Exercise Posture Suggestion System using Deep Learning Techniques

Paulo Mendoza
Computer Engineering Student
Technological Institute of the Philippines – Quezon City
Bulacan, Philippines
qpdcmendoza@tip.edu.ph

Benedick Labbao
Computer Engineering Student
Technological Institute of the Philippines – Quezon City
Rizal, Philippines
qbdlabbao@tip.edu.ph

Roman Richard
Computer Engineering
Technological Institute of the Philippines – Quezon City
Quezon City, Philippines
rrichard.cpe@tip.edu.ph

Maintaining proper posture during exercise is crucial for reducing the risk of injuries and maximizing the effectiveness of workouts. This paper presents an Exercise Posture Suggestion System utilizing deep learning techniques to analyze and recommend correct exercise posture in real-time. The system incorporates pose estimation models, human activity recognition algorithms, and error detection mechanisms to provide personalized feedback to users. Pose estimation models including ResNet-50, YOLOv8, and YOLO-NAS are evaluated based on their performance metrics and execution times. Human activity recognition models such as Conv(2+1)D with ResNet (Residual Network), 3D CNN, and CNN with Bidirectional LSTM are trained and tested using the UCF-101 dataset. An error detection algorithm, focusing on key body angles for specific exercises, enhances the accuracy of posture suggestions. The software interface is developed using Unity, providing an intuitive platform for users to receive feedback and visualize correct posture examples. The system demonstrates promising results in optimizing workout routines and reducing injury risks, with further refinement and validation necessary for widespread adoption in fitness settings. Integration of advanced deep learning techniques and human-computer interaction will continue to enhance the system's capabilities, contributing to improved health outcomes and enhanced quality of life.

Keywords—Posture Suggestion, Deep Learning, Pose Estimation, Human Activity Recognition, Injury Prevention, Unity

I. Introduction

Exercise is a cornerstone of a healthy lifestyle, contributing to physical fitness, mental well-being, and quality of life. From cardiovascular workouts to strength training and flexibility exercises, the benefits of regular physical activity are well-documented and far-reaching. Engaging in exercise not only helps maintain a healthy weight and improve cardiovascular health but also boosts mood, reduces stress, and enhances cognitive function.

In today's sedentary society, where technological advancements often lead to prolonged periods of sitting and decreased physical activity, prioritizing regular exercise is more important than ever. Whether it's a brisk walk, a yoga session, or a gym workout, finding enjoyable and sustainable ways to stay active is key to promoting longevity and vitality.

Although exercise is important, It is also important during exercise to have proper breathing and good posture, this helps the body to function and will cut muscle strain and injury[1]. But many individuals struggle to maintain correct posture, leading to suboptimal results and increased risk of injury. Proper body posture has been associated with a reduction in incidence of injuries[2,3]. This correlation shows the importance of correct posture in mitigating the risk of exercise-related injuries.

In response to the problem that we encountered, we propose an exercise posture suggestion system that aims to analyze and suggest correct exercise posture to help cut the risk of injury during exercise.

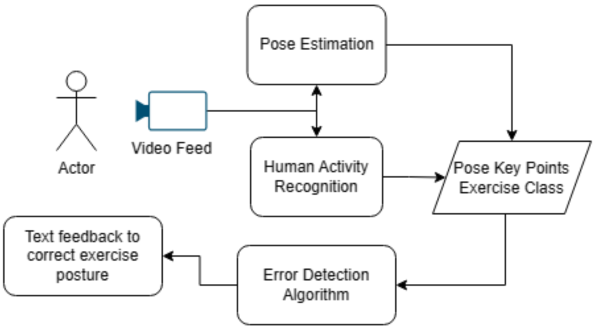
Pose Estimation
Actor
Video Feed
Human Activity Recognition
Pose Key Points Exercise Class
Text feedback to correct exercise posture
Error Detection Algorithm

Fig. 1. Posture Suggestion System Overview

In Fig. 1, this shows the process of the system, the system is composed of a device with a camera and a software. The software contains deep learning models and an algorithm used for suggestions in the posture.

II. Methodology

A. Dataset and Data Pre-processing

The methods used to create the system are grouped into deep learning models and error detection algorithm. The datasets we used was UCF-101 dataset, and we only used the data with the label Push Ups, Lunges, and Squats.

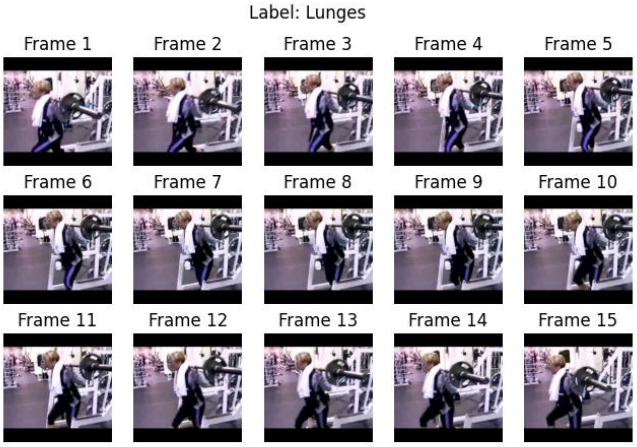

Fig. 2. Frames generated from Lunges label

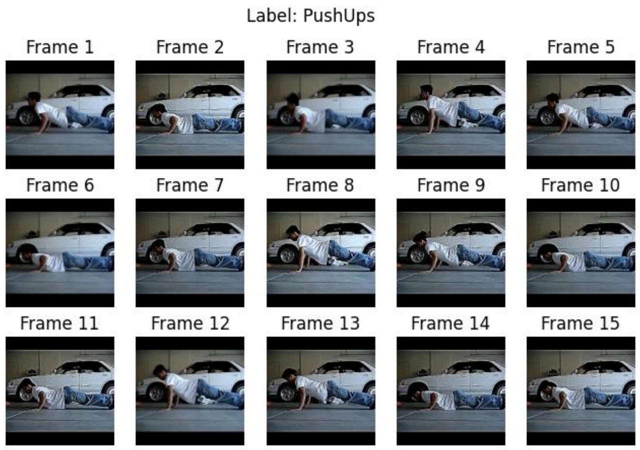

Fig. 3. Frames generated from Push Ups label

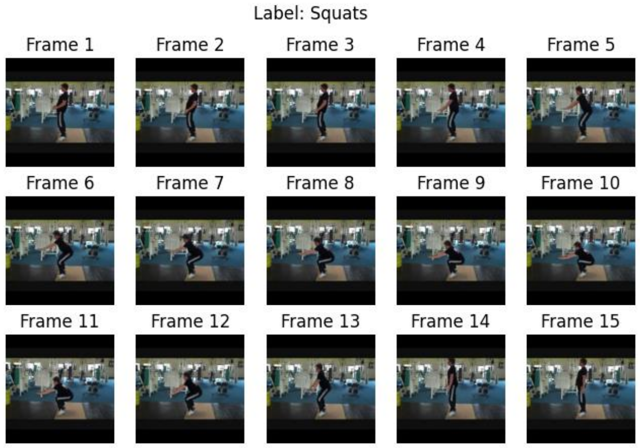

Fig. 4. Frames generated from Squats label

The dataset that we used was then split into each individual exercise repetitions and then we extracted 15 frames from each individual exercise, as seen in fig. 2, 3, and 4. Data augmentation like Gaussian blur, Sharpen, Affine and Flip was applied to each exercises randomly.

B. *Pose Estimation*

The pose estimation models that we tried were ResNet-50 (Residual Network), YOLOv8, and YOLO-NAS models. The ResNet-50 model was train using dataset acquired in the MPII Human Pose Dataset containing images annotated with 16 key body joint locations.

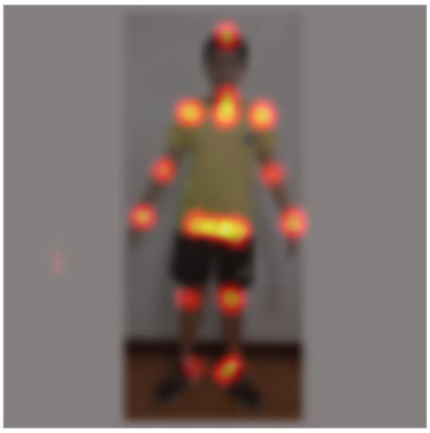

Fig. 5. Output of the ResNet-50 Model

In the fig. 5 we can see the example of output of the pose estimation model, from this we selected the joints as the pixels with red color.

The YOLOv8 model, by ultralytics was trained on COCO 2017 Dataset which has 17 key points. The model's input shape is a 640x640 image with 3 channels and the output is consisting of 17 key points. Next we have the YOLO-NAS model, which uses Neural Architecture Search and it was also trained using COCO 2017. This results in models that potentially have better performance compared to manually designed YOLO models. These models was then evaluated by obtaining their Mean Average-Precision Scores (mAP), execution time and model size, which are all crucial in using the model in production[4].

C. *Human Activity Recognition (HAR)*

Human Activity Recognition (HAR) it is a novel approach used for accurately recognizing and human activities using various signals. Most activity recognition systems are developed using datasets obtained through vision-based and sensor-based devices. In this study, we will use vision-based datasets to create three different model architectures, which are Conv(2+1)D using Residual Network, CNN with LSTM and 3D CNN.

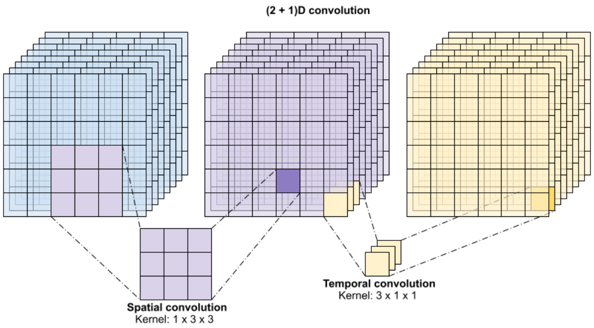

Fig. 6. Conv(2+1)D Layers

First we have the Conv(2+1)D with ResNet model, this deep residual learning is a technique used to train very deep neural networks by addressing the vanishing gradient problem. In fig. 6, we can see that the networks are constructed using residual blocks, which contain shortcut connections that bypass one or more layers. These shortcuts allow the network to learn the difference (residual) between the input and the desired output. This reformulation simplifies the learning process, making it easier to optimize deeper networks and improving performance, as evidenced by significant accuracy gains in image recognition tasks using ResNets[6].

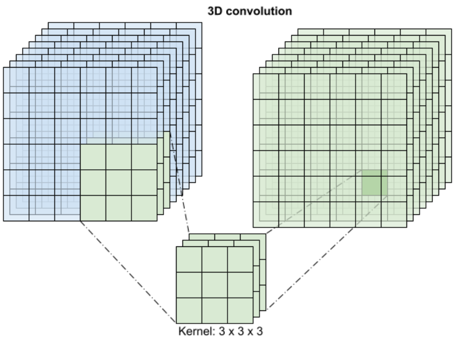

Fig. 7. 3D CNN Layers

The 3D Convolutional Neural Networks (3D CNNs) extend traditional 2D CNNs to three dimensions, allowing them to process volumetric data such as video sequences or medical imaging scans. By using 3D filters that slide over height, width, and depth, 3D CNNs can capture spatial and temporal information simultaneously. This makes them particularly effective for applications like action recognition in videos, where they analyze multiple frames at once, and medical imaging, where they can examine 3D structures within scans. We can the architecture in the fig. 7, which typically includes 3D convolutional and pooling layers that help in learning complex features across all three dimensions. The 3D CNN and the CNN with Bidirectional LSTM model both contains 4 ConvBlock, 2 Fully Connected Layers and the LSTM layers. each ConvBlock contains Conv3D which analyzes the spatiotemporal features of the data along with MaxPooling3D[7].

The models all underwent training on the training dataset, while monitoring via metrics such as loss, accuracy validation loss and validation accuracy on the training and validation set. Upon completion of training, the model's efficacy is evaluated on the test set, employing metrics such as accuracy.

D. *Error Detection Algorithm*

In the Error Detection algorithm, we used a combination of key points for each poses, we determined that at every pose there are only certain parts of the body that needed to be corrected.

TABLE I. IMPORTANT ANGLES

Exercise Classes	***Angles***
Squat	Torso, Legs
Lunges	Left Leg, Right Leg
Planks	Legs, Arms

Fig. 8. Important angles for each exercise classes.

In the fig. 8 we can see the key angles that we need to monitor in each exercises. For the squat, we monitor the angles in the torso and the legs of the user. For lunges, both the legs are monitored and for planks, we monitor the angle of the arms and legs. In the algorithm that we used, it can be seen how the angle of the three points is calculated. We first calculated the two vectors of the angle based on point 1 and point 2, then point 2 and point 3. Next we performed dot product in both vectors, and calculated both magnitude of the vectors. Using the dot product and magnitude of the vectors, we can get the cosine of the angle, and finally we can use inverse cosine function to get the angle theta.

III. TESTING AND RESULTS

A. *Pose Estimation*

The testing was conducted by using sample images and using camera, we split the data into testing and training data for the validation of the model.

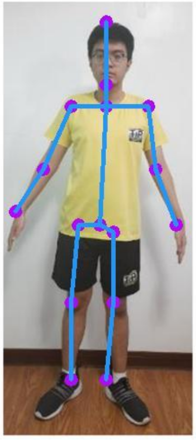

Fig. 9. Sample Output of ResNet-50 Model

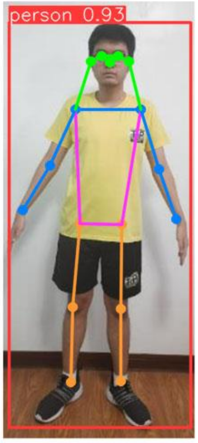

Fig. 10. Sample Output of YOLOv8 Model

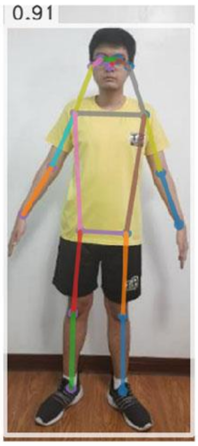

Fig. 11. Sample Output of YOLO-NAS Model

The Fig. 9 to Fig. 11 shows the example output of the pose estimation model.

TABLE II. POSE ESTIMATION MODEL COMPARISON

Model	***mAP Scores***	***Execution Time***	***Model Size***
ResNet-50	96.351	265.8ms	132.7 MB
YOLOv8	95.329	600.5ms	45.7 MB
YOLO-NAS	88.989	722.8ms	60.2 MB

Fig. 12. Metrics for Pose Estimation Model

According to fig. 12, we can see that ResNet-50 performed better than YOLOv8 and YOLO-NAS in terms of the performance and speed, but this is because of the gap between the model size, we can see that both YOLOv8 and YOLO-NAS have significantly lower size compared to ResNet, this is because YOLO models boasts in have smaller size while maintaining good performance.

B. Human Activity Recognition (HAR)

For our Human Activity Recognition model, we used the testing data from the dataset and performed predictions and these are the results:

TABLE III. ACCURACY RESULTS OF HAR MODELS

Model	*Accuracy*	*Execution Time/Step*	*Model Size*
Conv(2+1)D using ResNet	92.40%	36.65ms	395.6 MB
CNN with LSTM	85.00%	30.14ms	751.3 MB
3D CNN	99.04%	35.20ms	2821.2 MB

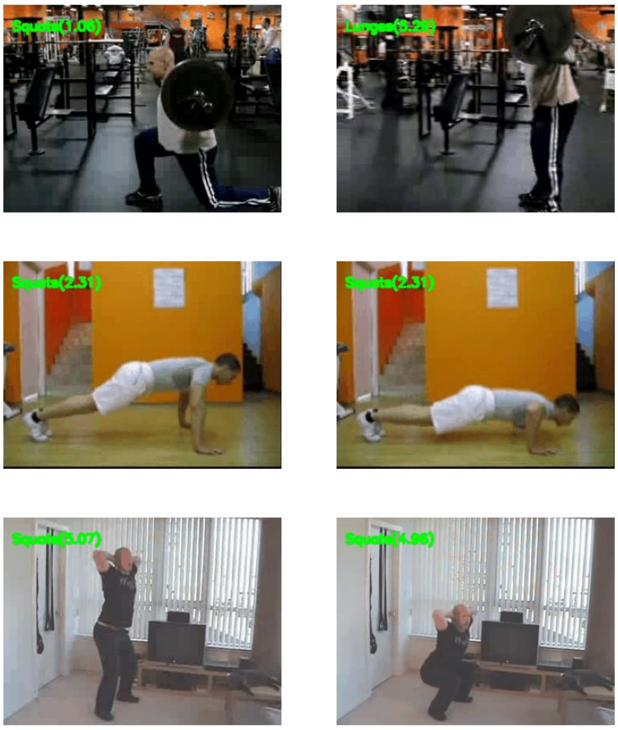

Fig. 13. Conv(2+1)D using ResNets Results

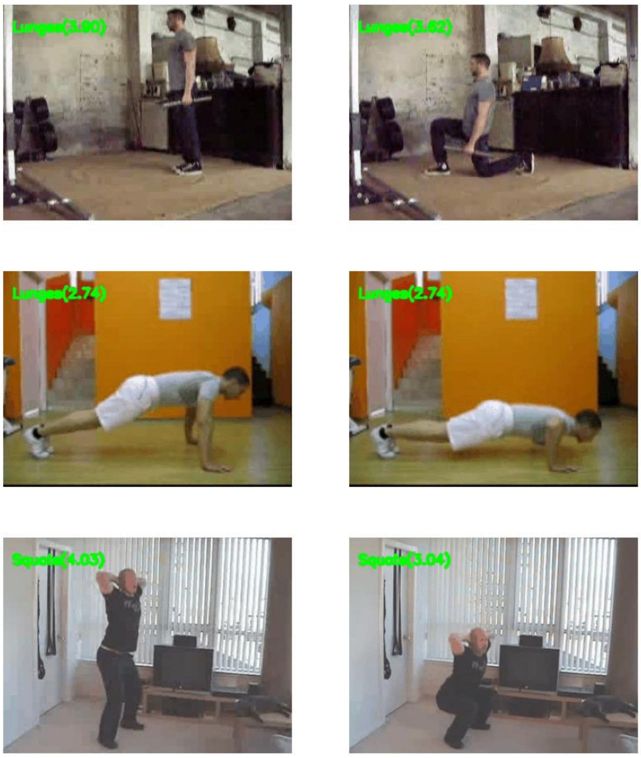

Fig. 14. CNN with LSTM Results

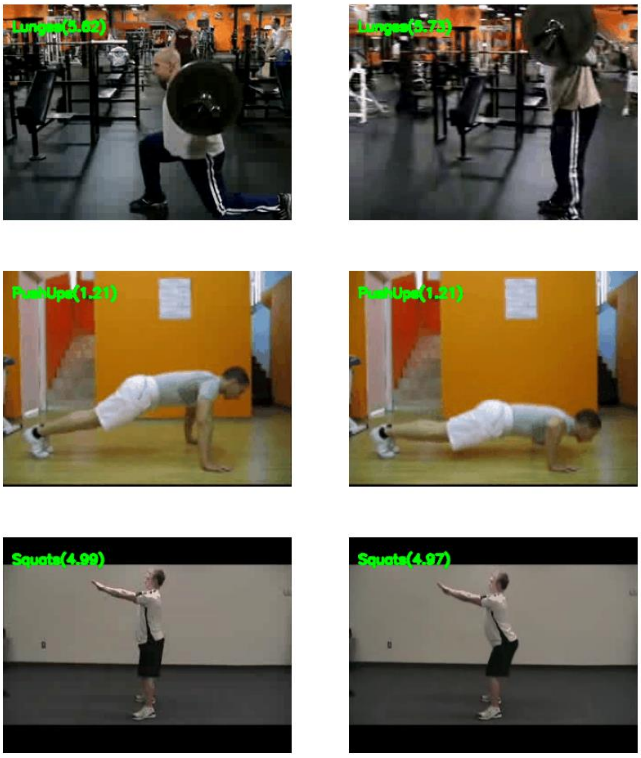

Fig. 15. 3D CNN Results

IV. DISCUSSION

Three Pose Estimation models were evaluated: ResNet-50, YOLOv8, and YOLO-NAS. While ResNet-50 showed superior performance, YOLO models offered smaller sizes with acceptable accuracy, making them viable alternatives.

Three HAR model were also tested using the UCF-101 dataset. Our Conv(2+1)D + ResNet model for HAR has achieved 92.40% accuracy, we can see the accuracy of 3D CNN shows the steady increase in accuracy and has also managed to achieved 99.04% accuracy. The CNN with Bidirectional LSTM also shows steady increase but it has the lowest accuracy of the 3 models, with an accuracy of 85%. The 3DCNN model achieved highest accuracy, and zigzag pattern was observed in all models, emphasizing the need for an even smaller learning rate.

An algorithm based on key points and specific body angles was used for error detection. This targeted approach enhances posture suggestions tailored to individual exercises, reducing injury risks. Unity was chosen for its user-friendly interface. The software integrates deep learning models for real-time posture analysis and feedback, enhancing user engagement and understanding.

V. Conclusion

The exercise posture suggestion system offers a promising solution for individuals seeking to optimize their workout routines while minimizing the risk of injury. Continued refinement and validation through user feedback will be essential for ensuring the system's effectiveness and widespread adoption in fitness settings. As technology continues to evolve, integrating advancements in deep learning and human-computer interaction will further enhance the system's capabilities, ultimately contributing to improved health outcomes and enhanced quality of life.

Acknowledgment

We would like to express our gratitude to the online communities and resources that have contributed to the success of this project. Our work would not have been possible without the availability of online data and the wealth of knowledge shared by individuals and organizations online. We acknowledge and appreciate the valuable contributions of these sources in shaping our understanding and implementation of advanced techniques in our research.

References

[1] D. Rellinger, “Regular breathing and proper posture when exercising is important,” *MSU Extension*, Dec. 22, 2016. https://www.canr.msu.edu/news/regular_breathing_and_proper_posture_when_exercising_is_important

[2] Dawid Koźlenia and Katarzyna Kochan-Jacheć, “The Impact of Interaction between Body Posture and Movement Pattern Quality on Injuries in Amateur Athletes,” *Journal of clinical medicine*, vol. 13, no. 5, pp. 1456–1456, Mar. 2024, doi: https://doi.org/10.3390/jcm13051456.

[3] Audrey, “Why Having Proper Form & Exercise Techniques is Important,” *Jack City Fitness*, Jan. 21, 2021. https://jackcityfitness.com/why-having-proper-fitness-form-technique-is-important/

[4] B. Xiao, H. Wu, and Y. Wei, “Simple Baselines for Human Pose Estimation and Tracking,” *arXiv:1804.06208 [cs]*, Aug. 2018, Available: https://arxiv.org/abs/1804.06208

[5] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, “A Closer Look at Spatiotemporal Convolutions for Action Recognition,” arXiv:1711.11248 [cs], Apr. 2018, Available: https://arxiv.org/abs/1711.11248v3

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv.org*, Dec. 10, 2015. https://arxiv.org/abs/1512.03385

[7] J. You and J. Korhonen, "Deep Neural Networks for No-Reference Video Quality Assessment," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 2349-2353, doi: 10.1109/ICIP.2019.8803395.