IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133122

MOTION TUTOR: ANIMATED MOTION USING DEEP LEARNING

Mr. Annappa Swamy D.R[1], Akshira[2], Arghyashree[3], Ashvitha Shetty[4], Gajesh Naik[5]

Associate Professor, Dept. of Computer Science and Engineering, Mangalore Institute of Technology & Engineering, Moodabidre, India[1]

Student, Dept. of Computer Science and Engineering, Mangalore Institute of Technology & Engineering, Moodabidre, India[2,3,4,5]

Abstract: In a world that values personalized, interactive, and easily accessible learning, our project stands at the intersection of art and technology, offering an innovative solution. We aim to revolutionize the understanding and teaching of complex movements by providing a tailored and immersive learning experience, departing from traditional tutorials. In the domain of Motion Knowledge, where movement's beauty meets learners' enthusiasm, our project represents a groundbreaking approach.

Driven by the belief that tutorials should be inclusive, we leverage cutting-edge technology like PoseNet and CGAN to deconstruct tutorials into digestible steps, simplifying the learning journey. Our primary objective is to empower individuals of all skill levels to explore, learn, and excel in the art of movement without unnecessary complications. Our project provides dynamic and highly personalized learning experiences accessible to individuals from diverse backgrounds, whether they're novices or seasoned practitioners.

Users actively shape their motion education narrative, fostering creativity, skill mastery, and a profound connection with their movements. Our unique solution marries art and technology to meet the demand for engaging and personalized learning experiences.

Keywords: PoseNet, CGAN, Motion Knowledge, personalized learning experiences, immersive learning experience.

I. INTRODUCTION

In today's fast-paced world, the importance of maintaining a healthy and active lifestyle cannot be overstated. Unfortunately, much of the available instructional content tends to adopt a one-size-fits-all approach. Our innovative solution seeks to blend art and technology seamlessly, fundamentally changing how we teach and learn about motion. Instead of simply copying movements onto digital avatars, we've embraced a novel approach that harnesses technology and computer vision to reshape the learning process, ensuring accessibility and engagement for all. Dance, exercise, and fitness have evolved into immersive experiences driven by the demand for personalized, innovative, and effective approaches.

Deep learning, a dynamic aspect of artificial intelligence, is emerging as a transformative tool in redefining how we approach physical well-being. It offers an unprecedented opportunity to infuse precision, personalization, and adaptability into dance, exercise, and fitness routines.

At the core of our project is an interactive learning journey where learners are empowered to take control. We utilize advanced technologies like Convolutional Neural Networks (CNN) for precise pose recognition, deep learning for tailored animation, and Conditional Generative Adversarial Networks (CGAN) for seamlessly integrating character and motion.
Rather than simply mirroring poses onto digital characters, our emphasis is on instructional videos. By transforming input videos into a series of moving images, we simplify complex motion sequences into manageable steps.

Our commitment is to ensure that our tutorials are accessible to everyone, catering to both experienced enthusiasts seeking mastery and newcomers venturing into the world of movement. This approach allows individuals to craft personalized animations that reflect their unique movements, making the learning and refinement of motion an inclusive, enjoyable, and personalized experience for all.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133122

II. LITERATURE SURVEY

In [1] It introduces a novel Virtual Reality Dance Training System that leverages motion capture and virtual reality (VR) technologies. Inspired by traditional learning methods involving imitation and feedback from a teacher, the system aims to provide an immersive learning experience. A prototype of the proposed system allows students to mimic motions demonstrated by a virtual teacher projected onto a wall screen. Simultaneously, the system captures and analyzes the student's movements, providing real-time feedback. User studies indicate that the system effectively enhances students' skills, with participants finding it engaging and motivating for learning.

In [2] It introduces an innovative multimodal convolutional autoencoder that integrates 2D skeletal and audio data through an attention-based feature fusion mechanism, facilitating the generation of varied dance motion sequences. Initial testing on skeletal data revealed optimal performance with 500 input poses in unimodal settings. Subsequently, the multimodal architecture underwent training using teacher-forcing and self-supervised curriculum learning to address autoregressive error accumulation. Evaluation results showcased enhancements in realism, motion diversity, and multimodality, leading to a notable 0.39 reduction in the Fréchet Inception Distance (FID) metric. Subjective assessments, involving 45 evaluators and 1800 responses, confirmed a 6% increase in preference for style consistency.

In [3] Utilizing deep learning-based feature extraction has enabled the synchronization of audio and body movements, opening up promising avenues in research with vast applications such as generating sign language, computer animations, and dance sequences. Previously, computer-generated choreography was restricted to simplistic stick figure representations. This paper introduces an image translation technique into dance generation, allowing for the creation of realistic dance movements. Through this technique, it becomes feasible to generate a dance video featuring an amateur individual dancing at a professional level. The transformation from stick figure to lifelike image is achieved using Generative Adversarial Network (GAN) technology.

In [4] The objective is to produce Labanotation scores using motion-captured data derived from real-world dance performances. To address challenges such as diverse dance movement patterns, varying dancer shapes, and data noise, a novel feature is introduced that remains consistent despite anthropometric variations and body orientations. Subsequently, notations are generated for both lower-limb movements and upper-limb gestures.

For lower-limb movements, a hidden Markov model (HMM) is employed to analyze temporal dynamics and map each movement to its corresponding dance notation. For upper-limb gestures, a multi-class classifier is trained based on extremely randomized trees to identify the appropriate notations.

Ultimately, Labanotation symbols are generated based on the analyzed movements, resulting in Labanotation scores. This method can generate spatial symbols describing directions and levels in both the support column and arm column based on motion-captured data. The produced scores are precise and dependable. Experimental results indicate an average recognition accuracy exceeding 92% for the generated notations, a significant improvement over prior research effort.

III. SCOPE AND METHODOLOGY

3.1 Aim of the project

The scope of this project is to develop a user-friendly platform to provide service across the world enabling users to upload MP4 format videos containing human movements and generate corresponding animated character movements in real-time. Leveraging Convolutional Neural Networks (CNNs) for pose estimation, particularly utilizing the PoseNet algorithm, and Conditional Generative Adversarial Networks (CGANs), specifically the Pix2Pix architecture, the project encompasses several key components.

This includes the development of a web-based user interface that facilitates seamless video upload, ensuring intuitiveness and responsiveness. Backend systems will be created for real-time processing of uploaded videos, with efficient algorithms to extract human poses within one minute. PoseNet will be utilized to accurately estimate poses from the videos, extracting keypoint coordinates and confidence scores for each frame.

The Pix2Pix architecture within the CGAN framework will then convert these estimated poses into realistic animated character movements, synchronized with the original human movements. Robust error handling mechanisms will be implemented to address issues during video processing, and the accuracy and quality of the generated animations will be validated against the source videos.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.133122

3.2 Objectives

The main objectives of the "MotionTutor" project center around the detection, extraction, transformation, and deconstruction of input sources. These objectives include:

- **Identity Protection**: Our main goal is to provide identity protection for the professional tutors while uploading their tutorials on any social media platform by enabling them to convert their tutorials into animated video.

- **Video Transformation**: Develop a system capable of converting source videos into a sequence of animated images, thereby breaking down complex routines into individual steps.

- **Video Deconstruction**: Utilize computer vision techniques like poseNet and CGAN to precisely capture and segment each step by removing unnecessary background and any other objects, enhancing clarity and simplification.

- **Technology Integration**: Integrate advanced technologies such as PoseNet and CGAN-pix2pix to ensure accurate pose detection, image generation, and transformation, enhancing the effectiveness and functionality of the platform.

3.3 Existing system

The existing literature highlights several approaches to dance training and motion generation using advanced technologies. The Virtual Reality Dance Training System proposed in [1] utilizes motion capture and VR technologies to enable students to imitate motions demonstrated by a virtual teacher, with feedback provided based on motion analysis. DanceConv, as described in [2], introduces a multimodal convolutional autoencoder that combines skeletal and audio data to generate diverse dance motion sequences, achieving improvements in realism and motion diversity.

Additionally, the paper on Beat Based Realistic Dance Video Generation using Deep Learning in [3] emphasizes the use of deep learning for synchronizing audio and body movements, enhancing the generation of realistic dance moves through image translation techniques using Generative Adversarial Networks (GANs). These approaches collectively demonstrate advancements in utilizing technology to enhance dance training and motion generation, with a focus on realism, diversity, and personalized learning experiences.

3.4 Proposed system

The PoseNet model in the suggested system is first used to track and detect keypoints in the video frames in order to create stick figure representations. The movements of the characters are based on these stick figures. Then, while preserving the original motion from the video, the pix2pix model enhances these stick figures into more intricate and animated characters. Convolutional neural networks (CNNs) are used by PoseNet and pix2pix for their respective tasks.

As demonstrated by PoseNet's analysis of body keypoints, CNNs are effective at processing spatial information, which makes them appropriate for applications like image processing and feature extraction. Furthermore, the pix2pix model enhances the stick figures into animated creatures using conditional generative adversarial networks (cGANs). The transformation process is facilitated by the model's ability to learn the mapping between input and output images thanks to cGANs. In order to create a complete method for video animation from real-life footage, the system combines CNNs for keypoint identification and refining with cGANs for animated character development.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133122

3.5 System Architecture

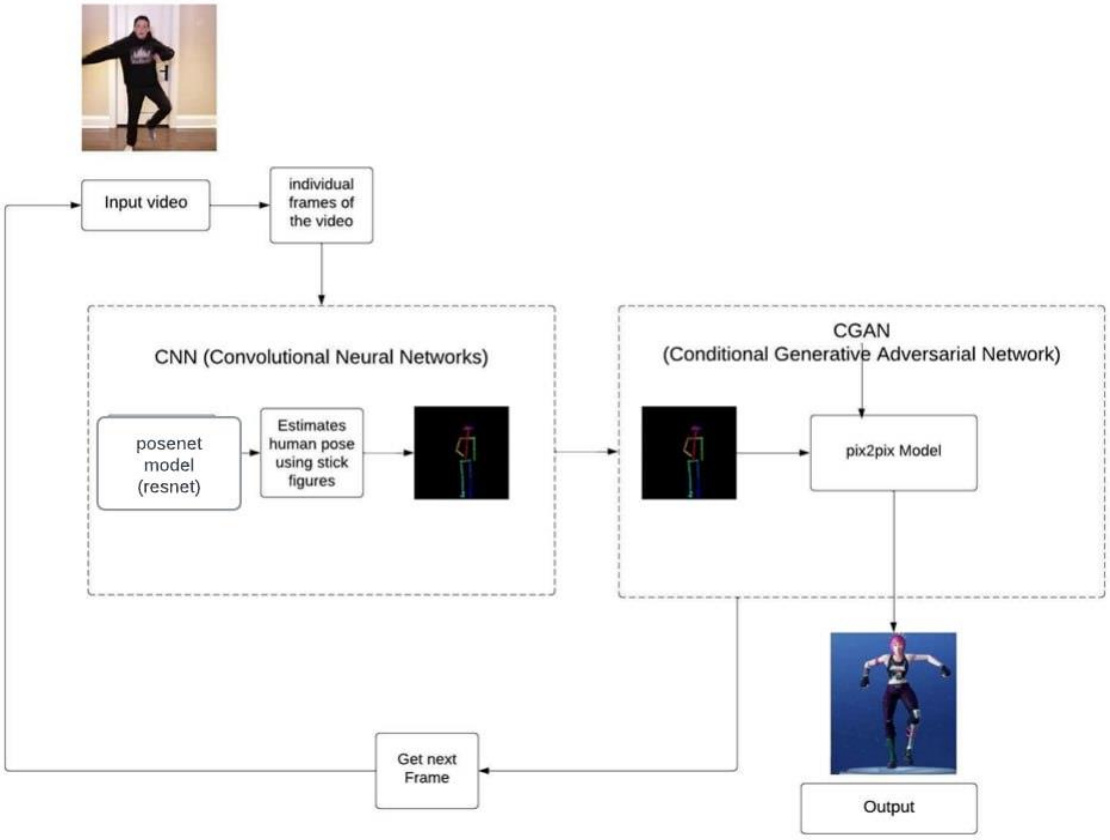

Fig 1 : System Architecture for Motion Tutor

Using a number of machine learning techniques, this technology makes it easier to convert uploaded video data into animated charcater. Through the platform's interface, users register and upload videos. A PoseNet model is then used to track and identify the body's keypoints in each frame, producing a stick figure depiction of the postures and movements.

These stick figures are then refined into more active and detailed characters by a pix2pix model, which learns to connect the original representations to their improved equivalents. These animated characters are then given the motion from the original video, creating a new video in which they mimic the motions of the person in the original clip. Lastly, the animated video is presented for watching and interaction on the user interface, providing streamlined method for users to create engaging animated content from their existing videos.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133122

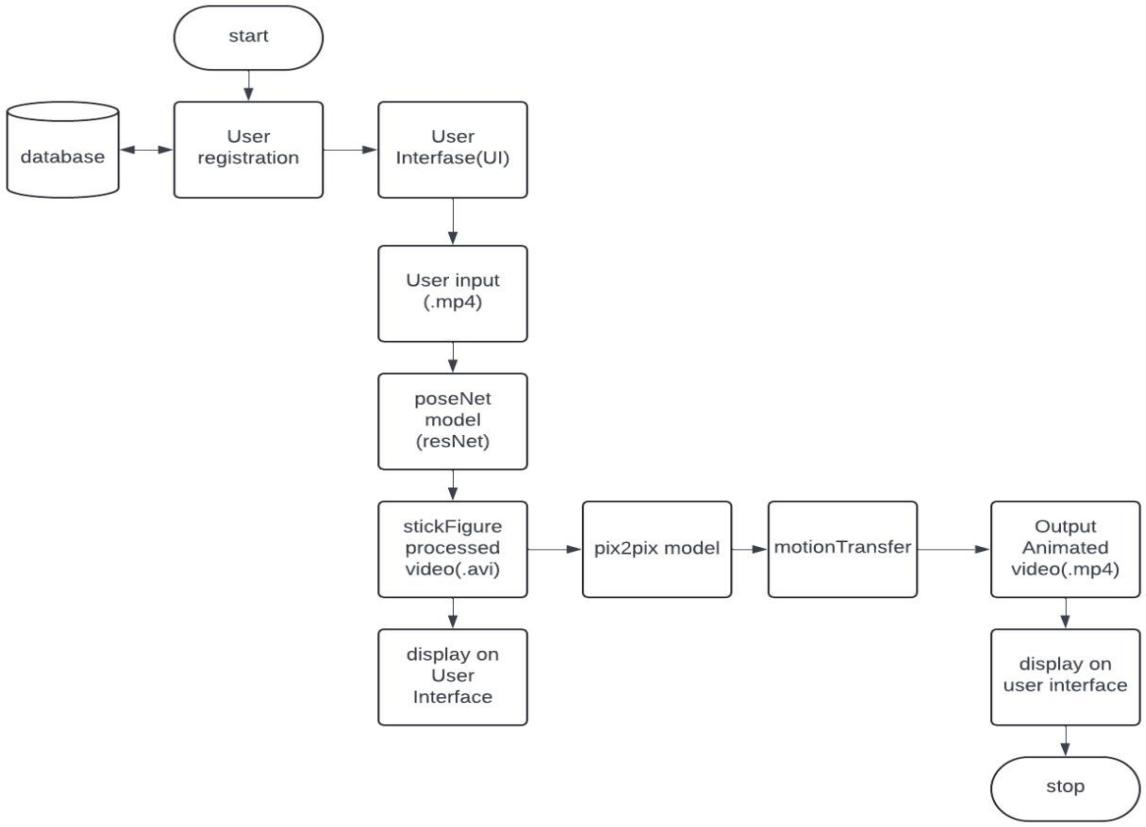

Fig 2: Flowchart

Users whoever wants to convert their video into animated video needs to register with us for using this service. A database is used to keep track of the user registration. Once the user register and sign in they will be redirected to home page where they ca n upload desired video which they want to convert into animated video. Input video will be fetched from the backend for further processing and output animated video will be displayed to the user from where they can download.

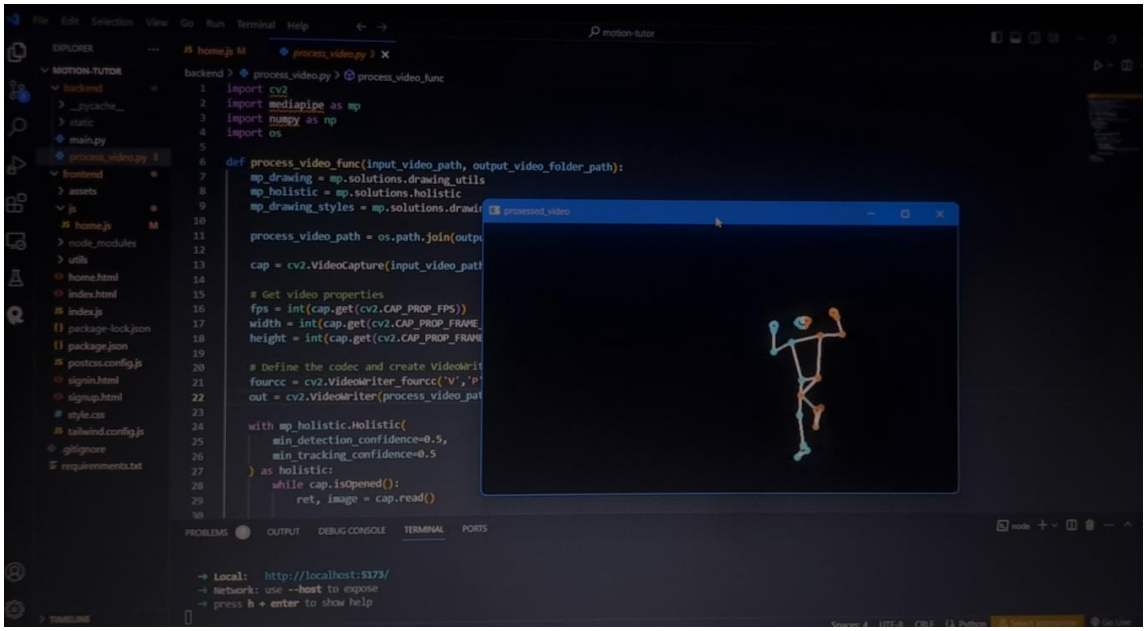

Fig 3: Snapshot of detecting the keypoints from the input video frame

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.133122

IV. CONCLUSION AND FUTURE WORK

In conclusion, the objectives of the "Motion Tutor" project were successfully met through the strategic integration of advanced technologies. The development of a system capable of video transformation facilitated the breakdown of complex routines into manageable steps, enhancing accessibility and understanding for users. Utilizing computer vision techniques like poseNet and CGAN enabled precise video deconstruction, resulting in enhanced clarity and simplification of dance movements. The creation of an interactive learning platform provided users with engaging step-by-step instructions and animated visuals, fostering immersive learning experiences. Through the integration of technologies such as PoseNET and CGAN-pix2pix, accurate pose detection, image generation, and transformation were achieved. Overall, the successful implementation of these technologies has contributed to the realization of the project's objectives, resulting in a comprehensive and innovative solution for learning dance, yoga etc.

Further enhancement can be done for detecting the pose for multi person video by using advanced pose detection model providing greater effectiveness and functionality of the platform.

REFERENCES

[1] A Virtual Reality Dance Training System Using Motion Capture Technology- Published by IEEE, 2010 and Issue: 2 - April-June 2011.https://ieeexplore.ieee.org/document/5557840
[2] DanceConv: Dance Motion Generation with Convolutional Networks- Published by IEEE, 2022. https://ieeexplore.ieee.org/document/9762306
[3] Beat Based Realistic Dance Video Generation using Deep Learning– Published by IEEE, 2019.https://ieeexplore.ieee.org/document/9087510
[4] Dance Movement Learning for Labanotation Generation Based on Motion-Captured Data https://www.researchgate.net/publication/337035713_Dance_Movement_Learning_for_Labanotation_Generation_Based_on_Motion-Captured_Data
[5] Pix2Pix Generative adversarial Networks (GAN) for breast cancer detection https://ieeexplore.ieee.org/document/10029087
[6] Yoga Pose Detection Using Posenet and k-NN https://ieeexplore.ieee.org/document/9776451
[7] Development of Human Pose Recognition System by Using Raspberry Pi and PoseNet Model https://ieeexplore.ieee.org/document/9590593