

Research on 2D Animation Simulation Based on Artificial Intelligence and Biomechanical Modeling

Fangming Dai¹ and Zhiyong Li^{1,*},

¹School of Digital Creation & Animation, ShenZhen Polytechnic University, ShenZhen 518055, China

Abstract

Animation techniques have been completely transformed by the union of Artificial Intelligence (AI) and biomechanical modelling, particularly in 2D animation. This study looks at a combination of AI and biomechanics to address the challenges of simulating 2D animation. Current approaches in 2D animation often struggle to achieve lifelike and fluid movements, especially when representing complex motion or interaction. These traditional techniques rely on manual keyframing or physics simulation, which may be time-consuming and do not provide the rich detail needed for realism in animations. To meet these aspects, this study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM). Our approach thus harnesses Deep Neural Network (DNN) for moving forecasts and improvement using biopsychological principles to help us imitate natural human actions better. In addition to character animation, it could apply to interactive storytelling and educational simulations. As a result, animators get more control over motion generation while drastically reducing the necessity for manual intervention through this fusion of AI and biomechanics, which smoothens the production pipeline for animations. This paper considers several important metrics to evaluate the proposed approach's effectiveness, including user satisfaction, computational efficiency, motion smoothness and realism. Comparative studies with classical animation methods showed that the method generates realistic movements on 2D characters while saving time during production. The numerical findings exemplify that the recommended 2D-AI-BM model improves an accuracy rate of 97.4%, computational efficiency ratio of 96.3%, motion control ratio of 95.4%, pose detection ratio of 94.8% and scalability ratio of 93.2% compared to other popular techniques.

Keywords: biomechanical modelling, artificial intelligence, character animation, deep neural network.

Received on 17 January 2024, accepted on 8 May 2024, published on 23 May 2024

Copyright © 2024 Dai *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5907

1. Introduction

The rapid development of animation has always been on the very basis of Artificial intelligence and computer technology in the era of digital animation, the emergence of 2D and 3D animation, and virtual animation [1]. Character animation is a comprehensive and heterogeneous form with applications in education, entertainment, medical and military contexts, not forgetting the newest and most innovative fields of immersive technologies, like augmented and virtual reality [2]. The diversity and complexity of the subject often make it difficult to identify

differences, advancements and challenges, such as autonomy, creative freedom, control, computational cost, and so on [3]. However, one thing to note is that due to the interdisciplinary importance of character animation (in robotics, medical analysis and video games), much synergistic research has led to interesting and imaginative new animation techniques [4]. Computer animation rapidly relies on character models based on human anatomy. In principle, progressive fidelity in biomechanical modelling should result in more realistic human animation [5]. Given realistic biomechanical models, however, it must confront various difficult motor control problems due to the complexity of human anatomy [6]. The challenge in

*Corresponding author. Email: daiwilliam135028@sina.com , jerome222330@sina.com

controlling biomechanical human models stems from anatomical intricacy, which complicates the character's kinematics, dynamics, and actuation [7]. Existing techniques for animating 2D computer characters can be time-consuming and detailed in the application, requiring complex and often unintuitive user interfaces [8]. For this reason, the character of the final animation is often compromised throughout the animation process [9]. Computer vision techniques were employed to develop a prototype desktop product and associated animation process, allowing an animator to control character animation through hand gestures [10]. Rather than relying on motion capture technology, a performance capture system will use hand gestures as its foundation to enable virtual manipulation [11]. This should provide a softer, more intuitive user interface for the animator that should improve the productivity of the animation workflow and the quality of the resulting animations [12].

Current narrative intelligence research emerges from several areas: interactive drama, cinema, virtual theatre, immersive storytelling, and emergent storytelling [13]. Authoring, narrative, and character-based models are the three main categories this type of investigation belongs to. They aim to address the problem of generating interactive narratives, and different narrative design approaches for user experience [14]. Character expression is the lifeblood of credibility in a narrative system that centres on characters. To improve the character-control system, this must prioritize three crucial tasks: 1) creating a model of general psychological instructions, 2) recognizing and classifying expressive body language utilized during tale performances, and 3) developing an input module for stories that can interpret narrative environments [15]. First, personality, emotion, self-motivation, social relationships, and behavioural capabilities are the fundamentals for providing high-level directives for autonomous character architecture [16]. To enhance the completeness of element animation and provide feedback to virtual customers, a user may use deep learning animation smoothing technology after choosing certain design materials on the platform. A potential approach might include using deep learning algorithms that have been taught to derive 3D joint estimates from 2D video with a smartphone. Three metrics were used to assess the deep learning network's accuracy: rotational velocity, marker location, and kinetic energy, all applied to various human body parts. To achieve smooth and realistic animation, it's essential to adhere to basic animation principles based on AI and biomechanical modelling, including proper spacing, timing, and easing in and out of movements. These principles ensure the animation appears natural and visually appealing to the audience [17].

The main contribution of the paper is

- Designing the 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation
- Evaluating the statistical model of deep neural network for efficiently controlling a dynamic musculoskeletal model.
- The experimental outcomes demonstrate high accuracy, motion control, pose detection and computational efficiency compared to existing techniques.

The remainder of the article is prearranged: Section 2 reviews significant prior work. Section 3 analyses the proposed 2D-AI-BM for the human musculoskeletal model. Section 4 overviews the experiment outcomes with virtual human models. Section 5 concludes the article with future works.

2. Related Study

In recent decades, academia and industry have paid great attention to human modelling and animation as essential subjects in contemporary computer graphics. Development of several methodologies is included in this field of study, including but not limited to task level, behaviour animation, skin deformation face animation, motion control of articulated figures, etc. These days, many academics are beginning to recognize sketching as a useful and natural way for regular people to start experimenting with 2D modelling and animation.

Akif Yasin Ayas et al. [18] suggested the AI-based two-dimensional video game in Unity 2D Game Engine, which can decide what the game characters will do within their abilities. The user takes control of a two-dimensional character inside game applications. The two-dimensional game plane that this character traverses is generated via procedural plane-generating methods. Furthermore, data like the number of opponents conquered and the level attained are recorded. An overall of eight experiments were conducted in the research. The author tracks how long the user can complete the game using various character-specific equipment in these tests. The duration of that time varies between 0.54 and 1.88 seconds. Kun Liu et al. [19] proposed the Decision Tree Classification (DTC) for Animated Character Style Investigation. The author noted whether exaggerations were toward augmentation or reduction of body parts compared to prototype actual human bodies in animated characters. Characters in animated stories are classified according to gender, country of origin, and story role using decision tree classification. The necessary body length criteria for this process were identified using this approach as well. According to the data, male cartoon characters from the United States and Japan tend to have disproportionately large heads and bodies, albeit the US characters' exaggerations are more notable. To differentiate between animated figures from the United States and Japan, the decision tree used just five head and chest length parameters (with an accuracy of 94.48% in the training group and 67.46% in the testing group).

Kumar et al. [20] recommended AI-3D Animation and VR Integrated Computer Graphics Imagery in modern

education. This article aims to provide universities with information that will be useful to them as they use a virtual reality (VR) e-learning platform to maintain and improve their education and learning system. The effectiveness of virtual reality (VR) in medical education was evaluated using machine learning (ML) and AI. The authors additionally looked at the suggested model with various parameters and projected student performance using several validation tests, including the T-test, P-test, and One-Sample Validation tests. Wei Ding and Wenfa Li [21] discussed the Improved Deep Convolution Neural Network (DCNN) for High Speed and Accuracy of Animation 3D Pose Recognition. The author created an abstract posture data structure from the human pose retrieved from the input picture. The author then used the modified dataset to construct the necessary character animation during runtime. This threatens the long-standing idea of monocular 3D pose assessment, notwithstanding its inherent difficulty. It can run at 384 frames per second, which is the pace of real-time. The findings showed that compared to other traditional algorithms, the enhanced algorithm greatly outperformed them regarding recognition accuracy (up 3.5%), performance (up 8-10 times), and overall quality.

A. S. Fangbemi et al. [22] deliberated the 2D and 3D Pose Estimation for Quadrupeds Using Synthetic Data. This study uses keyframe animations to present a new method for creating synthetic training data for 3D and 2D animal posture estimation. Utilizing synthetic data, the author trains several 2D and 3D pose estimation models and implements an end-to-end pipeline named ZooBuilder to automate the generation of animal animations. The pipeline receives an image of an animal in its natural environment and outputs the 2D and 3D positions of each skeletal joint. This method allows the author to generate motion capture data that may be used to generate animations of animals. Yong Wan et al. [23] presented the FluoRender architecture for 2D and 3D posture analysis. Notably, twenty movies were used for input and categorized into four groups: body, head, feet and tail. The classification was based on the viewing angles and camera locations concerning the balancing beam, which were 90° and 45°, respectively. This research identified critical mouse feature points to monitor posture in still video frames. A weighted average of the collected walk cycles was used to construct the standard walk cycle (SWC), which the author obtained by concentrating on foot motions. The author utilized the association between every walk cycle and the SWC as the weight. The process for posture analysis enhances traditional behavioural testing and analysis, enabling the identification of minor yet noteworthy variations in motor coordination and vestibular function.

Abhishek Kumar et al. [24] introduced the Artificial Intelligence integrated Internet of Medical Things (AIoMT) for medical students to learn about the human brain using 3D animation with virtual reality technology. The effects of VR on student engagement and performance in the classroom are investigated here using structural

equation modelling (SEM) and the ARCS model. The findings demonstrate that VR positively influences motivation and knowledge of the concept-to-execution procedure via practice and simulation-based training. A user-feedback-based and 3D-simulation-based design was created utilizing the suggested study technique to evaluate students' learning, analysis, and comprehension of the objects of analysis. This article claims that medical practitioners may benefit from using a smartphone VR application to practice the concept-to-execution process. Li Siyao et al. [25] offered the AnimeRun for 2D Animation Visual Correspondence from Open Source 3D Movies. The results of these studies validate that the projected dataset is more similar to the image composition of actual anime and has motion patterns that are both more complex and richer than those found in current datasets. The author builds a thorough benchmark using this dataset by assessing numerous current optical flows and segment-matching algorithms and analyzes the inadequacies of these approaches to animation information. This is done to produce a comprehensive benchmark.

Marion Mundt et al. [26] suggested animating 3d motion data to enlarge 2d video databases. The author takes a new approach to the problem of sparse data in professional sports by registering 3D marker trajectory to generic 3D body forms (hulls) and then using a 2D pose approximation technique to forecast anatomical landmarks' key spots in the generated 2D video streams. The author examines the impact of 1) changing anthropometrics and 2) the 2D camera perspective on the accuracy of keypoints estimates using 3D long jump information as an actual example. It was shown that body form influences the accuracy of 2D keypoint determination. Sagittal plane camera views are more accurate than frontal plane views. Syed Muhammad Hazry Asraf et al. [27] proposed the Hybrid animation for motion capture. The transmission of information is a challenge in terms of accuracy and interactivity, making it more appealing and easy for consumers while increasing their exposure. Thus, the most effective methods currently available are motion capture and animation. Consequently, this project's output has the potential to significantly enhance the animation industry's offerings in terms of engaging video and animation. This animation is very interactive because it uses lighting and colour. Designers may use any specialized computer language to organize and code multimedia. Before utilizing the program, one should learn as much as possible about a subject's software evaluation criteria.

Based on the survey, there are numerous problems with existing techniques in attaining high accuracy, motion control, pose detection, computation efficiency and scalability. Hence, this study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation.

3. Artificial Intelligence with Biomechanical Modeling (2D-AI-BM)

Character animation in 2D entails making static pictures move around on a screen. The process may use digital software or more conventional methods, including hand-drawn procedures. 2D character animation provides games, movies, advertisements, or online content with a distinct visual personality. Computer graphics character animation is becoming more realistic using ever-more-complicated physics-based models, yet these models provide formidable obstacles to motor control. This is particularly true when directing the actions of a biomechanically well-modelled virtual human with a complex network of contractile muscles that mimics the human body. Researchers in the field of graphics have attempted to apply machine-learning techniques to neuromuscular control. Yet, regarding complicated biomechanical models and the high-dimensional training datasets that go together, typical neural network learning methods have their restrictions. This study demonstrates that DL is valuable for training neuromuscular controller for biomechanical character animations. To avoid the tedious and intensive work mode and instead concentrate on creative innovation, animation art creation is slowly shifting towards the field of intelligence, made possible by the unprecedented technological changes and breakthroughs brought about by the era of artificial intelligence. This study makes an effort to explore the relationship between artificial intelligence and virtual animation. Hence, this study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation. The virtual skeleton was designed based on biomechanical models of the human body. An autonomous animation of a realistic virtual human head monitoring an unrestricted array of visual input is described in this paper, which applies neurobiological models of visual attentions and head/eye motions in prelates.

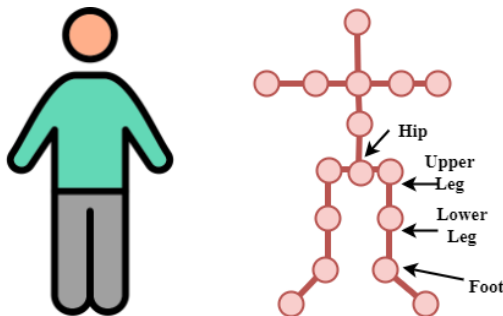


Figure 1: Depiction of a humanoid skeleton for animating virtual characters

Figure 1 depicts a humanoid skeleton used to animate virtual characters. The growing need for lifelike virtual characters in industries like entertainment, VR, and the metaverse has encouraged significant advancements in data-driven character animation. Although triangle meshes are often used to create virtual characters, it is impractical

to define animations as the movement of triangles. Joints stand in for the human body's articulated portions. All of the vertices in a mesh are affected by the movement of any given joint. When defining the character space, the virtual root joint is an absolute must possess. A hip joint is projected onto a floor plane to construct it, and the projected forward direction of the hip, the vertical world vector, and the cross-product of these three parameters determine its coordinate frame. After that, the hip joint is converted into the root space of the virtual environment. To determine the foot's location, this research multiplied the hip joint transformation matrix by the lower leg, the upper leg, and the foot.

Figure 2 shows the proposed 2D-AI-BM model. The data are taken from the Anime Dataset 2023 Kaggle Dataset [28]. Data pre-processing using data cleaning and normalization is integral to any image processing workflow.

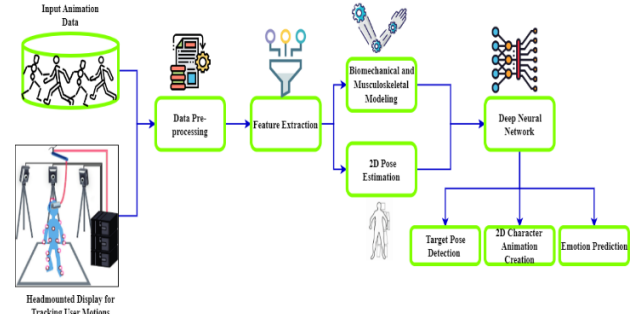


Figure 2: Proposed 2D-AI-BM model

It is possible to improve neural network learning performance using appropriately pre-processed input images. In terms of the classification system, this research must find a way to simplify animation image inputs without sacrificing classification information since the dataset would include cartoon images with varying patterns or styles. When features are propagated to the deep layer of a DNN, some shallow features like textures and lines are lost in the process of pooling and other processes. These features are not suitable for psychological feature extraction in anime development. One definite aspect of animation production is the use of brush lines, an important tool for expressing the painter's internal existence. Under varied painting methods and influenced by emotions, these lines will reveal a wide range of textures. As a subfield of biomechanics, neuromusculoskeletal simulation may either model the forces acting on muscles in response to a given movement or directly produce movement based on these forces. The relationship to physics-based animation is simple: a particular motion may be accurately achieved by applying an ideal set of muscle forces to a musculoskeletal model, considering all other needs. As an initial step in processing the input frames, the pose estimation stage applies the learnt 2D pose estimator to each frame to derive the actor's 2D posture. Estimating a 2D posture involves predicting where the body's critical points are positioned in 2D spaces. The model approximates the X and Y coordinates for every joint

localization phase. The 2D posture estimate additionally employs the expected joint location and an additional Z-axis to infer the spatial position. This study evaluates our method quantitatively by projecting a motion capture walk sequence to 2D using a known camera. The projected poses are the target poses for the evaluation. This allows for the accurate and immediate transformation of the input image's pixels into the output's spatial coordinates corresponding to the target's posture.

A. Biomechanical and Musculoskeletal Modeling

A set of line segments represents the geometry of each muscle arrangement. The line segments connect the muscle's insertion location, origin point, and a few via points. The correct segment coordinate system fixes each point. This means that the angle of the joint determines how the muscle's moment arm varies. The multi-rigid-link system represents the body model, which is driven by the individual muscles' force responses to signals from the neurological system. Contrasted with more complex muscle models like wrapping and finite elements, this one with fixed via points performs poorly in accuracy. Nevertheless, the fixed-viewpoints model may outperform the muscle geometry model when considering the trade-off between the computational cost and the precision of the synthesized motion. This work primarily aims to synthesize locomotor patterns by utilizing a limited range of joint movements. This expression describes the mechanical features of the force-length and force-velocity relationships shared by all muscle models for accurate motion prediction:

$$F_n = B_n \eta f_k(k_n) f_u(k_n, u_n) w_n \quad (1)$$

As shown in equation (1), where F_n denotes the n th muscular tension, B_n indicates the physiological cross-sectional region of muscles, η represents the coefficient signifying the extreme muscular tension per unit physiological cross-sectional area, $f_k(\cdot)$ is the force-length association, $f_u(\cdot)$ is the force-velocity connection, k_n signifies the muscle length, u_n symbolizes the muscular contraction velocity standardized by its maximum contraction velocities and w_n represents the normalized neuronal character stimulus provided by neuronal models.

The generation of muscle force F_n of musculotendon units, i can be shortened as the sum of the passive and contractile force for evaluating computational efficiency by:

$$F_{ni} = [f_q(\bar{k}_n) + b_i \cdot f_k(\bar{k}_{ni}) \cdot f_u(\bar{k}_{ni})] \cdot F_{0i} \quad (2)$$

As inferred from equation (2), where f_q denotes passive force relationships, b_i is the muscle activations, f_k indicates force-length relationships, f_u the force-velocity relationship, F_{0i} represents the maximum isometric forces, and $\bar{k}_{ni}$ denotes the regularized length of muscle units (typically, the muscle's length at rest is used

for normalization). Numerous models have been suggested to estimate the f_k and f_u the relationship concerning experimental information.

The tendon forces f_{it} , the musculotendon unit output, is determined by considering the pennation angles:

Comprehensive dynamics of the musculotendon units include activation dynamic, meaning that there are non-linear temporal relationships amongst the neural excitations v_i and the efficient activations of muscles.

$$f_{it} = F_{ni} \cdot \cos \beta_i \quad (3)$$

This non-linear association is estimated by second-order differential equations, showing dissimilar time constants for activations and deactivations:

$$b_i = \begin{cases} e_i = (v_i - e_i)/\tau_{ne} \\ \frac{e_i - b_i}{\tau_{act}}, & e \geq b \\ \frac{e_i - b_i}{\tau_{deact}}, & e < b \end{cases} \quad (4)$$

As discussed in equation (4), where v_i denotes the neural excitations, b_i indicates the muscle activations, e_i represents the intermediate variables, τ_{ne} signifies the neural excitation constant period (often ignored), and τ_{act} and τ_{deact} denotes the deactivation and activation time constant correspondingly. In 2D character animation, the activation and deactivation time constant are equivalent to simulating the activation dynamics.

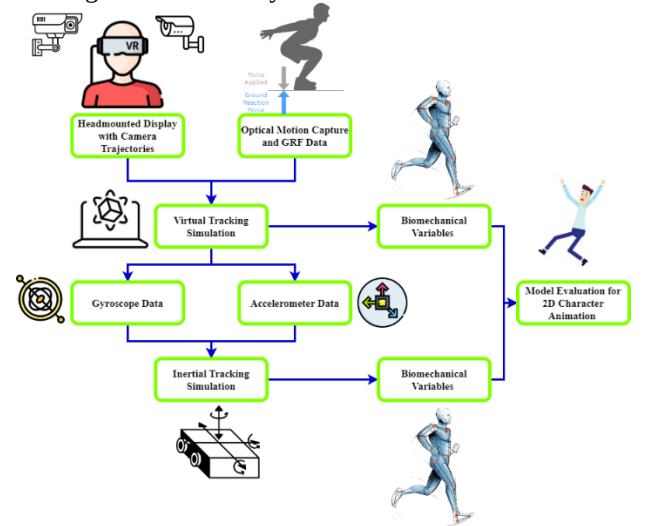


Figure 3: Biomechanical Modeling for 2D Animation Evaluation

Figure 3 shows the biomechanical modelling for 2D animation evaluation. Using inertial information from sensors for straight running and walking, this study demonstrated that optimal control simulation of sagittal-plane lower-body musculoskeletal models could precisely estimate kinetics and kinematics, with kinetics and kinematics, respectively, having Pearson correlation coefficients higher. In addition, the research discovered that sparse sensor setups, which do not include inertial

sensors on every body segment of interests, yielded equivalent accuracy when utilizing the same technique. Using a minimizing sum-of-squared difference between the mean and simulated signals standardized by the measured signals' variance, the optimum control problems in both cases followed the mean gyroscope and accelerometer signals throughout numerous gait cycles. Using normalization of the data, this study may avoid tracking signal elements that do not exhibit reproducibility over different gait cycles. In addition, a task restriction allowed the simulation to create a periodic straight gait cycle to promote convergency. This study, therefore, focused only on the kinematics of the lower body in two dimensions, as well as the kinetics of upright running and walking.

The optimization issue in this study was quite complicated due to the many unknowns introduced by simulating 2D full-body musculoskeletal models. To examine variations across trials, this article rebuilt individual trials rather than following the mean of many trials adjusted by the variance. Without knowing the movement beforehand, this study attempted to recreate random running motions without using any task constraints or beginning states. This study intended to assess the theoretical viability of such reconstructions by re-creating motions from a virtual inertial sensor signal, where the target solution was already known, to determine whether inertial sensor information lacking an initial condition adequately describes the kinematics and kinetics. Virtual data that does not include noise or soft-tissue artefacts may be used to test the suggested method accurately. This work used optimum control simulations, which utilized ground reaction force (GRF) data and optical motion capture to monitor marker trajectories. Based on this marker tracking simulation, our research calculated virtual gyroscope and accelerometer data. Using the virtual inertial information, this research created inertial tracking simulations for three trials, one for straight and one for curved running. Lastly, this research contrasted the marker and inertial tracking models' inertial signals, GRFs, angles, muscle forces and joint moments. As a result of the inverse relationship between changes in tendon length and changes in ultimate muscle force, musculotendon unit modelling is still possible. The length of the muscles should be measured correctly. In addition to the mechanical model, other models of muscles have evolved, including visual features like deformities in the muscles or a combination of the two. Three methods may be used to classify the models: data-driven, physically-based, and geometrically-based. In methods that rely on geometry, the skeleton configuration determines how the muscles will deform. During contraction, physically-based methods account for contractile forces and muscle shape changes. Lastly, data-driven approaches use surface data acquired from participants to forecast the skin shape deformed by the underlying muscle directly. These models provide an even higher degree of accuracy. However, because of the computational expense, they are often not used to control virtual characters.

B. Estimating 2D Character Animation Pose Analysis

Using a bottom-up method, the 2D pose estimator is a framework for determining human postures from pixel-level image evidence. To solve this, a multi-stage classifier is used, with each step enhancing the findings of the preceding one. The network in question is a two-branch, fully convolutional neural network. The feed-forward network sequentially forecasts a set of 2D heat map G_m of body part location and set of 2D vector field Q of part affinities describing the degree of relationship between parts. The final operation of the neural networks returns a matrix-vector for key points with x, y coordinate and their visibility scores.

$$\begin{aligned} \text{Heat map}(G_m): \mathbb{R}^2 &\rightarrow \mathbb{R} \\ \text{Affinity}(Q): \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \end{aligned} \quad (5)$$

The top branch DNNs (σ) predict the likelihood of perceptibility on every body part, and the bottom branch DNNs (Φ) predict part affinity fields. The system takes input video frame F of size height (H) $\times$ width (W) and feeds this frame to DNNs. In this way, the system mimics human output by creating 2D locations of important anatomical points for each frame. The first step involves using the input picture (frame) to generate a heat map that indicates the potential locations of each significant point in the image and a confidence score. A confidence matrix in a heat map represents the probability that a given pixel includes a certain network node. Each heat map should ideally reach its maximum when the related portion is visible, especially if a person is in the displayed area. The location values K_G in G_m is described as

$$G_m^*(K_G) = e^{-\frac{\|K_G - Y_i\|_2^2}{\rho^2}} \quad (6)$$

As inferred from equation (6), where ρ control the distribution over the peaks and $Y_i \in \mathbb{R}^2$ be the ground-truth positions of every body parts i for the human in frames. It then utilizes some joint involved via limbs to present another DNN branches, which forecasts a 2D vector field termed Part Affinity Q . Part Affinity Field is a matrix that provides joint pair orientation and position data. They come in pairs: having a Q in the y directions and Q in the x direction for every element.

A limb is generated by linking two parts, and Q encode the path from one part to another; every limb is entitled to a field of affinity between the parts of the body. If points O lie in the limbs, then its values Q_{limb} in the Q is unit vectors pointing from the starting points of the joint to the ending points of that limb; if it is outside the limbs, the values are zero.

$$\begin{aligned} Q_{limb}^*(O) &= \begin{cases} \vec{U} & \text{if limb on } O \\ 0 & \text{otherwise} \end{cases} \\ \text{Where } \vec{U} &= \frac{Y_{i2} - Y_{i1}}{\|Y_{i2} - Y_{i1}\|_2} \end{aligned} \quad (7)$$

Therefore, the network generates a set of confidence scores $G_m \in \mathbb{R}$ for every key points and a set of affinity vector fields $Q \in \mathbb{R}$ over the successive phases for scalability, $t \in 1, T$ as

$$\begin{aligned} G_m^t &= \sigma^t(F, G_m^{t-1}, Q^{t-1}), \forall t \geq 2 \\ Q^t &= \varphi^t(F, G_m^{t-1}, Q^{t-1}), \forall t \geq 2 \end{aligned} \quad (8)$$

As shown in equation (8), where σ and φ represent the DNNs for inference at every phase. At every phase, the forecasts from both branches are convolved with the image feature F for the subsequent phase. The network's last operation consists of joining two vectors, one for heat maps and the other for part affinity fields.

C. Deep Neural Networks for Model Construction

The convolutional neural network is one type of deep neural network that sees the most action. The convolution layer is the most crucial operation in a DNN. It uses a fixed-size convolutional kernel to extract local image information from a specified picture region. A single convolutional kernel convolves an ensemble of input feature maps in a convolutional layer, producing an output set of feature maps. Therefore, much like convolution kernels, the number of feature map produced as output is fixed. Supposing $A(k)$ is the outputs of the l convolution layers of the DNN ($A^{(z)}$ the video frame input images); then, the $j(k < j < n(k))$ feature maps can be articulated by:

$$A_j^k = \sum_{i=1}^{n^k} A_i^k \otimes a_j^k \quad (9)$$

Both random initialization and transferring the results of convolutional layers trained on a similar network topology on large-scale datasets may be used to produce the weights of the convolution layers. The input feature maps are convolved using similar convolution kernels, a feature with shared weight to get each element of a shared feature map.

Nonlinearity mapping, which occurs after the convolution layer, is the activation function often given the output data. To boost the feature transformation capacity, the activation function augments the output of linear operations with non-linear operations. Therefore, in theory, DNNs may approximate models of any function, including complicated non-linear ones. The Sigmoid function (S-shaped functions) transfers the values from 0 to 1 on the input domain. Supposing that the input is y , the activation functions are described.

$$\text{sigmoid}(y) = \sum \frac{1}{\sqrt{e^y - 1}} \quad (10)$$

The Sigmoid function is a good representation of how biological neurons work: when the sum of all incoming signals reaches a particular point, the neuron is stimulated and activated; otherwise, it is inhibited.

Data from a stream is often processed by adding a pooling layer after it has been through the activation and

convolution layers. Extracting significant features and reducing data size by subsampling are the goals of the left-hand pooling layer. Three examples of common pooling techniques are stochastic pooling, mean pooling, and max pooling. The input image's features are pooled, activated, and convolved across several layers. Then, the semantic features are elevated from the original pixel value to higher levels. Finally, the fully connected layers translate these to the label space. Condensing feature maps into feature vectors makes it easier to enter into the classifier; this is its principal use.

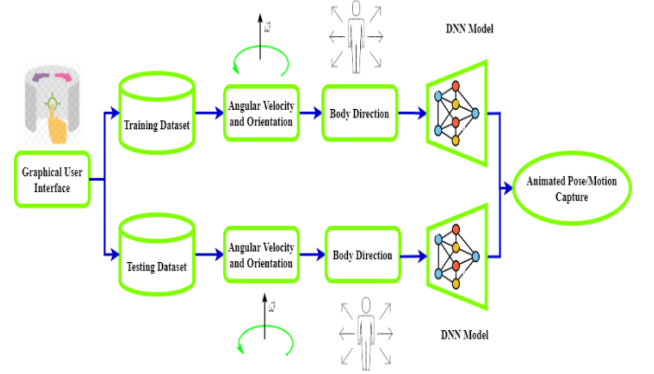


Figure 4: Training and Testing process for the body orientation using Deep neural network.

Figure 4 shows the Training and Testing process for the body orientation using a Deep neural network. Most branding and marketing campaigns now use 2D animation due to its scalability and widespread availability. A Graphical User Interface (GUI) has been created to make this model easier. By connecting this model to a VR environment, this study can see the theoretical position of each sensor as it moves. Consequently, it provides an opportunity for a more comprehensive comprehension of the vestibular system. Virtual reality simulation provides adaptable and multipurpose tools to enhance data acquisition, collection, and analysis. The training is carried out conventionally using feed-forward neural networks. However, the network would not be able to learn from its errors and improve its orientation prediction capabilities if it were to rely only on ground truth data for orientation predictions. Following the processing of anime styles for data augmentation, a deep neural network is constructed to successfully categorize anime creative styles. To include a new motion into the system, one must first design a new gesture that a unique set of tokens can recognize. Furthermore, one must establish a suitable set of states and keyframes to produce parameterized versions of the desired motion. Compared to other popular models, the proposed 2D-AI-BM model increases the accuracy, pose detection, motion control, computational efficiency and scalability ratio.

4. Simulation Findings

This study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation. The data are taken from the Anime Dataset 2023 Kaggle Dataset [28]. This dataset provides information for studying and understanding other animated movies' features, popularity, ratings, and audience size. Among the many analyses that can be performed on this dataset are the following: finding the anime with the highest ratings, investigating the most popular genres, seeing how ratings are distributed, and developing insights into viewer preferences and trends. Furthermore, the dataset allows for the development of recommendation algorithms, time series analysis, and clustering, further investigating patterns in anime and user behaviour. On a personal computer with a 1:8 GHz ATHLON (AMD) CPU and 1 GB RAM, the algorithms were executed using C and C++. The real-time capabilities and motion estimate process were evaluated using simulated and real-world movements. This study compared the results with a traditional method of reconstruction that relies on direct linear transformation to see how accurate it was. The performance of the recommended 2D-AI-BM model has been examined based on metrics like accuracy, motion control, pose detection, computational efficiency, and scalability ratio compared to existing methods DTC [19], DCNN [21], and AIoMT [24].

4.1. Accuracy Ratio

Using a variety of sensors, virtual reality applications monitor the user's bodily motions and then synchronize them with their avatar's actions in a virtual environment. Systems for motion capture, depth sensors, trackers, etc., are all examples of devices. A proposed VR device is an accurate motion capture system. In particular, this assistive technology could be a valuable assistant to the patient in performing the exercises/motions of interest more precisely and accurately. The original imagery is lost when the input image is transformed into an abstract 2D posture via the 2D pose network. Consequently, using data like pixel values to train a 2D posture estimate network to be more accurate is pointless.

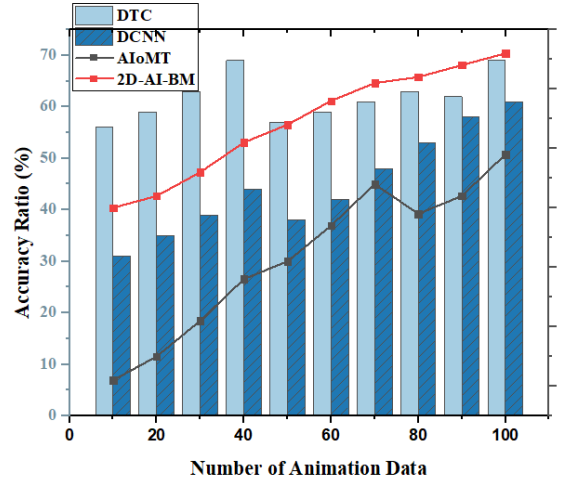


Figure 4: Accuracy Ratio

Preserving accuracy while attaining high operating speed was the design aim of our network. So, this study suggests a novel design for 2D platform detection networks that can multitask and have several levels. A more advanced level of research contrasting the impacts of control complexity and modelling accuracy on final motion quality would be noteworthy. Based on equation (1), the accuracy ratio has been analyzed. Figure 4 shows the accuracy ratio.

4.2. Motion Control Ratio

The field of character animation gives life to digital images by developing tools like motion controllers that can mimic and even create whole new actions. Simple periodic signal generation and tuning are two of the simplest methods for controlling motion. The primary function of these controllers is to mimic the oscillatory motions shown by creatures like fish, worms, snakes, and even human chest motions. The number of variables may be reduced by grouping similar muscles and applying these signals to them. Various nontrivial sensorimotor tasks were tested in this work, including creating natural limb-reaching gestures to touch moving visual objects and writing and sketching with a finger. Our unique technique performed well in all of these activities. Gravity makes it difficult to do these activities, which call for controlled eye movements for saccadic foveation and smooth pursuit of visual objects, as well as controlled, realistic head motion, visually directed dynamic limb movement, etc. Based on equation (6), motion control has been determined. Figure 5 demonstrates the motion control ratio.

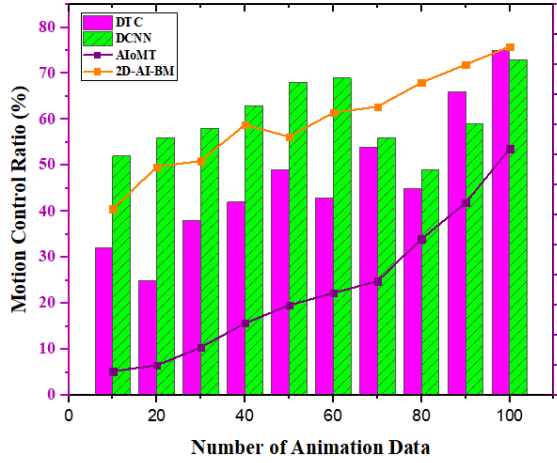


Figure 5: Motion Control Ratio

4.3. Pose Detection Ratio

Ensuring real-time posture estimation and visualization may be challenging, particularly for low-performance VR devices with comparatively slow network speeds since users' hardware performance and network speeds differ. Using visual data like images and movies, 2D human posture estimation may estimate where important parts of the human body are in space. Conventional 2D human posture estimate relies on part-by-part feature extraction methods. This study uses video image analysis technologies to analyze the motion of weightlifting during weight training. To expedite the training process for designers, this study provides a way for automatically extracting the important posture frames of the motion process, understanding the motion technology, and enhancing training. Training efficiency and development are far-reachingly important. Based on equation (5), animation pose has been detected. Figure 7 illustrates the pose detection ratio.

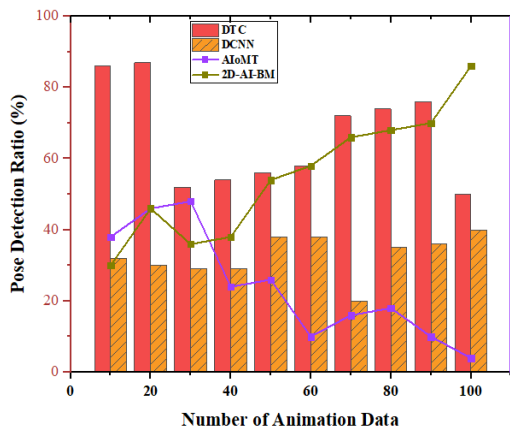


Figure 7: Pose Detection Ratio.

4.4. Computational Efficiency Ratio

For animated applications to seem realistic, animators work hard to process human motion computationally. Inverse kinematics and online motion capture algorithms are often used for posture estimation in virtual reality avatars. This study strives for a steady process that uses less computation in a small region, which differs from previous techniques. Consequently, in networked multi-user applications, the suggested method offers minimal latency for users of virtual reality devices, ranging from high performance to low performance.

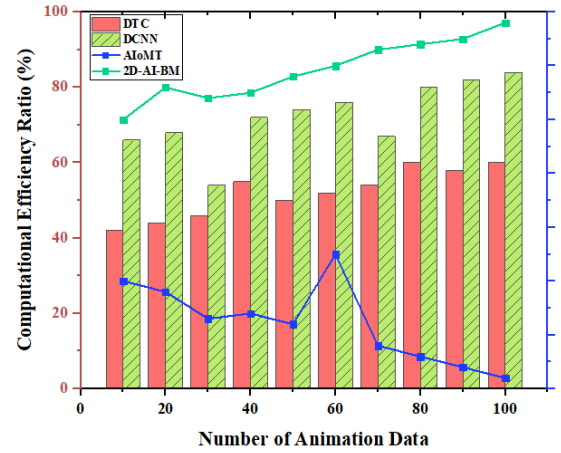


Figure 8: Computational Efficiency Ratio

This research employs a DL algorithm to assess the user's upper body position in real time while utilizing virtual reality. Instead of estimating poses using inverse kinematics, this work accomplishes pose-mimicking in real-time with reduced processing needs via data-driven methods based on motion capture datasets and DL. Based on equation (2), the computational efficiency has been analyzed. Figure 8 signifies the computational efficiency ratio.

4.5. Scalability Ratio

Neural networks, deep learning, and other trained networks enable objective-based autonomous controller development. In earlier iterations, scalability was a major issue; the complexity of characters did not scale well. Creature motion animations that are both striking and flexible have been created using this control method. The suggested method for emotion recognition is both generalizable and scalable. Through the integration of emotions and the provision of unforeseen possibilities, the system becomes scalable. One example is the ability to include moods and mixed emotions into the system via deep learning. This includes combinations such as happiness and surprise, rage and fear, contempt and rage, and so on. This study delves into the societal standing, cultural background, or emotional reactions of fictional characters in this study. Different motion and gesture datasets and text inputs for conversations impose

limitations on the system. Based on equation (8), the scalability of the suggested model has been identified. Figure 9 denotes the scalability ratio.

This study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation.

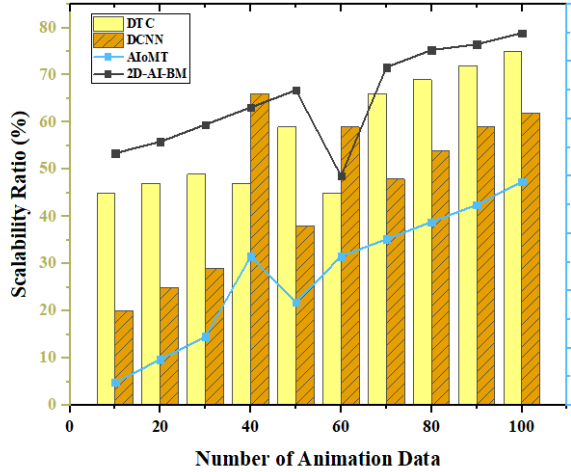


Figure 9: Scalability Ratio

The numerical findings exemplify that the recommended 2D-AI-BM model improves the accuracy rate of 97.4%, computational efficiency ratio of 96.3%, motion control ratio of 95.4%, pose detection ratio of 94.8% and scalability ratio of 93.2% compared to other popular techniques.

5. Conclusion

This study suggested 2D animation using Artificial Intelligence with Biomechanical Modeling (2D-AI-BM) for motion capture and character animation. Using reference videos of performers, this research demonstrates a method to assist artists in creating and animating 2D characters. Our tool is especially useful for artists who want to create a character that captures an actor's personality by assisting them in selecting typical hand and arm positions. This technique improves the immersion and realism of virtual reality experiences by making the user's upper body character seem more real. The highest level of accuracy for precisely capturing human movements and displaying a matching virtual character is on-site motion capture technology. The first steps in preparing the video data for processing using a motion estimation tool started with implementing a pre-processing stage. DNNs undergo their training using massive amounts of data generated by the biomechanical model of the human musculoskeletal system. Consequently, backpropagation DNN performs high-dimensional non-linear regression while learning from the synthesized data. Lastly, after the DNNs are trained correctly, they provide input-output controllers for biomimetic neuromuscular motors. The three parts of our method should be understood separately. First, when the

DNNs are trained offline, the training data may be removed; second, trained DNNs do not need training data to be retained or cached to function online. Simplified context usage while defining motion and 2D input devices may make the system more scalable. Adding more controllability would result in a more convoluted interface. The numerical findings exemplify that the recommended 2D-AI-BM model improves the accuracy rate of 97.4%, computational efficiency ratio of 96.3%, motion control ratio of 95.4%, pose detection ratio of 94.8% and scalability ratio of 93.2% compared to other popular techniques. This suggested learning technique has limits in forecasting the precise pose of this abnormal pose. In the future, this study wishes to add gestures for various falling motions and interaction with the environment.

References

- [1] Farahani, J., Soltani, P., & Roberts, R. E. (2020). Expertise differences in a 2D animation simulation decision-making task: The influence of presentation speed on performance. *Progress in brain research*, 253, 87-100.
- [2] Schreck, C., & Wojtan, C. (2022, May). Coupling 3D liquid simulation with 2D wave propagation for large scale water surface animation using the equivalent sources method. In *Computer Graphics Forum* (Vol. 41, No. 2, pp. 343-353).
- [3] Untari, R., Kamdi, W., Dardiri, A., Hadi, S., & Nurhadi, D. (2020). The development and application of interactive multimedia in project-based learning to enhance students' achievement for 2D animation making. *International Journal of Emerging Technologies in Learning (IJET)*, 15(16), 17-30.
- [4] Morgenroth, D., Reinhardt, S., Weiskopf, D., & Eberhardt, B. (2020, December). Efficient 2D simulation on moving 3D surfaces. In *Computer Graphics Forum* (Vol. 39, No. 8, pp. 27-38).
- [5] Ejimonye, J. C., Ugwuanyi, C. S., Okeke, C. I. O., & Nwoye, M. N. (2020). Two-dimensional animation and students' achievement in mathematical economics: implications for science teaching. *International Journal of Engineering Research and Technology*, 13(6), 1220-1230.
- [6] Farahani, J., Soltani, P., Rezlescu, C., & Walsh, V. (2020). Assessing decision making using 2D animations in elite academy footballers. *Progress in Brain Research*, 253, 71-85.
- [7] Verma, M., Thapliyal, V., Mishra, A., & Rani, S. (2024). Efficacy of 3D Monte Carlo Simulations vis-à-vis 2D in the Estimation of Pi: A Multifaceted Approach. *International Journal of Applied and Computational Mathematics*, 10(2), 1-15.
- [8] Even, M., Bénard, P., & Barla, P. (2023, May). Non-linear Rough 2D Animation using Transient Embeddings. In *Computer Graphics Forum* (Vol. 42, No. 2, pp. 411-425).
- [9] Alherz, A., Tsuruta, N., & Kondo, K. (2020). Multiple-character animation systems for children at different educational stages. *International Journal of Asia Digital Art and Design Association*, 24(1), 1-7.
- [10] Wang, J., & Gao, J. (2023). Research on 3D Animation Technology of Human and Scene Based on Virtual Reality. *Art and Performance Letters*, 4(1), 12-17.
- [11] Bishop, P. J., Cuff, A. R., & Hutchinson, J. R. (2021). How to build a dinosaur: Musculoskeletal modeling and

- simulation of locomotor biomechanics in extinct animals. *Paleobiology*, 47(1), 1-38.
- [12] Feldotto, B., Morin, F. O., & Knoll, A. (2022). The neurorobotics platform robot designer: modeling morphologies for embodied learning experiments. *Frontiers in Neurobotics*, 16, 856727.
- [13] Ling, X., Zhu, Y., Liu, W., Liang, J., & Yang, J. (2023). The Generation of Articulatory Animations Based on Keypoint Detection and Motion Transfer Combined with Image Style Transfer. *Computers*, 12(8), 150.
- [14] Li, J., & Wang, J. (2023). Digital animation multimedia information synthesis based on mixed reality framework with specialized analysis on speech data. *International journal of speech technology*, 26(1), 63-76.
- [15] Babadi, A., Van de Panne, M., Liu, C. K., & Hämäläinen, P. (2021). Learning Task-Agnostic Action Spaces for Movement Optimization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 4700-4712.
- [16] Liu, Y., Zheng, C., Xu, F., Tong, X., & Guo, B. (2020). Data-driven 3d neck modeling and animation. *IEEE Transactions on Visualization and Computer Graphics*, 27(7), 3226-3237.
- [17] Wu, Y., Chen, Z., Liu, S., Ren, Z., & Wang, S. (2022). Casa: Category-agnostic skeletal animal reconstruction. *Advances in Neural Information Processing Systems*, 35, 28559-28574.
- [18] Ayas, A. Y., Aydin, H., Çetinkaya, A., & Güney, Z. (2023). Artificial Intelligence (AI)-Based Self-Deciding Character Development Application in Two-Dimensional Video Games. *Bilgi ve İletişim Teknolojileri Dergisi*, 5(1), 1-19.
- [19] Liu, K., Chang, K. M., Liu, Y. J., & Chen, J. H. (2020). Animated Character Style Investigation with Decision Tree Classification. *Symmetry*, 12(8), 1261.
- [20] Kumar, A., Jilani Saudagar, A. K., AlKhathami, M., Alsamani, B., Abul Hasanat, M. H., Khan, M. B., ... & Singh, K. U. (2022). AIAVRT: 5.0 Transformation in Medical Education with Next Generation AI-3D Animation and VR Integrated Computer Graphics Imagery. *Traitement du Signal*, 39(5).
- [21] Ding, W., & Li, W. (2023). High speed and accuracy of animation 3D pose recognition based on an improved deep convolution neural network. *Applied Sciences*, 13(13), 7566.
- [22] Fangbemi, A. S., Lu, Y. F., Xu, M. Y., Luo, X. W., Rolland, A., & Raissi, C. (2020). ZooBuilder: 2D and 3D pose estimation for quadrupeds using synthetic data. *arXiv preprint arXiv:2009.05389*.
- [23] Wan, Y., Edmond, M. A., Kitz, C., Southern, J., & Holman, H. A. (2023). An integrated workflow for 2D and 3D posture analysis during vestibular system testing in mice. *Frontiers in Neurology*, 14, 1281790.
- [24] Kumar, A., Saudagar, A. K. J., Alkhathami, M., Alsamani, B., Khan, M. B., Hasanat, M. H. A., ... & Srinivasan, B. (2023). Gamified learning and assessment using ARCS with next-generation AIoMT integrated 3D animation and virtual reality simulation. *Electronics*, 12(4), 835.
- [25] Siyao, L., Li, Y., Li, B., Dong, C., Liu, Z., & Loy, C. C. (2022). Animerun: 2d animation visual correspondence from open source 3d movies. *Advances in Neural Information Processing Systems*, 35, 18996-19007.
- [26] Mundt, M., Oberlack, H., Morris, C., Funken, J., Potthast, W., & Alderson, J. (2021). No dataset too small! Animating 3D motion data to enlarge 2D video databases. *ISBS Proceedings Archive*, 39(1), 25.
- [27] Asraf, S. M. H., Abdullasim, N., & Romli, R. (2020, February). Hybrid animation: implementation of motion capture. In *IOP Conference Series: Materials Science and Engineering* (Vol. 767, No. 1, p. 012065). IOP Publishing.
- [28] <https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset>