

Enhancing Online Yoga Instruction: Evaluating the Effectiveness of Visual Augmentations for Performance Assessment

Ajit Gopal

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Industrial and Systems Engineering

Sol le Lim, Chair

Samantha M. Harden

Nathan K.C. Lau

September 25, 2024

Blacksburg, Virginia

Keywords: Online yoga, computer vision, pose estimation, yoga instruction

Copyright © 2024, Ajit Gopal

Enhancing Online Yoga Instruction: Evaluating the Effectiveness of Visual Augmentations for Performance Assessment

Ajit Gopal

ABSTRACT

Yoga is a mind-body practice known for its substantial psychological and physiological benefit, contributing to a healthy lifestyle. However, without professional guidance, individuals may experience reduced performance and increased risk of injury. While online yoga classes on platforms like Zoom have grown in popularity, tools to support instructors in accurately assessing and monitoring student performance remain insufficient. For certain populations, this lack of real-time professional guidance poses safety risks and limits the effectiveness of the practice.

This study examined the effectiveness of using computer-vision-based visual augmentations in enhancing instructors' ability to assess student performance and ensure safety. Specifically, we investigated the effectiveness of various visual augmentations in aiding instructors' visual search for unstable or unsafe poses. Eleven certified yoga instructors (8 female, 3 male), each holding 200 to 500 RYT certifications, participated in the study. Instructors completed eight trials assessing 12 yoga poses using four different visual augmentations—Raw Video, Skeleton (joint locations overlay), Contour (participant outlines), and Contour + Skeleton—across two camera views (Single vs. Multiple Views). During each trial, eye-tracking data was collected as instructors identified potentially unstable (unsafe) poses, and they subsequently completed a usability questionnaire and NASA - TLX rating. Upon finishing all trials, instructors provided overall feedback on the usability of the visual augmentations and camera views

Instructors showed no significant difference in their assessment performance across different visual augmentations and camera views. The Skeleton augmentation led to increased cognitive workload, as indicated by larger pupil diameters. The Contour alone augmentation was less effective for visual search based on the usability ratings, and combining Contour with Skeleton did not offer notable improvements. Simpler visualizations, such as Raw and Skeleton, received

higher usability ratings, and instructors preferred Single View layouts over Multiple Views for their ease of use and lower cognitive demand.

In conclusion, while Skeleton augmentation increased cognitive load, it did not significantly enhance visual search performance. Future research should explore alternative visual augmentation techniques and configurations to better assist instructors on performance assessment which increases overall performance while not substantially increasing cognitive workload.

Enhancing Online Yoga Instruction: Evaluating the Effectiveness of Visual Augmentations for Performance Assessment

Ajit Gopal

GENERAL AUDIENCE ABSTRACT

Yoga is a great way to improve both mental and physical health. However, practicing yoga without proper guidance can sometimes lead to injuries or mistakes. With more people attending yoga classes online, like through Zoom, it's harder for instructors to closely monitor how their students are performing, which can reduce the safety and benefits of the practice.

This study looked at whether certain computer tools could help instructors better see and correct their students' poses during online yoga classes. Eleven experienced yoga instructors tried out different visual aids while watching students perform yoga poses. These aids included a simple video, a video with lines showing where the students' joints were (called Skeleton), a video that showed just the outline of the student (Contour), and a mix of both (Contour + Skeleton). The instructors were asked to identify any unstable or unsafe poses while using these aids.

The results showed that none of the visual aids helped the instructors spot mistakes better than regular video. While the Skeleton aid made the instructors work harder mentally, it didn't actually help them perform better. The instructors preferred using simple video over the more complex tools and found that using a single camera view was easier to work with.

In short, more complex visual tools didn't help instructors improve their performance. Future studies should explore other ways, like using different camera angles or adding sound, to help instructors in online yoga classes.

Dedication

*To my mother, for her constant support and love,
which have been my greatest source of strength.*

Acknowledgments

I am immensely grateful to my advisor, Dr. Sol Lim, for her unwavering support and exceptional guidance in every aspect of this journey. I would also like to express my sincere thanks to my committee members—Dr. Samantha M. Harden and Dr. Nathan Lau—for their invaluable insights, guidance, and encouragement throughout this process.

A special thank you goes to Dr. Hyunkyung Kim from KwangWoon University, Seoul, Korea, for her mentorship and to her team for their outstanding work in developing the User Interface for this study.

Additionally, I am deeply thankful to Sakshi Taori, Shafiqul Islam, Saman Zahabi and Utkarsha Mohan for their assistance and support in the completion of this thesis. Finally, my heartfelt gratitude goes to all the participants, whose time and effort made this study possible.

Table of Content

List of Figures ix

List of Tables x

List of Abbreviations xi

1. Introduction 1

2. Methodology 3

 2.1 Participants 3

 2.2 Experimental Procedures 3

 2.2.1 Yoga Performance Videos 3

 2.2.2 Experimental Tasks 3

 2.2.3 Equipments 6

 2.3 Dependent Variable 6

 2.3.1 Eye Tracking Metrics 6

 2.3.2 Subjective Responses 7

 2.4 Data Processing 8

 2.5 Data Analyses 8

3. Results 10

 3.1 Assessment Performance 10

 3.2 Eye Tracking Metrics 10

 3.2.1 Target AOI 10

 3.2.2 Selected AOI 10

 3.2.3 Average Pupil Diameter 13

 3.3 Subjective Responses 13

 3.3.1 Usability Questionnaire 13

3.3.2 Perceived Workload	15
3.3.3 Usability Feedback	15
4. Discussion	18
4.1 Visual Augmentations	18
4.2 Camera Views	19
4.3 Limitations	19
4.4 Future Research Direction	21
References	23
Appendices	28
Appendix A. A list of Yoga Poses Used in the study with example pictures of and potentially stable and unstable poses.	28
Appendix B. Usability Questionnaire	31
Appendix C. Usability Feedback	32
Appendix D. Average Pupil Diameter	33
Appendix E. Extended Thematic analysis	34

List of Figures

2.1	Interface showing a 2x2 tile format of performance videos.	4
2.2	Four types of visual augmentation; left to right: Raw, Skeleton, Contour, Contour + Skeleton.	4
2.3	Single (Left) and Multiple Camera View (Right) with a Skeleton visual augmentation.	5
2.4	Experiment setup with a monitor equipped with a screen-mounted eye-tracking device.	6
2.5	AOIs (green boxes) around videos A-D. Blue circles show fixations, with larger ones indicating longer durations, and numbers in circles marking the fixation sequence.	7
3.1	Normalized mean number of fixations across different Visual Augmentations and Camera Views.	11
3.2	Normalized mean number of glances across different Visual Augmentations and Camera Views.	12
3.3	Normalized mean number of saccades across different Visual augmentations. Conditions with different letters are significantly different.	12
3.4	Average pupil diameter across Visual Augmentations. Conditions with different letters are significantly different.	13
3.5	Mean subjective response for usability ratings (Usefulness, Confidence, and Satisfaction) across different Visual Augmentations and Camera Views.	14
3.6	Overall NASA-TLX Score across different Visual Augmentations and Camera Views.	15
3.7	Illustration of all three categories (round boxes), 9 secondary level themes (capsule shaped boxes), and related tertiary level themes (italicized text in in rectangular boxes).	17

List of Tables

3.1	Summary of ANOVA results [F-value (p-value)] for the Effects of Visual Augmentations and Camera Views on eye movements in Target AOI.	10
3.2	Summary of ANOVA results [F-value (p-value)] for the Effects of Visual Augmentations and Camera Views on eye movements in incorrect AOI.	11
3.3	Summary of ANOVA results [F-value (p-value)] for the Effects of Visual Augmentations and Camera Views on Usability Ratings.	14
3.4	Summary of ANOVA results [F-value (p-value)] for the Effects of Visual augmentation and Camera view on Average Pupil Diameter.	33

List of Abbreviations

ANOVA: Analysis of Variance

AOI: Area of Interest/s

CV: Camera Views

I-VT Filter: Velocity-Threshold Identification Gaze Filter

IRB: Institutional Review Board

NASA-TLX: NASA Task Load Index

REML: Restricted Maximum Likelihood

SD: Standard Deviation

SPA : Social Physique Anxiety

SUS: System Usability Score

TAM: Technology Acceptance Model

UEQ: User Experience Questionnaire

VA: Visual Augmentations

χ : Chi is the 22nd letter in the Greek alphabet.

In statistics, chi is used to represent the characteristic function and chi-square distribution.

1. Introduction

Yoga offers numerous physical and mental health benefits (Raub, 2002), such as improved strength, endurance, and mental well-being (Islam, Frazier, et al., 2024). Since the Covid-19 pandemic, online yoga has emerged as a popular, convenient, and affordable alternative to in-person classes (Brinsley et al., 2021), offering benefits like stress relief and pain reduction (Sharma et al., 2020).

However, online yoga presents substantial challenges for instructors. Issues like ineffective camera angles, poor video quality, and unreliable internet connections hinder instructors' ability to effectively assess participants' performance (Islam et al., 2024). These limitations make it difficult to provide real-time feedback, leading to increased cognitive demands for instructors, who report heightened stress and anxiety (Guo & Fussell, 2022). Additionally, participants may hesitate to share their video feeds due to privacy concerns or fear of evoking Social Physique Anxiety (SPA) (Brunet & Sabiston, 2009), further complicating assessments.

To address these challenges, several visual augmentations offered by advanced computer vision techniques, such as Skeleton [joint overlays using OpenPose (Cao et al., 2019)] and Contour [body outlines via Canny Edge (Canny, 1998)] have been proposed as tools to enhance online yoga instruction (Cao et al., 2018; Fujiyoshi & Lipton, 1998). These augmentations could guide instructors' attention to important joint locations or body outlines by highlighting them with visual overlays, aiding instructors in quickly assessing participant performance and potential safety concerns. Research suggests that such augmentations can reduce cognitive load by directing attention to the most relevant visual elements (Brunyé et al., 2019). Similar approaches have been successfully applied in physical training and therapy, where skeleton-based algorithms have been used to detect and correct posture with high accuracy (Rishan et al., 2020). Moreover, the use of Contour overlays can help maintain privacy while still allowing instructors to assess body movements, addressing concerns around SPA (Brunet & Sabiston, 2009). Simplifying the visual scene by using Contour augmentations has been shown to improve the efficiency of visual search tasks by reducing distractions (Henderson et al., 2009). This approach mirrors methods used in some medical training, where visual overlays enhance task performance by making key information more accessible (Skaramagkas et al., 2023).

Visual augmentations, despite their potential, may introduce challenges by increasing visual clutter, which can overwhelm instructors and heighten cognitive workload (Henderson et al., 2009;

Rosenholtz et al., 2007). Visual attention becomes crucial when navigating complex visual fields, such as augmented video feeds in online yoga. Eye-tracking metrics like fixations, saccades, and pupil dilation could be effectively used to thoroughly investigate the effectiveness of visual search and also assess cognitive load (Holmqvist et al., 2011). Longer fixations and frequent saccades often signal increased mental effort during visual search tasks (Poole et al., 2001; Van Orden et al., 2001). Pupil sizes are known to be associated with task complexity and stress, and increased size reflects heightened mental workload and stress (Beatty & Kahneman, 1966). While visual augmentations could enhance assessment efficiency by focusing instructors' attention on the important features in the videos, the complexity of camera views could interact with these anticipated effects. In online teaching, participants often position their cameras based on their home environments, resulting in diverse visual feeds (e.g., camera perspectives) (Islam, Harden, et al., 2024). This could increase cognitive demands on instructors, as the visual search becomes more complicated with the various camera views (e.g., front, side, back). Single-camera views simplify the task by reducing visual complexity but may lack sufficient information for comprehensive performance assessment (Henderson & Ferreira, 2004). Conversely, multiple-camera views provide better coverage of movements but increase the cognitive demands on instructors (Barron et al., 2014). Instructors managing these complexities could experience cognitive overload. The Feature Integration Theory (Treisman & Gelade, 1980) further suggests that focused attention is needed to manage complex visual features in augmented video feeds, and the task's difficulty dictates where attention is directed (Yarbus, 1967). Thus, we will explore how these visual complexities affect instructors' assessment performance, and whether additional visual augmentations could potentially interact with camera views.

In this study, we investigate how different visual augmentations (Raw, Skeleton, Contour, and Skeleton + Contour) and camera views (Single View, Multiple View) impact instructors' performance and cognitive load during online yoga assessments. Using eye-tracking metrics and subjective measures of cognitive workload (e.g., NASA-TLX), we aim to evaluate whether visual augmentations improve pose assessment. We hypothesize that while Skeleton and Contour + Skeleton visualizations will aid instructors in pose evaluation, they will also raise cognitive workload, especially when combined with complex camera views. As for the camera views, we hypothesize Single View will be preferred and will result in less cognitive load. This research seeks to bridge gaps in understanding how technology can optimize remote yoga instruction and enhance safety without overwhelming instructors.

2. Methodology

2.1 Participants

Individuals with 200 or 500 registered yoga teacher certifications (meeting Yoga Alliance standards) were recruited via local flyers, word of mouth, and through the listserv of a local yoga studio with over 50 instructors. Eleven registered yoga instructors [3 male, 8 female; mean age (SD) = 34.0 (12.5) years; yoga teaching experience = 4.6 (3.7) years] completed the study. To qualify, participants had to have normal vision with no history of eye-related surgeries, such as cataract procedures. The study protocol was approved by the Institutional Review Board (IRB) at Virginia Tech (IRB 22-703). Written informed consent was obtained from all participants prior to any data collection.

2.2 Experimental Procedures

2.2.1 Yoga Performance Videos

Two certified yoga instructors selected 12 common yoga poses (with 20 different pose variations; see Appendix A) that are widely performed in yoga classes and range from easy to challenging for online sessions. Five yoga instructors were recruited to generate videos of the 12 yoga poses by simulating both stable and unstable (potentially unsafe) postures. For the unstable posture videos, instructors were informed about the purpose of the study and asked to simulate typical unstable postures common to each pose, which they often observe in their own instruction. Posture, alignment, range of motion, and balance were manipulated to create unstable pose videos. For the video recordings, we played pre-recorded verbal instructions to synchronize the timing of the poses performed by the different yoga instructors.

2.2.2 Experimental Tasks

Participants completed 8 visual assessment trials (4 Visual Augmentations x 2 Camera Views), in which they had to select one unstable pose demonstration out of four presented videos (Figure 2.1). Our custom interface displayed four randomly selected videos for each pose in a 2 x 2 tile format, simultaneously presenting three stable pose demonstrations and one unstable pose demonstration. Within each trial, 12 randomly selected poses were presented in a random order with the same visual augmentation and camera view condition. The location of the unstable demonstration video was randomized for each pose. Participants were asked to submit their choice

(unstable pose demonstration) using the selection button at the bottom of the interface (Figure 2.1). The study took approximately 2.5 hours. A detailed description of visual augmentations and camera views is provided below.

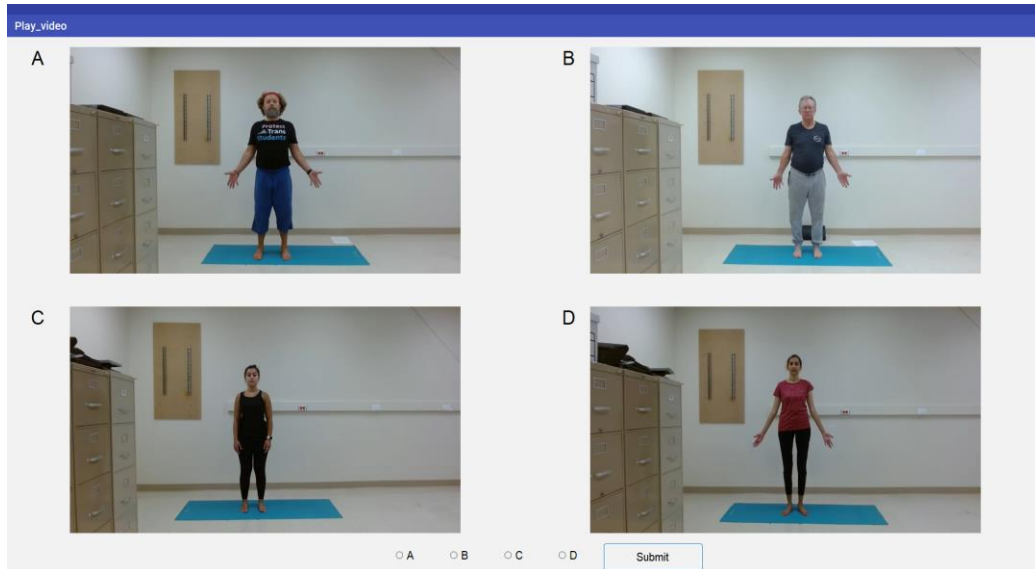


Figure 2.1. Interface showing a 2x2 tile format of performance videos.

Visual Augmentations: Four types of visual augmentation (Figure 2.2) were implemented: Raw, Skeleton, Contour, and Contour + Skeleton. Raw served as the control condition, presenting the original video without any visual augmentation. Skeleton overlaid 25 key joint locations, including the torso, hips, shoulders, upper and lower extremities, and facial features, onto the original video. Joint locations were detected using OpenPose (BODY25 model; Cao et al., 2018). Contour utilized a Canny edge detector within OpenCV software (OpenCV, 2015) to detect the edges of the human body, displaying outlines and boundaries of the human body silhouette without the original video.

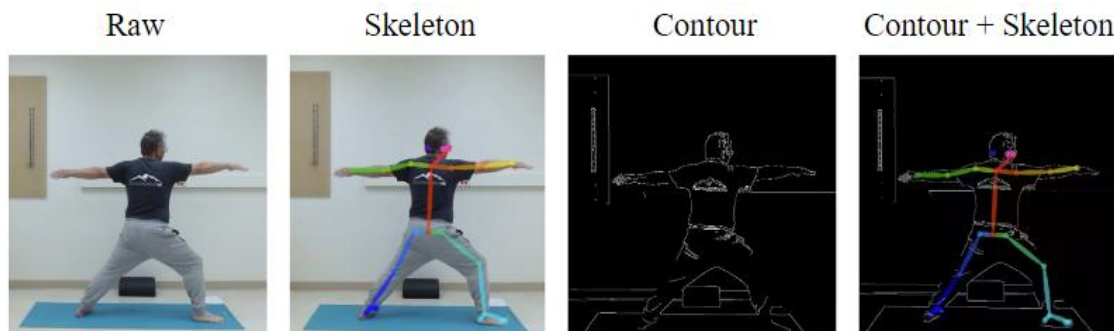


Figure 2.2 Four types of visual augmentation; left to right: Raw, Skeleton, Contour, Contour + Skeleton.

This condition was created to simulate video sharing without invading the privacy of the participants. The Contour + Skeleton condition merged the Contour and Skeleton features without the original video.

Camera Views: Two distinct camera view conditions (Figure 2.3) were simulated to mimic the variability encountered in online instruction settings. Yoga students often position their cameras in their own way, based on the layout of the room or personal preference, which can increase instructors' cognitive workload during assessment. A single camera view showed the same orientation, with the students' frontal sides facing the cameras. A multiple camera view randomly presented two frontal views, one side view, and one 45° angled view.

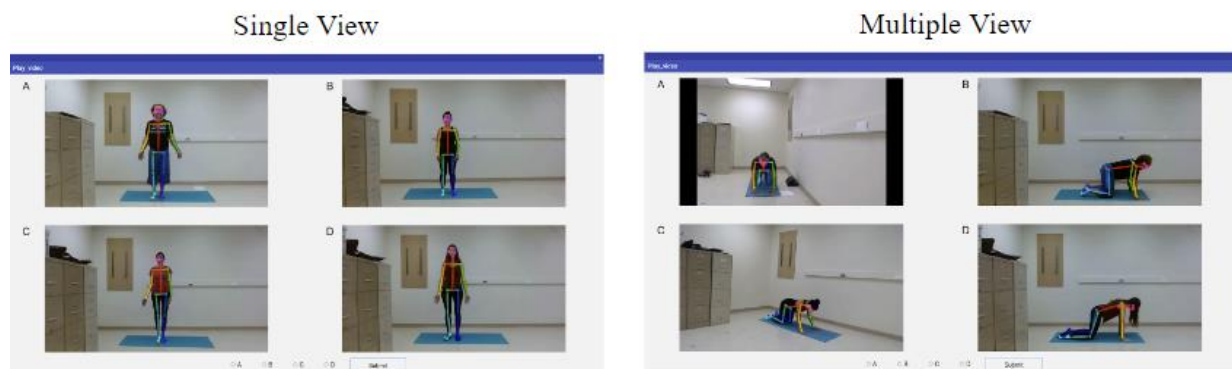


Figure 2.3 Single (Left) and Multiple Camera View (Right) with a Skeleton visual augmentation.

Following each trial, participants were prompted to assess the usability of the system by rating how useful, confident, and satisfied they felt using the specific visual augmentation and camera view condition. The NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) was used to assess participants' perceived workload.

Upon completion of all eight trials, participants were directed to complete a post-study questionnaire, composed of a mix of closed- and open-ended questions (see Appendix B). We asked participants to rank their preference for different types of visual augmentations and identify which visual features or characteristics were particularly useful in recognizing unstable poses in each visual augmentation condition. Open-ended questions inquired about their overall online teaching experience (e.g., availability in providing feedback and challenges), reasons for preferring specific visual augmentations based on their earlier responses, and any additional features or suggestions for future development to enhance their online instruction.

2.2.3 Equipment

We mounted the Tobii Pro Nano (Tobii AB, Danderyd, Sweden), an infrared-based eye tracker, at the lower edge of the monitor screen to collect eye-tracking data (Figure 2.4). The eye tracker used binocular tracking and sampled eye metrics at a 60 Hz sampling frequency. Participants used a standard RGB monitor (21-inch screen, 75 Hz refresh rate, Dell, USA) along with a wireless mouse and keyboard. Before data collection, we adjusted the chair's height, backrest, monitor angle (typically between 15° and 20°), and the distance between the monitor and the participants' eyes (65–80 cm) based on the recommended setup guidelines from the manufacturer (Tobii Eye Tracker Manager, Tobii AB, Danderyd, Sweden).



Figure 2.4 Experiment setup with a monitor equipped with a screen-mounted eye-tracking device.

Calibration was performed using both the Eye Tracker Manager software (Tobii Eye Tracker Manager, Tobii AB, Danderyd, Sweden) and Tobii Pro Lab software (Tobii Pro Lab, Tobii AB, Danderyd, Sweden). Baseline pupil diameter was obtained for each visual augmentation condition by showing static images of the same pose in each visual augmentation condition for 5 seconds. This baseline pupil diameter was later used to normalize the average pupil diameter measured during each trial.

2.3 Dependent Variable

2.3.1 Eye Tracking Metrics

Six eye-tracking metrics (i.e., number of fixations, glances, and saccades; peak velocity of saccades during entry and exit; and pupil diameter) were collected and filtered using the I-VT

Fixation gaze filter, following Tobii Pro Lab guidelines. For analysis, we created four Areas of Interest (AOIs), representing the areas for each video (see green boxes in Figure 2.5). The data extracted and normalized from the AOI displaying the unstable pose (target answer) were labeled as the “Target AOI.” The AOI selected by an instructor was labeled as the “Selected AOI.” If the instructor correctly identified the unstable pose, their Target AOI and Selected AOI were identical.



Figure 2.5 AOIs (green boxes) around videos A-D. Blue circles show fixations, with larger ones indicating longer durations, and numbers in circles marking the fixation sequence.

2.3.2 Subjective Responses

Usability Questionnaire: A hybrid questionnaire (Appendix B) was created to capture a broader range of user experiences, specifically asking questions related to usefulness, confidence, and satisfaction by combining items from well-established sources. Items related to effectiveness and satisfaction were derived from the System Usability Scale (SUS) (Brooke, 1995). Questions addressing cognitive aspects, such as users' ability to comprehend exercise poses and their confidence in decision-making, were adapted from the Technology Acceptance Model (TAM) (Davis, 1989). Additionally, elements of user enjoyment and overall satisfaction were included from the User Experience Questionnaire (UEQ) (Laugwitz et al., 2008). Each category consisted of 4-5 sub-questions. To ensure consistency and comparability, all responses were measured on a scale from 0 to 100 (0 = Strongly Disagree, 100 = Strongly Agree), in line with widely accepted

usability measurement practices.

Perceived Workload: Following each trial, participants were asked to evaluate their perceived workload using the NASA Task Load Index (NASA-TLX) (Hart et al., 1988), which consists of six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Participants rated each of these dimensions on a scale, allowing for a comprehensive assessment of the demands imposed by the tasks.

Usability Feedback: To gather more detailed usability feedback, we developed a mix of closed- and open-ended questions (Appendix C) to explore participants' preferences and experiences with the visual augmentations used during the study. Specifically, participants were asked to rank their preferred visual augmentation and to elaborate on their selection, including which visual features were particularly helpful or unhelpful in identifying unsafe poses with the specific visual augmentation (e.g., angle, body alignment, instability). They were also asked to suggest additional features that could enhance the performance evaluation process in online yoga instruction.

2.4 Data Processing

Normalizations were performed for three eye-tracking metrics due to differences in assessment duration. The number of fixations, glances, and saccades in each pose assessment was normalized by dividing the total count (either in the Target or Selected AOI) by the average of the three remaining AOIs.

The average pupil diameter was normalized using the baseline pupil diameter for each visual augmentation, which was collected at the start of the experiment for each participant.

2.5 Data Analyses

A chi-squared test of independence was performed to compare assessment performance between different visual augmentation and camera view conditions. Separate repeated measures analyses of variance (ANOVAs) were performed for eye-tracking measures, usability questionnaire ratings, and NASA-TLX ratings, with visual augmentation and camera views as fixed effects and pose as a blocking effect. Where relevant, significant main and interaction effects were assessed using test slices for pairwise comparisons. The model assumptions of ANOVA were tested before performing the analyses, and none were violated. Normality was confirmed using the Shapiro-Wilk test ($p > .05$), while homogeneity of variances and independence of observations were verified. All

statistical analyses were performed using JMP Pro (JMP® 17.2.0, NC: SAS Institute Inc., USA) with the restricted maximum likelihood (REML) method and a significance level of $p < .05$. For statistical analysis of eye-tracking data, we excluded the Mountain pose (see Appendix A) from further analysis due to the absence of noticeable differences between its stable and unstable forms in the videos presented.

Usability questionnaire ratings were calculated by averaging the ratings of 4-5 sub-questions from each category: usefulness (the practical utility of the visual augmentations), confidence (the instructor's belief in the accuracy of their assessment), and satisfaction (the overall satisfaction with the visual augmentations and camera views provided). Similarly, the overall workload score was calculated by averaging the ratings of six subscales from the NASA-TLX (Hart & Staveland, 1988).

Descriptive statistics were calculated for the close-ended usability feedback questions. To interpret the results of the open-ended questions, a thematic analysis approach was undertaken following the methodology outlined by Lochmiller (2021). Initially, all open-ended responses were segregated and meticulously examined to gain familiarity with the respondents' varied answers. An inductive approach was adopted to conduct thematic analysis, a widely used qualitative research method (Thomas, 2006). The coding process involved carefully labeling each theme to accurately capture the essence of the participants' feedback. These codes were then reviewed for coherence and consistency to ensure the accuracy of the thematic representation. Responses with similar themes were grouped to form three-level themes. Subsequently, these thematic clusters were systematically coded to identify overarching themes based on the content of the responses. This hierarchical structuring allowed for a comprehensive understanding of the major themes emerging from the usability feedback data.

3. Results

3.1 Assessment Performance

Participants showed no performance differences between the different visual augmentations [$\chi^2(3) = 0.77, p = .855$] and camera views [$\chi^2(1) = 2.78, p = .095$]. The mean (SD) rates of correct assessment were 76.4 (1.5), 79.3 (1.6), 77.7 (1.2), and 79.8 (1.3)% for Raw, Skeleton, Contour, and Contour + Skeleton, respectively.

3.2 Eye Tracking Metrics

3.2.1 Target AOI

No significant main or interaction effects were found for the eye tracking metrics calculated for the Target AOI (Table 1).

Table 3.1 Summary of ANOVA results [F -value (p -value)] for the Effects of Visual Augmentations and Camera Views on normalized eye tracking metrics in Target AOI.

Normalized Eye Tracking Metrics Target AOI	Effects F (p)		
	Visual Augmentations	Camera Views	Visual Augmentation x Camera Views
Number of fixations	1.39 (.243)	0.19 (.658)	0.13 (.941)
Number of Glances	0.36 (.781)	1.85 (.173)	1.19 (.311)
Number of Saccades	2.24 (.081)	1.45 (.227)	0.10 (.959)
Peak Velocity of Entry Saccade	1.68 (.168)	0.93 (.333)	0.34 (.798)
Peak Velocity of Exit Saccade	0.18 (.909)	0.22 (.632)	0.46 (.708)

3.2.2 Selected AOI

Significant interaction effects were found for the normalized number of fixations and glances (Table 2). However, we did not find any significant pairwise comparisons from the post hoc analysis performed on the normalized number of fixations (Figure 3.1). The normalized number of glances (Figure 3.2) was higher for the Single Camera View in the Contour + Skeleton condition.

The normalized number of saccades (Figure 3.3) had a significant main effect. Instructors showed a significantly greater normalized number of saccades with Contour + Skeleton condition

compared to Skeleton.

Table 3.2 Summary of ANOVA results [*F*-value (*p*-value)] for the Effects of Visual Augmentations and Camera Views on normalized eye tracking metrics in Selected AOI. Significant differences are shown in bold text. R = Raw; S = Skeleton , C = Contour , C+S = Contour + Skeleton.

Normalised Eye Tracking Metrics Selected AOI	Effects <i>F</i> (<i>p</i>)		
	Visual Augmentations	Camera Views	Visual Augmentations x Camera Views
Number of fixations	0.38 (.761)	0.24 (.617)	0.24 (.617)
Number of Glances	0.87 (.454)	1.32 (.250)	3.03 (.028) C+S: Single > Multiple
Number of Saccades	2.70 (.044) C+S > R : C > S	0.33 (.568)	0.88 (.446)
Peak Velocity of Entry Saccade	0.72 (.536)	1.08 (.298)	0.56 (.639)
Peak Velocity of Exit Saccade	0.77 (.506)	0.43 (.513)	0.36 (.785)

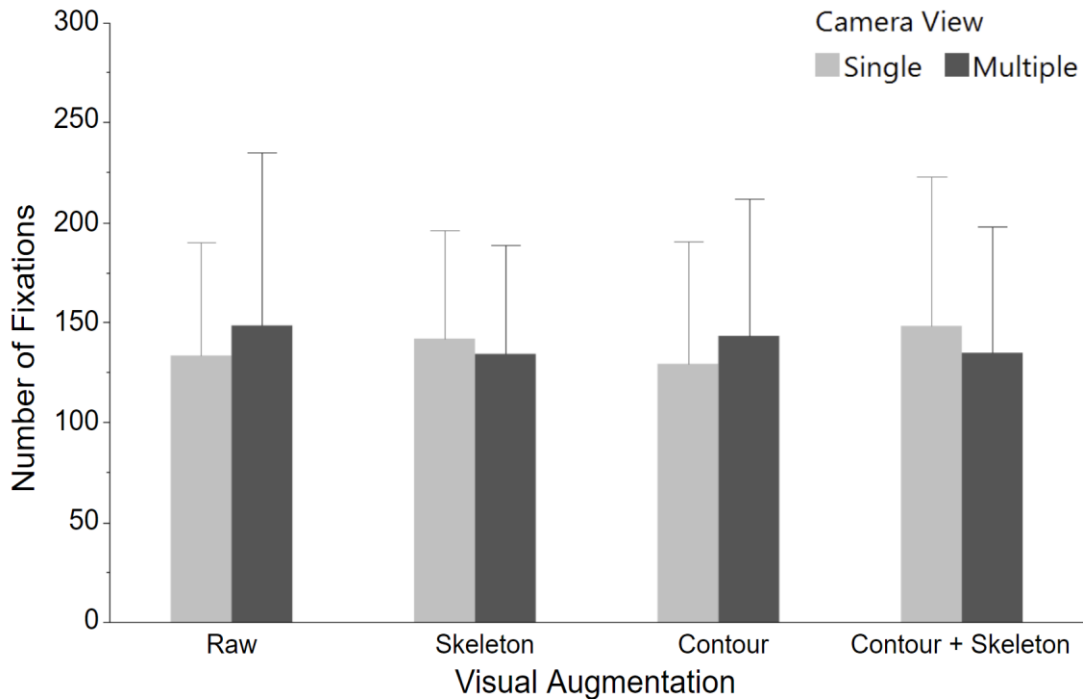


Figure 3.1 Normalized mean number of fixations across different Visual Augmentations and Camera Views. Error bar = standard deviation.

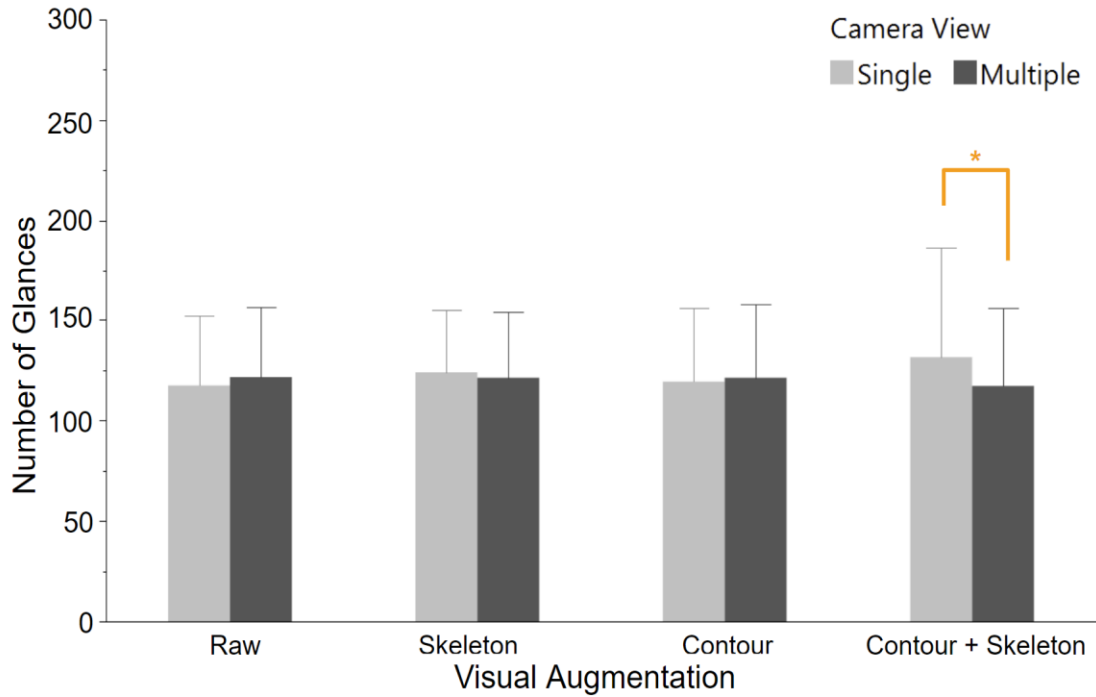


Figure 3.2 Normalized mean number of glances across different Visual Augmentations and Camera Views. “*” indicates pairwise significant interaction effect Error bar = standard deviation.

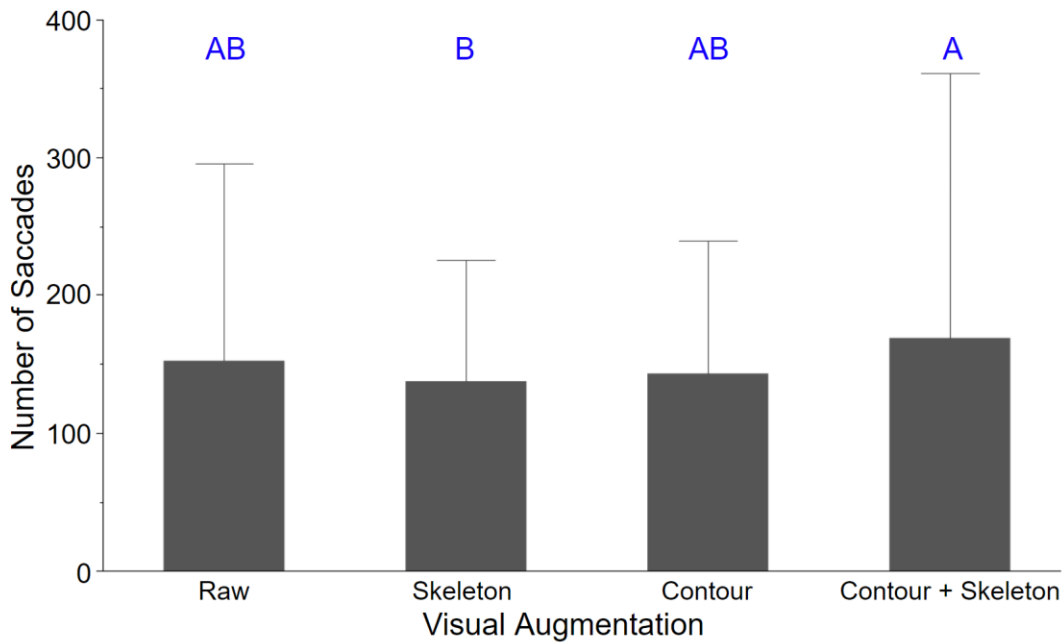


Figure 3.3 Normalized mean number of saccades across different Visual augmentations. Conditions with different letters are significantly different. Error bar = standard deviation.

3.2.3 Average Pupil Diameter

A significant main effect of visual augmentation was found for average pupil diameter [$F(3, 940) = 16.37, p < .0001$]. The Skeleton and Contour + Skeleton conditions had a higher average pupil diameter followed by Raw and Skeleton.

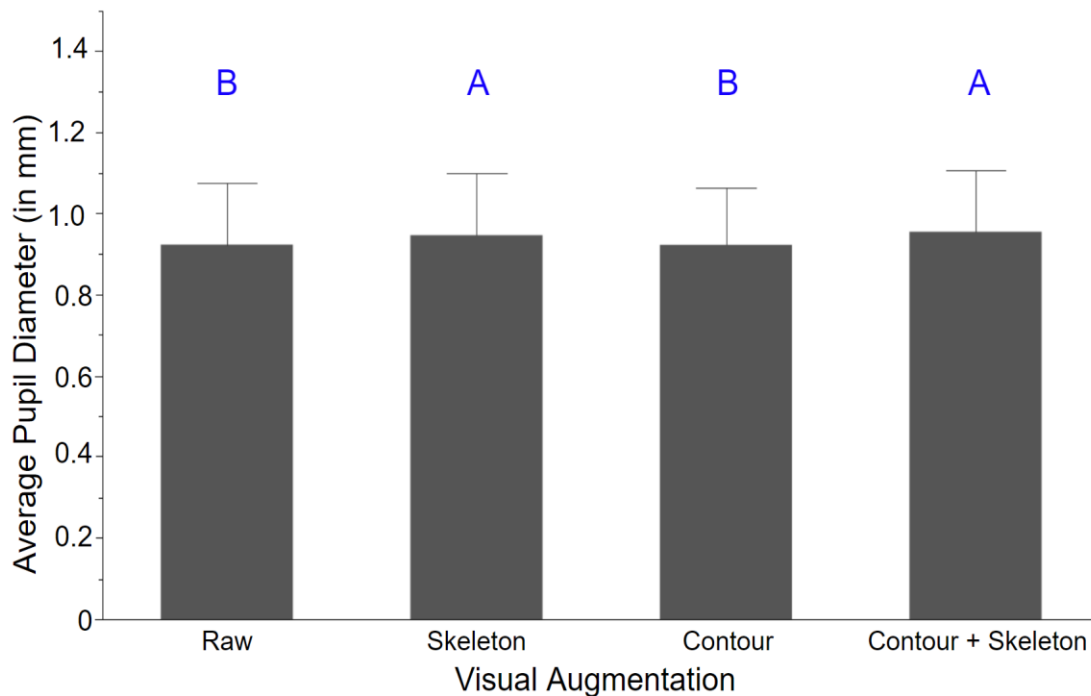


Figure 3.4 Average pupil diameter across Visual Augmentations. Conditions with different letters are significantly different. Error bar = standard deviation

3.3 Subjective Responses

3.3.1 Usability Questionnaire

All three usability categories (usefulness, confidence, satisfaction) showed similar main effects for both visual augmentations and camera views, with post-hoc results indicating that the satisfaction category had a significant interaction effect. Across all usability measures, participants rated Raw as the best option compared to the other visual augmentations. The only exception was in the usefulness category, where Raw and Skeleton showed no significant difference. Skeleton was rated better than Contour + Skeleton, which, in turn, was rated better than Contour for all three usability categories. Participants found the Single View better compared to the Multiple View.

Table 3.3 Summary of ANOVA results [F -value (p -value)] for the Effects of Visual Augmentations and Camera Views on Usability Ratings. Significant differences are shown in bold text. R = Raw; S = Skeleton; C = Contour; C+S = Contour + Skeleton.

Usability Ratings	Effects F (p)		
	Visual Augmentations	Camera Views	Visual Augmentations x Camera Views
Usefulness	159.91 (<.0001) R, S > C+S > C	36.57 (<.0001) Single > Multiple	1.0 (.39)
Confidence	259.08 (<.0001) R > S > C+S > C	80.59 (<.0001) Single > Multiple	1.99 (.112)
Satisfaction	229.88 (<.0001) R > S > C+S > C	32.39 (<.0001) Single > Multiple	3.05 (.027) R, C+S, S: Single > Multiple

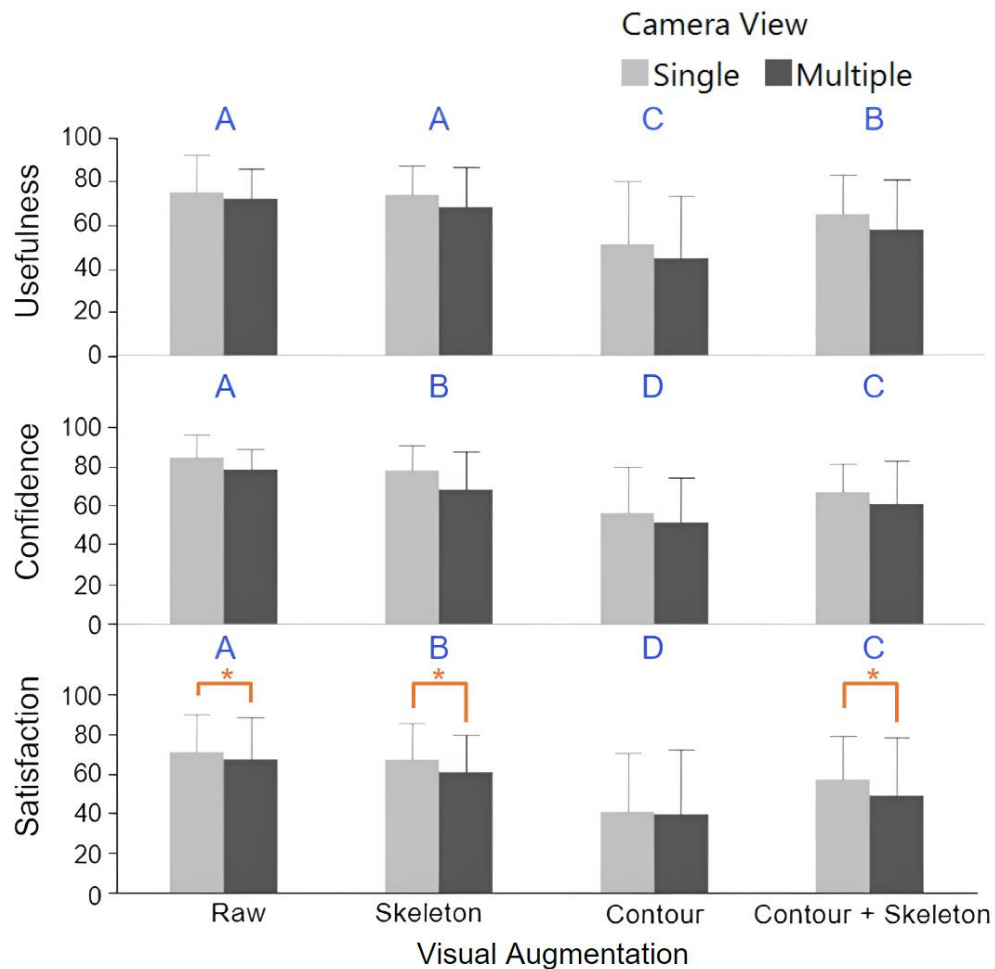


Figure 3.5 Mean Usability Ratings (Usefulness, Confidence, and Satisfaction) across different Visual Augmentations and Camera Views. “*” indicates a significant paired difference ($p < .05$). Conditions with different letters are significantly different. Error bar = standard deviation

3.3.2 Perceived Workload

A significant interaction effect between visual augmentations and camera views was observed for perceived workload [$F = 5.85, p = .0006$] (Figure 3.6). Within the Skeleton and Contour + Skeleton conditions, participants reported a greater workload with the Multiple View. However, this was the opposite for the Raw video condition.

Visual augmentations had a main effect [$F = 81.74, p < .0001$], with all visual augmentation conditions being statistically different from each other. Participants rated the Raw video as the least demanding, followed by Skeleton, Contour + Skeleton, and Contour.

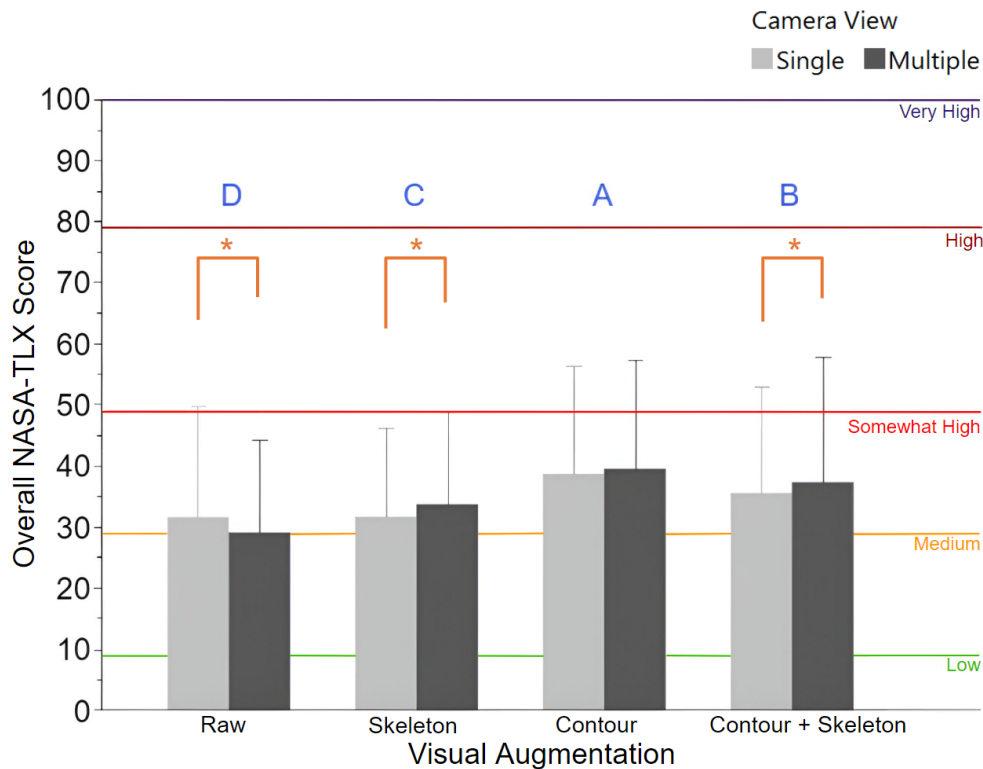


Figure 3.6 Overall NASA-TLX Score across different Visual Augmentations and Camera Views. “ * ” indicates a significant paired difference ($p < .05$). Conditions with different letters are significantly different. Error bar = standard deviation

3.3.3 Usability Feedback

The Raw video was selected as the most preferred visual augmentation type (64%), followed by Skeleton (27%). Both Contour (10%) and Contour + Skeleton (0%) were the least preferred. About 80% of the participants selected the single camera view as their first preference. Participants found augmented visual features helpful in distinguishing unsafe poses because they showed body alignment, pose, and angle. Participants reported that gestures and facial expressions were helpful, but noted that these were only available in the Raw videos. The Skeleton visual augmentation was

reported as very helpful, as it offers deeper insights into body structures and is not limited to just pose alignment. Some participants commented that Skeleton was helpful when students wore baggy clothes, which hindered accurate assessment of body alignments and joint angles.

We identified three broad first-level themes from the thematic analysis: 1) Challenges in assessing performance; 2) Helpful/unhelpful visual features; and 3) Potential improvements. The first, second, and third-level themes are depicted in Figure 3.7. A complete list of themes and associated quotes is summarized in Appendix E.

Challenges in Assessing Performance: Three second-level themes were identified from the challenges associated with assessing performance in online yoga teaching: camera positioning and visibility, color contrast and background, and increased cognitive load. Participants found the ineffectual use of cameras or the lack of visual feeds to be the primary challenges in online performance assessment. Even when visual feeds are available, it is still challenging to assess pose and alignment if there are distractions in the background or visual clutter, which increase cognitive load.

Helpful/Unhelpful Visual Features: Participants mentioned that Raw video was their preferred mode because it allowed them to observe hands and feet clearly, while adding visual augmentations like Skeleton or Contour increased visual clutter. Skeleton was perceived as helpful for inspecting pose and alignment, while Contour was considered useful for focusing on pose details, as it removed visual distractions from the background.

Potential Improvements: Due to the limitations of visual augmentations, participants found the flickering images distracting and fatiguing. They also suggested displaying visual information that is not easily perceived through the videos, such as showing weight distribution to assess the balance and safety of a pose. Providing instructions on effective camera use for safety was also recommended, which could encourage students to use their cameras more.



Figure 3.7 Illustration of all three categories (round boxes), 9 secondary level themes (capsule shaped boxes), and related tertiary level themes (italicized text in rectangular boxes).

4. Discussion

4.1 Visual Augmentations

We found that instructors' performance was not enhanced by augmentations like Skeleton and Contour when compared to the Raw video feed, and in some cases, they introduced new cognitive challenges. Instructors may be accustomed to assessing students with standard video feeds (Raw), rendering the additional visual augmentations less beneficial. Despite the benefits of highlighting pose alignment and reducing distractions, the increased visual complexity seemed to be distracting to instructors rather than aiding their performance, potentially overloading their visual attention. The ratings on the usability questionnaire supported this, with Raw consistently selected by the instructors as the most preferred option across three usability categories. Conversely, Contour was rated as the least effective and most challenging to use, confirming its higher cognitive demands and lower usability.

Eye-tracking metrics revealed that both Contour + Skeleton and Skeleton conditions led to an increase in the normalized number of fixations, glances, and saccades compared to the Raw condition. These increased visual search behaviors suggest that instructors had to exert more cognitive effort navigating the augmented visuals, as frequent saccades and longer fixations reflect increased mental effort (Van Orden et al., 2001). Glances and saccades were more frequent in the Contour + Skeleton condition, indicating difficulty in quickly processing the visual information. The increased pupil diameter observed in the Contour + Skeleton and Skeleton conditions further reinforces the notion of heightened cognitive load in these augmented visual conditions. Pupil dilation is a well-established indicator of mental effort, and larger diameters in these conditions suggest that instructors were experiencing greater processing load as they attempted to evaluate participants' postures (Beatty, 1982; Kahneman & Beatty, 1966; Laeng et al., 2012). This result confirms the hypothesis that complex visual augmentations like Skeleton and Contour + Skeleton increase cognitive demands, requiring more mental effort to interpret and assess body movements. However, despite Contour being rated as the least preferred option in subjective responses, pupil diameters during the Contour condition were as low as in Raw, suggesting lower cognitive demands compared to Skeleton or Contour + Skeleton.

The Raw visual augmentation was preferred for its simplicity and lower cognitive demands. While further improvement is needed, Contour could serve as an alternative to no videos when privacy is a concern for the participating students. This is supported by the fact that assessment performance was not significantly affected by this visual augmentation.

4.2 Camera Views

The hypothesis that Single Camera View would lead to lower cognitive load and be preferred by instructors was confirmed, based on the analysis of usability ratings and perceived cognitive load. In contrast, the Multiple Camera Views provided more comprehensive coverage but led to visual clutter, as instructors had to manage and switch between multiple visual streams. This was reflected in the increased frequency of glances, which are indicators of heightened cognitive demand (Van Orden et al., 2001). This is likely due to the limited perspective provided by a single camera view, which made it harder for instructors to get a complete view of the participant's pose (Henderson & Ferreira, 2004). This could also indicate that instructors were more engaged with the Single View, potentially focusing more on the task simply due to the ease of comparing the visual feeds in a 2 x 2 tile format (Rensink, 2000).

However, the limited field of view in Single View may have hindered instructors' ability to thoroughly assess certain poses (e.g., High Plank, Sphynx), leading them to make more glances to compensate for the restricted perspective (Liversedge & Findlay, 2000; Henderson & Ferreira, 2004). Camera view yielded more mixed results compared to visual augmentation, with only usability ratings and perceived cognitive load favoring Single View, while other metrics showed the opposite.

The only exception was the Contour + Skeleton augmentation, in which the Single View led to a higher mean (SD) normalized number of glances [131.8 (54.6)] compared to Multiple View [117.5 (38.9)]. This finding supports earlier research that showed while multiple views provide more information, they also require more mental effort to process, potentially increasing the likelihood of errors (Barron et al., 2014).

4.3 Limitations

A limitation of this study is the small sample size ($n = 11$), which may have impacted the power of the statistical testing. In our a priori testing using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), the required sample size was calculated as 18. Due to the study's inclusion criteria for extracting eye-tracking data and selecting eligible yoga instructors, recruiting a larger sample was not feasible. Despite efforts to distribute flyers across various yoga studios around the Virginia Tech campus, including Blacksburg and Christiansburg, recruitment challenges persisted.

Another drawback is the use of the I-VT Fixation gaze filter, which is better suited for screen-based visual stimuli where head movement is more restricted, as opposed to studies involving real-world or dynamic stimuli (Salvucci & Goldberg, 2000; Olsen, 2012). In this study, dynamic videos were presented in a 2x2 tile format. Using a gaze filter designed for dynamic stimuli, such as the I-VT Attention filter, which accounts for rapid eye and head movements, may have uncovered significant interactions between visual augmentation and camera views on metrics like saccades.

An important constraint of the study is the design of the Multiple View condition, which departs from standard practices in the field of pose estimation and performance guidance (Rodriguez-Criado et al., 2024). In most studies, multiple views focus on a single participant viewed from different camera angles, providing a comprehensive assessment. However, this study utilized a 2x2 tile format, where participants were shown in varied orientations. This restricted the yoga instructors' ability to fully assess performance and provide consistent feedback based on typical pose estimation methods. While this setup mimics real-world online yoga classes, it introduces complexity that limits direct comparisons with prior studies. This also raises concerns about whether less experienced instructors would be able to effectively interpret these diverse visual inputs, potentially reducing the overall effectiveness of the visual augmentation tools in real-world applications.

Furthermore, a limitation of this study was the absence of audio cues and direct interaction between the yoga instructors and the students. We used pre-recorded videos to control the experimental settings, including the background and participant performance (one unsafe video among four options). However, this approach might have limited the realism and ecological validity of the instruction scenarios. In real-world settings, instructors rely heavily on verbal feedback and real-time interaction to guide and correct students' postures. This lack of interaction might have limited the instructors' ability to accurately assess the poses, as assessments often depend on verbal feedback and real-time interaction to provide context and depth to visual information.

Additionally, the Mountain Pose was excluded from the eye-tracking data due to its lack of noticeable instability between different forms, which may reduce the applicability of the findings across all yoga poses. The reliance on self-reported measures of cognitive load (NASA-TLX) also introduces potential biases, as instructors may not fully perceive or accurately report their cognitive effort during the task. Additionally, instructors are typically positioned much farther away from their camera or screen during pose demonstrations, making it significantly more challenging for them to evaluate performance accurately from small visual feeds. Future studies should consider

incorporating audio cues and live interaction to better simulate a true teaching environment and more accurately assess the impact of visual augmentations and camera views.

Lastly, the current study used a 2x2 video tile format, which is still less than the usual number of participating students in online yoga classes. Instructors' performance and eye-tracking movements could be more notable with an increased number of visual feeds. Furthermore, while eye-tracking metrics such as fixations and saccades provide insights into visual attention and cognitive load, the real-world applicability of these findings in dynamic online environments remains to be fully explored.

4.4 Future Research Direction

Future studies should explore alternative visual augmentation techniques that simplify rather than complicate the visual field. Heat maps or dynamic overlays that provide real-time feedback on specific body parts or movements could be more effective in guiding instructors' attention without overwhelming them. Additionally, integrating real-time motion detection algorithms could offer automated assistance in identifying unstable poses, thus reducing the instructors' cognitive load. Further research could also examine privacy-preserving augmentations that balance participants' privacy with instructors' ability to assess performance, addressing concerns about Social Physique Anxiety (Brunet & Sabiston, 2009).

Research should also investigate varying camera angles to understand how they impact the performance evaluation of yoga poses. Ideally, adaptive camera views for each pose (e.g., a side view for the Warrior pose and a frontal view for the Tree pose) could enhance performance assessment. Allowing instructors to choose from different visual augmentations and camera views could offer valuable insights into their preferences and adaptive strategies. Incorporating audio components could also improve instructional flexibility and reduce cognitive load. Additionally, expanding the scope to include multiple visual feeds from students could provide a deeper understanding of how different visual augmentations and camera setups influence assessment accuracy. Moreover, shifting the research focus toward identifying specific types of errors rather than solely assessing pose safety within a limited timeframe could reveal the trade-offs between efficiency and thoroughness in online yoga instruction. This approach would help optimize the use of visual tools and improve overall instructional effectiveness.

5. Conclusion

This study tested the hypothesis that overlaying human joint locations on raw videos (Skeleton) would enhance visual search effectiveness, despite potentially increasing cognitive workload. We also hypothesized that displaying only the contour of participants' videos (Contour) would be less effective, but combining joint locations with contour (Contour + Skeleton) might improve visual search.

While the Skeleton augmentation showed a trend toward increased cognitive workload, as indicated by larger pupil diameters, it did not significantly enhance visual search effectiveness or improve assessment accuracy compared to other augmentations. Contour alone was less effective for visual search, aligning with our hypothesis, but the combination of Contour and Skeleton did not demonstrate a clear improvement in visual search or assessment performance and proved to be more cognitively demanding. Usability ratings indicated that simpler visualizations, such as Raw and Skeleton, were preferred, with the Single View layout favored over Multiple Views for ease of use and reduced cognitive strain.

Overall, while Skeleton augmentation increased cognitive load, it did not significantly enhance visual search performance as hypothesized. Future research should explore additional visual augmentation techniques and configurations, including varied camera angles and audio components, to better understand their impact on visual search effectiveness and cognitive workload in online yoga instruction.

References

- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3), 191–204. [https://doi.org/10.1016/0025-5564\(75\)90075-9](https://doi.org/10.1016/0025-5564(75)90075-9)
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371–372. <https://doi.org/10.3758/BF03328444>
- Brinsley, J., Smout, M., & Davison, K. (2021). Satisfaction with Online Versus In-Person Yoga During COVID-19. *The Journal of Alternative and Complementary Medicine*, *27*(10), 893–896. <https://doi.org/10.1089/acm.2021.0062>
- Brunet, J., & Sabiston, C. M. (2009). Social physique anxiety and physical activity: A self-determination theory perspective. *Psychology of Sport and Exercise*, *10*(3), 329–335. <https://doi.org/10.1016/j.psychsport.2008.11.002>
- Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications*, *4*(1), 7. <https://doi.org/10.1186/s41235-019-0159-2>
- Canny, J. (1998). *OpenCV: Canny Edge Detection*. Retrieved September 11, 2024, from doi: 10.1109/TPAMI.1986.4767851
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields* (arXiv:1812.08008). arXiv. <https://doi.org/10.48550/arXiv.1812.08008>
- Collins, C. (1998). Yoga: Intuition Preventive Medicine Treatment. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, *27*(5), 563–568. <https://doi.org/10.1111/j.1552-6909.1998.tb02623.x>
- Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., Catena, A., & Pannasch, S. (2010). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, Space, and Environmental Medicine*, *81*(4), 413–417. <https://doi.org/10.3357/asm.2579.2010>
- Ericsson, K. A. (2006). The Influence of Experience and Deliberate Practice on the Development

- of Superior Expert Performance. In *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.038>
- Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. (2007) G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, 39, 175-191. <http://dx.doi.org/10.3758/BF03193146>
- Guo, J., & Fussell, S. R. (2022). “It’s Great to Exercise Together on Zoom!”: Understanding the Practices and Challenges of Live Stream Group Fitness Classes. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1–28. <https://doi.org/10.1145/3512918>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504. <https://doi.org/10.1016/j.tics.2003.09.006>
- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, 9(1), 32. <https://doi.org/10.1167/9.1.32>
- Henderson, J. M., & Ferreira, F. (2004). Scene Perception for Psycholinguists. In *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). Psychology Press.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In *Eye guidance in reading and scene perception* (pp. 269–293). Elsevier Science Ltd. <https://doi.org/10.1016/B978-008043361-5/50013-4>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide To Methods And Measures*.
- Islam, M. S., Frazier, M. C., Harden, S. M., & Lim, S. (2024). Barriers and Benefits of Online Group Exercise Programs for Older Adults. *Journal of Applied Gerontology*, 07334648241240599. <https://doi.org/10.1177/07334648241240599>
- Islam, M. S., Harden, S., Lee, S. W., & Lim, S. (2024). Verbal and Nonverbal Communication Differences between In-Person and Live-Streamed Group Physical Activity: A Specific Investigation into Yoga Instruction. *SSRN Electronic Journal*.

- <https://doi.org/10.2139/ssrn.4802390>
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(1), 18–27. <https://doi.org/10.1177/1745691611427305>
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14. [https://doi.org/10.1016/S1364-6613\(99\)01418-7](https://doi.org/10.1016/S1364-6613(99)01418-7)
- Lochmiller, C. (2021). Conducting Thematic Analysis with Qualitative Data. *The Qualitative Report*, 26(6), 2029–2044. <https://doi.org/10.46743/2160-3715/2021.5008>
- Lohrenz, M. C., & Beck, M. R. (2010). Evidence of Clutter Avoidance in Complex Scenes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(18), 1355–1359. <https://doi.org/10.1177/154193121005401811>
- Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of Mental Workload during En-route Air Traffic Control Task Execution Based on Eye-Tracking Technique. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 592–597). Springer. https://doi.org/10.1007/978-3-642-21741-8_63
- McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual Skills in Airport-Security Screening. *Psychological Science*, 15(5), 302–306. <https://doi.org/10.1111/j.0956-7976.2004.00673.x>
- Olsen, A., & Matos, R. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 317–320. <https://doi.org/10.1145/2168556.2168625>
- OpenCV. (2015). Open Source Computer Vision Library.
- Palinko, O., Kun, A. L., Shyrov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, 141. <https://doi.org/10.1145/1743666.1743701>
- Poole, A., Ball, L. J., Poole, A., & Ball, L. J. (1 C.E., January 1). *Eye Tracking in HCI and Usability Research* (eye-tracking-hci-usability-research) [Chapter]. <https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-59140-562-7.Ch034>; IGI Global. <https://doi.org/10.4018/978-1-59140-562-7.ch034>

- Raub, J. A. (2002). Psychophysiologic Effects of Hatha Yoga on Musculoskeletal and Cardiopulmonary Function: A Literature Review. *The Journal of Alternative and Complementary Medicine*, 8(6), 797–812. <https://doi.org/10.1089/10755530260511810>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (2006)*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6(1), 31–43. <https://doi.org/10.1037/1076-898X.6.1.31>
- Rishan, F., De Silva, B., Alawathugoda, S., Nijabdeen, S., Rupasinghe, L., & Liyanapathirana, C. (2020). Infinity Yoga Tutor: Yoga Posture Detection and Correction System. *2020 5th International Conference on Information Technology Research (ICITR)*, 1–6. <https://doi.org/10.1109/ICITR51448.2020.9310832>
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 17. <https://doi.org/10.1167/7.2.17>
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00*, 71–78. <https://doi.org/10.1145/355017.355028>
- Sharma, N., Modi, D., Nathwani, A., Pandya, B., & Joshi, J. (2020). *Tele- Yoga therapy for Patients with Chronic Pain during Covid-19 Lockdown: A Prospective Nonrandomized Single Arm Clinical Trial*. <https://doi.org/10.1101/2020.07.16.20154229>
- Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes. *IEEE Reviews in Biomedical Engineering*, 16, 260–277. *IEEE Reviews in Biomedical Engineering*. <https://doi.org/10.1109/RBME.2021.3066072>
- Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748>
- Tobii Pro Nano, Tobii AB, Dandyard, Sweden.
<https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-nano>
- Tobii Pro Lab, Tobii AB, Dandyard, Sweden.

<https://www.tobii.com/products/software/behavior-research-software/tobii-pro-lab>











Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive*











Psychology, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)





Woodyard, C. (2011). Exploring the therapeutic effects of yoga and its ability to increase quality of life. *International Journal of Yoga*, 4(2), 49–54. <https://doi.org/10.4103/0973-6131.85485>

Appendices

Appendix A. A list of Yoga Poses Used in the study with example pictures of stable and potentially unstable poses.

No.	Pose Name	Stable Pose	Potentially Unstable Pose
1	Kneeling		
2	Cow		
3	Cat		
4	High Plank		
5	Downward Facing Dog		

6	Tree (Kickstand)		
7	Half Moon		
8	Warrior 1		
9	Warrior 2 (Modified)		
10	Extended Side Angle		

11	Sphinx		
Excluded from Data Analyses for Eye Tracking Metrics			
12	Mountain		

Appendix B. Usability Questionnaire

Please read the statements and respond to them on a scale of 0 to 100. 0 = Strongly Disagree ; 100 = Strongly Disagree

Useful

1. It helps me to be more effective
2. It is useful
3. It makes the things I want to accomplish easier to get done
4. It meets my needs

Confidence

5. I can comprehend trainee's exercise pose
6. I can easily perceive change in the trainee's exercise pose
7. I am confident on my decision
8. I am so confident that I do not want to change my answer

Satisfaction

9. I am satisfied with it
10. I would recommend it to a friend
11. It is fun to use
12. It is wonderful
13. It is pleasant to use

Appendix C. Usability Feedback

1. Do you teach yoga online?
2. If you teach online, can you describe the difficulties you face while assessing trainees' poses?
For example, are you facing difficulties in assessing poses while you need to monitor multiple video feeds from different participants, which is cognitively demanding? Or are you unable to assess due to participants turning off their videos?
3. Do you often provide feedback to the group as a whole or to a specific individual?
4. If you provide feedback for pose correction or encouragement, does it differ by online vs. in-person instruction?
5. Please rank your preference on the following visual feeds: [Raw], [Skeleton], [Contour], [Contour + Skeleton]
6. Any reasons for making the above choice?
7. Do you think the additional visual augmentation (e.g., skeleton, contour) helps assess participants' poses? Why?
8. Any other functions you would like to add to help with visual inspection of the poses besides the visual features we provided?
9. It is easier to identify unsafe poses when the camera angle is : a. Same in all feeds b. Different in all feeds
10. Select which visual feature/characteristics were useful in identifying unsafe poses in the specific video condition (Multiple selections are allowed): Raw, Skeleton, Contour, and Contour + Skeleton
 - a. Angle; b. Body alignment; c. Participants facial expression; d. Pose; e. Instability

Appendix D. Average Pupil Diameter

Table 3.4. Summary of ANOVA results [F-value (p-value)] for the Effects of Visual augmentation and Camera view on Average Pupil Diameter. Significant differences are shown in bold text. R - Raw; S - Skeleton , C -Contour , C+S - Contour + Skeleton, SV- Single View, MV- Multiple View

Eye Tracking Metric	Effects (F, p)		
	Visual Augmentation	Camera View	Visual Augmentation x Camera View
Average Pupil Diameter	16.37 (< .0001) C+S : S > R : C	0.04 (0.847)	2.09 (.099)

Appendix E. Extended Thematic analysis

First-Level Theme	Second-Level Theme	Tertiary-Level Theme	Quotes
Challenges in Assessing Performance	Camera Positioning and Visibility	Ineffective Camera Positioning	<p>P1: “I think the biggest challenge is camera angle, lighting, ...”</p> <p>P3: “Students not being centered with the camera or you can only see part of them when they are standing ...”</p> <p>P2: “It's easiest for me to assess if the long edge of the mat is parallel with the camera, but then of course they can't see me as well.”</p> <p>P3: “... students turning to look at the screen in poses more so than when in the studio, distractions in the background that divert your attention, not being able to move to look at them and get a better angle ...”</p> <p>P4: “... it is easier to request that their cameras are on and we work with them to adjust their cameras to provide the best angle to see their shapes.”</p> <p>P4: “The most challenging is when the students don't situate their cameras so that you can't see their full form ..., which makes it actually impossible to assess their shapes. It changes the postures I choose to teach during an online class.”</p>
		Limited Full Body Visibility	<p>P2: “It is hardest when students are facing me at the front or short side of the mat with their bodies lined up away from me...”</p> <p>P4: “The most challenging is when the students don't situate their cameras so that you can see their full form. It is nearly impossible to support them if you cannot see their bodies....”</p> <p>P8: “... usually when they are on, it is an incomplete frame of their body and form ...”</p>

			P1: “In person there is the ability to get a 360 view if needed for visual inspection as well as an opportunity to give hands on assists/use of props.”
		Small Video Feed	P5: “Yes, and the challenges are primarily the size of the image, and not being familiar with each person's practice and body.” P6: “... and the challenges are primarily the size of the image”
		Participant Behavior (Turning off Cameras)	P1: “Often times participants turn there video off” P2: “Some students don't turn the video on at all, so I don't even know they are there. I usually have only one or two students with videos during class.” P4: “Many students do practice with their videos off, which makes it actually impossible to assess their shapes” P4: “... it is easier to request that their cameras are on and we work with them to adjust their cameras to provide the best angle to see their shapes.” P5: “People often leave their cameras turned off on Zoom.”
	Color Contrast, Background	Pose Assessment	P1: “... clothing color against background color, clothing choices (baggy clothing limits visibility of joint orientation)” P3: “Yes because I can better see the lines of the body (with the Skeleton) especially through baggy clothing.”
		Distractions in Background	P3: “... distractions in the background that divert your attention, not being able to move to look at them and get a better angle.”
Increased Cognitive Load	Managing Multiple Video Feeds	P7: “It's difficult to see everyone if you have more students than Zoom screens that fit in one configuration.”	
	Visual Clutter	P8: “I think that it can, when showing how joints are aligning (e.g. somebody's foot directly against their knee in tree pose) but sometimes it was a distraction or wholly inaccurate when picking up on the wrong visual data.”	

Helpful/Unhelpful Visual Features	Raw Video	Observing Hands and Feet	P1: “The raw allows for observation of the hands and feet which are typically the first place that I observe a posture especially in closed chain pose.” P9: “I felt like I could see the best in the raw videos.”
	Skeleton	Pose/Alignment Inspection	P1: “While the skeleton does provide insight into the alignment of the bones, it does not allow for inspection of joint orientation or orientation of the feet and hands.” P7: “Skeleton allows for a faster interpretation of posture versus raw.” P3: “I can better see the lines of the body especially through baggy clothing or hair that is in the way. I noticed things like broken lines in poses that should be straight much faster with the skeleton and contour because it stood out so much more. It was also easier to filter out the extra activity like people moving or shaking their hands because it didn't really change their overall shape in each of the poses.” P10: “I can see body movements more naturally in the raw or skeleton format” P11: “The skeleton was a little hard to get used to at first...but then it really grew on me. it was easier to see alignment.”
		Visual Clutter	P3: “... I thought I would like the skeleton, but it actually made it hard to see the person behind it. “ P9: Sometimes the skeleton obstructs the important visual information, especially at certain angles where the nodes overlap.
	Contour	Focus on Pose Details	P4: “I like the contour because it gives more focus but it was so much harder than raw when the students were at inconvenient angles ...”
		Challenges with Visual Effect	P1: “... The contour often does not show the hands and feet or it is blurry...”

			<p>P4: “I really struggled with the contour because the lines were shaking and some parts of the body dropped off the screen.”</p> <p>P7: “ The contour makes me work harder visually to capture the pose we are looking for versus raw...”</p> <p>P9: “ ... The contour alone often lacked sufficient visual data to see the whole participant.”</p>
Potential Improvements	Stability in Visual Effects	Removing Flickering	<p>P3: “...The skeleton is most useful for assessing poses, but because of the flickering, I prefer it raw. ”</p> <p>P3: “The flickering of both skeleton and contour were very hard after a while.”</p>
	Additional Visual Effects	Weight Distribution	P4: “If the skeleton coloring lightened and darkened on a scale based on pressure, that could be helpful in the combination of contour/skeleton in assessing the movements of the individual's body.”
		Improved Adjustment Guidance	P9: “I would prefer that I could (choose to see between) raw and skeleton simultaneously.”
	Technical Issues	Camera Instruction for Safety	P1: “A tool kit and explanation for having cameras on for safety could be beneficial.”

