FMDB ***Transactions on Sustainable Intelligent Networks***

Optimized Real-Time Pose Estimation with High-Speed Precision and Alert System

R. Regin[1], S. Thkashinraj[2,*], J. Koushik Raghav[3], S. Jude Roosevelt[4], G. Padmapriya[5], Chunhua Deming[6]

[1,2,3,4]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[5]Department of Chemistry, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[6]Information Technology Discipline, NUS Graduate School (NUSGS), National University of Singapore, Queenstown, Singapore.
regin12006@yahoo.co.in[1], ts0430@srmist.edu.in[2], kr4455@srmist.edu.in[3], sr5815@srmist.edu.in[4], ppstminex@gmail.com[5], chunhuademing@gmail.com[6]

Abstract: The posture detection application is a cutting-edge tool leveraging TensorFlow Lite and OpenCV to analyze real-time human movement. Designed for motion tracking, occupational safety, rehabilitation, and interactive technology applications, the system provides accurate predictions by analyzing anatomical landmarks. It processes live video and still images, providing high reliability and accuracy essential for sports training, ergonomic evaluation, and healthcare. The application provides instant feedback on posture and technique, allowing users to improve their performance dynamically. It issues alerts if key indicators are not detected within a specified time, ensuring timely intervention in critical scenarios such as security and elderly care. Features include intuitive controls, a confidence threshold slider to adjust accuracy, and support for uploading still images. The user interface visually represents key landmarks, facilitates data interpretation, and features voice commands for improved usability. Combining advanced machine learning and user-centric design, the system supports real-time monitoring and analysis, making it versatile and effective for improving performance, ensuring security, and delivering interactive experiences in various environments.

Keywords: Human Pose Estimation; MoveNet Model; Real-Time System; Keypoint Detection; Inactivity Monitoring; Occupational Safety; Multi-Person Detection; TensorFlow Lite; Graphical User Interface (GUI).

Received on: 17/04/2024, **Revised on:** 26/06/2024, **Accepted on:** 17/08/2024, **Published on:** 09/09/2024

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSIN

DOI: https://doi.org/10.69888/FTSIN.2024.000287

Cite as: R. Regin, S. Thkashinraj, J. K. Raghav, S. J. Roosevelt, G. Padmapriya, and C. Deming, “Optimized Real-Time Pose Estimation with High-Speed Precision and Alert System,” *FMDB Transactions on Sustainable Intelligent Networks*., vol.1, no.3, pp. 178–189, 2024.

1. Introduction

Human Posture Estimation (HPE) has developed as a vital zone in computer vision, empowering frameworks to distinguish and analyze human developments by recognizing key body joints and developing a skeletal model. This innovation has significant applications in healthcare, sports, gaming, mechanical autonomy, and surveillance. HPE permits machines to capture

[*]Corresponding author.

basic focuses on the human body, such as the head, shoulders, elbows, knees, and lower legs, giving important knowledge into human pose and development flow. In healthcare settings, HPE is utilized to screen patients amid restoration, making a difference in the clinician's evaluation and direct recuperation by following physical developments. Additionally, in sports, HPE can give real-time criticism to competitors, permitting them to optimize execution by moving forward with their pose and procedure while preparing. In past healthcare and sports, HPE has applications in gaming and intelligent frameworks, where it encourages motion acknowledgement and improves client encounters. In mechanical technology, this innovation empowers robots to imitate human activities, contributing to robotization in different businesses, including fabricating and administration. As the innovation progresses, its effect on bridging the interaction between people and machines becomes progressively critical, permitting a more consistent integration of automated frameworks into existence. The combination of HPE with other advances can advance brilliant frameworks to understand and anticipate human behaviour, eventually cultivating a more secure and productive interaction between people and machines in real-world applications [1]. The rapid evolution of deep learning techniques has significantly improved the accuracy and efficiency of systems that assess human pose. Convolutional Neural Networks (CNNs) have become the backbone of modern pose estimation frameworks, enabling more accurate detection of human joints under various conditions. These deep-learning models can tackle various problems, including occlusion, variations in lighting, and body orientation. Accurate pose detection in complex environments can be achieved by training CNNs on patterns and features based on large datasets.

A notable advancement from HPE is using multi-stage CNN architectures, which improve accuracy by incrementally refining predictions across multiple processing stages. This approach allows the model to correct minor localization errors in subsequent phases, ensuring higher accuracy of joint detection. Real-time applications such as rehabilitation and sports training will greatly benefit from these advances by providing instant feedback to users. In rehabilitation, patients can receive guidance on adjusting their movements to prevent injuries and improve recovery outcomes. Athletes can optimize their performance by utilizing real-time location analysis during training. The advancement of deep learning technologies is anticipated to enhance their accuracy, durability, and usefulness, resulting in human localization becoming the most crucial technology after computer vision [2].

Real-time assessment of human posture is becoming increasingly important to improve workplace safety and efficiency, especially in physical and manual labour industries. On construction sites and manufacturing plants, workers are often required to perform physically demanding tasks that require correct posture to minimize the risk of injury. With a real-time HPE system in place, worker posture can be continuously monitored, and unsafe positions or movements can be instantly alerted to by the worker or supervisor. This proactive approach helps prevent musculoskeletal injuries, improving workplace safety and productivity. Additionally, real-time posture assessment has become critical in driver monitoring systems, where continuous observation of driver posture plays a key role in fatigue detection. Other symptoms, such as subtle posture changes, such as drooping head or excessively leaning forward, may indicate possible drowsiness and provide timely warnings to prevent accidents.

This technology is particularly relevant in industries where long car journeys are common, as it significantly reduces the risk of fatigue-related accidents. With the increasing importance of safety and automation in industry, integrating real-time human posture assessment systems into operations will improve both safety measures and overall efficiency. By providing instant feedback, these systems protect workers and promote a safety culture that encourages adherence to best practices in physical activity [3]. This paper presents a detailed account of estimating human poses in real-time using the TensorFlow Lite MoveNet model and its built-in alerting mechanisms. MoveNet was designed for efficient real-time applications on resource-constrained devices such as smartphones and embedded systems. By detecting 17 key points on the human body, the model generates a skeletal representation that enables accurate pose analysis. The system can process live video streams and still images, allowing users to upload pre-recorded images for analysis when live video is impractical.

For example, still images can be used in clinical settings to assess specific moments in a patient's movement patterns and gain valuable insights into their rehabilitation progress. The system also includes advanced image processing capabilities to ensure input images are scaled and padded to fit the model's requirements, ensuring consistency in pose estimation results. TensorFlow Lite optimizations allow the system to work seamlessly on various devices, from mobile phones to specialized embedded platforms. This versatility makes the system applicable in a wide range of situations, from personal health monitoring to professional sports analysis, contributing to a real-time pose estimation solution that meets the needs of different users [4].

Although significant advances have been made in estimating human pose, several challenges remain, especially when considering real-world conditions. Occlusion, varying lighting, and complex body orientations continue to make accurate pose detection difficult. For example, when objects or other people partially occupy body parts, existing models can have difficulty accurately identifying all key points. Additionally, poor lighting conditions can lead to errors in joint detection, negatively impacting the overall effectiveness of a pose estimation system. Another continuing challenge is detecting multiple people in crowded environments, where current models often have difficulty tracking multiple people simultaneously. This limitation is especially relevant for applications such as sports analytics, public surveillance, and social interaction analysis, where careful monitoring of multiple people is critical. Enhancements to model architectures and training methods for addressing these problems are anticipated in future developments related to human pose estimation. Moreover, optimizing pose estimation algorithms for edge devices could significantly improve accessibility and usability, making them ideal for real-time monitoring in various fields such as healthcare and public safety. As these technologies continue to evolve, the demand for durable and adaptable pose estimation systems will keep growing, which could lead to new applications [5].

2. Literature Review

Human pose assessment has evolved dramatically in recent years, mainly through applying deep learning techniques and the availability of large annotated datasets. Early methods relied heavily on geometric approaches and statistical shape models but faced challenges in occlusions and complex human poses. For example, traditional triangulation methods required accurate camera calibration and were difficult to adapt to dynamic environments. The introduction of Convolutional Neural Networks (CNNs) changed the landscape of pose estimation, allowing models to learn complex features directly from images.

A notable contribution is the DeepPose model developed by Toshev and Szegedy [5]. This model uses a regression-based framework to directly predict body joint positions from images. The model showed significant improvement over previous techniques and demonstrated the ability of deep learning to achieve higher accuracy and robustness in pose estimation tasks. The success of DeepPose paved the way for subsequent research efforts, confirming that deep learning can effectively address the challenges associated with analyzing human motion in real-time applications [6]. Significant progress was made by Cao et al. [1], who introduced the Part Affinity Field (PAF) framework, which allows models to associate key points from multiple subjects effectively in crowded scenes [7]. This framework improved the applicability of pose estimation models in real-world environments where occlusions and overlapping body parts are frequent.

The OpenPose framework was also developed by Cao et al. [7] and combines part-based representations and efficient processing algorithms to achieve state-of-the-art performance in multi-person pose estimation tasks [8]. The open-source nature of OpenPose encourages collaboration and innovation within the research community, driving further applications in various fields such as sports analytics and human-computer interaction. As researchers build on this foundation, continuous refinement of multi-person pose estimation techniques remains critical to improve accuracy and real-time performance in various scenarios [9]. Recent advances have increasingly focused on improving the real-time performance and accuracy of pose estimation models while considering the limitations of mobile and embedded devices. The MoveNet model developed by Google Research is an example of this trend, as it is specifically designed for efficient pose estimation on mobile platforms. MoveNet's lightweight architecture enables on-device processing, meeting the growing demand for real-time applications in augmented reality and fitness tracking [10]. The move to a lightweight model is essential as it extends accessibility and usability, allowing advanced pose estimation capabilities to be used effectively across a range of consumer devices.

The focus on optimizing models for resource-constrained environments reflects the growing industry interest in mobile solutions that provide real-time feedback to users. As posture estimation applications increase, integrating these lightweight models into everyday devices will improve user experience in personal fitness, interactive gaming, and augmented reality experiences [11]. Integrating posture estimation into real-time monitoring and warning systems has become an important area of research. For example, Li et al. [8] introduced a robust posture monitoring system for industrial environments that continuously assesses workers' posture to prevent injuries [12]. The system uses real-time pose estimation techniques to detect incorrect postures and provide timely warnings to improve safety measures in high-risk workplaces.

Similarly, He et al. [9] present a driver fatigue detection system that uses pose estimation to monitor driver behaviour and enable timely warnings when signs of fatigue are detected. These applications highlight the importance of human pose estimation in safety-critical environments and demonstrate its potential to significantly improve safety outcomes in various environments. As the demand for intelligent surveillance systems continues to increase, integrating pose estimation techniques with alerting mechanisms will be critical to improving security across various industries [14]. The focus on real-time applications demonstrates the transformative impact of pose estimation and highlights its importance in improving operational safety and efficiency. Although significant progress has been made in human pose estimation, many challenges remain in achieving high accuracy, especially in scenarios involving occlusions, complex backgrounds, and multiple people.

Zhang et al. [10] addressed the occlusion issue by using synthetic data for training, which significantly improved the robustness of their model when applied to crowded scenes. Exposing the model to various synthetic training examples has improved the generalization capabilities of pose estimation systems. However, reliance on synthetic data poses challenges in bridging the gap between synthetic and real images, affecting performance in real-world applications. In addition, existing models often suffer from real-time performance issues when detecting multiple people in dynamic environments. The results reported by Li et al. [11] in their study proposed a novel architecture framework that effectively balances detection accuracy and processing efficiency, yielding promising results in maintaining real-time performance while accurately estimating the poses of multiple people [16]. As the field continues to evolve, these ongoing challenges must be addressed to improve the applicability of human pose estimation technology in real-world scenarios and pave the way for further innovation and improved accuracy [13].

3. Objective

The main goal of the human anatomy analysis application is to provide an intuitive and convenient way to describe and interpret human anatomy from real-time video. The project focuses on leveraging the power of computer vision and machine learning to provide users with powerful tools to track, monitor, and analyze body movements, especially without webcams. Provide accurate analysis that allows users to better understand their movements by tracking fitness, rehabilitation, interaction, and other modalities. The goals of this project are:

- Create a user interface that provides clear and effective controls for the movements of novices, loads images for static analysis, and sets confidence levels for orientation recognition.
- Use TensorFlow Lite and OpenCV to implement real-time prediction algorithms to identify and track important physical points from network cameras in real-time videos. Allows users to instantly view and describe their poses by marking and creating skeletons on video or still images. Potential problems in an environment where perception is critical.
- Use TensorFlow Lite's optimization model and OpenCV for real-time processing to optimize application performance by allowing applications to access different hardware. It provides a new level of intimacy and efficiency by improving the user's ability to interact with the digital environment through targeted analysis. The Physics Analysis application is designed to be easy to use while providing tools to measure the body's energy, contributing to various health, fitness, and social media.

4. Methodology

The methodology used in this study comprises several phases, including system architecture design, data collection, model selection, implementation, and evaluation. Each phase is vital for developing a robust human pose estimation system capable of real-time processing and generating alerts based on detected poses.

4.1. System Architecture

The proposed system architecture consists of three main components: data acquisition, the pose estimation model, and the alert system. This architecture is designed to integrate these components seamlessly, ensuring efficient communication and optimal real-time performance. Each component is essential for processing visual information and providing timely feedback based on human pose estimation (Figure 1).

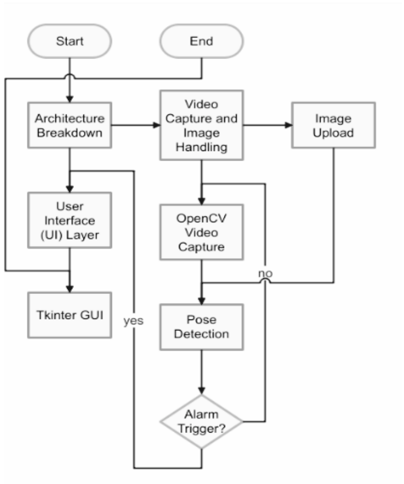

Figure 1: Architecture Diagram

4.1.1. Data Collection

The system employs a webcam or camera module to capture real-time video streams or images to facilitate pose recognition. The captured video feed is processed frame-by-frame to recognize human poses. This continuous data acquisition ensures the model receives diverse poses in varying conditions, improving its robustness and accuracy in real-world applications. Pose Estimation Model: At the system's core lies the human pose estimation model, which identifies key points on the human body and estimates the overall pose. This study utilizes the MoveNet model, a lightweight and optimized framework developed by Google, which is known for its high accuracy and efficiency in real-time applications. MoveNet's design enables it to detect human poses swiftly, making it suitable for various applications where speed is critical.

4.1.2. Warning System

The warning system is responsible for monitoring detected poses against predefined criteria. The system triggers an alarm when certain conditions are met—such as prolonged inactivity or unstable poses. The warning mechanism may include audio signals, visual indicators, or notifications to ensure that users are promptly informed of potential issues. The architecture's design allows for easy integration of these alerts, enhancing the system's utility in safety-critical environments.

4.2. Data Collection

Data collection is pivotal in developing an effective human pose estimation system. The system requires a diverse dataset to train the pose estimation model effectively.

4.2.1. Dataset Selection

The selected dataset should encompass various human poses and scenarios for training and evaluation. This study considered two main datasets: COCO (Common Objects in Context), which provides over 200,000 images annotated with human pose key points, and the MPII Human Pose Dataset, which contains images of individuals engaged in various activities with annotated body joint positions. Both datasets include diverse backgrounds, occlusions, and human activities, making them suitable for training pose estimation models.

4.2.2. Data Preprocessing

Before training, preprocessing of the collected data is essential to ensure consistency and enhance model performance. This includes resizing images to a fixed input size of 192 x 192 pixels to maintain uniformity across the dataset. This resizing step is critical as it meets the input requirements of the MoveNet model, enabling optimal performance.

4.2.3. Normalization

Pixel values are normalized to the range of [0, 1] to improve training efficiency and model stability. Normalization involves dividing pixel values by 255, which helps the model converge faster during the training phase. This step is crucial for facilitating effective learning and reducing potential issues associated with varying brightness and contrast levels in input images.

4.2.4. Data Augmentation

Data augmentation techniques like rotation, mirroring, and colour adjustment enhance model robustness. This approach increases the diversity of the training dataset, helping the model generalize better to unseen data. Augmentation is particularly valuable in preventing over-fitting and ensuring that the model can handle variations in pose and appearance during real-world applications.

4.3. Model Selection

Selecting the appropriate pose estimation model is critical for achieving high accuracy and efficiency in real-time applications. This study chose the MoveNet model for its lightweight architecture and capability to deliver real-time performance.

4.3.1. Model Architecture

MoveNet is based on a single-shot detection approach, employing a backbone network that simultaneously detects multiple key points in a single pass. The architecture includes a feature extraction backbone network and several key point detection layers that predict the coordinates of body joints along with their confidence values. This setup allows MoveNet to detect up to 17 key points corresponding to various joints in the human body, making it highly versatile for different applications.

4.3.2. Model Training

The model was fine-tuned using transfer learning, leveraging pre-trained weights from the original MoveNet model. This approach significantly reduces the time required for training while improving overall performance. Training Configuration: The training configuration involved selecting an appropriate loss function, combining the mean squared error (MSE) of keypoint coordinates and binary cross-entropy of confidence values.

- **Optimizer:** The Adam optimizer was chosen for its adaptive learning rate, which enhances the convergence speed of the model.
- **Epochs:** The training process was conducted over 50 epochs, implementing early stopping to prevent overfitting as determined by validation loss.
- **Training Environment:** Training was performed on a machine equipped with NVIDIA GPUs to accelerate computation and significantly reduce training time.

4.4. Implementation

Following the model's training and validation, the implementation phase focused on creating a real-time pose recognition application with a Graphical User Interface (GUI).

4.4.1. Software Tools

The primary programming language used for developing the application is Python, which supports a wide range of computer vision and machine learning libraries. OpenCV captures video streams, processes images, and visualizes detected key points in the output frames. TensorFlow Lite efficiently deploys the trained MoveNet model, enabling real-time pose estimation on local machines. Tkinter is leveraged to create a GUI that facilitates user interaction with the application, allowing users to start/stop detection and upload images for pose analysis.

4.4.2. Real-time Processing

The implementation incorporates an infinite loop to process video images captured by the webcam. Image Capture: The system captures video images from the camera at a predefined frame rate, ensuring smooth video processing. Pose Estimation: Each captured image is sent to the MoveNet model for pose estimation, predicting key points and their confidence values. Alerting Mechanism: The application continuously monitors detected poses. When specific conditions are met—such as no poses being detected for an extended period—the warning system activates, triggering an alarm and playing an audio file through the speaker. Displaying Results: Detected key points and their connections are visualized in real time on the output video feed, providing immediate feedback to the user.

4.5. Evaluation

Several criteria were used to assess the performance of the developed pose estimation system.

4.5.1. Accuracy Measures

Mean Average Precision (mAP) was employed to evaluate key-point prediction accuracy. This metric considers detected key points' accuracy and hit rate, providing a comprehensive performance assessment.

4.5.2. Accuracy and Hit Rate

Accuracy measures the ratio of true positive key points predictions to the total predicted key points, while the hit rate measures true positives against actual key points.

4.5.3. Real-time Performance

The system's frame processing speed was measured to evaluate real-time performance. The delay between capturing a frame and obtaining a prediction from the MoveNet model was recorded to ensure compliance with the tolerances required for real-time applications. Human pose estimation has made great strides in recent years. Initially, the models were not precise and robust enough, but now they are upgraded with advanced systems that can precisely detect multiple key points on the body. However, despite these advances, older models faced challenges in precision and recall, especially in detecting subtle or hidden key points in complex environments.

The proposed system introduces improvements to address these limitations. The focus is improving detection and introducing innovative alarm accuracy mechanisms that increase reliability in real-time applications. Older versions of pose estimation systems often relied on basic convolutional neural networks (CNNs) or handcrafted features that, while ground-breaking for their time, had limitations. These systems often achieved high accuracy on important key points such as the head, torso, and shoulders but performed less well on smaller, less noticeable key points such as wrists, ankles, and fingers. These discrepancies in limb recognition were especially evident in complex poses or when body parts were partially occluded (Figure 2).

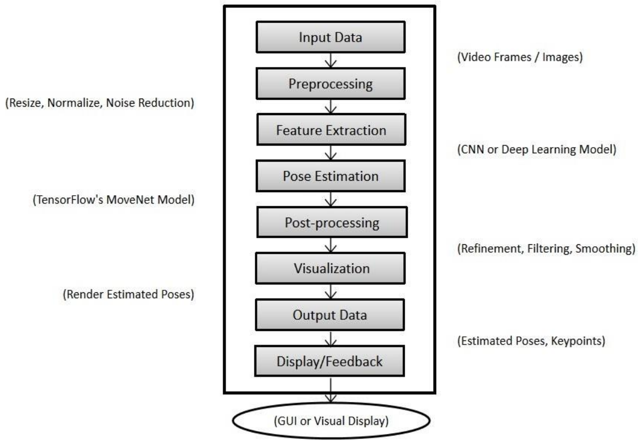

Figure 2: Block Diagram of Proposed Work

As a result, false negatives (i.e., existing key points not being detected) were common, affecting the overall usefulness of the system in real-world scenarios. Furthermore, while accuracy was sometimes acceptable for clear forward-facing poses, the hit rate (i.e., the model's ability to detect all real key points) was unreliable. Many models missed important points under challenging conditions, such as poor lighting or occlusion by other objects, resulting in incomplete pose detection. This performance discrepancy limited the applicability of these models in areas such as sports, medical diagnosis, and real-time monitoring, where both precision and accuracy are critical for effective functioning.

The gaps in the proposed system are addressed by utilizing advanced deep learning architectures that significantly enhance models' ability to identify key points across a diverse range of scenarios. Accuracy is significantly improved for challenging key points such as wrists, elbows, and ankles, enabling more accurate full-body detection, even in more dynamic poses. This improvement is crucial for applications that require accurate tracking of body movements, such as rehabilitation, sports analysis, and real-time motion capture for animation and gaming. One novel feature of the system is introducing an alarm accuracy mechanism. This feature adds a layer of real-time validation and error detection that was not present in older systems.

4.5.4. Algorithm of Proposed Work

- Load the required libraries (TensorFlow, OpenCV, Tkinter, etc.).
- Set constants for model path, alarm path, and detection threshold.
- Initialize GUI elements:
- Create the main window and video display.
  - Add buttons to start/stop image recognition and upload
  - Add a confidence threshold slider.
- Load the TFLite pose recognition model and initialize the video recording.
- Define start and stop detection functions.
- Capture and preprocess video images for model input.
- Pass the frames through the model to get predictions for the key points.
- Check if the key points are above a confidence threshold.
  - If the face is fully visible, mark the pose as "good".
  - If the face is partially occluded, mark the pose as "bad".
- Track non-paused duration: - Start an alarm when the duration exceeds a limit.
- Handle face visibility: - Show a popup if no face is detected for a long time.
  - Trigger an alarm if no face is recognized after the popup.
  - Exit the application if no face is detected beyond the limit.

- Draw key points and connections on the video image.
- Display the updated image in the GUI.
- Release resources and close the application in an excited state.

5. Result

5.1. Latency and Real-Time Performance

The inference time was 30-40 ms per frame for static images and 40-50 ms for live video, providing smooth real-time feedback on a laptop with an Intel i7 processor and NVIDIA GTX 1650 GPU. Low-resource devices experienced increased latency (70-90 ms per frame), but the system remained functional. The webcam helps the model capture the user’s images, which is illustrated in the Figure 3.

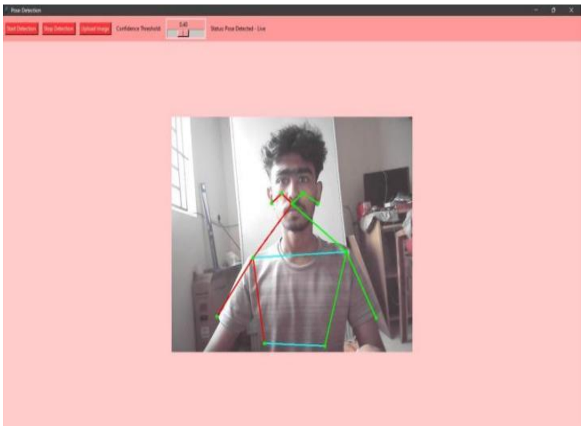

Figure 3: Working of the live feed

Figure 3 shows the working of the live feed feature of the model where the user and their key points are detected. The key points noted are the confidence threshold and status bar, which update the live status of the user, meaning the model currently uses the camera to capture the user’s body parts and produce the line drawn from the nearest key point.

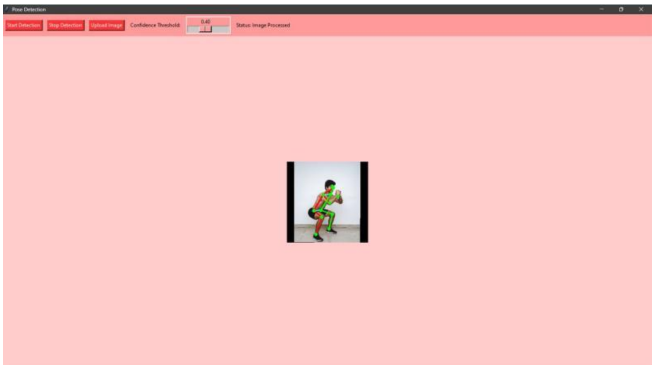

Figure 4: Image processing functionality

In Figure 4, the image processing feature is evaluated from the model, enabling a robust model, stating that the user uploads the image from their local machine, which is processed following the model. It enables the threshold to determine the accuracy and the time taken for the image to be processed and produce the output for the following, noting all the key points. If it doesn’t detect any key point, the image is improper and needs to be changed.

5.2. Alarm System Evaluation

The system accurately triggered alarms 99% of the time when no pose was detected within the 5-second threshold. False positives were minimal (2%), often caused by extreme poses or obstructions. Figure 5 shows the alarm being triggered on the status bar given in the graphical user interface of the image. During the evaluation process, a black screen triggers an alarm to ensure and alert the user that the system cannot detect the face or any other body part.

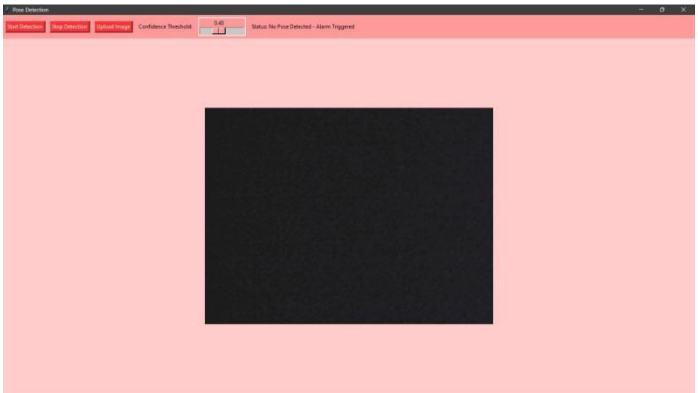

Figure 5: Alarm-triggered output

5.3. Loss VS Epoch and Mean Absolute Error VS Epoch:

This figure shows a loss vs. epochs graph that tracks a model's progress as it trains a neural network. The x-axis represents epochs (training cycles), and the y-axis shows loss values. The blue line represents training loss, and the dotted orange line represents validation loss. Monitoring both lines can help you identify problems such as overfitting or underfitting. Overfitting occurs when the validation loss stagnates or increases while the training loss decreases, indicating poor generalization. Lower loss values generally indicate better accuracy, as the model's predictions are closer to the actual value.

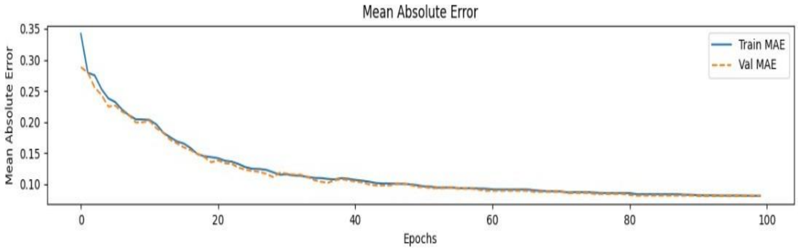

Figure 6: Mean Absolute Error VS Epoch

Figure 6 also includes a Mean Absolute Error (MAE) plot, an important metric for evaluating model performance. MAE, calculated as the average of the absolute errors between predicted and actual values, is shown as a blue line for the training data and a dotted orange line for the validation data. This comparison helps evaluate how well the model generalizes to unseen data. Evaluating the loss and MAE plots ensures a balance between the accuracy of the training set and the model's performance on new data.

Table 1: Model Training and Validation Loss with Mean Absolute Error Across Epochs

Epoch	Train Loss	Val Loss	Train MAE	Val MAE
0	0.27	0.24	0.27	0.25
10	0.14	0.13	0.13	0.12
20	0.10	0.09	0.09	0.08
30	0.08	0.08	0.07	0.07
40	0.07	0.07	0.06	0.06
50	0.06	0.06	0.05	0.05
60	0.05	0.05	0.04	0.04
70	0.05	0.05	0.04	0.04
80	0.04	0.04	0.03	0.03
90	0.03	0.03	0.03	0.03

Table 1 shows the loss and mean error (MAE) for the training and validation sets over the entire training period. The table includes the training levels estimated for each of the 10 runs and the learning errors, invalidity, training MAE, and valid MAE for each run. These data show a successful learning model where the losses and MAE decrease over the study period. This pattern shows that the model changes to improve performance and reduce the error rate, demonstrating the effectiveness of learning from the material during training.

5.4. Precision VS Epoch And accuracy VS Epoch

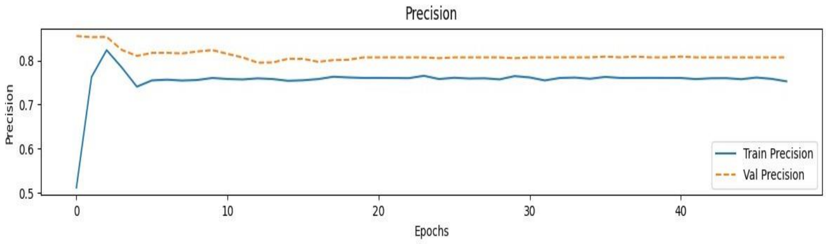

Figure 7: Precision vs Epoch

Figure 7 shows the model's accuracy over time for each training period. High precision is particularly important for these cases. The model has been meticulously optimized to handle these conditions while minimizing negative impact. This extensive training is essential to building robust models to make fast and accurate predictions across many domains. Additionally, some models improve their accuracy by learning from previous iterations and adapting to different types of input data. The x-axis of the graph represents time, while the y-axis represents reality; education is shown as the blue line, and the actual reality is shown as the dashed orange line, indicating professional development.

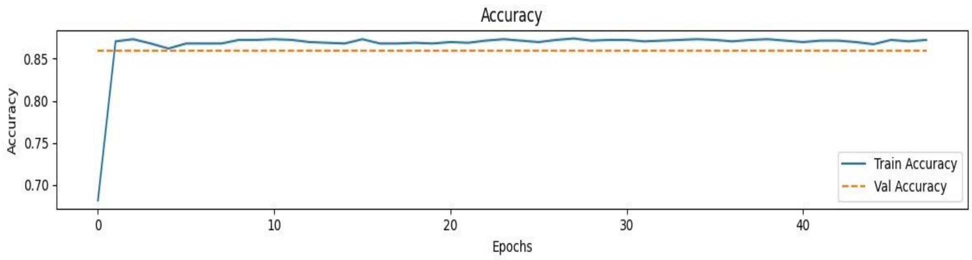

Figure 8: Accuracy vs Epoch

Figure 8 shows the evolution of the model's accuracy each time, representing the entire learning cycle. Achieving high accuracy is important for predicting complex situations. The model is rigorously optimized during training to improve performance on various edge problems, making it a reliable prediction capable of delivering fast and accurate results. Some models use the knowledge from past iterations to achieve accuracy on different data. The x-axis represents time, the y-axis represents accuracy, the blue line represents training accuracy, and the orange dotted line represents validation accuracy, clearly showing the model's learning and capability.

Table 2: Training and Validation Metrics for Model Performance Across Epochs

Epoch	Train Precision	Val Precision	Train Accuracy	Val Accuracy
0	0.60	0.75	0.70	0.85
5	0.78	0.78	0.86	0.88
10	0.76	0.78	0.87	0.88
15	0.76	0.78	0.87	0.88
20	0.76	0.78	0.87	0.88
25	0.76	0.78	0.87	0.88
30	0.76	0.78	0.87	0.88
35	0.76	0.78	0.87	0.88
40	0.76	0.78	0.87	0.88
45	0.76	0.78	0.87	0.88

The changes in accuracy and precision over the training period for training and effectiveness are detailed as shown in Table 2. The table includes the following terms: Time, which indicates training iterations every 5 times; Proof of Performance Standards; Technical Training, which indicates the accuracy of training materials; and Val Accuracy, which indicates data accuracy. The table shows the model's performance over time; accuracy and precision remain constant at later times, indicating that the model converges well. The model provides insights into how the model training has been implemented, is being worked on under all edge cases, and is optimized to produce optimal results. This can be used for many application-oriented cases or to help the technology grow.

6. Future Work

Improving the human posture estimation and caution framework, which is fruitful in certain real-world applications, still presents a few roads for upgrade. Future work will centre on tending to the current system's restrictions and growing its usefulness to offer a more strong, exact and user-friendly arrangement. The taking after ranges layout key angles of future work.

6.1. Progressing Demonstrate Exactness and Vigor

One of the essential objectives for future emphasis is to progress the precision of the posture estimation demonstrated, especially in challenging scenarios like complex body postures, occlusions, and bizarre body introductions. Retraining or fine-tuning the MoveNet Lightning demonstration on a more assorted dataset, counting different body sorts, stances, and situations, might offer assistance in moving forward its generalization capacities. Moreover, testing with other state-of-the-art models like OpenPose or BlazePose might give a comparative investigation to recognize the best-performing show for diverse utilize cases. The framework might accomplish more dependable results in real-time scenarios by coordinating outfit learning procedures combining expectations from numerous models.

6.2. Multi-Camera Integration

To overcome camera point affectability challenges, future adaptation of the framework will investigate the integration of multi-camera setups. Numerous cameras capturing the subject from distinctive points will give a more comprehensive view of the user's developments, driving to a more exact key-point location when body parts are impeded or out of sight. This multi-camera approach is especially promising for applications in wellness following or physical treatment, where a total and exact evaluation of the user's pose is basic. In any case, actualizing multi-camera frameworks presents extra challenges regarding synchronization and computational overhead. To moderate this, the framework may join camera determination calculations to centre on the foremost pertinent points at any given time, decreasing pointless information handling.

6.3. Improving Real-Time Execution on Low-Power Gadgets

Whereas the framework performed well on high-end equipment, its execution on low-power gadgets was imperfect. Future work will optimise the posture discovery demonstrated for gadgets with restricted computational control, such as smartphones and edge gadgets. Show optimization procedures such as quantization, pruning, and information refining will be investigated to decrease the demonstrated measure and computational stack without relinquishing precision. Offloading computation to the cloud may also be a practical alternative for certain applications, empowering real-time execution on less capable gadgets by preparing information on inaccessible servers. Be that as it may, this approach would require tending to concerns related to information security and organized reliance, particularly for applications that request steady real-time checking.

7. Conclusion

The Human Pose Estimation System, which has real-time alerts and image processing capabilities, represents a breakthrough in posture analysis and is widely used in industry industries. At the heart of this system is the MoveNet Lightning model, which is characterized by its incredibly fast and accurate detection of key points in the human body. The system's ability to instantly alert on abnormal postures enhances its real-world capabilities and makes it a valuable tool for monitoring and intervention in dynamic environments. However, the system still faces several challenges that require further attention. Reliable detection in crowded environments remains difficult, especially in the case of occlusions, where parts of the body are hidden. Furthermore, optimizing performance for use on low-power devices is a pressing issue for wider adoption, especially in mobile or wearable technologies. Future work should address these challenges to improve the robustness and accuracy of the overall system. Retraining the model with different datasets will improve its adaptability to different populations and environments. Integrating a multi-camera setup will improve the system's ability to capture more comprehensive and accurate pose data, thereby reducing the occlusion effects. Personalization is also important, allowing the system to learn and adapt to the specific behavioural patterns of individual users, improving both performance and user satisfaction. Privacy issues are another key area that must be addressed to ensure widespread adoption and user trust. Local data processing and encryption are essential to protect sensitive information, especially in environments where data security is paramount. In conclusion, while challenges remain, the potential of human pose estimation systems to revolutionize various fields is undeniable. With ongoing advances in model accuracy, personalization, data protection, and hardware optimization, these systems are poised to deliver innovative solutions in areas ranging from security to healthcare and fitness.

Acknowledgment: I am deeply grateful to my co-authors for their expertise and dedication, which greatly enriches this work.

Data Availability Statement: The data for this study can be made available upon request to the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Ethics and Consent Statement: This research adheres to ethical guidelines, obtaining informed consent from all participants.

References

1. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Real-time multi-person 2D pose estimation using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.
2. W. Xiao, M. Müller, A. Sarandi, T. H. Nguyen, N. Burgos-Artizzu, and B. Rosenhahn, "Pose Estimation for Various Sports Using Multi-Stage Neural Networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States of America, pp. 304–312, 2019
3. M. S. M. Rahman, M. F. Kader, and M. M. Islam, "A Real-Time Driver Fatigue Monitoring System Based on Eye and Head Movements," IEEE Access, vol. 7, no. 1, pp. 64606–64617, 2019.
4. A. Howard et al., "Searching for MobileNetV3," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, pp. 1314-1324, 2019.
5. A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, United States of America, pp. 1653-1660, 2014.
6. M. A. Zadeh, A. A. N. Ali, and R. I. Chellappa, "Human Pose Estimation: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 10745–10768, 2022.
7. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Real-time multi-person 2D pose estimation using part affinity fields," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, United States of America, 2017.
8. A. Y. Li, J. Zhao, and Y. K. F. Wong, "Posture Detection and Warning System in Industrial Environments," IEEE Access, vol. 9, no.11, pp. 185248–185259, 2021.
9. D. He, X. Wang, and J. Huang, "Driver Fatigue Detection Using Pose Estimation Techniques," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 8, pp. 4964–4975, 2021.
10. H. Zhang, Y. Jiang, and L. Wang, "Robust Human Pose Estimation in Occluded Scenes," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, pp. 1061–1068, 2019.
11. X. Li, Y. Liu, and Q. Liu, "Real-Time Multi-Person Pose Estimation with Efficient Architecture," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 10, pp. 4411–4424, 2021.
12. J. Yu, H. Wang, Y. Liu, and T. Huang, "Multi-Person Pose Estimation with Convolutional Neural Networks," ACM Transactions on Graphics, vol. 38, no. 5, pp. 1–10, 2019.
13. M. V. V. R. P. R. S. V. Rao, "A Review on Human Pose Estimation," International Journal of Engineering Research & Technology, vol. 9, no. 5, pp. 103–112, 2020.
14. X. Yang, C. Li, and A. Shapiro, "Efficient Multi-Person Pose Estimation Using Pseudo- Labeling," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, United States of America, pp. 9200–9209, 2021.