Research on the Application of Artificial Intelligence in the Film Industry

Yaxing Li[1], [*]

[1]Fuzhou University, China, 350000

ABSTRACT: With the rapid development of technology today, artificial intelligence is inevitably involved in numerous fields including the film production industry, and helps the industry from all periods of the production cycle. This paper focuses on three areas: script writing, special effects production and video restoration, and highlights the application of technologies such as Benjamin, flux system and ESRGAN in these three areas. By analyzing the positive impacts and potential risks, this paper points out that AI technology and human artistic creation are indispensable for the healthy development of the film industry only when they complement to each other, hoping to offer some references for the future research of the the application of artificial intelligence in the film industry.

1. INTRODUCTION

Modern computer science has very large branches, and artificial intelligence, one of the most emerging technologies, is gaining more and more attention from every aspect. As a technical science that studies, develops and simulates human intelligence, artificial intelligence has gained widespread application and achieved fruitful results in many disciplines. Films has been a fusion of art, culture, entertainment and high technologies since its inception, and the combination of the two is inevitable due to films' high exposure to the public, strong market influence and demand for high technologies, which provides a vast scope for the practical application of AI in the film and television industry. Currently, AI is capable of participating in the pre-production budget estimation, script writing, casting, being actors, special effects production, post-production editing, image restoration and publicity development of films, but AI is not yet capable of doing such complex tasks as directing.

This paper delves into how AI is being used in film-making, focusing on how the new technological revolution of AI is impacting the film industry. Film-making is divided into pre-production scripting, mid-production shooting, post-production special effects editing and picture restoration. This paper is also divided into three parts to introduce how AI is involved in script writing, how AI is used in filming and special effects production, and how AI is used in picture restoration. This paper focuses AI on the film production industry in both two sides to consider the positive impacts and potential risks of AI, and analyses the possible future of artificial intelligence in the film-making industry, so as to offer some references for the researches of AI in in film industry.

2. AI IN FILM INDUSTRY

2.1. Script Writing

In the pre-production phase of a film, the most important and complex aspect is script writing. Scripts are the lifeblood of films, and their excellence or not almost determine success or failure. On the one hand, the script needs to be innovative to attract the audience's attention. Stories that fall into cliché are often not excellent, and in some cases may even be caught up in legal disputes over plagiarism. On the other hand, the script needs to be written with both narrative and emotional expression in mind, requiring the creator a keen insight into life and strong literary skills that can be able to distil the issues in life and incorporate them into the plot in an ingenious way [1]. Both of these are very difficult to achieve and often require a great deal of effort and time on the part of the creator. Artificial intelligence can solve most of these problems. It can access the vast amount of information available on the internet and analyse it to select suitable stories and references, and can compare them with existing works in databases to avoid duplication. At the same time, AI can complete a script at a much faster rate than a human scriptwriter.

Script-writing AI already exists, in 2016 AI expert Andy Herd developed an AI automated script-writing software via Google's open source machine learning toolkit tensorflow, into which he fed the entire script of *Friends* into the program that collected and analyzed it and automatically generated a new episode [2]. But in some parts the logic was still a bit confusing and poorly worded. In June of the same year, the sci-fi London 48 hour challenge, Benjamin, an AI script-writing program created

---

*Corresponding author. Email: YAXING.LEE@outlook.com

by a team led by Oscar Sharp and Ross Goodwin, emerged as a top ten finalist with his script *sun-sprint* [3]. And two years later it won a special mention at the festival for his *zone out*. This paper will present Benjamin's algorithm.

Benjamin used a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM). To train Benjamin, Goodwin fed the AI dozens of science fiction movie scripts he found online, including Star Trek, Truman's World, X-Men, and more. After the input was completed, Benjamin used the algorithm to break down the films in great detail, learning and predicting which letters tended to be linked together and which words and phrases tended to appear together. Over time, Benjamin learns to imitate the structure of the scripts, create stage directions and generate well-formed lines, and quickly generates scripts by following what it has learned. RNN compares to traditional neural networks, has memory capability. It carries a loop on the model that points to itself and can be used to pass information processed in the current moment for use in the next moment, rather than focusing only on the current. But general RNN cannot solve the long-time dependency problem, which is fatal when it comes to learning script writing. LSTM algorithm however, as a special kind of RNN can solve this problem very well. Its centre is a four-layer structure. The first layer is used to decide what information can be passed through

the cell state. This part is used to selectively filter information that is not currently needed in the process of learning the script. The second and third layers are used to update new information. The second layer is an input gate that determines which information is used to update by sigmoid function. The third layer is a tanh function that generates new candidate values which are added together to obtain the final candidate values. The fourth layer is used to determine the output of the model, first by using the sigmoid function to obtain an initial output, then by using the tanh function to scale the values to between -1 and 1, then by multiplying them pair by pair with the output obtained from the sigmoid to obtain the output of the model, which is the script. The output of the sigmoid function does not consider the information learned at the previous moment. The tanh function is a compression process for the previously learned information, which plays a role in stabilizing the value. The combination of the two can realize a long-term dependent recurrent neural network [4]. The advantage of the LSTM algorithm over Markov chains is that it can sample much longer strings of letters, so it can predict whole paragraphs rather than just a few words. The LSTM is also good at generating raw sentences rather than cutting and pasting sentences from a corpus.
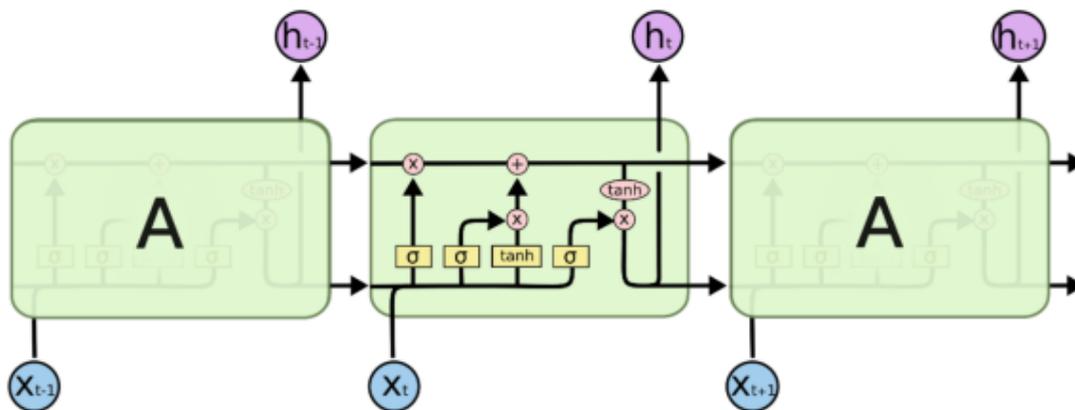

Figure 1. The repeating module in an LSTM contains four interacting layers [4]

The creative speed of artificial intelligence screenwriters is completely unattainable for humans. Even mediocre scripts can take months or even years to write, whereas AI can reduce the time to a few days, which can certainly speed up the production process. And with its vast database, AI can create new ideas that creators may not have thought of. At the same time, there are still some inevitable flaws in AI script-writing, such as confusing language that does not make sense, absurd plots that don't have the depth to stand up to scrutiny. Human scriptwriters have complex and realistic emotions, decades of literary skills, unique and versatile thinking outside the box, and a cultural spirit rooted in tradition, all of which AI does not currently possess. This is why there is still a very large gap between the scripts written by AI and those polished over the years by human. However, at present, human scriptwriters can use AI to open up their minds and inspire them, while at the same time adding the polish of the

human mind to create better stories in less time, which may be a better option now to cooperate with AI.

2.2. Special Effects

The most important aspect of post-production film-making is shooting and production of special effects, where the most high-tech technology is required. One of the most challenging aspects of special effects, ageing and de-aging techniques, have been around since the 1930s and have evolved with time and technology. With the development of computer technology, films have become progressively digital and computers have the ability to do all sorts of incredible special effects. Digital image enhancement techniques achieves aging in the picture at the pixel levels by changing the pixel values, but overuse can lead to the loss of facial details and changes in the colour of the film. The digital field team in *The Curious*

Case of Benjamin Button used the mova contour filming system to create a 3D database of the protagonist's facial expressions by setting up two arrays of cameras in a light-enclosed room. Use fluorescent make-up on the protagonist's face to record the changes in appearance and shape of his face under multiple expressions at the same time. The production team then created high-resolution 3D models of the protagonist at different ages, and finally used the AI to control the data from the 3D facial expression database to drive the head models to age [5]. The result is very realistic. In films such as *Gemini Man* and *Blade Runner 2049*, the production team used motion capture technology to track and record the protagonist's face data and then reconstruct the 3D face in a computer, before modifying the data to achieve age reduction. This method is very accurate and flexible but requires high qualities of hardware and a lot of money.

Martin Scorsese's film *The Irishman* took the age-reversal technique to a higher level. Instead of using marker dots or fluorescence to capture the face, the team captured the face by parsing the light texture of the frame. For this purpose, *Industrial Light and Magic* created the *Three-Headed Monster* camera and *Flux* system [6]. The *Three-Headed Monster* camera consists of a main camera and two infrared cameras that record the actors' faces from multiple angles with no shadows and a higher density of spatial stereoscopic information. After filming, the data is transmitted to the F*lux* system for processing. They also use the *Medusa* system to record the actor's expressions in a given environment and use light stage to capture the skin texture, pores and other features before refining the models according to age requirements to form the final 3D face model of the digital body asset library, which is then transferred to the *Flux* system for processing. The team also spent nearly two years creating and training an artificially intelligent called *face finder*, containing thousands of screenshots of actors from this film at all ages of their careers. The *Flux* system receives the screenshots from the T*hree-Headed Monster* camera and then deforms them to fit the actor's face to obtain a mesh model of the actor's face. This is followed by finding the corresponding expressions from *Medusa*'s library of digital double assets to obtain a 3D facial model of the character. The *Flux* system then calculates the light and texture information from the infra-red data which the *Three-Headed Monster camera* captured and parses the actor's face model to add more facial details. This is how the generator works in *Flux*. The generated image then goes into the discriminator to be calibrated with the age-appropriate screenshots of the actor in the *face finder* to see if the light angles and facial details are consistent. If they are not almost the same, they will be passed back to the generator for further adjustment, and if they are almost the same, they will be rendered for composite output.
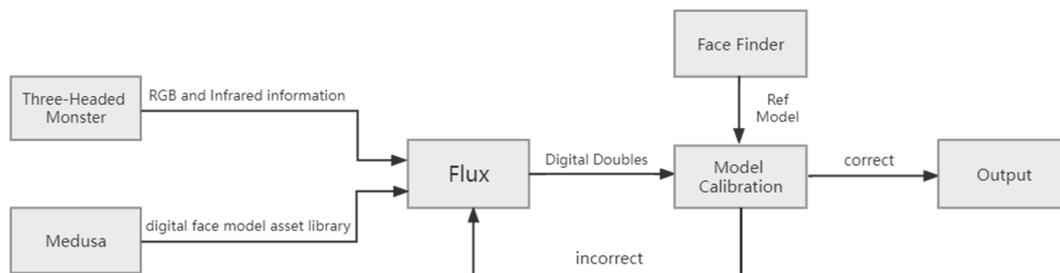

Figure 2. The process of Flux system

This system has distinct advantages over traditional face capture and digital enhancement, as for the first time it allows the actor to be more comfortable and natural during their performances by getting rid of the special helmet and facial tracking points, while at the same time restoring as much light texture and facial detail as possible, so that the face is not stiff and more realistic. However, there are obvious limitations to this technology. The virtual assets of the entire software are almost impossible to be reused in other films, and all the development and training of the models takes a lot of money and time. Even if facial de-ageing is achieved, they still cannot change the fact that they look old due to their old-man gestures. That is why there is still room for the development of special effects for ageing, and researchers need to work on more convenient and appropriate techniques.

2.3. Video restoration

In the past, the main form of movie storage was films. As time passes, film stock tends to suffer from various problems such as dirt, colour loss, scratches, shaking, blurring, flickering, etc. Coupled with damage caused by use and handling, a large number of film stocks have resulted in serious damage. And as one of the most important cultural carriers, film restoration was created to reappear the classics. Digital film restoration is the process of transferring film to tape and then storing it as serial frame files, which are then imported into a computer to digitally restore the image. After restoration, the film can be restored to a resolution of up to 4K, with smoother detail and more natural lighting.

The main technique used for film image restoration is image super-resolution restoration. This method is divided into three types, namely interpolation-based, reconstruction-based and learning-based. The interpolation-based method involves filling in the empty shortcomings of a magnified image with the corresponding pixel values, thereby restoring the image content and achieving an increase in image resolution [7]. Interpolation-based methods are simple to implement and have been widely used, but the linearity of these models limits the detail of their ability to recover high frequencies.

Reconstruction-based methods involve the process of aligning multiple low-resolution images of the same scene with sub-pixel accuracy in space, obtaining the motion offsets between the high- and low-resolution images and constructing the spatial motion parameters in the observation model to obtain a high-resolution image, but these methods are computationally complex and require significant computational resources.

With the advancement of technology and the development of deep learning, learning-based super-resolution restoration is gradually being used. The learning-based super-resolution approach refers to the direct learning of the end-to-end mapping function from low-resolution images to high-resolution images through neural networks, using the prior knowledge acquired by the model to obtain high-frequency details of the image, thus obtaining better image recovery results. One of the most classic is the Super-Resolution Convolutional Neural Networks (SRCNN), which first used CNN in Single-Image-Super-Resolution (SISR) and used only three layers of the network. In the learning phase, a painting is divided into a high-resolution version and a low-resolution version, and the first convolutional layer in the CNN network extracts the low-resolution version of the feature maps. The second convolutional layer transforms the feature maps into a high-resolution image, and the last layer reconstructs it into a high-resolution image, which is then compared with the high-resolution version, then updated and iterated for the next learning [8]. Similar to the traditional method, it actually uses low-resolution images to fill in the gaps, thus improving the resolution and quality. The simplicity of the structure is its biggest advantage, but as it has only one convolutional layer to extract features, there is not enough detail. And the image can appear unrealistic at magnifications of 4 and above.
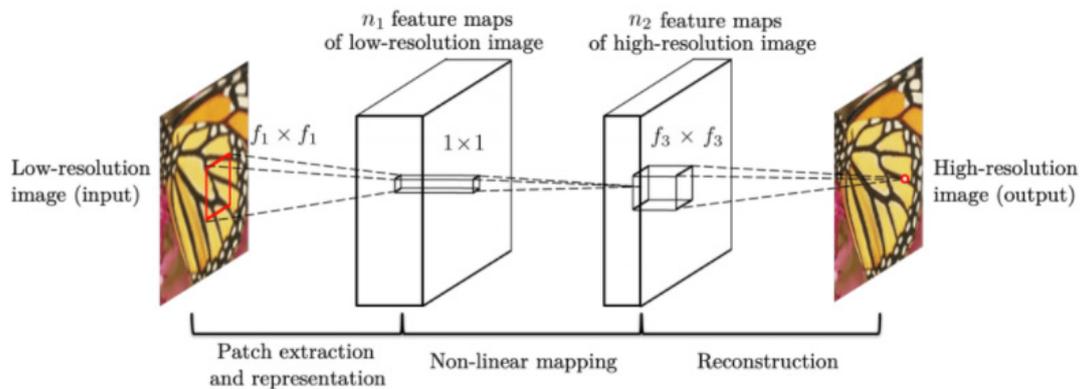

Figure 3. Network structure of SRCNN [8]

The most popular application is Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), which is based on Super-Resolution Generative Adversarial Networks (SRGAN), the principle of which is to take a low-index image as the noise input and fit the noise probability distribution space as closely as possible to the distribution space of the real image through the generator. The generator also uses the incomplete block, which effectively ensures that the gradient information can be effectively transmitted and enhances the robustness of the generative adversarial network. SRGAN is the first framework to support a realistic image at 4x magnification, thanks to the concept of perceptual loss function, which consists of an adversarial loss and a content loss [9]. The adversarial loss uses a discriminator network to determine the difference between the authenticity of the output image and the original image. The content loss is driven by perceptual similarity, rather than pixel space similarity. The introduction of the perceptual loss function allows SRGAN to generate realistic textures for a single image to complement the lost detail when performing super-resolution restoration of the image.
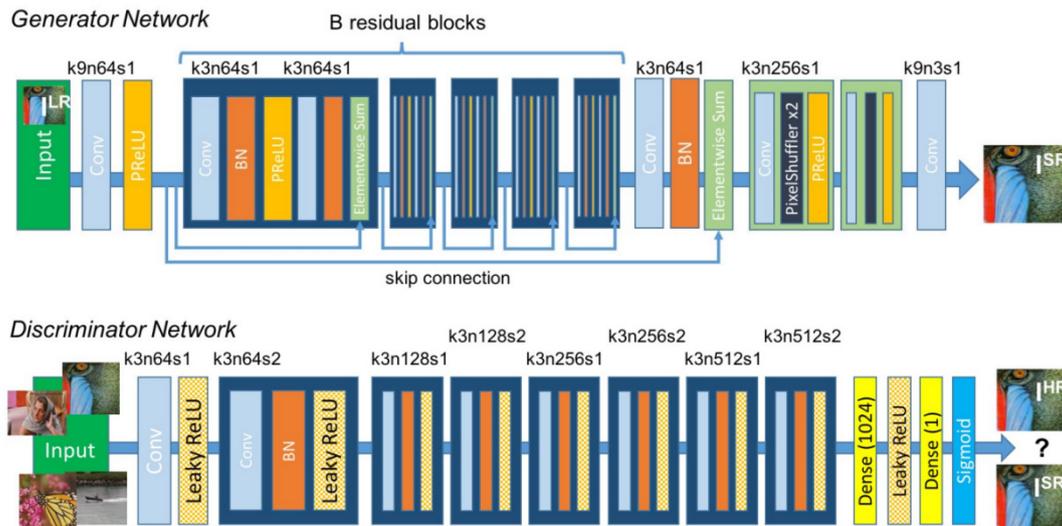
Figure 4. Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer [9]

ESRGAN takes one step further by improving the network structure, perceptual loss and adversarial loss. It introduces the larger and easier-to-train Residual-in-Residual Dense Block (RRDB) to improve the network structure, removes the Batch Normalization (BN) layer and uses residual scaling and smaller initialization to improve training of deep networks. It also uses the RaGAN to improve the discriminator which predicts the relative realism between the high-resolution image and the original image instead of the absolute value, thus allowing the generator to recover more realistic texture details of the original image. The perceptual loss is improved by changing the VGG features in SRGAN from being executed after the activation to before the activation [10]. This enhances edge sharpness and texture realism in output images. The improved perceptual loss provides more visually appealing results. That's why it is now widely used.
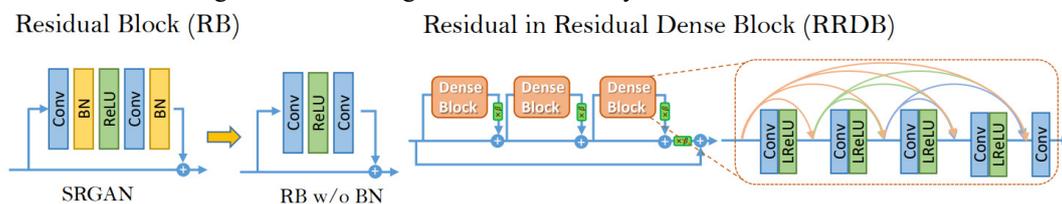

Figure 5. Removing the BN layer and using RRDB block [10]

There are many advantages of using artificial intelligence in digital film restoration. The restoration of films that used to take a lot of time and equipment can now be done directly by artificial intelligence, freeing up human and material resources and increasing efficiency. In addition, existing deep learning restoration algorithms have far surpassed the recognition and generation capabilities of the original algorithms, allowing for greater resolution and more detail in the image. At the same time, however, existing AI film restoration has certain drawbacks. It cannot follow artistic considerations, fail to accurately restore the historical texture of the period. Film restoration still requires the appropriate amount of noise and style to be retained by hand.

3. CONCLUSION

This paper details the application of artificial intelligence in the film production industry in three areas: script writing, special effects production and video restoration. At the time when technology is in full swing, artificial intelligence has been integrated into the film and television industry and benefits to the industry. Whether in each area, AI can make a difference, dramatically reducing the film production cycle and making a significant contribution to the special effects images of a film. The fusion of technology and art is inevitable nowadays, and it is important that AI technology is used in a way that complements technology and art to create better works for the film industry. Artificial intelligence provides more ways of thinking and technical support for the development of films, but they still need to be supported by more subconsciously produced creative content, more moving complex and realistic emotions and certain artistic attainments, one without the other. Only the combination of the two can lead to the healthy development of the film-making industry. This paper offers some references for the research of AI in the film industry.

AUTHORS' CONTRIBUTIONS

This paper is dependently completed by Yaxing Li.

ACKNOWLEDGEMENTS

REFERENCES

1. G. Wu, A study on the function of artificial intelligence machine learning to assist script creativity [J]. Arts Management, 2020, (02):57-63.

2. T. Sun, Application of artificial intelligence in the field of film and television media [J]. Journal of Journalism Research, 2020,11(21):253-254.

3. Y. Chai, The opportunities and challenges of artificial intelligence for film and television industry [J]. Modern Film Technology, 2020(10):51-55.

4. C. Olah, Understanding lstm networks, 2015.

5. H. Mo, J. Zhao, The application of face aging and reverse aging technology in film special effects production [J]. Modern Film Technology, 2021(2):9-14.

6. D. Dong, B. Li, The development and application of age-defying visual effects technology in The Irishman [J]. Modern Film Technology, 2021(6):18-24.

7. X. Bai, An introduction to the application of artificial intelligence in film restoration [J]. Modern Film Technology, 2020(5):52-55. DOI:10.3969/j.issn.1673-3215.2020.05.010.

8. C. Dong, C. C. Loy, K. He and X. Tang, Image Super-Resolution Using Deep Convolutional Networks, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, Feb. 1, 2016, doi: 10.1109/TPAMI.2015.2439281.

9. C. Ledig, L. Theis, F. Huszár, et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681-4690.

10. X. Wang, K. Yu, S. Wu, et al. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp. 0-0.