PosePerfect: Integrating HRNet and Gemini Vision Pro for Enhanced Yoga Posture Classification and Posture correction Analysis

Khushi Vasant Habib
KLE Technological University
Hubballi, India
khushisvh@gmail.com

Sumit Meharwade
KLE Technological University
Hubballi, Karnataka, India
sumitmeharwade1@gmail.com

Kushalgouda S Patil
KLE Technological University
Hubballi, Karnataka, India
kspatil.ksp@gmail.com

Sameer M Nadaf
KLE Technological University
Hubballi, Karnataka, India
sameernadaf787@gmail.com

Sneha Varur
KLE Technological University
Hubballi, Karnataka, India
sneha.varur@kletech.ac.in

Padmashri D Desai
KLE Technological University
Hubballi, Karnataka, India
padmashri@kletech.ac.in

***Abstract*—This study, named "PosePerfect," introduces the integration of yoga with advanced computer vision techniques, specifically emphasizing precise pose categorization using HRNet and intelligent analysis of posture correction using Gemini Vision Pro. The application of HRNet to the varied YOGA2022 dataset results in a remarkable accuracy of 95.23% for both pose estimation and classification. PosePerfect goes beyond simple categorization by incorporating Gemini Vision Pro to offer immediate, tailored recommendations for correcting posture, representing a notable progression. The meticulously evaluated comprehensive solution not only efficiently categorizes yoga poses but also improves accessibility and customization through web-based applications. The study confirms HRNet's efficacy in pose classification and underscores the enormous benefits of combining Gemini Vision Pro for intelligent posture correction analysis, improving the entire yoga experience.**

***Index Terms*—HRNet, key point, yoga pose estimation, yoga pose classification,Gemini Vision Pro.**

I. Introduction

Ensuring correct alignment in yoga postures is essential for optimizing practice outcomes and reducing potential health hazards. Our research utilizes sophisticated Deep Convolutional Neural Networks (DCNNs), particularly HRNet, to accurately detect human body positions. The distinguishing characteristic of HRNet is its capacity to preserve high-resolution representations, hence improving precision and facilitating communication between humans and machines. This capability has been showcased through its performance on benchmark datasets such as YOGA2022.

Our primary objective is to enhance key point recognition, real-time processing, and network efficiency in order to address the issues in yoga posture estimation. The potential of these advances is revolutionary, particularly in the development of improved monitoring tools and interactive systems.Figure 1 shows the overview of proposed method.

Integrating HRNet [1] with digital platforms allows users to demonstrate yoga positions and receive prompt, personalized feedback. Our research incorporates Gemini Vision Pro to achieve intelligent posture correction, surpassing the scope of pose estimation. This innovative integration merges technology with yoga, offering customized and efficient approaches for practice and instruction, thereby boosting general well-being.

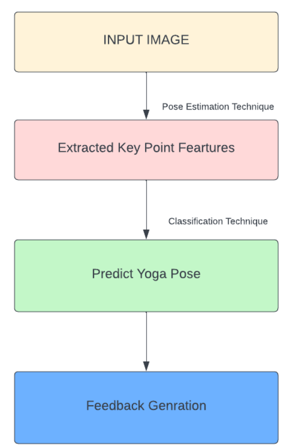

Fig. 1: Overview of the proposed method

The article is organized into different sections. Section II provides a detailed description of HRNet and Gemini Vision Pro [2] architecture. Section III explains the methodology used in the research. Finally, Section IV presents the conclusions of the investigation. Sections V and VI address the future possibilities and provide a conclusive summary of the study, ensuring a logical and well-structured presentation.

II. Related Work

A. HRNet

In the realm of human pose estimation, various methodologies have been employed to address the challenge of identifying precise keypoints or body components within an image. One common approach involves transforming the problem into the estimation of heatmaps, where each heatmap corresponds to the confidence level of a specific keypoint. Among the advanced techniques, the use of convolutional networks is prevalent. HRNet is a notable method it utilizes a unique convolutional neural network (CNN) structure that preserves high-resolution feature responses. This research specifically concentrates on HRNet because it has the most advanced performance among CNN networks. Human pose estimation,

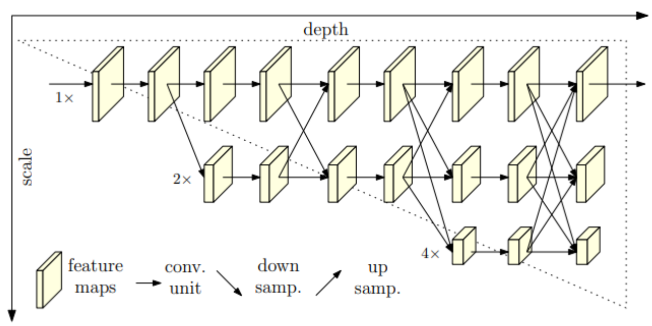

Fig. 2: Illustrating the architecture of the proposed HRNet. It consists of parallel high-to-low resolution subnetworks with repeated information exchange across multi-resolution subnetworks (multi-scale fusion). The horizontal and vertical directions correspond to the depth of the network and the scale of the feature maps, respectively.

also known as keypoint detection, is the task of identifying the precise locations of K keypoints or body components (such as the elbow or wrist) from an image I with dimensions $W \times H \times 3$. The most advanced techniques convert this problem into the estimation of K heatmaps, each with dimensions $W' \times H' \{H_1, H_2, \ldots, H_K\}$. Each heatmap H_k represents the level of confidence in the position of the k^{th} keypoint. [1] We utilize a commonly-used pipeline to forecast human keypoints using a convolutional network. This network is comprised of a stem that includes two convolutions with strided outputs, resulting in a decreased resolution. The main body of the network produces feature maps with the same resolution as the input feature maps. Lastly, a regressor is employed to estimate the heatmaps, which determine the positions of the keypoints and are transformed to the full resolution. Our primary attention is on the design of the main body. The High-Resolution Net (HRNet), is illustrated in Figure 2. [3]

1) Sequential Multi-Resolution Subnetworks: Current posture estimation networks are constructed by linking subnetworks of varying resolutions in a sequential manner. Each subnetwork, referred to as a stage, consists of a succession of convolutional operations. Additionally, there is a down-sample layer between adjacent subnetworks that reduces the resolution by half.

Let N_{sr} be the subnetwork in the s^{th} stage and r be the resolution index. The high-to-low network with S stages can be denoted as:

$$N_{11} \rightarrow N_{22} \rightarrow N_{33} \rightarrow N_{44}. \tag{1}$$

2) Parallel Multi-Resolution Subnetworks: Our approach begins with a high-resolution subnetwork as the initial stage. We then systematically incorporate other subnetworks of decreasing resolution, one at a time, to create a comprehensive network.

Introduce additional phases and establish connections between the subnetworks operating at different resolutions simultaneously. Consequently, the resolutions for the parallel subnetworks in a subsequent stage comprise the resolutions from the previous stage, along with an additional lower resolution. [4] The provided network structure consists of 4 parallel subnetworks.

$$\begin{array}{r} N_{11} \rightarrow N_{21} \rightarrow N_{31} \rightarrow N_{41} \\ \searrow N_{22} \rightarrow N_{32} \rightarrow N_{42} \\ \searrow N_{33} \rightarrow N_{43} \\ \searrow N_{44} \end{array} \tag{2}$$

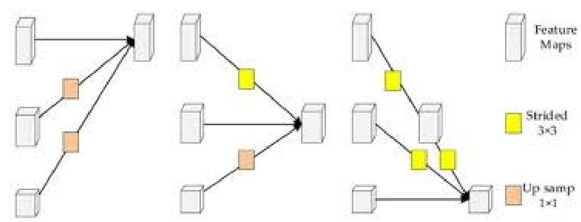

Fig. 3: Illustrating how the exchange unit aggregates the information for high, medium and low resolutions from the left to the right, respectively. Right legend: strided 3×3 = strided 3×3 convolution, up samp. 1×1 = nearest neighbor up-sampling following a 1 × 1 convolution.

3) Repeated Multi-Scale Fusion: The implementation exchange units within parallel subnetworks to facilitate the repeated transmission of information between them. Here is an illustrative demonstration of the process of transferring information. The third stage was partitioned into many exchange blocks, specifically three. Each block consists of three parallel convolution units, with an exchange unit connecting the parallel units. The exchange unit is defined as follows:

$$\begin{array}{cccccccccccc} \mathcal{C}_{31}^1 & \searrow & & \nearrow & \mathcal{C}_{31}^2 & \searrow & & \nearrow & \mathcal{C}_{31}^3 & \searrow & & \\ \mathcal{C}_{32}^1 & \rightarrow & \mathcal{E}_3^1 & \rightarrow & \mathcal{C}_{32}^2 & \rightarrow & \mathcal{E}_3^2 & \rightarrow & \mathcal{C}_{32}^3 & \rightarrow & \mathcal{E}_3^3, \\ \mathcal{C}_{33}^1 & \nearrow & & \searrow & \mathcal{C}_{33}^2 & \nearrow & & \searrow & \mathcal{C}_{33}^3 & \nearrow & & \end{array} \tag{3}$$

The symbol C_{sr}^b denotes the convolution unit in the rth resolution of the bth block in the sth stage, while E_{sb} indicates the corresponding exchange unit. Figure 3 depicts the exchange unit, while the formulation is presented below. We omit the subscript s and the superscript b for the sake of ease in our discussion. The inputs consist of a set of response maps denoted as X1, X2, ..., Xs. The outputs consist of a set of response maps, denoted as Y1, Y2, ..., Ys, which have the same resolutions and widths as the input. The output is a summation of the input maps, represented as $Y_k = \sum_{i=1}^{s} a(X_i, k)$. The interstage unit has an additional output mapping $Y_{s+1} : Y_{s+1} = a(Y_s, s+1)$. The function $a(X_i, k)$ involves either increasing or decreasing the resolution of X_i from level i to level k. We utilize strided 3×3 convolutions to achieve downsampling. For example, a 3x3 convolution with a stride of 2 can be applied on a 2x Resuming the narrative from the point where it was previously interrupted: [5]

...for reducing the resolution and applying two consecutive convolutions with a stride of 2 for a 2× reduction in resolution. For upsampling, we utilize a straightforward method called nearest neighbor sampling, which involves using a 1×1 convolution to align the number of channels. If i is equal to k, then the function a(·, ·) represents an identity link, where a(Xi, k) is equal to Xi.

4) Heatmap Estimation: We do a regression on the heatmaps using the high-resolution representations generated by the final exchange unit, which has been observed to be effective. The absence or reduction of anything valuable or significant. [6]

The function, known as the mean squared error, is utilized to compare the predicted heatmaps with the groundtruth heatmaps. The groundtruth heatmaps are created by using a 2D Gaussian distribution with a standard deviation of 1 pixel, centered on the groundtruth position of each keypoint. [7]

5) Network instantiation: The following text pertains to the concept of "Network instantiation" as discussed in the paper:

Creation of a network. To create the network for estimating keypoint heatmaps, we adhere to the design principle of ResNet, which involves allocating the depth to each stage and the number of channels to each resolution. The primary component, known as our HRNet, consists of four stages, each containing four parallel subnetworks. These subnetworks continuously decrease in resolution by half while simultaneously doubling in breadth, which refers to the number of channels. The initial phase consists of 4 residual units, each resembling the ResNet-50 model. Each unit is comprised of a bottleneck structure with a width of 64, followed by a 3×3 convolution operation that decreases the width of the feature mappings to C. The second, third, and fourth phases consist of one, four, and three exchange blocks, respectively. A single exchange block consists of four residual units, with each unit including two 3×3 convolutions at each resolution level, as well as an exchange unit that operates across different resolutions. There are a total of 8 exchange units, meaning that 8 multi-scale fusions are carried out. [4]

B. Gemini Vision Pro

The Gemini Vision Pro models' design expands on the fundamental concept of Transformer decoders, initially proposed by Vaswani et al. in 2017. The Transformer architecture has demonstrated remarkable efficacy in tasks related to natural language processing, and its application to image comprehension represents a significant change in the field of computer vision. Gemini Vision Pro improves this architecture by implementing various essential adjustments and optimizations, enabling it to effectively train on a significant scale and perform efficient inference on Google's Tensor Processing Units (TPUs) [2].The simplified architecture of Gemini Vision Pro is given below:

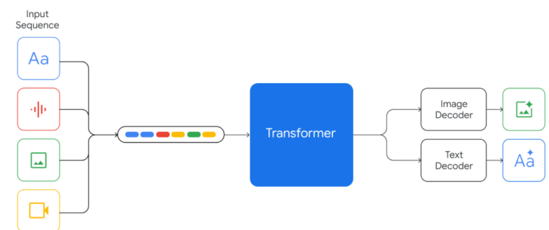

Fig. 4: Simplified architecture of Gemini Vision Pro [2]

Gemini Vision Pro stands out for its ability to handle a context length of 32k. This indicates that the model has the ability to take into account a large quantity of contextual information, enabling it to grasp nuanced nuances and connections within images. The increased context length is especially advantageous in tasks that require a thorough comprehension of the entire image, such as scene understanding, item detection, or image segmentation.

By integrating effective attention techniques, like the multi-query attention proposed by Shazeer in 2019, the model's capacity to concentrate on pertinent image characteristics is significantly improved during both training and inference. The utilization of this attention mechanism enables the model to assign varying degrees of importance to different sections of the input image, so allowing it to focus selectively on crucial regions and enhancing its overall comprehensibility.

Gemini Vision Pro has three primary sizes—Ultra, Pro, and Nano—which allow for a flexible and adaptable solution to cater to a wide variety of applications. The Pro model is particularly renowned for its intrinsic capacity to scale in both infrastructure and learning algorithms. This scalability guarantees both efficient model training and the capacity to finish pretraining in a few weeks, using only a small portion of the computational resources needed by the Ultra model [2].

Gemini Vision Pro's architecture makes it a potent tool for a range of computer vision jobs within the field of picture understanding. The model's Transformer-based design and improved attention processes enable it to excel in picture understanding, ranging from basic item and scene recognition to more intricate tasks such as comprehending context and relationships within images [2].

III. METHODOLGY

A. Data Description

The "Data Description" portion of the study article provides a comprehensive examination of two crucial datasets used to train and evaluate our yoga position estimation model. The initial dataset is the Common Objects in Context (COCO) dataset, which holds a prominent position in the field of computer vision. The dataset has more than 200,000 photos and 250,000 individual occurrences, each labeled with 17 specific keypoints, such as the nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. The work utilizes the COCO train2017 dataset, which has 118,287 pictures, to train the model. The model is then assessed on the val2017 dataset, which consists of 5,000 images. [8]

The YOGA2022 dataset is a specialized collection of data that complements the COCO dataset. It has been carefully curated to specifically highlight human yoga movement poses. This dataset is crucial in improving the model's comprehension and examination of yoga positions. The dataset comprises 15,350 photos, encompassing ten fundamental yoga poses: Warrior I Pose, Warrior II Pose, Bridge Pose, Downward Dog Pose, Flat Pose, Inclined Plank Pose, Seated Pose, Triangle Pose, Phantom Chair Pose, and Goddess Pose. The dataset comprises a total of five people, offering a varied array of body shapes and postures. The YOGA2022 dataset provides annotations for each image, consisting of 17 keypoints [9]. These keypoints correspond to the ones found in the COCO dataset and include precise location coordinates as well as visibility information. The uniformity in annotation enables a smooth incorporation of data from both datasets, guaranteeing a strong and thorough training and evaluation procedure. [8]

Fig. 5: Example of a Warrior Pose in the YOGA2022 dataset.

The integration of these two datasets, namely COCO for a comprehensive view of human poses and YOGA2022 for a specialized emphasis on yoga postures, establishes a strong basis for our research. This combination not only allows for accurate assessment of various yoga poses, but also makes a substantial contribution to the progress of human-motion analysis and posture estimation. The comprehensive annotations and diverse range of poses documented in these datasets offer a valuable resource for cultivating a sophisticated comprehension of human postural dynamics within the realm of yoga.model

B. Approach

The research methodology, illustrated in Figure 6, combines HRNet with Gemini Vision Pro to achieve accurate classification of yoga poses and immediate correction suggestion of posture in real-time. This comprehensive approach consists of two crucial stages: Yoga Pose estimation and Classification utilizing HRNet and Posture Correction Analysis employing Gemini Vision Pro.

Our system, at the early stage of Yoga Pose Classification using HRNet, exhibits versatility by taking photos as input, accommodating various techniques of photographing yoga positions. This input is subjected to thorough preprocessing to improve its quality and relevance. The notable characteristic of HRNet is its capacity to maintain high-resolution representations throughout its processing pipeline. This is accomplished by utilizing a parallel architecture that consists of many branches, with each branch processing the image at varying resolutions. These branches cooperate to communicate information in order to preserve intricate features that are essential for precise keypoint detection.

HRNet utilizes a multi-stage methodology to identify keypoints. The initial layers provide rough heatmaps that indicate the probability of each keypoint being present at various points in the image. Successive layers iteratively enhance these heatmaps, augmenting their resolution and accuracy. During the concluding phases, HRNet retrieves the precise coordinates of the 17 keypoints from the improved heatmaps, accurately identifying their specific positions on the image. The spatial configurations, angles, and distances among these keypoints are subsequently utilized to extract distinctive characteristics that embody the fundamental nature of the yoga position.

The retrieved features are inputted into a classification module that is equipped with a softmax function. This module analyzes the characteristic "fingerprint" and categorizes the image into the most likely yoga posture category, while also providing a confidence level for the classification. The classification procedure is essential to our revolutionary methodology, establishing the foundation for accurate identification of yoga poses.

During the following stage of Posture Correction Analysis using Gemini Vision Pro, the real-time determination of body position utilizes the 17 keypoints identified by HRNet. Gemini Vision Pro leverages these keypoints to accurately estimate angles and alignments in real-time throughout the user's yoga practice, providing immediate feedback on the user's posture in relation to the recognized yoga position.

Gemini Vision Pro subsequently compares the approximated body position to a repository of optimal reference poses for the recognized yoga category. The reference poses act as standards for correct alignment, enabling a detailed comparison with the user's actual posture. The system delivers individualized feedback for the user in real-time. This feedback incorporates visual indicators, such as heatmaps or skeleton overlays, that emphasize regions of misalignment, along with explicit and succinct voice directives that inform the user on the necessary modifications for achieving accurate alignment.

The Figure emphasizes the system's capacity to adapt to photos, ensuring versatility in collecting yoga poses. The 17 keypoints play a crucial role, acting as the basis for both pose

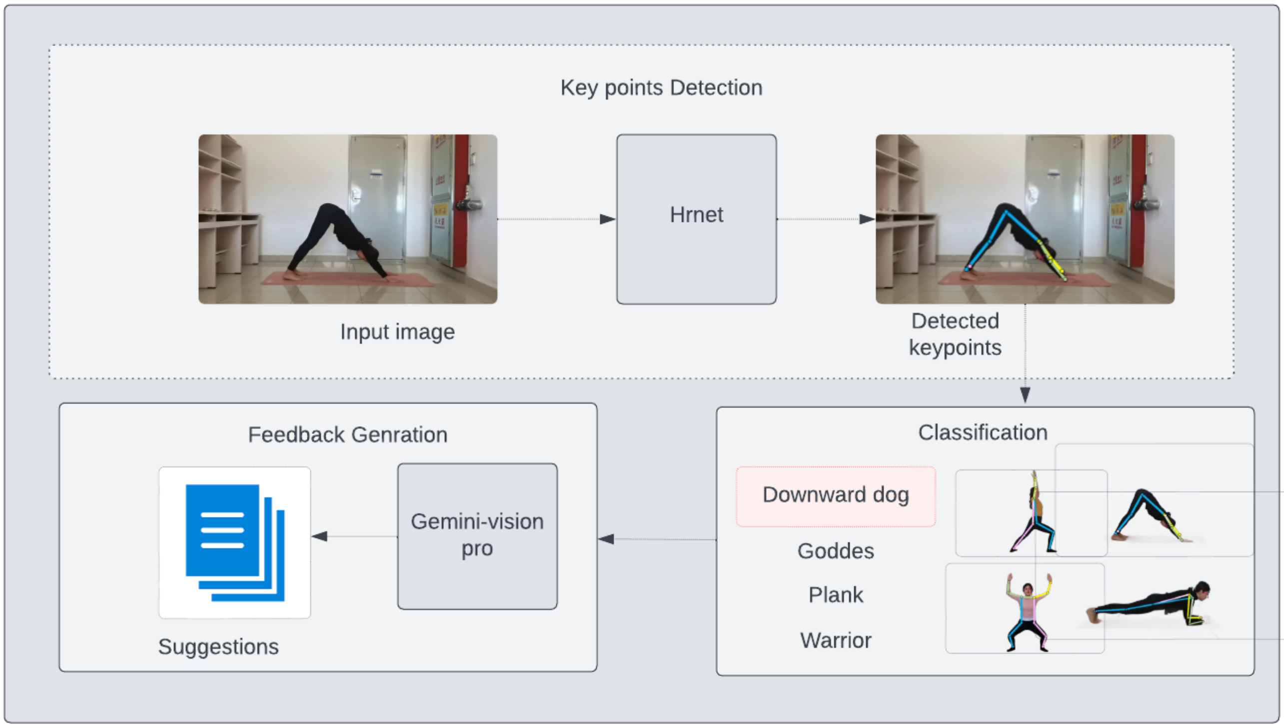

Fig. 6: A schematic diagram of the proposed approach for yoga pose estimation and feedback generation.

classification and real-time posture analysis. The HRNet and Gemini Vision Pro demonstrate an interconnected processing system that exemplifies an integrated approach, highlighting the smooth and uninterrupted flow of information across different stages.

The real-time feedback loop guarantees prompt corrections and modifications during yoga practice, so enhancing the learning experience. Practitioners gain advantages from receiving personalized coaching that is customized to their unique posture and requirements. This approach facilitates adaptive learning and effectively addresses the individual differences that exist in yoga practice. The technique described in these paragraphs highlights the careful design of our suggested system and its ability to meet the various requirements of yoga practitioners.

IV. Result

The Results section presents a thorough evaluation of the HRNet model's ability in accurately estimating yoga poses. This assessment incorporates essential measurements, including training accuracy and the Intersection Over Union (IoU) score, utilizing the YOGA2022 dataset. The results highlight the accuracy of the model in defining yoga positions, and its computational efficiency is compared to well-known methods. The discussion explores the consequences of these findings for the utilization of yoga instruction and practice, highlighting the model's capacity to greatly improve the area of movement analysis, well-being and classication of yoga poses. [10]

A. Intersection over Union (IoU) Score

This article presents the crucial discoveries obtained from the implementation and testing of our HRNet model, which is specifically designed for yoga pose estimation. The YOGA2022 dataset was rigorously evaluated to assess the model's efficacy and efficiency in accurately capturing complex yoga positions. Performance indicators, including training accuracy and the Intersection Over Union (IoU) score, provide vital information about the model's precision and ability to reliably estimate yoga postures. These results highlight the significant potential of the HRNet model in many applications related to movement analysis and interactive systems in yoga scenarios. [11]

TABLE I: Performance Metrics of the HRNET Model

Metric	Value
Training Accuracy	95.23%
Training IOU Score	0.667
Validation Accuracy	90.79%
Validation IOU Score	0.451

The HRNet model, specifically designed for yoga pose estimation and classification, exhibits exceptional performance, as shown in Table I. The model demonstrates remarkable precision in correctly detecting input photos in the YOGA2022 dataset, with an astounding training accuracy of 95.23%. The

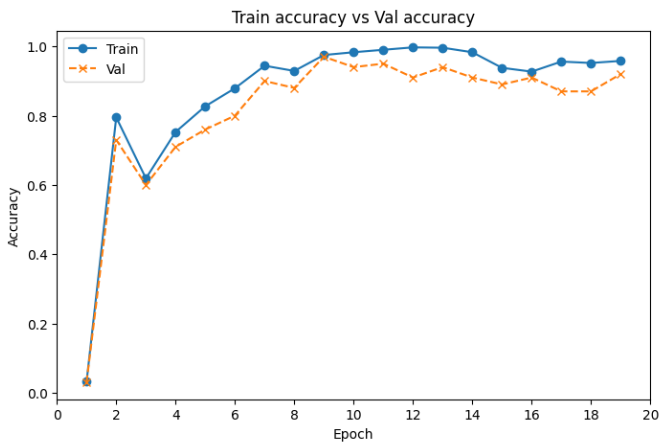

Fig. 7: Train accuracy vs Validation accuracy

utmost importance is in achieving a high level of accuracy, as it guarantees dependable observations of practitioners' postures. Consequently, this improves the overall quality of yoga instruction and feedback. In addition, the Intersection Over Union (IoU) score, as shown in Table 1, confirms the model's effectiveness in accurately defining yoga postures.

The Intersection over Union (IoU) score, presented in Table 1, is a crucial indicator for evaluating the effectiveness of yoga posture prediction models. The HRNet model achieves an IoU score of 0.667, indicating a moderate yet commendable level of accuracy in reliably classifying yoga postures. The visual depiction, in conjunction with the IoU score, is crucial for assessing the model's applicability in yoga instruction situations, where accurate position estimate is essential for appropriate coaching and performance analysis.

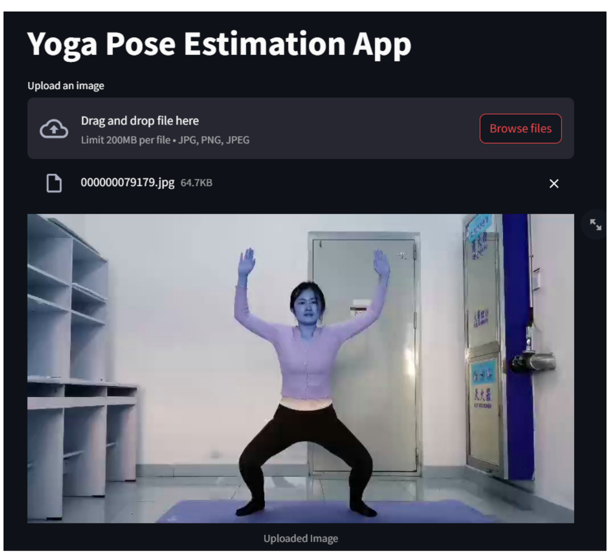

Fig. 8: Deployement of the hrnet model on the web page

The HRNet model, which underwent rigorous training for yoga position estimation and classification using the YOGA2022 dataset, has effectively progressed from the development stage to real-world implementation. The model is smoothly incorporated into a user-friendly web page as shown in Figure 8, allowing users to upload photographs of yoga poses. After the user submits the web page, it responds in real-time by presenting specific and thorough information about the key points of important bodily joints in the yoga pose. In addition, the model demonstrates its expertise in posture classification by appropriately identifying the submitted stance as shown in the figure 9. An outstanding characteristic of the implemented system is its capacity to classify the yoga posture as shown in figure 10. If the pose given by the user diverges from the correct form, the web page immediately instructs users to change their posture in order to achieve appropriate classification. The utilization of an interactive and feedback-driven method significantly improves the user experience by providing accessible and actionable information from the model. The integration of Gemini Vision Pro plays a vital role in our creative approach, especially when the system classifies a user's yoga position as negative class. During these occurrences, the system effortlessly transmits the data on body

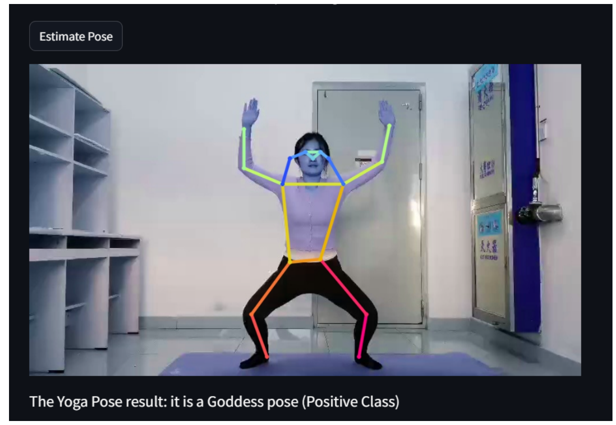

Fig. 9: Key point estimation of the given image as well as classification as positive class with its posture name using the keypoints estimated

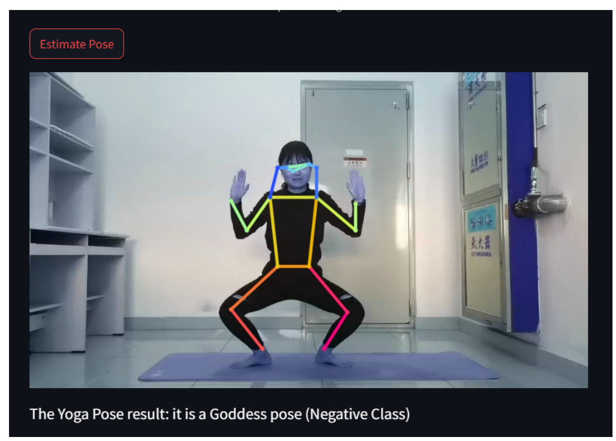

Fig. 10: Key point estimation of the given image as well as indication of wrong posture(negative class)

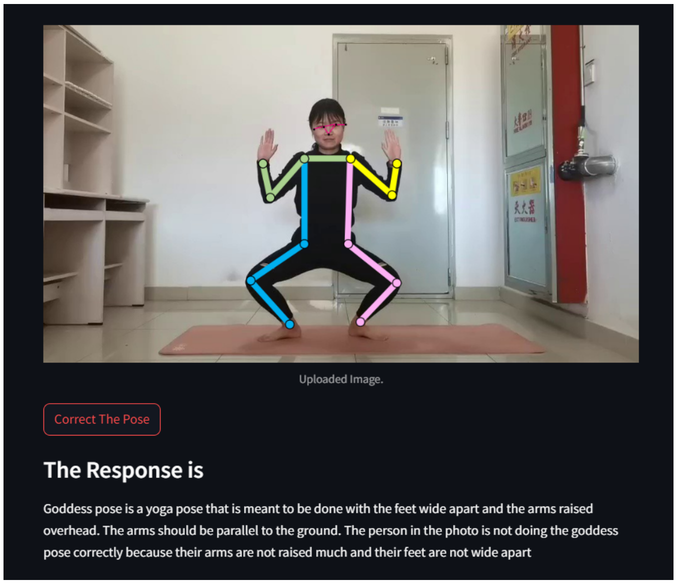

Fig. 11: Gemini vision pro analyses and recommendation of the given negative class image

position to Gemini Vision Pro, an advanced model specifically created for analyzing and correcting posture intelligently.

The integration signifies a dynamic and adaptable phase in our system, with Gemini Vision Pro assuming the responsibility of offering perceptive and valuable input to the user.Upon obtaining the posture designated as negative, Gemini Vision Pro utilizes its superior skills to study the precise subtleties of the position and detect areas that can be enhanced as shown in the figure 11. The model, equipped with its sophisticated comprehension of accurate yoga positions, produces practical suggestions on how the user might correct and improve their posture. The cooperative interaction between HRNet and Gemini Vision Pro not only guarantees the precise categorization of yoga poses but also actively contributes to the process of rectification.

V. Future Scope

The research endeavors to advance yoga pose estimation through HRNet and Gemini Vision Pro models, employing advanced deep learning techniques such as attention mechanisms and generative adversarial networks. This includes integrating exercise routines for personalized training and precise performance monitoring. Future efforts focus on optimizing HRNet for faster processing, enabling real-time applications, and deploying models on Android devices through a mobile application. Collaboration with yoga instructors is essential for model customization, with overarching goals of continuous improvement, exploring new applications, and promoting technology-driven wellness practices for enhanced fitness and yoga instruction.

VI. Conclusion

To summarize, this study emphasizes the significant use of HRNet for accu- rate assessment of yoga poses, offering potential progress in the fields of fitness and education. Potential future developments involve expanding HRNet's use to encompass exercise routines and positions related to sports, thereby transform- ing tailored training and the delivery of feedback. Effective collaboration with teachers is crucial for personalization, and Gemini Vision Pro improves the user experience by specifically addressing misclassified postures. The integration sig- nifies a notable transition in the utilization of technology in yoga, highlighting continuous enhancement and varied implementations in health practices.

References

[1] A-HRNet: Attention Based High Resolution Network for Human pose estimation. https://ieeexplore.ieee.org/document/9253142, sep 1 2020.

[2] Gemini Team. Gemini: A Family of Highly Capable Multimodal Models. https://arxiv.org/abs/2312.11805, dec 19 2023.

[3] Manisha Verma. Yoga-82: A New Dataset for Fine-Grained Classification of Human Poses. https://openaccess.thecvf.com/content_CVPRW_2020/html/w70/Verma_Yoga82_A_New_Dataset_for_Fine-Grained_Classification_of_Human_Poses_CVPRW_2020_paper.html, 2020.

[4] Jingdong Wang. Deep High-Resolution Representation Learning for Visual Recognition. https://arxiv.org/abs/1908.07919, aug 20 2019.

[5] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. Deep 3d human pose estimation: A review. https://doi.org/10.1016/j.cviu.2021.103225, sep 1 2021.

[6] Nikolaos Sarafianos, Bogdan Boteanu, Bogdan Ionescu, and Ioannis A. Kakadiaris. 3d Human pose estimation: A review of the literature and analysis of covariates. https://doi.org/10.1016/j.cviu.2016.09.002, nov 1 2016.

[7] Leonid Sigal. Human Pose Estimation. https://doi.org/10.1007/978-3-030-63416-2_584, jan 1 2021.

[8] Shubham Garg, Aman Saxena, and Ritu Gupta. Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application. https://doi.org/10.1007/s12652-022-03910-0, jun 3 2022.

[9] Tsung-Yi Lin. Microsoft COCO: Common Objects in Context. https://arxiv.org/abs/1405.0312, may 1 2014.

[10] A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning. https://ieeexplore.ieee.org/abstract/document/8920892?casa_token=JLt3FyZQOKgAAAAA:eijwalSIcmsZGLW71P6Gd6dXJ5Td1J16TxjGO03X4sJCT6Hl3DNOaZ3uWF8ZPVFNp7upbboi3k, nov 1 2019.

[11] Mykhaylo Andriluka. 2d Human Pose Estimation: New Benchmark and State of the Art Analysis. https://openaccess.thecvf.com/content_cvpr_2014/html/Andriluka_2D_Human_Pose_2014_CVPR_paper.html, 2014.