multi-person detection.	Computationally intensive; requires powerful hardware.
HRNet	2D	Maintains high-resolution representations through the network for precise pose estimation.	High accuracy; robust to occlusions.	High computational cost; complex architecture.
BlazePose	2D	Designed for mobile devices; provides real-time pose estimation.	Optimized for speed; efficient on mobile platforms.	May sacrifice some accuracy for efficiency.

TABLE 1: Comparison of Pose Estimation Models in Fitness Applications

AR Applications

AR has been identified as ideal for enhancing the fitness experience since it incorporates real-world views with information superimposition [23]. In fitness, AR applications use pose estimation models to provide users with immediate feedback by placing reference poses or skeleton videos on the user's screen. In the study by Srijeyanathan et al. [24], a real-time application of skeletonization is presented, where 3D pose estimation, termed human pose estimation, is used to give real-time feedback on users' movements. The system calculates the discrepancy between the user's joints and the target joints using a cost function:

$$D = \frac{1}{N} \sum \left((x_i' - x_i)^2 + (y_i' - y_i)^2 + (z_i' - z_i)^2 \right) \tag{4}$$

It plays a critical role in evaluating and refining human pose estimation models in fitness applications. This formula measures the Euclidean distance between the predicted and ground-truth joint positions in 3D space, providing an error metric that helps optimize the pose estimation model.

, y_i, z_i: The ground truth coordinates of the *i*-th joint in 3D space.

x_i', y_i', z_i': The predicted coordinates of the *i*-th joint in 3D space.

N: The total number of joints being predicted.

The 3D pose estimation model evaluates the discrepancies in all three spatial dimensions, helping to assess more complex, realistic movements during exercises.

Importance of cost function

Captures Full-Body Movements

Unlike 2D pose estimation [7], which only considers *x* and *y* coordinates, this function accounts for depth information (*z*), making it more robust for real-world movement assessment in fitness applications.

Precision in Fitness Form Correction

By quantifying the difference between the user's posture and the ideal form, the model can provide real-time feedback, allowing users to adjust their movements and minimize the risk of injuries.

Optimization of Pose Estimation Models

The function serves as an objective criterion for training deep learning models [1], enabling them to learn from errors and improve accuracy over time. Lower *D* values indicate that the model is performing well in tracking human movements accurately.

Generalization Across Exercises

Whether it is yoga, strength training, or rehabilitation, this function helps evaluate complex motions where alignment, depth, and joint coordination is crucial.

Review methodology

The process through which the present studies were selected and evaluated was developed to provide a clear and balanced picture of deep learning in fitness pose estimation and movement assessment.

Search Strategy

Searches were conducted on online databases such as as Google Scholar, IEEE Xplore, PubMed, and Scopus. The keywords used in the search were 'human pose estimation', 'deep learning fitness', 'real time feedback', 'AR fitness', 'positional analysis', and 'preventing injury through pose estimation'.

Inclusion Criteria

1) Papers published in journals and conferences within the last decade, between 2013 and 2023.

2) Current work revolved around deep learning algorithms used in human pose estimation and the evaluation of human movements, specifically for fitness.

3) Research articles that involve real-time feedback paradigms, AR applications, or mobile fitness systems using pose estimation methods.

4) Research papers and case studies describing experiments and applications more precisely for the accuracy of human pose estimation and its consequences for fitness or in cases of preventing an injury.

Exclusion Criteria

1) Works that focus on approaches other than deep learning for pose estimation.

2) Generally, papers not specifically related to fitness applications, such as those on gaming or robotics.

3) Peer-reviewed studies that could not be accessed in full-text searches or were behind paywalls.

4) Submissions that cannot be peer-reviewed, including white papers, abstracts, or conference papers.

By applying these inclusion/exclusion criteria, this review aims to include the most recent, high-quality studies that offer the most up-to-date analysis of recent developments in deep learning-based pose estimation and movement assessment for training exercises.

Critical analysis and expanded literature review

Despite being in its nascent stage, the literature for deep learning in fitness pose estimation and movement assessment has surged. In this section, the descriptions of the research studies themselves will be offered to those readers who seek more nuanced criticism of the studies reviewed herein.

Deep Learning Models for Pose Estimation

The review highlights several related studies demonstrating that CNN-based models are efficient for pose estimation in fitness applications. For instance, PoseNet and OpenPose have become the most conventional structures for 2D pose estimation, achieving accuracies significantly better than traditional models. However, many of these models are limited by short training sequences and cannot effectively apply to specific exercises such as yoga or dancing [7]. It is suggested that future work should address these limitations by incorporating larger datasets that include various exercise forms and body types.

Real-Time Feedback Systems

Several papers (for instance, Wang et al. [6]) show the effectiveness of applying real-time feedback for the correction of posture in fitness environments. However, one of the issues of these systems is the conflict that can exist between accuracy and performance. Some of the models offer very good feedback, but they have high resource utilization and hence not advisable for portable user devices. The future models have to strike a balance between high accuracy and real-time necessity.

Applications of deep learning in fitness

Deep learning has been applied in the fitness industry and has proven to enhance the practice and documentation of exercise routines. This section discusses five major deployment areas: correction of movement [25] in fitness, real-time feedback applications, AR, and mobile fitness. All of these areas adopt deep learning-based pose estimation and assessment models.

Fitness Movement Correction

Perhaps one of the most typical uses of deep learning advancements in fitness is movement correction on autopilot. When it comes to exercise performance, pose estimation models integrated into deep learning systems can identify small distortions in a user's posture during exercises and provide immediate feedback on correct performance. This can be especially useful for those training at home or without a trainer.

The Fitness Tutor system shows how pose estimation can assist users in adopting the correct posture. The system takes a picture of the user, then superimposes the preferred posture [26] on it, highlighting areas that need correction in real-time. This system is facilitated by a 2D pose estimation model named PoseNet, which identifies key points on the body and checks whether they are correctly aligned with the reference posture.

As a second use case, CNN-based models are applied to recognize movements and assign a score to the user's form during exercise. This score helps users identify gaps in their fitness movements that need improvement and provides a fact-based evaluation.

Real-Time Feedback Systems

Real-time feedback is crucial in fitness apps as it enables users to adjust their form during specific workouts, reducing the chances of injury and increasing the effectiveness of their workouts. Real-time analysis, tracking of body movements, and providing feedback in the same live feed [27] are tasks that deep learning models can accomplish. Real-time fitness evaluation systems operate similarly to the system described in the work by Chen et al. [2], comparing the user's posture frame by frame to the ideal movement using deep learning algorithms.

Also, users' posture during yoga and workout exercises is tracked with the help of real-time pose estimation in the Fitness Tutor system. Since such systems constantly compare the user's pose to a reference and prompt necessary corrections, workout accuracy is improved even without human guidance.

Mobile Fitness Applications

The use of mobile devices has continued to advance due to their portability and flexibility, spurring the creation of fitness applications with deep learning models to estimate a person's pose and the quality of their movements. These applications allow users to monitor their fitness, track their progress, receive real-time feedback, and correct their mistakes with the help of their smartphones.

Models such as BlazePose [28] and ConvNeXtPose have been adapted for mobile devices, enabling pose estimation algorithms to be executed without the need for substantial computational resources. Fitness applications on smart mobile devices can record a user's motion in real-time and make corresponding recommendations [29]. This has made deep learning-based fitness tools accessible to many people, especially those who prefer exercising at home or outdoors.

Injury Prevention and Rehabilitation

There are also methods using deep learning models to protect people against injuries and assist with rehabilitation by ensuring that exercises are performed with proper body positioning. Poor mechanics while exercising can lead to overuse injuries, which is why addressing these issues early is crucial. Many applications have built-in pose estimation [30], allowing them to inform users about their posture and suggest modifications to minimize stress on joints and muscles.

In rehabilitation environments, deep learning models can monitor patient movements and ensure that physical therapy exercises are performed correctly. These systems provide feedback and adjustments to help patients recover from injuries and avoid causing further harm to themselves.

Challenges and limitations

Although deep learning has brought significant improvements to fitness applications through pose estimation and real-time feedback, several problems and drawbacks remain. These concerns impact the efficacy, usability, and feasibility of such systems, especially in real-world conditions. The section below discusses some of the major challenges related to data requirements, efficiency, accuracy, and more.

Data Requirements

Pose estimation, in particular, is a vital application of deep learning models that require extensive volumes of accurately labeled data with diverse variations. Human pose estimation models need large amounts of data containing multiple poses and variations in body structure, background, and illumination. For instance, research papers discussing the training of these systems often highlight that they work well with curated images of fitness models, which have been prepared with considerable data enhancement and labeling.

Gathering such accurate datasets is time-consuming and expensive, and the data must incorporate as many human motions as possible to achieve generalization. Datasets like COCO or Human3.6M [12] are more suitable for 2D and 3D pose estimation [10]; however, a particular fitness movement or the number of samples can be insufficient, thereby limiting the model's application across different exercises.

Generalization Across Body Types and Movements

One of the key challenges in fitness applications is ensuring that deep learning models generalize well across different body types, ages, and fitness levels. Most pose estimation models are trained on datasets with limited variability, meaning they may not perform equally well for all users. Variations in height, weight, body shape, and flexibility can affect how well the model detects key body joints, particularly in complex fitness movements.

The study on human pose estimation also notes that achieving high accuracy in recognizing diverse fitness poses remains a challenge. Fitness apps that rely on these models may fail to provide accurate feedback if the model is not fine-tuned for various body types or if the user performs non-standard movements. For example, exercises like yoga or pilates, which involve intricate postures, may be harder to assess accurately without extensive training data for those specific movements.

Accuracy vs. Efficiency Trade-Offs

There is often a trade-off between the accuracy of a deep learning model and its computational efficiency. Highly accurate models, such as those used for 3D pose estimation [10], require more processing power and memory, making them less suitable for real-time use on mobile devices. A fitness movement evaluation system that incorporates efficiency in its CNN-based system may be highly effective but not as accurate for detailed evaluations, as it may take a lot of time to complete complex assessments. Meanwhile, lightweight models such as PoseNet excel in speed performance but are less accurate when identifying specific poses or working with occlusions or low-resolution inputs. Achieving the right balance between these factors remains a considerable challenge, especially in mobile and real-time applications where users expect speedy results in a small memory space.

Future scope

The future of deep learning in fitness holds significant promise, with a lot of scope for potential advancements:

Enhanced Data Collection and Diversity

The next steps should focus on expanding datasets to cover a greater diversity of fitness movements, human body types, and conditions. Expanding new fitness-specific datasets that provide detailed description of exercises such as yoga, pilates, and strength training will help improve model feedback across a wider range of activities.

Integration With Wearable Devices

Additional physiological data from wearable sensors mounted on the human body can enhance any pose estimation models by including muscle activation, joint angles, and heart rate data. Integration of deep learning [11] with the movement data from sensors could help in the detailed analysis of movements, which could assist in the prevention of injuries and better rehabilitation.

Personalization and Adaptation

Future fitness applications should use adaptive learning to adjust according to the user's progress and exercise goals. This probably means that these systems could provide more personalised and thus potentially more efficient workouts by adjusting training schedules and correction feedback in real time.

Extended Use of AR and VR in Fitness

With deep learning integrated with AR and VR [8], it is possible to enhance fitness by providing engaging experiences. Future systems can create virtual environments where users interact with virtual trainers, receive feedback on their positions, and correct them on the spot. Additionally, users can track their progress in a virtual gym.

Expanding the Scope of 3D Pose Estimation

Although significant progress has been made in 3D pose estimation, providing increased accuracy in fitness movement analysis, future research could focus on incorporating 3D motion capture and biomechanical analysis techniques to enhance the understanding of movements. Additional advancements that could intensify these improvements include the creation of real-time 3D models that occupy reasonable space on mobile devices while maintaining high accuracy and efficiency.

Conclusions

The use of deep learning has changed the way people assess and transform human movements in fitness applications. With advanced models such as CNNs, PoseNet, and ConvNeXtPose, pose estimation enhancements have enabled real-time workout feedback and posture adjustments. These technologies have afforded people the opportunity to engage in a more specific and specialized mode of training without the need to hire a physical trainer. While correcting the subject's movements in real-time and enabling applications in AR, the role of deep learning in fitness tracking and assessment has emerged. However, challenges remain, including data requirements, computational complexities, and generalization. For feedback to be effectively implemented across a range of environments, body types, and exercises, precision is critical. Also, the problem of achieving an optimal combination of speed and reliability, particularly in correctly identifying the search query in mobile applications, represents a significant issue for future advancements. Current deep learning applications in fitness are still developing, and further research will be needed to eliminate these drawbacks and optimize the functionality of these technologies.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Dhanashree A. Phalke, Varsha Kotipalli

Critical review of the manuscript for important intellectual content: Dhanashree A. Phalke, Varsha Kotipalli, Priya Ranjan, Yash Pawar, Pranjal Bharat

Supervision: Dhanashree A. Phalke

Acquisition, analysis, or interpretation of data: Varsha Kotipalli, Priya Ranjan, Yash Pawar, Pranjal Bharat

Drafting of the manuscript: Varsha Kotipalli, Priya Ranjan, Yash Pawar, Pranjal Bharat

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Chang W-J, Su J-P, Chen C-H, Liu S-C, Chang C-C, Tsai M-L: iFitness: A deep learning-based physical fitness motion. 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan. 2021, 458-459. 10.1109/GCCE53005.2021.9621944
2. Chen K-Y, Shin J, Hasan MAM, Liaw J-J, Yuichi O, Tomioka Y: Fitness movement types and completeness detection using a transfer-learning-based deep neural network. Sensors. 2022, 22:5700. 10.3390/s22155700
3. Ji H, Githinji KS, Terada K: AI fitness coach at home using image recognition [PREPRINT]. 2022, 10.21203/rs.3.rs-2047283/v1
4. Wang Y, Zhu X, Qu C: Fitness movements recognition and evaluation based on LSTM. Journal of Physics: Conference Series. 2019, 1302:032057. 10.1088/1742-6596/1302/3/032057
5. Chen Y, Liu X: Design of fitness movement detection and counting system based on MediaPipe. Communications in Computer and Information Science. 2024, 1880:77-92. 10.1007/978-981-99-5971-6_6
6. Wang J, Qiu K, Peng H, Fu J, Zhu J: AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. MM '19: Proceedings of the 27th ACM International Conference on Multimedia.

2019, 2228-2230. 10.1145/3343031.3350609

7. Jeon H, Yoon Y, Kim D: Lightweight 2D human pose estimation for fitness coaching system. 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Jeju, Korea (South). 2021, 1-4. 10.1109/ITC-CSCC52171.2021.9501458
8. Ueta M: Improving mirror fitness through augmented reality technology. 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China. 2022, 186-194. 10.1109/ICBAIE56435.2022.9985796
9. Papandreou G, Zhu T, Chen L-C, Gidaris S, Tompson J, Murphy K: PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model [PREPRINT]. arXiv:1803.08225. 10.48550/arXiv.1803.08225
10. Newell A, Yang K, Deng J: Stacked hourglass networks for human pose estimation [PREPRINT]. arXiv:1603.06937. 2016. 10.48550/arXiv.1603.06937
11. Mahendran N: Deep learning for fitness [PREPRINT]. arXiv:2109.01376. 10.48550/arXiv.2109.01376
12. Toshev A, Szegedy C: DeepPose: Human pose estimation via deep neural networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. 2014, 1653-1660. 10.1109/CVPR.2014.214
13. Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y: OpenPose: Realtime multi-person 2D pose estimation using part affinity fields [PREPRINT]. arXiv:1812.08008. 10.48550/arXiv.1812.08008
14. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition [PREPRINT]. arXiv:1512.03385. 10.48550/arXiv.1512.03385
15. Xu P, Du R, Zhang Z: Predicting pipeline leakage in petrochemical system through GAN and LSTM. Knowledge-Based Systems. 2019, 175:50-61. 10.1016/j.knosys.2019.03.013
16. Ionescu C, Papava D, Olaru V, Sminchisescu C: Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014, 36:1325-1339. 10.1109/tpami.2013.248
17. Ba JL, Kiros JR, Hinton GE: Layer normalization [PREPRINT]. arXiv:1607.06450. 10.48550/arXiv.1607.06450
18. Xiao B, Wu H, Wei Y: Simple baselines for human pose estimation and tracking. Computer Vision - ECCV 2018. Ferrari V, Hebert M, Sminchisescu C, Weiss Y (ed): Springer, Cham; 2018. 11210:466-481. 10.1007/978-3-030-01231-1_29
19. Higgins JP: Smartphone applications for patients' health and fitness. The American Journal of Medicine. 2016, 129:11-19. 10.1016/j.amjmed.2015.05.038
20. Su C, Li J, Peng Y, Chen Y: Chinese metaphor sentiment computing via considering culture. Neurocomputing. 2019, 352:33-41. 10.1016/j.neucom.2019.03.087
21. Flores A, Hall B, Carter L, Lanum M, Narahari R, Goodman G: Verum fitness: An AI powered mobile fitness safety and improvement application. 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI). 2021, 980-984. 10.1109/ICTAI52525.2021.00156
22. Kanase RR, Kumavat AN, Sinalkar RD, Somani S: Pose estimation and correcting exercise posture. ITM Web of Conferences. 2021, 40:03031. 10.1051/itmconf/20214003031
23. Win S, Thein TLL: Real-time human motion detection, tracking and activity recognition with skeletal model. 2020 IEEE Conference on Computer Applications (ICCA). 2020, 1-5. 10.1109/ICCA49400.2020.9022822
24. Srijeyanthan K, Thusyanthan A, Joseph CN, Kokulakumaran S, Gunasekara C, Gamage C: Skeletonization in a real-time gesture recognition system. 2010 Fifth International Conference on Information and Automation for Sustainability. 2010, 213-218. 10.1109/ICIAFS.2010.5715662
25. Rawat W, Wang Z: Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation. 2017, 29:2352-2449. 10.1162/neco_a_00990
26. Chu X, Ouyang W, Li H, Wang X: Structured feature learning for pose estimation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 4715-4723. 10.1109/CVPR.2016.510
27. Yang W, Ouyang W, Li H, Wang X: End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 3073-3082. 10.1109/CVPR.2016.335
28. Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, Grundmann M: BlazePose: On-device real-time body pose tracking [PREPRINT]. arXiv:2006.10204. 10.48550/arXiv.2006.10204
29. Hendrycks D, Gimpel K: Gaussian error linear units (GELUs) [PREPRINT]. arXiv:1606.08415. 10.48550/arXiv.1606.08415
30. Woo S, Debnath S, Hu R, et al.: ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders [PREPRINT]. arXiv:2301.00808. 10.48550/arXiv.2301.00808