CSC490A1- Group 8

Isaac Abell - 1008831329, Thomas(Yehyun) Lee - 1008992217,
Zeling (Zoey) Zhang - 1007714779, Jessica Zhang- 1008494130

Part One: Interest Statements

Despite the massive J-pop and K-pop dancing community, many introductory dancers struggle to begin due to the difficulty to learn, and professional dancers lack tools to discover details they are missing. Thus, we hope to build a “dance coach that catches what the eye misses.” A **beat-synchronous video analysis system that compares a user’s dance clip to a reference clip of the same song using 2D/3D pose and music audio alignment**. Instead of “move similarity,” the app extracts tacit micro-technique signals: micro-timing, groove phase, attack/decay control, isolation, and grounding, and turns them into timestamped coaching and drills.

Isaac: I am excited to be a part of this project as it poses an interesting and challenging problem of creating a vision model that can compare two users' movement patterns to find the flaws and differences with extreme accuracy. Since dance movements require high precision, we need to ensure our model has a very high level of accuracy in order for it to be useful to potential users.

Thomas: As a typical CS undergrad, I’m a person very far from a dancer. As someone who seeks to work on new projects outside my comfort zone, I’m curious to learn about how dancers practice, small details dancers care about, and generally how the industry works. From a business perspective, with the recent progress in generative AI, video analysis techniques, and from my research, that the industry is being hyper-focused on productivity application, I believe there are small fruits from the dance domain that have not been explored yet that our team can capture and build useful applications for dancers.

Zoey: I am excited to contribute to the project with domain knowledge of dancers’ pain points in learning, experience conducting various user studies, and technical expertise in exploring new use cases for vision models. I want to build a product that helps dancers overcome the idea that talent is the key to success in dancing by capturing areas for improvement that are not easily spotted by the human eyes.

Jessica: I am excited to contribute to this project as I am a dancer myself, and I hope to have a product that is integrative of all the functionalities that the community found to be helpful, and explore functionalities with machine learning to help dancers succeed.

Part Two: Landscape Analysis

Technological supports		
Relevant Item	Description	Commentary
MimicMotion	Open source motion transfer framework. It maps a person's likeness from a single photo onto a target motion video to create a personalized reference video.	This could be a useful tool for users practicing, allowing them to track their progress, which may improve learning.
Ultralytics YOLO26	Provides high-frequency, 17-point skeletal tracking and persistent multi-person identification.	This is useful for processing movements locally on the user's device, which eliminates cloud latency and allows for the real-time analysis of technique without expensive server costs.
DEVA	State-of-the-art video object tracking and segmentation framework developed to separate spatial (image-level) segmentation from temporal (frame-to-frame) tracking.	This is a theoretical starting point for us, as it is a current object tracking in video data.
librosa 0.11.0 documentation	A Python audio analysis library for extracting features (e.g., onsets, chroma, MFCCs) from sound.	For this project, librosa can be used to process and compare audio from the reference and user videos so temporal alignment (e.g., beats or onset patterns) can be computed.
Market investigation		
STEEZY Studio	Filmed and created by instructors (this app is fully human); Key features: digital studio(you can see all angles of a move), **loop specific sections**, adjust the **playback speed** without changing the pitch of the music, and **virtual mirror** to dance alongside the instructor.	This is one of the most widely used platforms for online dance learning. Revenue of $19 million in June 2024 is a strong indicator of market and need.
Dalea - AI	AI-powered dance practice app where students record their dance and is compared pre-uploaded teacher videos. Features: Teacher Video Uploads, Student Practice Recording, AI Pose Comparison, Accuracy Scoring, Timing Analysis, Step-by-Step Progress.	Relevant to our product as it has similar AI-driven features. However no customer reviews. We tested the product, and the video analysis ouputed timeout error. It also seems mainly targeting Bharatanatyam, leaving room for our product.
SyncTrainer by Techie Ray	An app that rates dancers' choreography frame by frame on the following dimensions: group sync, difficulty, formation changes, and reliability	We can refer to their user workflow and user feedback on this app.
Datasets		
AIST++ Dataset & AIST Database	10,108,015 frames of 3D keypoints with corresponding images; 1,408 human dance motion sequences on 10 dance genres with music; 9 views of camera intrinsic and extrinsic parameters.	We plan to use this as one of the main datasets for its high quality and variety of angels. The AIST database is the parent dataset of AIST++. We can use this as extra data if needed
ImperialDance-Dataset	69,300 seconds of dance motions and music, 5 genres, 20 pieces of music and 20 choreographies with dancers of 3 different expertise levels	We plan to use this as one of the main datasets for tuning the feedback with different expertise levels.

Part Three: Project Outline

Problem Statement

Traditional dance practice relies heavily on mirrors and manual video review. Mirrors provide immediate feedback but do not allow frame-by-frame analysis; Manual video comparison is time-consuming and struggles to identify subtle yet critical song-specific techniques (in posture, timing, and control) that reference dancers develop over years of experience and musicality.

Proposed Solution

An intelligent application that automates the feedback loop. Dancers upload a clip of themselves dancing to a song, along with their learning reference video. Using computer vision, the app extracts skeletal data to identify specific technical flaws. It provides the user with visual overlays and performance metrics, allowing for a data-driven approach to improving form and execution.

High-level Technical Approach (For more details, please refer to Appendix B.)

1. **Pose Estimation Integration:** implement a computer vision model (such as YOLO or MediaPipe) to extract skeletal coordinates and joint angles from video frames.
2. **Temporal Synchronization:** develop a synchronization layer to align the timing of the user performance with the reference video, even if the user starts early or late.
3. **Feedback Logic:** build a comparison engine that calculates "deltas" (differences) in body positions to highlight errors in real time.
4. **Deployment and Optimization:** package the solution into a user-friendly mobile interface with a hybrid processing backend.

Milestones for MVP

1. Preliminary user study to guide design decisions.
2. Outline estimation for overlay
3. Spatial Normalization
4. manual Temporal Synchronization(user aligns the timing)
5. Deployment as a web app
6. Other basic features like looping, found in market/product analysis.

List of Unknowns to Investigate

- **Precision and Fidelity:** Is YOLO accurate enough to catch the high-speed and precise micro-movements found in technical dance styles?
- **Synchronization Logic:** What is the most robust method to sync a user's movement with a reference image when the tempo or start time varies?
- **Perspective Handling:** How do we programmatically handle significant differences in body size or camera angles between the user's environment and the reference video?
- **Hardware Constraints:** Which parts of the analysis are performant enough to run locally on a user's mobile device, and which processes require more compute power to run in a reasonable time via AWS?

Part Four: Project Press Release

The AI-Powered Dance Coach That Sees What the Eye Misses

Intermediate dancers can now master complex choreography with beat-synchronous video analysis that provides professional-level feedback at home.

Summary: A team at the University of Toronto has announced TempoFlow, a mobile application that uses computer vision to help dancers improve their technique. By comparing a user's practice video to a professional reference clip, the app identifies timing and form mismatches and provides immediate feedback indicating where and how they deviated from the reference.

Problem: Many dancers find themselves in a frustrating situation: the choreography looks "right" in the mirror, yet something still feels off compared to the professional reference. That missing "something" is often style-level execution: song-specific techniques in posture, timing, and control that master dancers refine through years of experience and musicality. Traditional practice still relies heavily on mirrors and manual video review. Mirrors provide immediate feedback, but they don't support frame-by-frame analysis, and the human eye can't reliably detect tiny timing offsets or subtle sequencing differences. Manual side-by-side video comparison is time-consuming and still subjective, making it hard to isolate the small mismatches that drive the largest perceived performance gap.

Solution: TempoFlow automates the feedback loop. When a user uploads a clip of themselves dancing to a song, the app uses pose estimation to sync their movements with a professional reference, accounting for differences in body size and camera angles. The app then provides the dancers with a visual overlay and timestamped coaching on specific elements, explaining exactly how to adjust their moves to match the pro.

Leader Quote: "We wanted to build a tool for the hours dancers spend practicing alone between classes," said the project lead. "A mirror can't tell you you're consistently late on the hit by a few frames, or that your weight is floating during the chorus. Our goal was to take the expertise of a professional coach and turn it into a vision model that can catch the timing and control details that experts internalize over the years and make them visible."

Customer Quote: "I used to record myself for hours and still feel like my covers looked 'sloppy' compared to the originals," said Joe, a student who tested the app. "The first time I used this, it pointed out that my weight was too high during the chorus. Once I fixed my grounding based on the app's drill, the move finally clicked. It's like having a private instructor in my pocket."

Call to Action: TempoFlow is launching in beta for dancers who want structured feedback from their solo practice sessions. Visit TempoFlow.com to upload your first practice clip and receive your first automated technique breakdown.

Appendix

A. Press Release Iteration

Iteration 2: **The AI-Powered Dance Coach That Sees What the Eye Misses**

Intermediate dancers can now master complex choreography with beat-synchronous video analysis that provides professional-level feedback at home.

Summary: The team at the University of Toronto has announced a new mobile application that uses computer vision to help dancers improve their technique. By comparing a user's practice video to a professional reference clip, the app identifies technical errors in timing and form and provides immediate feedback indicating where and how they deviated from the reference.

Problem: Learning to dance is difficult without a professional coach. Most students rely on mirrors or recording themselves on their phones, but these methods don't provide objective feedback. It is hard for a dancer to see exactly why their movement looks different from an instructor's. Many dancers struggle to identify subtle errors in their timing or posture, leading to fewer improvements over time and the development of bad habits.

Solution: This new application automates the feedback loop. When a user uploads a clip of themselves dancing to a song, the app uses pose estimation to sync their movement with a professional reference while accounting for differences in body size and camera angles. The app then provides a visual overlay and timestamped coaching on specific elements explaining to the dancer exactly how to adjust their body to match the pro.

Leader Quote: "We wanted to build a tool for the hours dancers spend practicing alone between classes," said the project lead. "Most beginners get frustrated because they can't tell what they are doing wrong. Our goal was to take the expertise of a professional coach and turn it into a vision model that can catch the tiny details, the 'micro-techniques,' that even a mirror can't show you."

Customer Quote: "I used to record myself for hours and still feel like my covers looked 'sloppy' compared to the originals," said Joe, a student who tested the app. "The first time I used this, it pointed out that my weight was too high during the chorus. Once I fixed my grounding based on the app's drill, the move finally clicked. It's like having a private instructor in my pocket."

Call to Action: Dancers ready to level up their practice can download the beta version of the app today. Visit TempoFlow.com to upload your first practice clip and receive your first automated technique breakdown.

Iteration 1: **The AI-Powered Dance Coach That Sees What the Eye Misses**

Intermediate dancers can now master complex choreography with beat-synchronous video analysis that provides professional-level feedback at home.

Summary: The team at the University of Toronto has announced a new mobile application that uses computer vision to help dancers improve their technique. By comparing a user's practice video to a professional reference clip, the app identifies technical errors in timing and form. Unlike standard video players, this tool analyzes "micro-techniques" like grounding and isolation, giving users specific drills to fix their mistakes instantly.

Problem: Learning to dance is difficult without a professional coach. Most students rely on mirrors or recording themselves on their phones, but these methods don't provide objective feedback. It is hard for a dancer to see exactly why their movement looks different from an instructor's. Many dancers struggle to identify subtle errors in their timing or posture, leading to "plateauing" where they stop improving despite hours of practice.

Solution: This new application automates the feedback loop. When a user uploads a clip of themselves dancing to a song, the app uses 2D/3D pose estimation to sync their movement with a professional reference. It automatically accounts for differences in body size and camera angles. Instead of just saying "you're off," the app provides a visual overlay and timestamped coaching on specific elements like "micro-timing" and "attack control," showing the dancer exactly how to adjust their body to match the pro.

Leader Quote: "We wanted to build a tool for the hours dancers spend practicing alone between classes," said the project lead. "Most beginners get frustrated because they can't tell what they are doing wrong. Our goal was to take the expertise of a professional coach and turn it into a vision model that can catch the tiny details, the 'micro-techniques,' that even a mirror can't show you."

Customer Quote: "I used to record myself for hours and still feel like my covers looked 'sloppy' compared to the originals," said Joe, a student who tested the app. "The first time I used this, it pointed out that my weight was too high during the chorus. Once I fixed my grounding based on the app's drill, the move finally clicked. It's like having a private instructor in my pocket."

Call to Action Dancers ready to level up their practice can download the beta version of the app today. Visit TempoFlow.com to upload your first practice clip and receive your first automated technique breakdown.

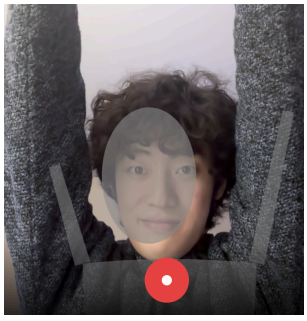
B. In-depth Technical Research

Detection and Overlay

For our app, the core part of the system requires detecting the dancer's body outline. There's few different approaches to make this happen. One way we plan on taking is using the latest YOLO or MediaPipe BlazePose. BlazePose is more mobile-friendly but focuses on tracking only one person. The latest YOLO model have benefits of detecting multiple people. Depending on the benefits of BlazePose, we may use both where we use YOLO to track multiple dancers, crop by frame, and use BlazePose to process once again, or we could just choose to use a single library. We'll conduct experiment and choose the right approach.

An important aspect is that we want to overlay the dancers' outline, not a skeleton of a person's body. Like the following:

and we want to avoid the skeleton overlay like the following:



Adding on, as a bit of detail to improve the UX, we would like to smooth out the overlay detection frequency and latency. Some of the latest detection models visualize new detection immediately and disregard the previous detection. While this is fast, it has two problems: one, it is visually unappealing. New detection every frame, looks a bit laggy. Two, when the model detects wrong and overlays in the wrong frame, and a few frames later, detects the correct frame, there's a phantom phenomenon that happens where the overlay jumps back and forth with different frames. As our team's major focus is on good UI/UX, we want to smooth this out, so the outline looks very smooth.

Reliability

Secondly, we need to make this happen extremely fast, on mobile, without lag. We aim to achieve this by conducting major computations on the device without relying too much on a server. As a case study from an earlier section, Spark had problems reported by users that the app often lags and pauses. We want to avoid this happening.

Synchronization Logic

Thirdly, there's a question of synchronization logic. Are we going to sync the user's dance move against the reference dance video purely based on when the music starts, or based on the minimum relative dance overlay difference? What is the most robust method to sync a user's movement with a reference image when the tempo or start time varies? While syncing the video simply based on when the music starts is easy and seems the best case, there might be occasional cases where, when we provide feedback to the user, we might want to crop some part of the video and sync based on the video and not the sound, and give feedback to the user based on that pure specific range of moment. This might be useful for beginners, since most times, their dance moves will be slower than the dance moves from the reference video, and dance move timing will likely vary with a lot of variance over time.

Effective Way to Give Feedback to the User

Fourthly, we need to come up with clever ways to give feedback to the user. Instead of relying on one method, we are thinking of using a variety of different styles of feedback to the user.

One, we may be using a Gemini Live or other native AI systems with video processing capability to let it analyze and give feedback either in real time, or uploaded video after the dance is done.

Two, we will use our dancers' outline detection system and compare users' against the reference dancer's overlay. Here, there are different ways to give visual feedback, which will come down to a matter of UI/UX. But importantly, we want to rely on visual feedback, like using the circle spotlight effect, lightbox/mask effect and avoid simply displaying numeric values. Another effect we could include is the use of pointers/arrows and animation in specific gestures or moves.

Third, we could use a generative AI video to suggest better moves in certain moves. This could be a generation of whole dance videos of users performing correct dance movement, or just a specific part of dance, where we generate a video to provide an example move.
Importantly, we would like to combine some of these animations/features in a visually attractive way and a good UX. We may as well use an agent that determines at which point we should give feedback to the user, using which feature/style of feedback.