CATWALKGRADER: A CATWALK ANALYSIS AND CORRECTION SYSTEM USING MACHINE LEARNING AND COMPUTER VISION

Tianjiao Dong[1] and Yu Sun[2]

[1]Northwood High School, Irvine, CA 92620
[2]California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

In recent years, the modeling industry has attracted many people, causing a drastic increase in the number of modeling training classes. Modeling takes practice, and without professional training, few beginners know if they are doing it right or not. In this paper, we present a real-time 2D model walk grading app based on Mediapipe, a library for real-time, multi-person keypoint detection. After capturing 2D positions of a person's joints and skeletal wireframe from an uploaded video, our app uses a scoring formula to provide accurate scores and tailored feedback to each user for their modeling skills.

KEYWORDS

Runway model, Catwalk Scoring, Flutter, Mediapie

1. INTRODUCTION

It’s no secret that fashion supermodels make a lot of money. [1, 8] In fact, most high fashion models’ salaries reach astronomical figures. For instance, Kendall Jenner’s net worth is estimated at a mind-boggling $22.5 million. This is a rare case, but the average annual salary for a professional model is $147,126.00, three times the average income of US citizens. Every year, thousands of young, beautiful, tall, ambitious people enter the modeling industry in hopes of becoming rich and famous. Just look at how many people audition for the “Fashion Model Reality TV Show’ or “American’s Next Top Model,” and how many viewers these TV shows have.

One of the determining factors for becoming a supermodel is a good runway walk, and there are hundreds of modeling training classes every year to teach this skill and others related to the industry.

There are modeling training classes online and in person to train models to execute a good runway walk. These classes are only effective when taught by experienced modeling instructors, which increases their price, making them too expensive for most just joining the industry since most beginners are also college students. Classes are also a limited resource, since trainers only have a certain amount of energy and time to focus on each student. Other alternatives, such as pre-recorded modeling training videos, cannot provide tailored feedback to each student, and are thus not the most effective way to train.

In this paper, we follow the same line of research given in the paper “Real-time Human Gesture Grading Based on OpenPose.” [13] Our goal is to develop an accessible mobile app that scores model walk videos and provides feedback for users to improve. Our method is inspired by the same paper. They presented a real-time, 2D human gesture grading system from monocular images based on OpenPose. [2, 3, 4] After capturing 2D positions of a person's joints and skeletal wireframe, the system computes an equation for the motion trajectory of every joint. Similarity metric was defined as the distance between motion trajectories of standard and real-time videos. A modifiable scoring formula was used for simulating the gesture grading scenario. We believe that this method provides a solution for our goal by providing cheap, accessible, and tailored feedback for each user.

In two application scenarios, we demonstrated how the above combination of techniques increases the accuracy, practicality, and efficiency of the application. First, through a random experiment comparing the accuracy and the efficiency of each method, we demonstrated that Mediapipe is better than OpenPose. [5, 6, 7] Second, we analyzed the accuracy of the two most practicable scoring formulas based on a controlled experiment.

The rest of the paper follows this structure: Section 2 provides details on the challenges we encountered during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges mentioned in Section 2; Section 4 presents the relevant details of the experiment; and Selection 5 presents related works. Finally, Section 6 provides concluding remarks and suggested future work of the project.

2. CHALLENGES

In order to design a real-time 2D model walk grading app based on Mediapipe, a few challenges were identified as follows.

2.1. Challenge 1: Variables Affecting How the Camera Collects Data

There is a lot of variation among runway walks. There could be exaggerated clothing that makes the motions of the body invisible to the camera; there could be long dresses that might be misinterpreted as a part of the leg or feet by the algorithm; there could be accessories that require the body to adjust (e.g., a purse the model needs to hold, or high heels that affect the gait, etc.). Even the placement of the camera could impact the interpretations of the algorithm, since front and side views can look completely different. All this indicates that we need to come up with a standard of collecting video feed so results remain consistent and accurate.

2.2. Challenge 2: Deciding Which Style Should Be Considered "Standard"

Another challenge in developing this catwalk grading algorithm was developing a set grading criterion. There are many different kinds of catwalks: the exaggerated style performed by Naomi Campbell and other supermodels of the 80s and 90s, which is to lean back or add an exaggerated swing of the hips; the "horse walk" pioneered by Gisele Bündchen, a stomping movement created when a model picks up her knees and kicks her feet out in front; or the most popular type presently, which is to walk "naturally." These styles are all very different, which begs the question of which the algorithm should perceive as its "standard." [10, 11]

2.3. Challenge 3: How to Capture and Evaluate Multiple Catwalk Components

There are many factors that determine the quality of a catwalk: keeping the back straight, keeping the stomach in, keeping the head straight, looking straight ahead, swinging hands to a certain height, swinging hands in a certain degree, straightening the legs, having the feet line up in a certain way, lowering the shoulders, relaxing the neck and facial expression, swinging the hips with the natural walk, etc. Which elements can be captured by video for analysis? How much should each of the components weigh in determining the quality of the catwalk? How should we evaluate each component?

3. SOLUTION

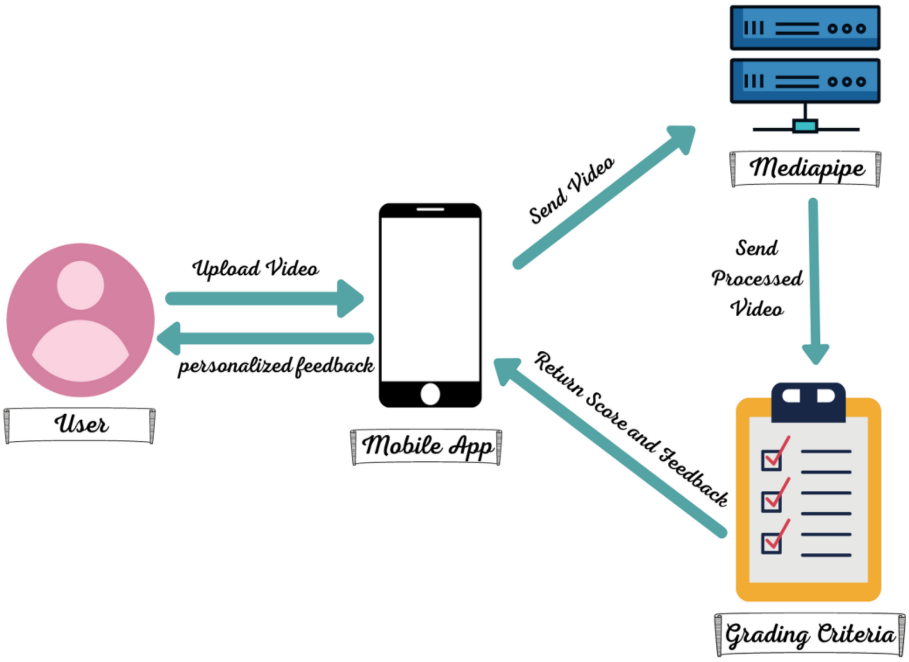

Figure 1. Overview of the solution

An overview of the solution is presented in Figure 1. Users upload videos they record onto a mobile app following the video guidelines. The video will then be sent for processing by Mediapipe, so that all joints and motions of the human body are detected and labeled in the video. Processed videos can then be used to provide the information necessary for grading concerning head, shoulder, hand, foot, arm, and hip movements. The grading system will thus calculate the score of the video and provide recommendations to the user for what aspects need improvement through the mobile app.

Mobile applications are created through the Android Studio and Flutter. After a splash screen, it would turn to three pages: Learn, Upload, and My Home. On the Learn page (Figure 2) are the video requirements, e.g., at which angle and what height the camera needs to be positioned, as well as the scoring criterion. The components that determine the final score of the video are presented to the user so even new users can understand how to use the app as well as the rating system. On the Upload page (Figure 3), users upload the videos from their device, where the app will send it to the Amazon Web Server for analysis. On the My Home page (Figure 4), users are able access the scores of previous videos so they can track their catwalk improvement.

The data sent to the Amazon Web Server will be processed by Mediapipe, so that all joints and motions of the human body can be detected and labeled in the processed video. Then, the information necessary for scoring will be extracted from the processed videos, which is mainly concerning head, shoulder, hand, foot, arm, and hip movements. The grading system will then send the final score and recommendations for what aspects need improvement back to the Amazon Web Server, which will then be presented to the user through the mobile app.

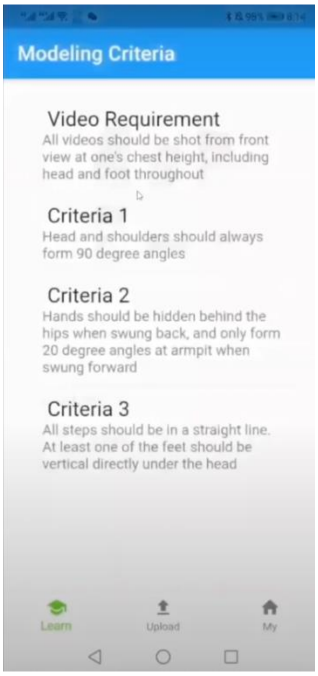

Figure 2. Screenshots of Criteria and Requirement

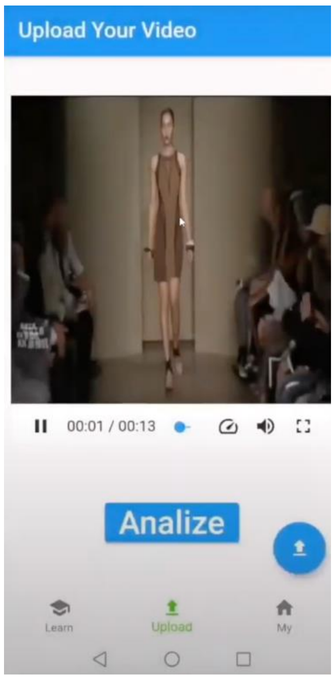

Figure 3. Screenshots of Video uploading

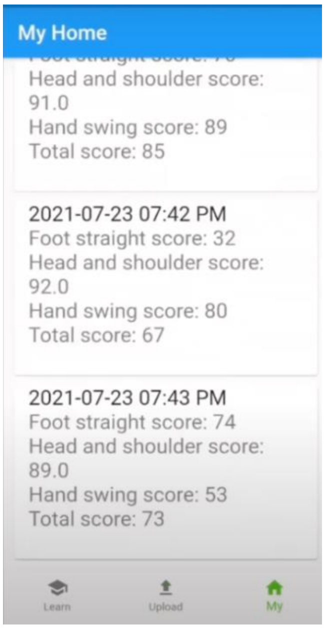

Figure 4. Screenshots of score history

4. EXPERIMENT

To evaluate the accuracy of our approach, we collected 30 great walks, 20 good other types of walks, and 5 bad walks for each grading component. With this data set, we conducted experiments to evaluate the accuracy of two components: the image library used to identify the human body and its movements and the different grading features.

Experiment 1: Comparison of Mediapipe or OpenPose

We conducted an experiment by grading several catwalk videos using both Mediapipe and OpenPose, while using the same grading criteria, and recording the score for each component. Compare the accuracy of each score to the standard score and view the processed video for any joint labeling error. The results are shown in the diagram below.

Show in Table 1, the results of Mediapipe and OpenPose are roughly the same when grading catwalk videos recorded following the guidelines. But OpenPose takes an additional 35 minutes to set up and connect to the library compared to the set-up time Mediapipe takes.

Table 1. Comparison of Mediapipe or OpenPose

	Mediapipe Scoring of shoulder	Openpose Scoring of shoulder	Mediapipe Scoring of arms	Openpose Scoring of arms	Mediapipe Scoring of feet	Openpose Scoring of feet	Mediapipe Overall Score	Openpose Overall Score
Video 1	87	88	67	65	66	69	70.15	74.45
Video 2	56	53	24	28	73	76	52.35	53.55
Video 3	72	67	88	87	68	69	75.40	73.7

Experiment 2: Comparison of different proportions of grading formula

Based on general knowledge of important components in the catwalk, we came up with 2 different proportions of grading formula: 1. shoulder score is 35% of the total score, arm score is 30% of the total score, and feet score is 35% of the total score; 2. shoulder score is 30% of the total score, arm score is 25% of the total score, and feet score is 45% of the total score. We grade 3 sample catwalk videos using these 2 different proportions of grading formula, while always using Mediapipe, and record the score for each component. Compare the accuracy of each score to the standard score. The results are shown in the diagram below.

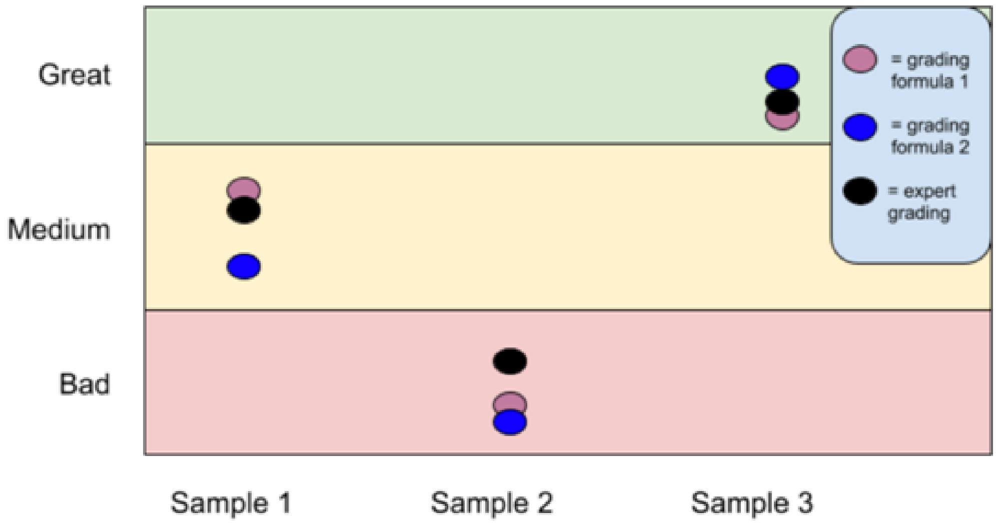

Figure 5. grading formula

Shown in Figure 5, grading formula 1 is more accurate compared to grading formula 3. Therefore, we use grading formula 1 as the grading components and scoring proportions.

5. RELATED WORK

Sen Qiao, et al. published a real-time 2D human gesture grading system from monocular images based on OpenPose. [13] After capturing 2D positions of a person's joints and skeletal wireframe, the system computes motion trajectory equations for every joint. A modifiable scoring formula was used for simulating the gesture grading scenarios. While our work and this paper both focus on grading human gestures, our work is focused on grading based on a standard of model catwalks, which is different from Qiao's modifiable scoring formula. While the modifiable scoring formula can accommodate more diverse conditions, it is not convenient for the general population to utilize. Our work provides a convenient mobile app and clear directions for how they can receive a score for their catwalks, and provides a grading system for users so that non-professional models can also benefit from it.

The work of Masato Nakai, et al. aims to build a posture analysis model using OpenPose skeletal recognition data and verifying the practicality of OpenPose by verifying the accuracy of the model. [14] As a posture analysis model, they adopt a logistic regression model that predicts the shooting probability of the basketball free throw with skeletal posture data as explanatory variables and whether the ball enters the basket or not as a binary target variable. Nakai's purpose is to verify the practicality of OpenPose through predictions based on Basketball Free Throw Shooting, while our work's purpose is to provide an accessible scoring app for people who want to learn catwalking and receive accurate feedback and recommendations to improve their catwalks.

Dae–Seong Go, et al. present a method to create robot motion data based on video-recorded human demonstrations. Their proposed method has two objectives: 1) the robot motion can be made through the existing performance video clips demonstrated by a human actor, and 2) the partial corrections for the robot motions can be made instantaneously when corrections are given by the performance director. To achieve these objectives, we first adopted OpenPose to extract the coordinates of human joints from a video clip of human performances. Then, the extracted coordinates were converted into joint angle values. Next, the joint angle values were renovated to generate robot motions satisfying the kinematics and hardware specifications of the robot. Finally, the efficacy of the proposed method was examined by applying the generated robot motions to a female android robot. A similarity between Go's work and ours is that we both use a key point identifying application to analyze video-recorded human motion. [12] Go's paper focuses more on creating robot motion that mimics human demonstrations, while our work focuses on identifying the correctness of human motion.

6. CONCLUSION AND FUTURE WORK

In this paper, we developed an accessible mobile app to score catwalk videos and provides personalized feedback for each user. Inspired by the Sen Qiao, et al.'s research "Real-time human gesture grading based on OpenPose", where they presented a real-time 2D human gesture grading system from monocular images based on OpenPose, we developed a real-time 2D model walk grading algorism based on Mediapipe. Using Mediapipe to capture the 2D positions of a person's joints and skeletal wireframes, the system computes a motion trajectory equation of for every joint, and a scoring formula was used to score the catwalk videos, providing tailored feedback on how each user could improve.

In two application scenarios, we demonstrated how the above combination of techniques increases the accuracy, practicality, and efficiency of our application. First, we demonstrated why Mediapipe is better than OpenPose through a random experiment comparing the accuracy and the efficiency of each method. Second, we analyzed the accuracy of the two most practicable scoring formulas based on a controlled experiment.

We completed a scoring system based on front view video, but some important features of certain model catwalks could not be accurately identified and graded using only a front view.

For now, there is only scoring for the most common types of catwalks, which might not be best for people who want to learn different types of catwalks. In addition, there is no scoring for model poses and model images, which are also important components of a modeling career.

As for future work, we will collect more data to develop grading criterion for different types of catwalks, model poses, and model images to satisfy the needs of people wishing to become models. After more experiments, we hope to find the best sideways video recording positions. Observing more catwalks of supermodels from a lateral direction is needed to create scoring criteria for sideways videos of catwalks.

REFERENCES

[1] Wissinger, Elizabeth. "1. Supermodels of the World." This Year's Model. New York University Press, 2015. 35-58.

[2] Osokin, Daniil. "Real-time 2d multi-person pose estimation on CPU: Lightweight OpenPose." arXiv preprint arXiv:1811.12004(2018).
[3] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." IEEE transactions on pattern analysis and machine intelligence 43.1 (2019): 172-186.
[4] Noori, Farzan Majeed, et al. "A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network." Scandinavian conference on image analysis. Springer, Cham, 2019.
[5] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).
[6] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).
[7] Lugaresi, Camillo, et al. "MediaPipe: A Framework for Perceiving and Processing Reality." (2019).
[8] Mears, Ashley. "Economics of the Catwalk." Pricing Beauty. University of California Press, 2011. 27-70.
[9] Peng, Dunlu, Lidong Cao, and Wenjie Xu. "Using JSON for data exchanging in web service applications." Journal of Computational Information Systems 7.16 (2011): 5883-5890.
[10] Hesselbein, Chris. "Walking the Catwalk: From Dressed Body to Dressed Embodiment." Fashion Theory 25.3 (2021): 367-393.
[11] BUNDCHEN, FASHION MODEL GISELE. "FASHION MODEL GISELE BUNDCHEN AND HER LEGS-BRAZIL."
[12] Malleson, Charles, John Collomosse, and Adrian Hilton. "Real-time multi-person motion capture from multi-view video and IMUs." International Journal of Computer Vision (2019): 1-18.
[13] Qiao, Sen, Yilin Wang, and Jian Li. "Real-time human gesture grading based on OpenPose." 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2017.
[14] Nakai, Masato, et al. "Prediction of basketball free throw shooting by openpose." JSAI International Symposium on Artificial Intelligence. Springer, Cham, 2018.
[15] Go, Dae–Seong, et al. "Andorid Robot Motion Generation Based on Video-Recorded Human Demonstrations." 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, 2018.