The Application of Artificial Intelligence-based Style Transfer Algorithm to the Design of Animation Special Effects

Xuebiao Niu[1, 2, *]

[1]School of Design Arts, Xiamen University of Technology , Xiamen 361024, Fujian, China
[2]Theory and history of art, Moscow State Stroganov Academy of Design and Applied Arts, Moscow 125080, Russia
18250750817@163.com

Abstract

Style transfer belongs to a new type of research on image tasks in the field of deep learning, which enriches and explores people's imagination of real life. In-depth research on style transfer can deepen the computer's understanding of images, which is conducive to accelerating the development of computers in the field of artificial intelligence. The purpose of this paper is to study the style transfer algorithm applied to the design of animation special effects based on artificial intelligence. First, the application status of animation special effects technology at home and abroad is systematically analyzed, and the importance of studying special effects technology is expounded. Secondly, the types and concepts of special effects technology, as well as the practicability and feasibility of special effects technology in animation production are revealed through objective analysis. The evaluation index of image style transfer results is given, and an experiment of special effects separation effect based on depth-aware special effects migration network is proposed and carried out. Finally, the indicators of each module are compared and analyzed. The experimental results show that the method in this paper can accurately find all the special effects in the scene.

Keywords: *Artificial intelligence, style transfer algorithm, animation special effects, animation design*

1 INTRODUCTION

With the advent of digital technology, virtual reality technology tends to be perfected. Because special effects digital technology can not only display traditional animations with powerful traditional animations, but also create natural and realistic images and graphics, and even combine game characters with real pictures, surpassing traditional special effects [1]. It not only made modern film and television more powerful, but also completely innovated the principles of shooting, editing, genre, printing and so on. Traditional imaging methods can only extract low-level parts from images, but cannot extract high-level parts of images [2]. Therefore, when dealing with images with complex colors and textures, the effect of post-style-transfer images created by traditional style-transfer methods is very sketchy and cannot meet real-world needs [3].

The application of special effects technology in animation will become more and more extensive, which will promote the development of animation production in a faster, better and stronger direction [4]. Iseringhausen J proposed four main models developed in the American animation business field. The various modes are called fusion, zip-crash, functional and poetic authentication. Each is used for a different aesthetic effect, with an ever-changing relationship to the image. The use of sound, music, sound effects and ambience and the way they are recorded, manipulated and mixed are all considered. In addition, the manner in which conventions flow from one period to the next is illustrated. Collectively, these proposed categories help to understand the historical and creative range of options available to animators outside the visual domain [5]. Purwaningsih D A uses observational and experimental methods. Observations were made by studying existing paper puppets from other kirigami shorts, tutorials, commercials and behind-the-scenes videos from filmmakers. The observations will be followed by an experiment that will use the data to explore paper puppet making, compare the pros and cons, and adapt the final method to suit the needs of a short animation project called "Spay & Neuter", in which the final designed paper puppet cat puppet will be applied [6].

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 1410-1415, 2023.
https://doi.org/10.2991/978-94-6463-040-4_210

Realizing the migration of special effects in the process of animation post-production can bring economic benefits to the animation industry that cannot be underestimated [7].

This paper studies and discusses the practical application of special effects technology in 3D animation projects. The development of computer graphics and the basic structure, working principle and classical model of convolutional neural network for deep learning are introduced. And build the network framework of the style transfer algorithm in this paper. The effectiveness of the proposed algorithm is verified by the style transfer evaluation index.

2 RESEARCH ON THE APPLICATION OF ARTIFICIAL INTELLIGENCE-BASED STYLE TRANSFER ALGORITHM TO ANIMATION SPECIAL EFFECTS DESIGN

2.1 *Computer Graphics*

Computer graphics is an important research branch and application direction of computer science. It is a science that uses computers for visual representation and can simulate two-dimensional and three-dimensional graphics. Computer graphics needs to use a variety of mathematical algorithms, physical laws, cognitive principles, etc., in order to convert two-dimensional and three-dimensional graphics into grids of computer displays [8].

The development of computer graphics research enables the function of the human right brain to be simulated in the computer, and realizes the computer from abstract symbol processing to figurative image representation. The right brain, known as the "image brain", is good at processing image information, and the visual representation of graphics is the focus of computer graphics. At present, its application fields are becoming more and more extensive, from 3D graphics modeling to model drawing, from static models to animation creation.

2.2 *Deep Learning*

The ultimate goal of deep learning is to enable computers to imitate human analysis and learning behaviors, so that computers can analyze and study data on the basis of truly understanding data such as text, images, and speech. Deep learning is subordinate to machine learning algorithms in the field of algorithms, but its outstanding capabilities in the research direction of image and speech have greatly surpassed previous related research.

Convolutional Neural Networks are one of the most successful application areas of deep learning algorithms. Convolutional neural network has strong image representation learning ability, can translate incoming information layer by layer according to its hierarchical structure, and has strong advantages in the fields of image classification, object recognition and image processing. A convergent neural network consists of three parts: an input layer, a hidden layer, and an output layer. Among them, the hidden layer can be divided into cohesive layer, concentrated layer and fully adhesive layer. In the lower grid layers of the network, the convergence and concentration layers work alternately, performing the function of outputting input image features and reducing parameters.

2.3 *Depth Perception Special Effects Mobile Network Framework*

The overall network structure of this paper consists of two parts: the separation module and the migration module, in which the focus and difficulty are the separation of the background and the special effects. The separation module consists of four levels, namely special effect category perception, special effect location perception, receptive field expansion and network cascade, which respectively deal with different challenges in special effect separation. This paper proposes a depth-sensing special effect migration network, which can extract its special effects from the input special effect image and migrate it to the input background image. First of all, according to the different semantics of special effects, this paper uses the special effect category-aware attention module to detect different special effects at the semantic level; for the different special effects direction, this paper uses the special effect position-aware attention module to detect the direction and position of special effects in space; The difference in the receptive field of special effects In this paper, atrous convolution is used on the basic network without downsampling to increase the receptive field of the network; for the difference in the transparency of special effects, this paper cascades multiple networks to increase the network depth to separate low transparency or no transparency effects.

In this paper, the downsampling step in traditional convolutional networks is abandoned, and atrous convolution is used, which solves the problems of information loss and limited receptive field well. By adjusting the size of the convolution kernel hole, the regulation of receptive fields of different sizes can be realized. The network convolution kernel in this paper is fixed at 3X3, and the hole convolution pores increase exponentially with the depth of the network at the base of 2, namely:

$$dilation = 2Adepth \tag{1}$$

A loop separation strategy is used to separate backgrounds and special effects.

The network is updated by taking the input of each cascaded network and the desired special effect label as the least square difference to calculate the loss. This cascades the network to remove the effects of varying opacities without damaging the background image too much. Finally, the separation loss function of this paper is defined as:

$$L_D = \sum_{k=0}^{n} (\hat{B}_k - B)^2 \tag{2}$$

S is the target background image.

3 INVESTIGATION AND RESEARCH ON THE APPLICATION OF ARTIFICIAL INTELLIGENCE-BASED STYLE TRANSFER ALGORITHM TO ANIMATION SPECIAL EFFECTS DESIGN

3.1 Experimental Setup

This article collects 20 videos with different backgrounds and 5 different special effects through the Internet, and each special effect is divided into five different levels of small, medium, large, sparse, and dense, with a total of 60 minutes of video. This paper combines different backgrounds and special effects through AE to synthesize the training data of this paper. Then this article uses OpenCV's video slice tool to slice the synthetic video, background video and special effect video into corresponding frame-by-frame pictures. At the same time, in order to avoid the redundancy of similar pictures, this paper cuts one picture every 2 frames, and then uses these pictures as the training data set of this paper. During this period, every 50 images will be used as the test set of this paper. Finally, this paper obtained a total of 1355 pairs of training pictures and 57 pairs of test pictures.

This article runs the experiments in this article on a server with CPUE5-2620v4 and graphics card GeForceGTX1080Ti. The data augmentation operations in this paper include: first, the input 1280*720 pixel image is first adjusted to 256*256, then randomly cropped to 224*224, then randomly flipped, and finally input to the network in this paper after subtracting the mean. This paper sets the batch size to 16 and the training epoch to 500, starts training from scratch, and tests all our results after the 500th epoch. The learning rate in this paper is set to 5xl0_3 and the network is updated with stochastic gradient descent (SGD).

3.2 Style Transfer Evaluation Metrics

In this paper, the peak signal-to-noise ratio (PSNR) is used to evaluate the content gap between a content image and the stylized image. The larger the PSNR, the clearer the content of the style-transferred image and can be preserved to the greatest extent content information. Given a content image I of size m×n and a stylized image k, PSNR (in dB) is defined as:

$$RSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{3}$$

The Structural Similarity (SSIM) index evaluates the similarity between the stylized image and the style image. The larger the SSIM, the smaller the difference between the two images, indicating that the effect of the style transfer of the image is good. Given two graphs x and y, the structural similarity of the two images is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4}$$

In formula (4), μxy is the mean value of x, μy is the mean value of y, σx is the variance of x, σ is the variance of y, σxy is the covariance of x and y, C1=(k1L), C2=(k2L) is a constant to maintain stability, L is the dynamic range of pixel values, k1=0.01, kz=0.03.

4 ANALYSIS AND RESEARCH OF ARTIFICIAL INTELLIGENCE-BASED STYLE TRANSFER ALGORITHM APPLIED TO ANIMATION SPECIAL EFFECTS DESIGN

4.1 Special Effects Location Awareness

Because the neural network is very easily affected by position and direction, this paper needs a reliable position and direction prediction module to assist the network to find the position and orientation of the special effect, as shown in Figure 1.

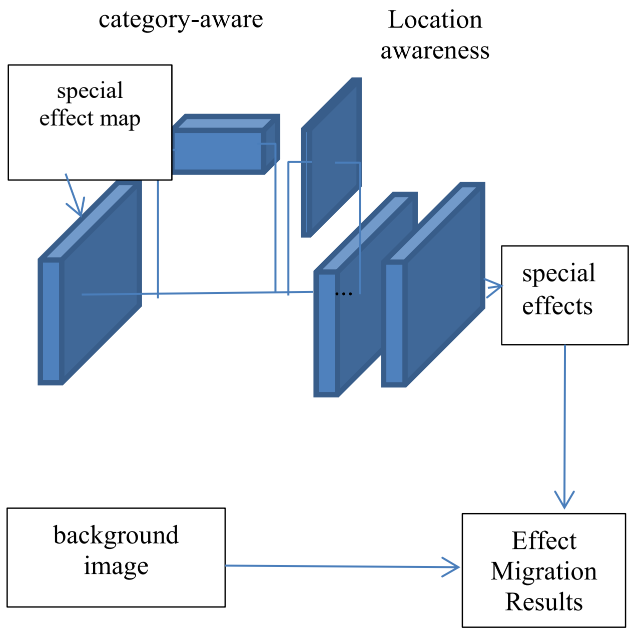

Figure 1. Network structure diagram with location awareness module added

This magnitude difference at different positions can be used as the input of an attention detection mechanism, which can be used as a position detector for special effects through simple calculations. Moreover, in order to ensure that the input is not distorted, the network in this paper does not downsample the image like the traditional direct-connected network, so it will hardly damage the input location information.

In this paper, maximum pooling and average pooling are performed on the feature maps in the channel dimension, respectively, and two feature maps with different perspective information are obtained. In this paper, the two feature maps are cascaded and input into a convolutional layer of a 7X7 convolution kernel to aggregate information from different perspectives into a feature map. Accurate judgment. The obtained feature map may have some large values that are inconvenient to calculate, so this paper needs to add a sigmoid function to limit the value to the range of 0~1 to obtain the final position-aware attention map of this paper. The feature map can capture different positions and orientations of different effects in the picture.

4.2 *Special Effect Separation*

This paper first uses four indicators to measure the effect of special effects separation from four different aspects. First, this paper uses PSNR peak signal-to-noise ratio to judge the degree of information loss between the separated special effects and the special effects marked in this paper; then this paper uses sLMSE to calculate the least squares distance of each local corresponding area of the two images to evaluate the local structure. Similarity; this paper then uses NCC to evaluate local structural similarity, but its limitation is that it is not sensitive to differences in overall intensity; this paper uses SSIM later to evaluate the global similarity of two images in terms of brightness, contrast and structure, and the results are as follows shown in Table 1.

Table 1. Verify the contribution of each module to the separation effect

module combination	PSNR	sLMSE	SSIM	NCC
Base	46	52	57	48
Basic + Category	53	58	66	51
base + category + location	78	82	91	79
base + category + location + cascade	91	95	93	90

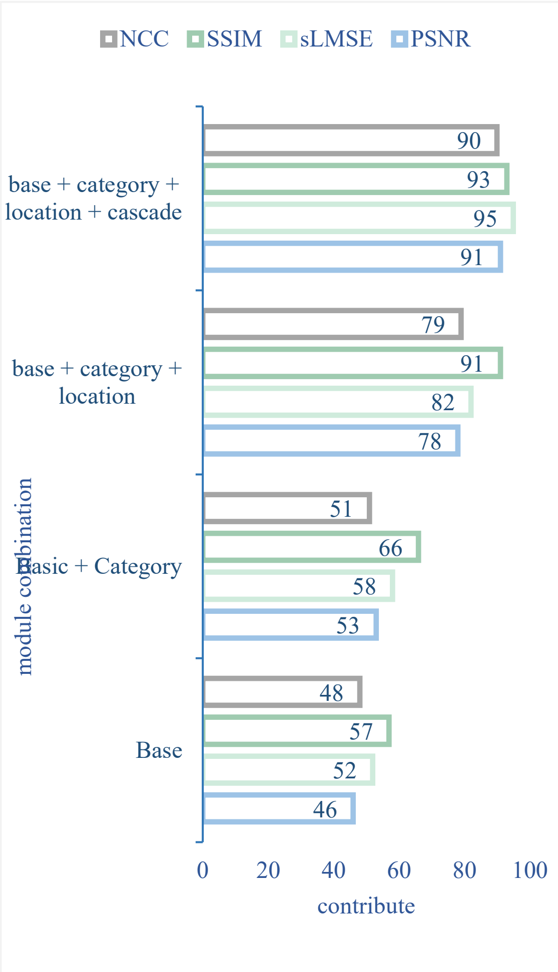

Figure 2. Special effect separation effect

This paper compares its three important modules respectively. First of all, this paper first adds a category-aware module to the basic hollow convolutional network. It can be seen that each index has increased, indicating that the structural similarity between the separated special effects and the labeling effects has been improved. Then this paper adds a location-aware module to the basic network, and each index has been significantly improved, especially the SSM and sLMSE have increased by 24 and 25 percentage points respectively compared with the basic network, indicating that the location-aware module is compared with the category-aware module. play a larger role, as shown in Figure 2. Finally, this paper combines all modules to form the final network structure of this paper. The experimental results show that in this case, whether it is global structural similarity or local structural similarity, or brightness and contrast, the method in this paper The index reaches the optimal situation, indicating that the method in this paper can accurately find all the special effects in the scene.

5 CONCLUSIONS

Due to the rise of deep learning, style transfer technology has also taken a major step forward. Compared with traditional methods, deep learning has become very simple in the processing of feature engineering. It no longer requires people to design data features based on a lot of experience and spend a lot of time, but automatically extract and train targets through the characteristics of neural networks. This paper introduces the related technologies used in this paper from the fields of traditional digital image processing and deep learning image processing. Among them, computer graphics is introduced in detail in traditional digital image processing; in deep learning image processing, the content and contributions of convolutional neural networks are mainly introduced, and depth perception special effects migration network is proposed. Subsequent work will focus on the following aspects: optimizing the network model to speed up network training and reduce the cost of learning. We can further improve the optimization algorithm to better eliminate the problems of incomplete preservation of image content details and distortion of stylized image information.

REFERENCES

[1] Bradshaw J. (2019). FMX 2019: Conference on Animation, Effects, Games and Immersive Media; 26th Stuttgart International Festival of Animated Film[J]. Afterimage, 46(3):1-8.

[2] Doan E M . (2019). The Role of Animation Cartoons in Value Education: Ice Age Example[J]. Social Sciences Studies Journal, 5(49):6142-6153.

[3] Evangelidis K, Papadopoulos T, Papatheodorou K, et al. 3D geospatial visualizations: Animation and motion effects on spatial objects [J]. Computers & Geosciences, 2018, 111(feb.):200-212.

[4] Hiraoka T, Hirota M. (2018). Generation of cell-like color animation by inverse iris filter [J]. ICIC Express Letters, 12(1):23-28.

[5] Iseringhausen J, Weinmann M, Huang W. (2020). Computational Parquetry: Fabricated Style Transfer with Wood Pixels [J]. ACM Transactions on Graphics, 39(2):1-14.

[6] Purwaningsih D A. (2020). JOINT AND SEGMENTATION DESIGN ON PAPER PUPPETS FOR CAT CHARACTERS IN CUT OUT STOP MOTION ANIMATION[J]. ULTIMART Jurnal Komunikasi Visual, 12(2):1-8.

[7] Suparmaniam C, Yatim M. (2020). EFFECTIVENESS OF INDIAN FOLKLORE ANIMATION ON MOTIVATION AMONG

YEAR FIVE PUPILS IN KULIM DISTRICT[J]. International Journal of Modern Education, 2(7):01-12.

[8] Taberham P. (2018). A General Aesthetics of American Animation Sound Design[J]. Animation, 13(2):131-147.