International Journal of Religion
2024
Volume: 5| Number 10 | pp. 1276 – 1288
ISSN: 2633-352X (Print) | ISSN: 2633-3538 (Online)
ijor.co.uk
DOI: https://doi.org/10.61707/3xe4r017

Computer Simulation and Innovation of Virtual Reality Technology in Film and Television Animation Scene Construction

Feng Long[1] and Wenyu Jiang[2]

Abstract

One of the artistic mediums of film and television is works of art. They represent a company's and a nation's cultural legacy in the modern period. They have long been the spirits that guide civilization towards affluence and advancement. The advancement of computer technology has led to extraordinary advancements in cinema and television technology, particularly in the field of special effects. When compared to more established real-world film and television technologies, virtual reality (VR) technology offers filmmakers and producers more creative freedom. This research propose novel technique in computer aided system in VR technology in film and television animation scene modelling based on artificial intelligence (AI). Here the virtual reality in film and animation scene has been constructed using machine learning. The 3D graphic design in animation scene modelling is carried out using convolutional generative fuzzy encoder face deformable model. Experimental analysis is carried out based on various animation scene construction dataset in terms of accuracy, precision, RMSE, robustness, F-1 score. This article designs a colour rendering system for 3D animation that is primarily separated into several modules based on various functionalities.

Keywords: *Computer Aided System, Virtual Reality Technology, Television Animation, Artificial Intelligence, 3D Graphic Design*.

INTRODUCTION

A new technique that is developing alongside computer hardware and software is three-dimensional animation, or 3D animation. Instead of relying solely on external graphic input, 3D animation is automatically created inside computer using a variety of inputs. Film production has entered digital era, entire process, from image creation to projection, has experienced a significant transformation due to the quick development of digital film method as well as computer graphics generation method. Digital cameras can be utilised in digital filmmaking to accomplish live-action shooting and produce high-quality, high-fidelity audio and video files. Additionally, virtual cameras, image graphics processing software, graphics video processing software, other tools can be used for digital modelling and adjustment, enabling process of combining live images with methods, synthesising and co-mingling dynamic images, seamlessly editing images to create virtual images [1]. Painting, photography, writing, and other artistic mediums are all combined into one entire art form: animation. Animation has the ability to more accurately depict people's inner thoughts and to bring to life concepts that were previously limited to fiction. When it comes to the methods used in the production of animation, the main types are computer animation, which uses a computer as its primary instrument, conventional animation created by hand, stop-motion animation created using photography technology. The still image is edited frame by frame by the camera, and it is then animated on the screen using a television broadcast system [2]. The term "virtual reality" (VR) refers to a technique that combines computer, electronic information, simulation technologies to create a virtual environment that mimics its fundamentals and allows users to experience it in real life. In tandem with ongoing advancements in science, technology, social productivity, VR has advanced significantly as well as gained widespread recognition. It is challenging to discern between the realism of the simulation environment and the actual world since users might feel the most true emotions in the virtual reality world [3]. A super simulation system facilitates human-computer interaction and allows users to freely operate and obtain their preferred environmental feedback. Through expert photography and visual processing technology, which integrates more artistic components, film and television production is a technique that presents new ideas as well as thoughts to audience [4]. Sound editing, special effects, and filming are often included in the production of motion pictures and television shows. Thus, the creation of films and television shows will heavily rely on

[1] Universiti Teknologi Mara, Selangor, 40450, Malaysia,Fuzhou University of International Studies And Trade, 350202, China, Email: longfeng7522665@outlook.com, (corresponding author)

[2] Universiti Teknologi Mara, Selangor, 40450, Malaysia

virtual reality and IoT technologies. Therefore, development of China's film as well as television industry greatly depends on how to use them to the creation of films and television shows as well as generate new production models and processes. The convergence of virtual reality (VR) and artificial intelligence (AI) has opened up previously unheard-of possibilities for animation scene systems. VR technology has demonstrated its distinct benefits in AI model training among them. use of VR through a specific case study to train AI-based animation scene systems, especially in the area of fall detection in animation behaviour. With the use of virtual reality technology, developers may create a very realistic environment in which to model different scenarios and hone AI models for precise behaviour identification. Conversely, with massive data learning and pattern recognition, artificial intelligence can enhance cognition and comprehension of intricate behaviours [5]. The two work together to make the animation scene system more dynamic, realistic, and intelligent. A character's behaviour changing is a frequent dynamic event in animation scenes. A system that can recognise and react in real-time to character fall behaviour can be built by fusing virtual reality and artificial intelligence technologies. Right now, the most widely used artificial intelligence science is machine learning. By studying the traits and outcomes of well-known data sets (training samples), it creates a prediction model (training model). It is possible to forecast and measure the properties and outcomes of unknown data by using the model. Classical ML methods include decision tree, NN, SVM [6].

Related works

Improvements in animation quality can be achieved with development of AI, particularly with the use of neural networks. Many studies have proven that Reinforcement Learning algorithms are able to generate better realistic movements. In work [7], an interactive method was employed to method animated trees. From photos and videos, intricate branches were rebuilt, Fourier transform was used to estimate parameters of leaves. These parameters were then applied to a reduced model to create tree animations. Author [8] proposed that teaching theories as well techniques are changed in order to improve teaching outcomes after analysing the animation design using computer graphics theory. In order to achieve real-time virtual filming with multiplayer motion capture, work [9] examined the effects of virtual filming on the creation of animation. It suggested using optical positioning methods to precisely locate camera, character positions and inertial pose sensors to capture multiple character poses. The author [10] suggested a 3D face modelling technique that reconstructs faces using convolutional neural networks and uses the Gabor features for feature point matching as well as LBF method for automatically detecting face features. Work [11] used computer vision and machine learning tools to analyse vast amounts of animated movie data, examining the visual expressions and patterns in the films. The author of [12] developed concept as well as techniques of animation design, examined the traits and evolution of digital media, established architecture of the animation system for digital media. Work [13] developed a gesture detection method based on a human-computer interface system, employed VR method as foundation, examined differences between animation design based on VR technology as well as traditional animated design. User happiness and immersion have been enhanced by [14] through natural and intuitive interaction, personalised customisation, real-time feedback, optimisation in flexibility and scalability. Scene construction is an essential skill in the production of films and television shows. Conventional scene modelling techniques frequently rely on hand modelling, which takes a lot of time and requires careful assurance of accuracy. Reverse modelling technique, based on multi-sensor measurement, has steadily emerged as a new solution with development of CAD film as well as television scenes. A thorough overview of the fundamentals, benefits, and uses of this technology in the creation of motion pictures and television shows was given by author [15]. Reverse engineering techniques are used in CAD film as well as television scene reverse modelling based on multi-sensor measurement, which transforms sensor data into a 3D CAD model. Reverse modelling technology may swiftly reconstruct accurate scene models for scenarios that are challenging to construct using traditional means, like expansive natural landscapes, historic structures, etc. Reverse modelling technology can assist in producing realistic virtual environments or objects in special effects creation, enhancing the realism of the effects. Reverse modelling technology can be applied to game creation to produce intricate game scenes and objects, enhancing the gaming experience. As a significant subfield of deep learning, Fully Convolutional Networks (FCN) offer fresh approaches to 3D scene texture restoration that yield excellent results. A thorough introduction to use of fully convolutional networks powered by computer vision algorithms in high-quality texture 3D scene reconstruction was given in work [16]. Fully convolutional neural networks, in contrast to conventional

convolutional neural networks, are able to receive input of any size and provide semantic data that is same size as input.

VR technology in film and television animation scene modelling

More frequently than not, we can give the image or sound interactive qualities so that viewers can choose the audiovisual language they want to hear. The audience can alter the audiovisual narrative's sequence through involvement, which alters the story's growth process and plot direction, even though the designer of VR environment provided majority of audiovisual materials to story beforehand. For instance, three TVs are positioned in the same indoor area, and viewers are free to switch them on at whim. Various TV shows will direct viewers to various plot points. While non-linear interactive audiovisual narratives of this type are not exclusive to VR animation, author argues that interactive audiovisual narratives in VR environments tend to resemble people's audiovisual perceptions more than interactive audiovisual narratives in other media. When developing interactive audiovisual language of VR animation, this allows us to watch and learn from people's real-world interactions. Although computer technology and film and television art are two very different fields, VR technology as well as film space scene design have one ultimate goal in common: to serve people. This goal can be expressed more clearly in the way that both technologies are presented, which is through visual as well as even human interaction with virtual environment to provide better services. As Figure 1 illustrates, the virtual reality technology used in films as well as television shows is more of a combination of technology and art. Use of software to create a realistic virtual environment for films as well as television shows that are produced based on a scripted fictitious film universe is known as virtual technology.

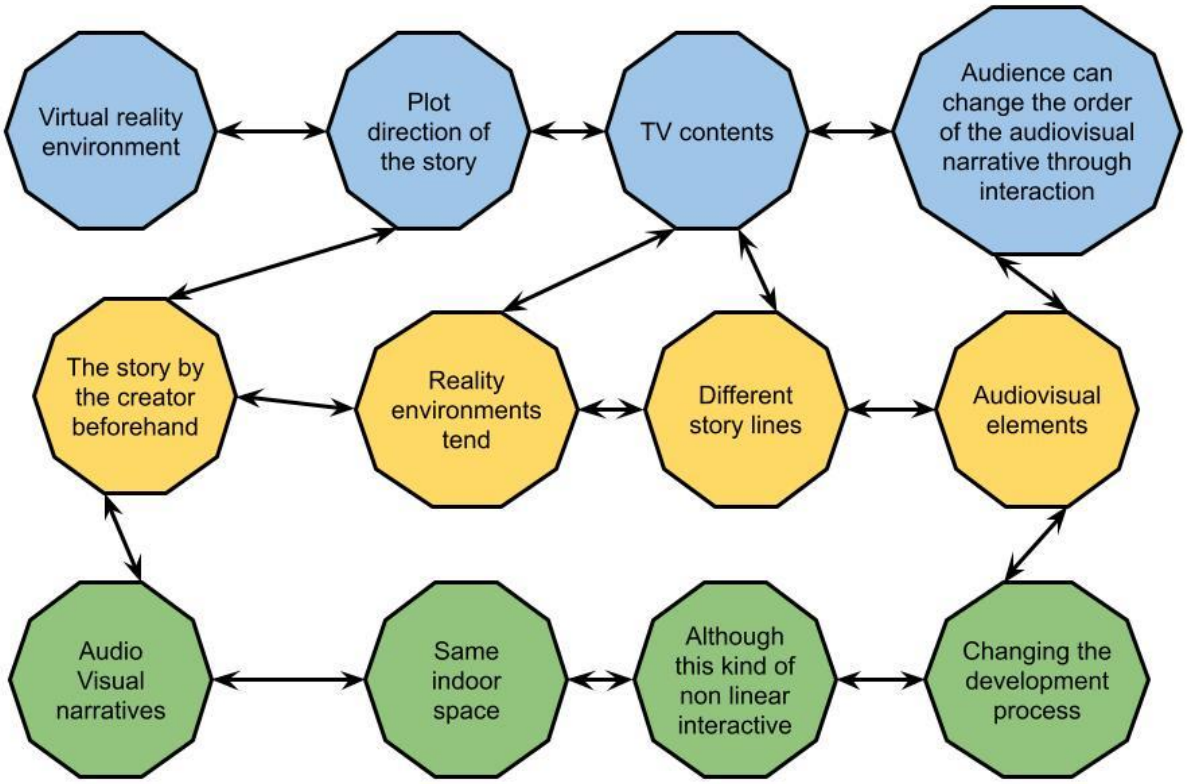

Figure 1: computer-aided virtual reality technology architecture

the plane's X, Y coordinate axis alone, or what is known as the "two-dimensional" plane. Pre-existing sceneries notwithstanding, the human brain processes visual information from the eyes to generate a three-dimensional sense. This sense is derived from objects in the picture that have distinct shapes, varying depths of field, shifting light and shadow, and the ability to create illusions. We are able to appreciate pictures and videos and the vibrant world we live in because of this illusion. We are able to create as well as receive analogous light changes from

3D environment, cinema screen is an active luminous body that allows us to create the illusion of reality. This helps to explain why we see two-dimensional films as able to produce a three-dimensional sense. Our eyes are also primarily used to judge and perceive the three-dimensional world in which we live. While the visual effects of virtual reality technology equipment and the former are very comparable, the latter offers greater potential for advancement in human interaction. VR technology allows users to see the three-dimensional world from the first perspective, even though watching 7D films and television shows will make it seem as though you are in the real world.

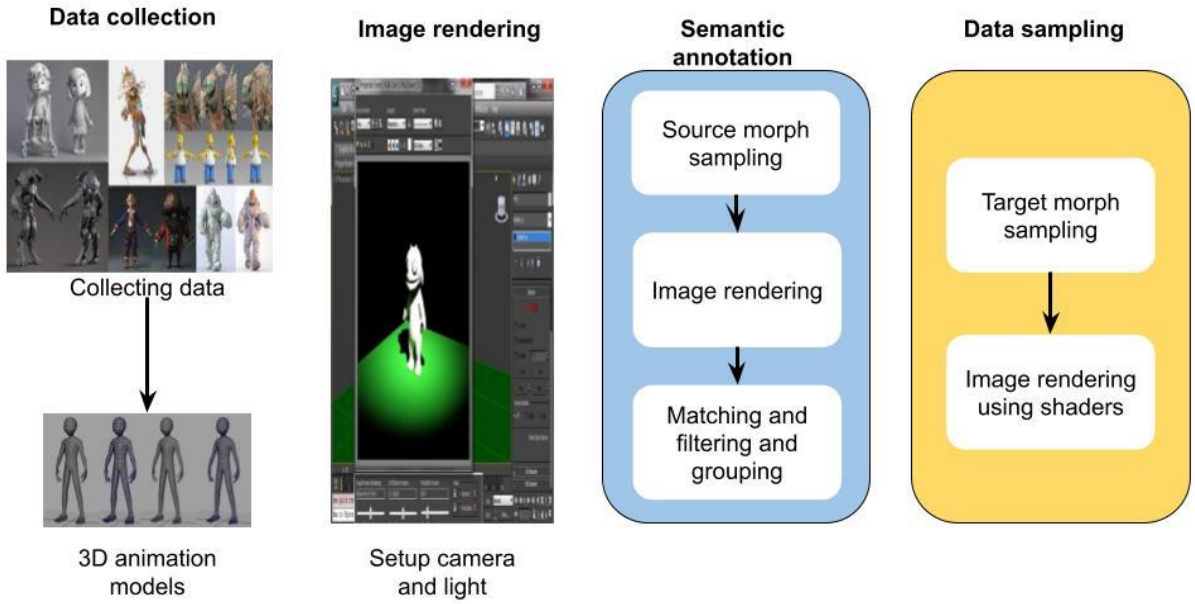

Figure 2: Dataset Creation Pipeline Overview.

Two distinct websites are used to gather 3D animation models (A). Next, a head portion of the gathered model is shown following the application of a maximally intense morph (B); they are subsequently employed for semantic annotation (C). Sampled target morphs are utilised in a data sampling process to create pose vectors, which act as requirements to generate multi-position images with a variety of head and facial expressions.

Image Rendering (B) We create a 2D head image production pipeline based on Blender, an open source 3D computer graphics programme that facilitates visualisation, modification, rendering of 3D animation methods, in order to accomplish an automatic sampling utilising 3D animation models. Three factors to produce animation head images in Blender: (1) camera position; (2) light condition; (3) image resolution.

In order to capture the head portion, we adjusted the camera position according to the location of a neck bone. We employ a directional light point along negative y-axis, representing the frontal direction of an animated character, in relation to the light situation (See Fig. 2 (B)). We first set image resolution to 256 x 256, which is the same resolution as prior head reenactment techniques, before rendering. Nevertheless, we are able to produce a better image resolution (1024 x 1024) because AnimeCeleb photos are created from a 3D vector graphics methods. In the supplementary material, we show several generated samples under diverse settings to illustrate its wide application. It should be noted that the rendered graphics include a transparent background with an alpha channel that allows the foreground animation character to be distinguished from the background.

Semantic Annotation (C) The amount of morphs in each 3D animation model varies greatly, ranging from 0 to even more than 100. Before accurately annotating the semantics of a particular morph, it is challenging to apply a uniform criterion because morph naming conventions vary depending on their originator. Finding expression-related morphs and annotating them in accordance with the unified naming system is one of the objectives of the semantic annotation. Crucially, this makes it possible to display a 3D animation model and

sample an expression-related source morph that is operating correctly. For instance, a morph あ associated with a particular 3D animation model may be designated as target morph if it is determined to indicate a semantic of speaking syllable "ah" with a mouth. Following annotation, target morph Mouth (A) is utilised to regulate mouth shape, using source morph あ of the 3D model. Our goal is to align the source and target morphologies. Set of source morphs that share the same name typically represent the same semantics. As a result, we employ a two-phase strategy: individual examination followed by group annotation. The latter is in charge of individually examining the matched source morphs to ensure that everything is functioning properly. The former matches a collection of source morphs that share same name to a target morph. In the course of group annotation, we tally number of source morphs present in the 3D methods as well as exclude those with fewer than fifty. The number of incorrect annotations at the group annotation is decreased by the individual scrutiny.

Data Sampling (D): A 3D animation model is applied with randomly chosen target morphs for every part during the data sampling process. Independent sampling from a uniform distribution, $\mathcal{U}(0,1)$, is used to calculate the magnitudes of the morphs. A 3D rotation matrix with respect to the head rotation is calculated using sampled yaw, pitch, roll values between -20∘ and 20∘. After applying morphs and rotation, we obtain a paired pose vector p∈ℝ20 and render a transformed head. Supplementary information contains a thorough explanation of the pose sampling procedure.

The altered photos and paired posture vectors are created using Blender's real-time rendering engine. As seen in Fig. 2, we use four different kinds of shaders during rendering to produce a variety of textured 2D images. Two image groups—a group of frontalized images with expression and a group of head-rotated images with expression (rotated-expression)—are included in AnimeCeleb because morphs and the head rotation are applied individually. Depending on how many annotated target morphs a 3D animation methods contains, a varied amount of photos are sampled from the 3D model. We produce 100 photographs from a 3D animation model if it has more than five annotated target morphs; if not, we only produce 20 images. While the reflection component is a little constant, light component is easily influenced by intensity of light. The reconstruction loss mostly dictates the loss function during the decomposition phase. According to eqn (1):

$$L = L_{recon} + \lambda_{ir} L_{ir} + \lambda_{is} L_{is} \tag{1}$$

The total is the weight of the structural smoothing loss and the reflection stability loss. Eqn (2) provides a definition of reconstruction loss.

$$L_{recon} = \sum_{i=low,\ nomal} \sum_{j=\ low,noma} \lambda_{ij} \left\| R_i \times I_j - S_j \right\|_1 \tag{2}$$

In order to limit the decomposition network to satisfy the Retinex constraint, original picture is recovered by reflection factor as well as light component that was extracted from methods using an inverse transformation. In order to guarantee that the dissociated reflection factor of ordinary illumination image and reflection component of decomposed low illumination image match, Lir loss is implemented. Eqn (3) provides a definition of stability loss of reflection component.

$$L_{ir} = \| R_{low} - R_{normal} \|_1 \tag{3}$$

It is vital to process smoothness of image by adding lighting smoothing loss to enhance output quality of image and lessen image degradation brought on by enhanced noise. The expression in mathematics is displayed in eqn(4):

$$L_{is} = \sum_{i=\ low,nomal} \nabla \left\| I_i \exp\left(-\lambda_g \nabla R_i\right) \right\|_1 \tag{4}$$

In terms of weight, Lis achieves a visually clearer light image by reducing control of extraction area of smooth big reflectivity, which is beneficial to image's structure. Entire codec structure must be received, illumination component and reflection factor must be sent, and the adjusted illumination value must be gained for the enhanced network portion. By multiplying its better lighting images, BM3D technology reduces noise of reflection components as well as produces final improved image. This portion of loss functions similarly to loss

during decomposition stage, which is primarily made up of structural and reconstruction loss. Eqn (5) illustrates how this portion of the loss is expressed.

$$L = L_{recon} + \lambda_{ij} L_{is} \tag{5}$$

More overhead performance and a richer model represented by voxels are associated with greater resolutions. The voxels discussed in this article can be found in scene models as axis-aligned bounding boxes that provide ambulable surfaces. Efficiency and compactness, or building boundary box whose points are closest to fitting in smallest amount of time, are two crucial indications in method of producing boundary box given a three-dimensional convex hull, let's say there are N triangles. Any triangle's area △ pkqkrk is expressed as Ak, and the convex hull's total area can be expressed as eqn (6):

$$A^M = \sum_{k=0}^{n-1} A^k A^M = \sum_{k=0}^{n-1} A^k \tag{6}$$

Eqn (7) provides the expression for weighted average value of centroid of complete convex polyhedron.

$$M^M = \frac{1}{A^M} \sum_{k=0}^{n-1} A^k m^k A^M = \sum_{k=0}^{n-1} A^k \tag{7}$$

Here, foot mark i denotes ith component, angle mark k represents K triangles, and i = 0, 1, 2 are in 3D scene. We first define angle mark M to represent full convex polyhedron. These specifications, as demonstrated by Eq. (8), make it simple to compute a 3 * 3 covariance array, which depicts distribution of point sets in 3D, and to represent the matrix using the eigenvector of the matrix. The eigenvectors of the bounding box's three axes are their direction vectors, which need to be normalised after computation.

3D graphic design modelling by convolutional generative fuzzy encoder face deformable model (CGFEFDM)

Channel of every convolution kernel is convolved with data of the corresponding channel of input data in classic convolution structure. Consider convolutional layer of a ∗ a, with an input channel of M and an output channel of N, the convolution can only be completed with N convolution kernels, each of which requires a ∗ a ∗ M parameters due to fact that convolution kernel must perform a convolution with all of input data channels. This results in an enormous computational complexity. Figure 3 shows depth separable convolution structure.

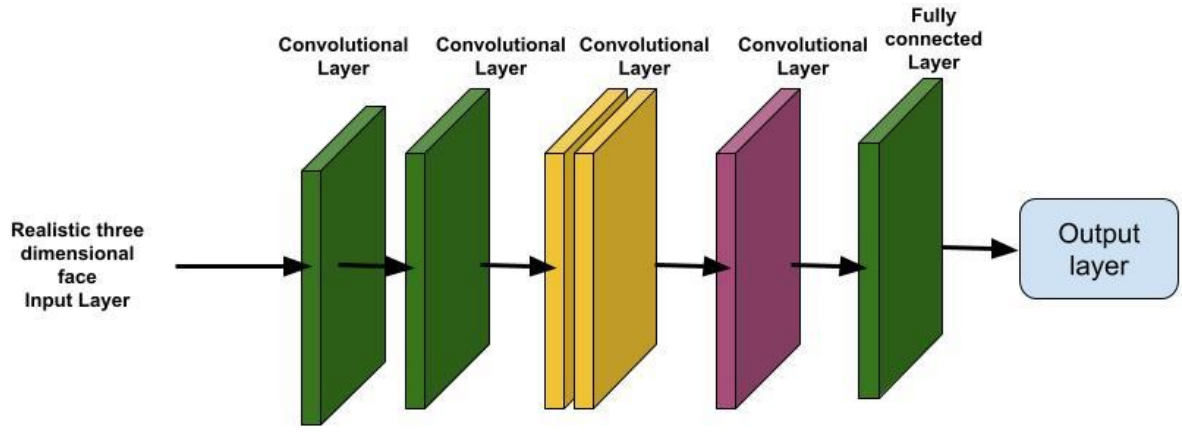

Figure-3 convolutional architecture

We contrast how many parameters are needed for depth separable convolution structure as well as conventional convolution structure. Depth separable convolution has far fewer parameters than the standard convolution, which results in a corresponding reduction in computation. Furthermore, depth separable convolution structure separates 2 parameters of channel and region, whereas the classic convolution structure frequently takes into account both at the same time. Prioritising consideration is given to the region, followed by the channel. Tests have demonstrated that this is how the building is constructed. The resulting model's accuracy is higher than that of the model created using conventional convolution, and it runs faster as a result by eqn (8)

$$\text{loss}\,(x_i, y_i) = \begin{cases} \frac{2}{3}\ln\,(x_i - y_i), & |x_i - y_i| < 1, \\ \ln\,\left(x_i - y_i - \frac{1}{2}\right), & |x_i - y_i| \geq 1. \end{cases} \tag{8}$$

Square loss in (−1, 1) interval and L1 loss in other circumstances represent the error of this function. This technique deftly sidesteps the gradient explosion issue. Using the GAN, images may be recognised. GANs are meant to create data that doesn't exist, kind of like giving artificial intelligence some creative freedom. The generator in an image GAN is in charge of creating false images. The discriminator's job is to discern between authentic and fake images, presuming that the training set contains true images. During "confrontation," the discriminator uses ongoing recognition to determine the validity of each image while the generator uses ongoing training to produce as many fake images as feasible. The generator goes through the following training steps. Initially, a label and a network are assigned. Next, adjust the input until the label that corresponds to the final image gets closer to the specified label. The discriminator is trained in a manner akin to a general neural network. Model uses a "Generator" to create an image, which it then feeds into a "Discriminator" together with the original image to be recognised. In the event that "true image and false generated image" are lost, the "discriminator" modifies the parameters. When the "generator" loses the "true generated image," it modifies the parameters. Figure 4 shows the GAN's structural layout.

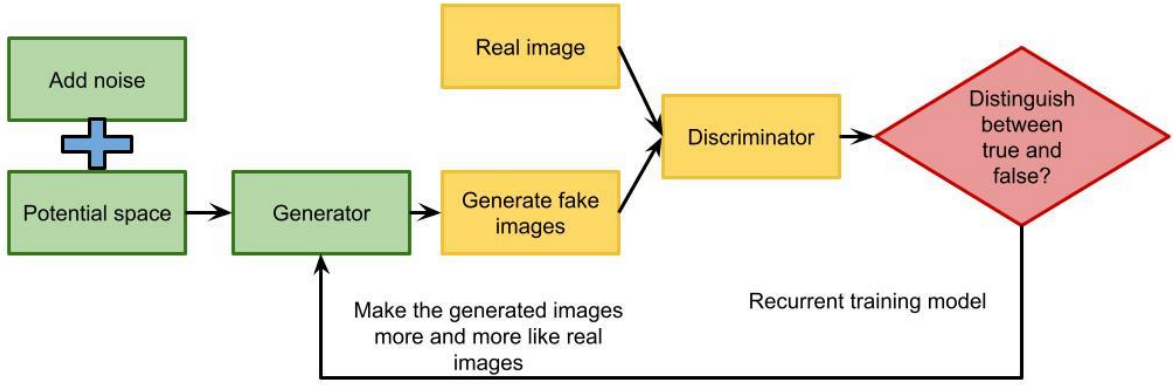

Figure 4 architecture of GAN

By minimising relative entropy between generated and true distributions, GAN solves the optimal parameter solution. Discriminator's goal function is stated as follows by eqn (9)

$$\max_D V(G, D) = E_{x \sim p_{\text{data}}}[\log\, D(x)] + E_{z \sim p_x}[\log\,(1 - D(G(z)))] \tag{9}$$

Equation (9) uses D to stand for the discriminator, G for the generator, z for generator G's input, z for random noise. E stands for anticipation. The random noise of prior probability is satisfied by px. The generator's objective function can be stated as follows by eqn (10)

$$\min_G V(G, D) = E_x \sim p_{\text{data}}\,[\log\, D(x)] + E_z \sim p_x[\log\,(1 - D(G(z)))] \tag{10}$$

Equation (10)'s primary goal is to gradually approach real data by decreasing generated value in the generator. Equation (11) thus displays the GAN training objective function.

$$\min_G \max_D V(G, D) = E_{x \sim p_r(x)}[\log\, D(x)] + E_{z \sim p_z(z)}[\log\,(1 - D(G(z)))] \tag{11}$$

Equation (11) denotes generator as G, discriminator as D, generator's input as z. G. A random noise with z that meets prior probability is called pz(z). R is the discriminator's input. D. A random noise with z that meets prior probability is called pr(x). Objective function is V (G, D). G seeks to produce as realistic an image as it can during this GAN training phase. The larger D(x), stronger D's ability. Equation (12) displays outcome of the objective function V (D, G) solution by eqn (12)

$$V(G, D^*) = 2D_{JS}\left(P_{\text{data}} \mid P_g\right) - \log 4 \quad (12)$$

Equation (12) denotes the created distribution data as Pg, genuine distribution data as Pdata, and the new discriminator parameter as D∗. The discriminator's relative entropy is represented by DJS. It is discovered that GAN has issues with gradient vanishing during training, which leads to unstable training. Similar form characteristics serve as the foundation for the various facial postures of same identity when using network regression parameters. Stated differently, the 3D face form of an item that recovers varied attitudes from its 3D face shapes should resemble its initial 3D face shape as much as possible. Method constraints are used to bring deviation as close to 3D face form as possible. As a result, the issue is changed to one of clustering. In addition to providing extremely accurate on-site coordinate data collection, 3D laser scanning is a great way to obtain measurement data for tasks like deformation detection, planning, mapping, data archiving. In terms of traffic safety, routine upkeep and inspection of metropolitan highways and overpasses are crucial duties. The state of natural environment, heavy traffic, sporadic accidents can all affect road bridges. Accurate and timely monitoring of tiny deformations is possible with 3D laser scanning technology. The entire animation creation system consists of both software and hardware components. User first selects the option to choose between an image and a video. following the system's window for local storage opening. System checks for supported image and video formats. After that, the chosen picture or video will be submitted and turned into a cartoon version. Following an image or video upload that is successful, the system processes the input and generates the output. The user can download the image or video by clicking the download button once the cartoonized output has been presented on screen. The outcome of the download will be kept locally. At last, the user has the option to exit the system. • The user can download the image or video by clicking the download button once the cartoonized output has been presented on screen. The outcome of the download will be kept locally. At last, the user has the option to exit the system. • When uploading a video, the file size must be less than or equal to 30 MB. Should the video exceed 15 seconds, it will be cropped to 15 seconds and transformed into a cartoon-style video. The cartoonized video will have audio added to it. We'll consider the first fifteen seconds of the video. The creation of a virtual reality world is significantly influenced by projection as well. The VR scene can be made to appear more realistic and organic by modifying the camera's position, orientation, field of view, perspective projection, and other settings. Additionally, the projection has an impact on how people engage with VR environment. To enhance accuracy of interactive experience, the projection diagram can be utilised to establish the handle's position and orientation in the virtual area while interacting with it.

RESULTS AND DISCUSSION

TensorFlow, an open-source deep learning framework from Google, is used to build the model and run the algorithm. A powerful DL platform, TensorFlow provides a wide range of tools and modules to help create complex and varied DL models. The experimental core consists of algorithmically combining the content image with style image to create a new image with a unique style. First, the content and style photos' respective aspects are separated to separate their content and style attributes. The difference between the content and style attributes is then computed using a designated loss function, which forms the basis for model training and optimisation. Finally, the model achieves the capacity to convert the content image's style into the intended target style, producing a new image with that aesthetic.

Dataset description: First dataset in wild with precise 3D poses for assessment is called 3D Poses in the Wild. There are other outside datasets, however they are all limited to a modest amount of recording volume. First one that use video from a moving phone camera is 3DPW.

Dataset has:

60 video sequences.

2D pose annotations.

3D poses obtained with technique introduced in paper.

Camera poses for every frame in sequences.

3D body scans and 3D people methods. Every sequence contains its corresponding methods.

18 3D methods in various clothing variations.

AnimeRun Dataset: Using computer graphics software, correspondence data can be gathered by creating high-quality frame sequences and capturing the corresponding ground-truth motions. But while there aren't many publicly accessible 2D anime resources, turning 3D films into 2D format offers an alluring substitute for simulating 2D cartoons. This part outlines the design decisions we took during the conversion process, including the technique of ground-truth correspondence label creation and our description of the 2D animation style and rendering parameters.

CelebHeads Dataset: gathered 3D animation models from DevianArt4 and Niconi Solid, two distinct websites. Since the authors of all 3D animation models have copyrights, we carefully verified extent of those rights as well as got permission from writers who could be reached. In the end, we obtained 3613 functional 3D animation models. To give credit to the artists, we will share a list of all 3D animation model artists together with AnimeCeleb. There are two key elements in the collection of 3D animation models. The morphs, which can change a 3D animation methods appearance on a face or other body part, are the initial component. A 3D animation model's look can be altered by varying a single morph's continuous value between [0, 1]; for instance, an animation head with an open mouth in a 0.3 proportion can be produced. The bones that regulate head angles make up the second. Specifically, a rotation matrix is applied to neck bone in order to control the head angles.

Table- 4 comparison for various animation scene construction dataset

Dataset	Techniques	Accuracy	Precision	RMSE	Robustness	F-1 score
3DPW	**GAN**	75	76	78	75	74
	FL	87	81	75	78	79
	CGFEFDM	91	85	70	82	84
AnimeRun Dataset	**GAN**	73	79	69	78	73
	FL	77	85	71	83	77
	CGFEFDM	95	90	68	86	85
CelebHeads Dataset	**GAN**	86	88	81	80	83
	FL	90	94	79	84	88
	CGFEFDM	97	96	75	93	95

Table-4 shows analysis for various sports player health monitoring dataset. Here the sports badminton player monitoring dataset analysed are 3DPW , AnimeRun Dataset and CelebHeads DATASET in terms of Prediction accuracy, Precision, RMSE, robustness, F-1 score.

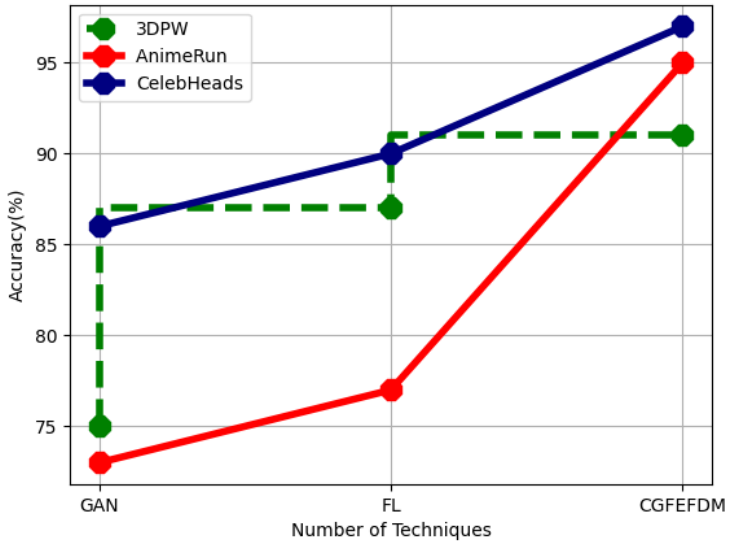

Figure-5 comparison of accuracy

Figure 5 analysis for accuracy is shown. Here proposed technique accuracy of 91%, existing GAN 75%, FL attained 87% for 3DPW dataset; for AnimeRun Dataset proposed technique accuracy of 95%, existing GAN 73%, FL 77%; proposed technique accuracy of 97%, existing GAN 86%, FL 90% for CelebHeads Dataset.

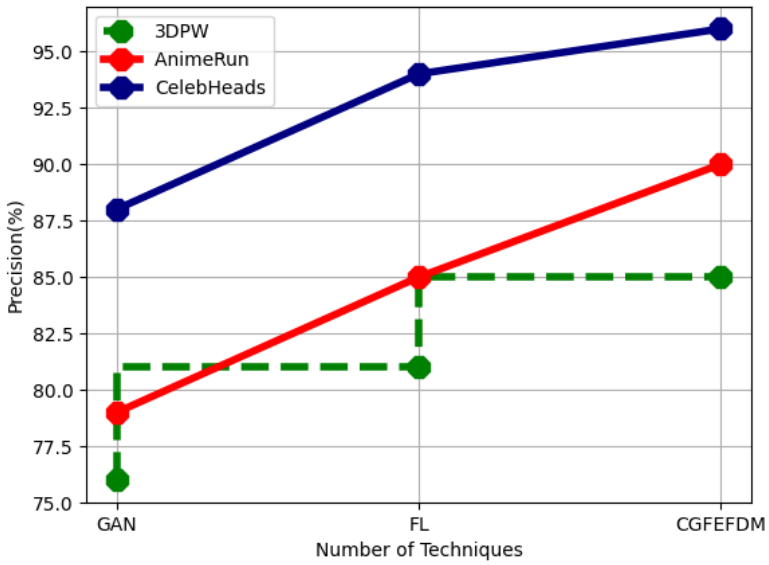

Figure-6 comparison of Precision

Figure 6 shows analysis in Precision. Here proposed technique Precision of 85%, existing GAN attained 76%, FL attained 81% for 3DPW dataset; for AnimeRun Dataset proposed technique Precision of 90%, existing GAN 79%, FL 85%; proposed technique Precision of 96%, existing GAN 88%, FL 94% for CelebHeads Dataset.

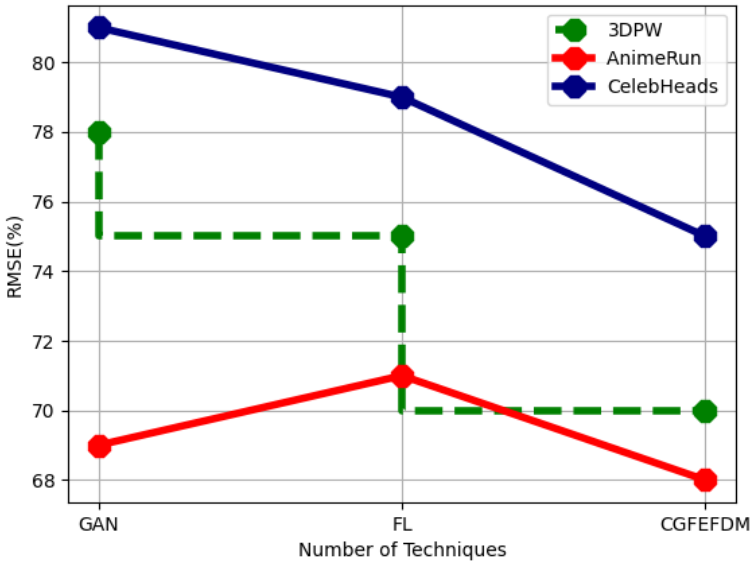

Figure-7 comparison of RMSE

Figure 7 shows analysis in RMSE. Here proposed technique RMSE of 70%, existing GAN 78%, FL attained 75% for 3DPW dataset; for AnimeRun Dataset proposed technique RMSE of 68%, existing GAN 69%, FL 71%; proposed technique RMSE of 75%, existing GAN 81%, FL 79% for CelebHeads Dataset.

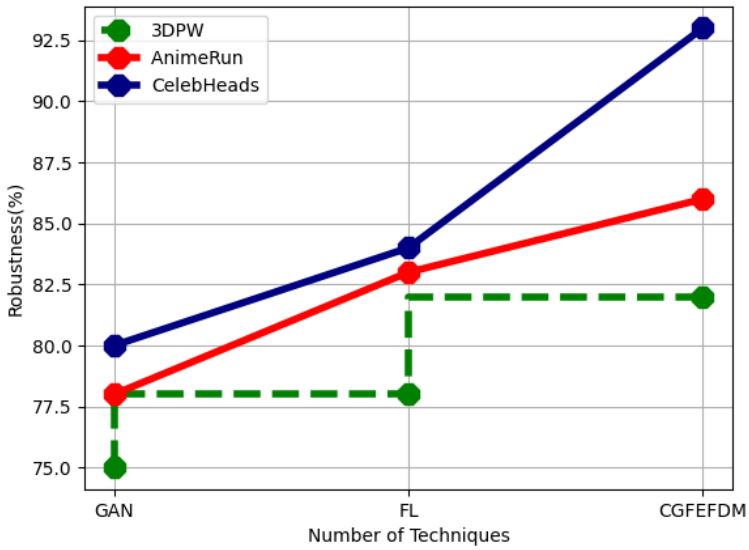

Figure-8 comparison of robustness

Figure 8 analysis for robustness is shown. Here proposed technique robustness of 82%, existing GAN 75%, FL attained 78% for 3DPW dataset; for AnimeRun Dataset proposed technique robustness of 86%, existing GAN 78%, FL 83%; proposed technique robustness of 93%, existing GAN 80%, FL 84% for CelebHeads Dataset.

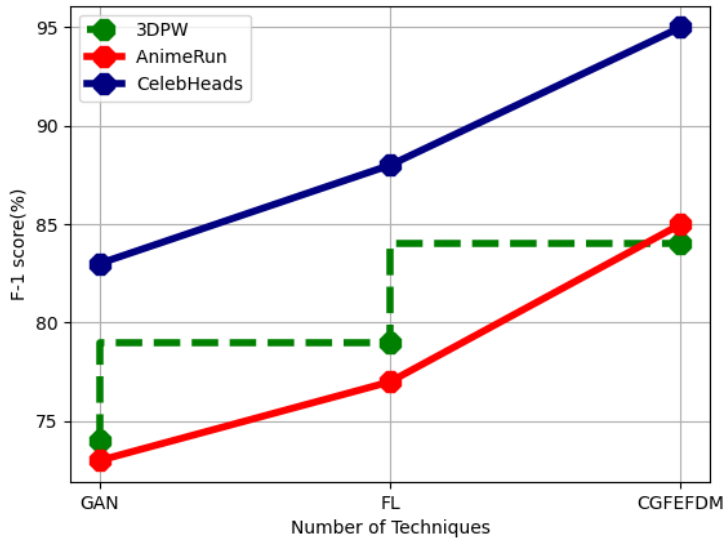

Figure-9 comparison of F-1 score

Figure 9 shows analysis in F-1 score. Here proposed technique F-1 score of 84%, existing GAN 74%, FL 79% for 3DPW dataset; for AnimeRun Dataset proposed technique F-1 score of 85%, existing GAN 73%, FL 77%; proposed technique F-1 score of 95%, existing GAN 83%, FL 88% for CelebHeads Dataset.

When used for front, side, and top views, the CNN error rate is as follows. The discriminator will not be able to trick the bogus data produced by the generator after they have reached an equilibrium point in the training process. This GAN feature comes in very useful while drawing three-dimensional animation. This is due to the fact that when processing three-dimensional data, the conventional CNN approach can only extract a limited number of characteristics from the input data, which may result in an inadequate understand of the complicated three-dimensional structure. But because GAN uses a generator to create images step-by-step, its creation process can be thought of as an interpretation of three-dimensional data. GAN is able to comprehend and use

the information in three-dimensional data more effectively thanks to this layer-by-layer generating approach. The ability of GAN to fully utilise each piece of data is another benefit. Pairs of data are employed in the GAN training step, allowing the model to fully use each data and produce higher outcomes with the same training samples. On the other hand, the conventional CNN technique may not perform as well as GAN as it can only use a portion of the input data. The benefits of GAN are also seen when looking at animation scene pattern optimisation. This is so that different aspects of the image, like object shape, position, illumination, and so forth, can be better understood and optimised. Moreover, the image produced by GAN can serve as a reference for the optimisation of animation scene pattern due to its layer-by-layer generation characteristics.

CONCLUSION

In order to construct animated scenes for films and television shows using artificial intelligence, this research proposes a revolutionary technique in computer-aided systems for virtual reality technology. ML has been utilized to create virtual reality in movie as well as animation scenes. The animation scene modelling process uses a convolutional generating fuzzy encoder face deformable model for 3D graphic creation. The suggested framework integrates spatiotemporal features that are saliency-relevant in order to reasonably infer the saliency of the video. In addition, a comprehensive eye fixation dataset is constructed to more accurately represent the video saliency performance. The experimental results demonstrate that the suggested architecture can compete with the most advanced video and image saliency models due to its high inference speed. It also achieves superior saliency prediction performance in terms of accuracy and computing speed. This method can only be applied to common cameras that are portable and does not place strict limitations on the acquisition apparatus. But this also has an immediate impact on the mapping effect, and it is challenging to generate face features like wrinkles by transplantation. In this sense, resolving the mapping effect of complex faces is an area that needs to be improved if you want to expand the application field of transplanting technology.

Acknowledgement

Science and Technology 2020 Project: Research on the innovative model of virtual studio course construction in applied universities; FBJG20200311, 2020 Provincial education and teaching reform research project "Innovation and Entrepreneurship Education Research and Practice of Animation Major" Information Technology + Chinese Fashion "from the perspective of New Liberal Arts".

REFERENCES

1. Du, N., & Yu, C. (2020, October). Application and research of VR virtual technology in film and television art. In Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education (pp. 108-114).
2. Zhang, M., Zhu, Z., & Tian, Y. (2020). Application research of virtual reality technology in film and television technology. IEEE Access.
3. Li, L., & Li, T. (2022). Animation of virtual medical system under the background of virtual reality technology. Computational Intelligence, 38(1), 88-105.
4. Gao, Z. (2023, March). Application of 3D Virtual Reality Technology in Film and Television Production Under Internet Mode. In The International Conference on Cyber Security Intelligence and Analytics (pp. 341-349). Cham: Springer Nature Switzerland.
5. Wang, Y. (2023). 3D Dynamic Image Modeling Based on Machine Learning in Film and Television Animation. Journal of Multimedia Information System, 10(1), 69-78.
6. Ronfard, R. (2021, May). Film directing for computer games and animation. In Computer Graphics Forum (Vol. 40, No. 2, pp. 713-730).
7. Reddy, V. S., Kathiravan, M., & Reddy, V. L. (2024). Revolutionizing animation: unleashing the power of artificial intelligence for cutting-edge visual effects in films. Soft Computing, 28(1), 749-763.
8. Zhang, J. Q., Xu, X., Shen, Z. M., Huang, Z. H., Zhao, Y., Cao, Y. P., ... & Wang, M. (2021, October). Write-An-Animation: High-level Text-based Animation Editing with Character-Scene Interaction. In Computer Graphics Forum (Vol. 40, No. 7, pp. 217-228).
9. García-Ortega, R. H., García-Sánchez, P., & Merelo-Guervós, J. J. (2020). StarTroper, a film trope rating optimizer using machine learning and evolutionary algorithms. Expert Systems, 37(6), e12525.
10. Xuan, D. (2023). Design of 3d animation color rendering system based on image enhancement algorithm and machine learning. Soft Computing, 1-10.
11. Xu, P., Zhu, Y., & Cai, S. (2022). Innovative research on the visual performance of image two-dimensional animation film based on deep neural network. Neural Computing and Applications, 34(4), 2719-2728.

12. Yuan, H., Lee, J. H., & Zhang, S. (2023). Research on simulation of 3D human animation vision technology based on an enhanced machine learning algorithm. Neural Computing and Applications, 35(6), 4243-4254.
13. Song, W., Zhang, X., Guo, Y., Li, S., Hao, A., & Qin, H. (2023). Automatic Generation of 3D Scene Animation Based on Dynamic Knowledge Graphs and Contextual Encoding. International Journal of Computer Vision, 131(11), 2816-2844.
14. Knight, J., Johnston, A., & Berry, A. (2022, June). Machine Art: Exploring Abstract Human Animation Through Machine Learning Methods. In Proceedings of the 8th International Conference on Movement and Computing (pp. 1-7).
15. Zhou, M. (2024). A proposed reconstruction method of a 3D animation scene based on a fuzzy long and short-term memory algorithm. PeerJ Computer Science, 10, e1864.
16. Tang, J. (2023). Graphic Design of 3D Animation Scenes Based on Deep Learning and Information Security Technology. Journal of ICT Standardization, 11(3), 307-328.