Journal of Computers Vol. 36 No. 1, February 2025, pp. 191-204
https://doi.org/10.63367/199115992025023601013

Application Methods of Deep Learning in Enhancing Visual Effects in Film and Television Post-production

Hao-Wen Wang*

Tangshan Polytechnic University,
Tangshan City 063299, Hebei Province, China
weiguanjvzu@126.com

Received 24 January 2025; Revised 4 February 2025; Accepted 11 February 2025

Abstract. Post-production of film and television is an important link related to the quality of film and television works. In response to the traditional method of relying on manual post-production of film and television works, this article uses artificial intelligence to improve visual enhancement and video editing of film and television works. Firstly, this article analyzes the current status of post-production in film and television. Three consecutive frames of images are selected from the video sequence image in film and television post-production, and the differential image of adjacent two frames is determined. Logical operations are performed on the two differential binary images to automatically obtain the ROI within the scene of the film and television post-production video image. Then, in order to achieve high-quality video restoration while automatically repairing the computational complexity during the video restoration process, a method of optical flow propagation based on global matching and Transformer encoder is proposed to effectively improve the accuracy of optical flow restoration in videos. Finally, this article takes a certain video clip from daily promotional activities as the experimental object, and uses the artificial intelligence post-production method proposed in this article to improve the visual effect of film and television works.

Keywords: AEPR, deep learning, video enhancement, color adjust

1 Introduction

In the process of film and animation production, post-production is undoubtedly a crucial link. It not only concerns the visual beauty of the final presentation of the work, but also deeply affects the smoothness of the story narrative, the immersive experience of the audience, and the artistic quality and market response of the entire work. In recent years, with the continuous development and improvement of digital media technology, its application in post-production of film and animation has become increasingly widespread and in-depth. The integration of digital media technology with traditional art in film and animation not only brings new challenges to artistic creation, but also greatly improves the quality and viewing experience of film and animation production. This article mainly explores the application of digital media technology in post-production of film and animation, aiming to reveal the impact of digital media technology on post-production of film and animation through an overview of the development and main functions of digital media technology, in order to promote the deep integration and application of digital media technology and post-production of film and animation [1].

Post-production is an indispensable part of the film and television production process, and its necessity is reflected in multiple aspects. Firstly, it is able to edit and integrate scattered materials from previous filming, ensuring the narrative coherence and logic of the film, allowing the audience to clearly understand the plot. Secondly, through special effects processing, color grading and other means, post-production can enhance the visual beauty of the film, create visual effects that match the plot atmosphere, and enhance the audience's viewing experience. At the same time, the addition of elements such as dubbing and music can better convey the emotional connotation of the film and resonate with the audience. In addition, post-production can also make necessary revisions and optimizations to the film, ensuring that it meets review standards and regulations, and improving the overall

* Corresponding Author

quality of the film. In short, post-production in film and television is an important means to enhance the artistic value of film and television works and enhance the audience experience. It plays a crucial role in the birth of a successful film and television work.

However, the current post-production of film and television has the following technical shortcomings:

1) The challenge of technological updates and iteration speed: With the rapid development of film and television post-production technology, new software, special effects algorithms, and production concepts continue to emerge, requiring post-production personnel to constantly learn new technologies to adapt to the needs of industry development. However, due to the rapid pace of technological updates, some production personnel may find it difficult to keep up with this pace, resulting in insufficient proficiency or limited application of the technology.

2) Post processing technology is relatively backward. Although high-end production software and technology are already quite mature, relatively backward editing techniques and packaging strategies may still be used in some TV programs or low-cost film and television works. This may be due to a lack of innovative spirit in television media or production companies, unwilling to take on the risks brought by new technologies and concepts, or unable to adopt the latest technologies due to limitations such as funding and time.

3) Software compatibility and stability issues, as well as compatibility and stability issues between different software, are also common technical challenges in film and television post-production. This may cause obstacles for production personnel when processing materials or creating special effects, affecting work efficiency and work quality.

4) The dependence on high-performance hardware often requires high-quality film and television post-production to rely on high-performance computer hardware, such as high-performance CPUs GPU, Large capacity memory and fast storage devices, etc. However, the cost of these hardware devices is relatively high, which may pose an economic burden for some small production teams or individual creators [2].

Therefore, the research direction of this article is to use deep learning based artificial intelligence methods to reduce the skill level requirements for personnel in film and television post-production and improve the efficiency of post-production. This article uses artificial intelligence techniques to improve post-production methods for film and television works

1) In terms of intelligent video cutout, threshold method is used to perform binary processing on differential images, and logical operations are performed on two differential binary images to obtain the regions of interest of the audience in the scene of post-production video images;

2) In terms of video restoration, after obtaining the matching optical flow in the video, it is used as the initial optical flow of the RAFT model, and RAFT is further trained as the optimizer to obtain the final refined optical flow. After obtaining the complete optical flow, the forward and backward optical flows are mapped to the input feature map through optical flow warp operation, thus completing the feature propagation of the optical flow.

3) Using short videos as experimental subjects, constructing an experimental environment, and verifying the effectiveness of the method proposed in this paper.

2 Related Work

There are not many related achievements in the application of advanced computer technologies such as artificial intelligence, deep learning, and reinforcement learning in film and television production. After searching, some articles that can be referenced are listed, which also provide guidance for the formation of the ideas in this article.

Kun Hu, focusing on the entire process of film production and combining advanced technology with practical applications, reviewed and organized the research progress of deep learning technology in various aspects of intelligent film production. Based on the current situation, he analyzed and looked forward to the development needs and future trends of the combination of deep learning and film production. The analysis showed that deep learning has been maturely applied in semantic segmentation and image enhancement, and future researchers should strengthen research in more creative aspects and pay more attention to copyright regulations [3].

Yu Hong discussed the application and challenges of AIGC empowerment in these two fields, introduced the basic principles and development background of AIGC technology empowerment, and focused on the application of AIGC in film and television technology. He attempted to explore the creative path and expression of AIGC technology empowerment in film and television color design [4].

Mengya Liu, taking the production process of traditional film visual effects as the starting point, explores the differences in the creation process between interactive engine real-time rendering and traditional offline rendering, and the changes brought by interactive engine based special effects technology to the visual presentation of films. She explores new processes for film visual effects creation and combines artificial intelligence technology in real-time visual effects to further explore the creation path and development inspiration of film visual effects in interactive engines [5].

Zhenxing Jie analyzed the current application status and existing problems of augmented reality technology in the animation field, and proposed an artificial intelligence based solution. Through technologies such as deep learning and computer vision, real-time video image recognition and tracking have been achieved, and virtual elements have been effectively integrated with the real environment. Meanwhile, by introducing technologies such as natural language processing and sentiment computing, the perception of users' emotions and intentions has been achieved, and corresponding adjustments have been made based on user feedback, enhancing the immersive experience of animation scenes. Finally, the feasibility and effectiveness of the method were verified through experiments [6].

Yongjie Pan, based on the development and utilization of industry historical data and the production needs of ultra high definition content, analyzed and proposed a video restoration and ultra high definition video re production technology method based on intelligent algorithms. He focused on introducing the principles and application effectiveness of AI denoising, scratch removal, coloring, and resolution enhancement algorithm technology, and elaborated on the exploration of relevant independent controllable technology system integration practices [7].

Yan Zhao from Hengshui TV believes that with the continuous development of artificial intelligence technology, its application in TV program production and broadcasting has gradually become popular. Through scene recognition and intelligent segmentation, sentiment analysis and content screening, automatic generation of narrative clues, personalized recommendation and intelligent optimization, and other technologies, the efficiency and quality of program production have been significantly improved. However, corresponding challenges such as technological bottlenecks, legal and ethical issues, changes in human resources, and data security and privacy protection have emerged. In order to address these challenges, the author proposes a series of measures, the implementation of which will help the television program production and broadcasting industry better cope with the challenges brought by AI technology, promote innovative development of the industry, and provide viewers with a higher quality and personalized program experience [8].

However, few scholars have conducted detailed research on how artificial intelligence can play a role in the entire process of video post-production, and provided relevant and effective experimental methods, to summarize the above research results, this article consists of the following contents. Firstly, the application of deep learning in video post-production process is introduced, including intelligent image cutout, intelligent video restoration and other links. Then, a chapter is used to verify the effectiveness of the proposed method through specific video post-production as experimental content. The conclusion section summarizes the work done in the article and also points out the shortcomings of the research.

3 The Application of Deep Learning in Film and Television Post-production

A complete video post-production includes the following main steps: editing, special effects processing, enhancing visual effects, sound and music design, color adjustment, video enhancement, format conversion, etc. In response to the above steps, this article uses artificial intelligence technology for post-production processing in video effect enhancement, video restoration, digital cutout, and digital color grading [9].

3.1 Artificial Intelligence Image Cutout Technology

Before post-production image segmentation, it is necessary to determine the Region of Interest (ROI) within the scene of the post-production video. In general, humans have varying levels of attention to different regions in an image during observation, and the areas that humans focus on are defined as ROI [10]. Accurately determining these ROIs will significantly improve the efficiency and accuracy of post-production image segmentation in film and television. During the post-processing of film and television, the three frame difference method is used to determine the ROI within the video image. Select three consecutive frames of images in the post-production video sequence, determine the differential image between adjacent frames, and apply thresholding to the differential image. Perform logical operations on two differential binary images to obtain ROI within the scene of post-production video images. This method is susceptible to noise and brightness fluctuations during actual operation.

Therefore, in order to suppress the probability of ROI region selection errors, it is necessary to analyze the depth information of ROI in the neighborhood. In post-production video images, depth information only fluctuates when the target is in motion. Therefore, the fluctuation value of depth information in adjacent frames of post-production video is set as the basis for determining ROI [11].

The main purpose of introducing prior information is to determine a region within the post-production video image, and after scaling and selection, the edge height is consistent. Due to the fact that the final post-production video image segmentation result is roughly the same, high precision of prior information is required in this process. Due to the fact that the information input by the user is only the approximate shape of the post-production video image, or even just a certain part of it, in the actual post-production image segmentation process, it is not only necessary to perform image segmentation based on the prior shape input by the user, but also to determine high-precision segmentation results based on the actual situation of the post-production video image [12].

For the ROI identified in post-production video images, an edge free active contour model is used for object segmentation. Closed boundary C can segment the post-production video image $I(i, j)$ with region $M_i(i, j)$ into target A_i and background B_i , where c_a and c_b are the grayscale values of target A_i and background B_i , respectively. Therefore, the energy function is expressed as:

$$Enery\left(C,c_a,c_b\right)=\beta\iint_{A_i}\left(I-c_a\right)^2 dxdy+\beta\iint_{B_i}\left(I-c_b\right)^2 dxdy+L\left(C\right) \quad \textbf{(1)}$$

In the formula, β and $L(C)$ respectively represent the weighting coefficient and the length of C, and the size of $L(C)$ can reflect the smoothness of C, The smaller the $L(C)$, the smoother the C . $\iint_{A_i}\left(I-c_a\right)^2 dxdy$ and $\iint_{B_i}\left(I-c_b\right)^2 dxdy$ are guarantee terms, and when they are located at the boundary of the target in the post-production video image, the sum of the two results in the minimum.

In order to establish the final result of image cutout in film and television post-production, this paper uses gradient descent method to calculate partial differential equations:

$$\frac{\partial\phi}{\partial t}=\theta\left(\phi\right)\left\{\nabla\cdot\frac{\nabla\phi}{\left|\nabla\phi\right|}-\mu\left[\left(I-c_a\right)^2-\left(I-c_b\right)^2\right]\right\} \quad \textbf{(2)}$$

In the equation, ϕ is used to describe C . When the result of this equation reaches its minimum value, the corresponding C value is optimal. And for the ϕ value, use its result to express C :

$$Enery\left(\phi,c_a,c_b\right)=\mu\iint_{M}\left[\theta\left(\phi\right)\left(I-c_a\right)^2+\left(1-\theta\left(\phi\right)\right)\left(I-c_b\right)^2\right]dxdy \quad \textbf{(3)}$$

In the equation, θ represents the derivative of the Heaviside function [13].

3.2 Deep Learning Algorithms for Repairing Post-production Videos

The success of video restoration largely depends on two key factors: a sufficiently large contextual receptive field and fine texture details. The former helps to restore the structural information of missing areas, while the latter ensures the authenticity of the repaired content. However, in practical operation, larger contextual receptive fields are usually established on the basis of downsampling to sufficiently small feature maps, while high-quality texture details mostly come from large-sized input images. Therefore, the repair results of missing areas often require a trade-off between better structural information and clearer texture details. Although patch based word vector (token) encoding can expand the receptive field and maintain the authenticity and rationality of the repair results, this often comes at the cost of sacrificing image clarity. Directly applying block based Transformers to large-scale feature maps or even original images to ensure texture details can lead to a significant increase in computational complexity, making it difficult to train and infer. Therefore, researching how to achieve high-quality video restoration while maintaining low computational complexity is an urgent problem that needs to be solved in current video restoration technology [14].

The classic method based on optical flow propagation regards video restoration as a pixel propagation problem, which repairs the video sequence by repairing the optical flow between adjacent frames, thus naturally maintaining temporal coherence through optical flow. However, current optical flow based methods often require three stages: optical flow completion, pixel propagation, and content generation. Due to manual operations, these three stages need to be executed separately, which can lead to errors in the previous stage gradually amplifying in subsequent stages, ultimately affecting the overall repair performance. As the first step of optical flow completion, optical flow estimation naturally requires high accuracy, because incorrect optical flow estimation can make all subsequent processes meaningless. Therefore, a core of optical flow method is to ensure the accuracy of optical flow estimation. Recent work has adopted more traditional optical flow estimation methods, such as Flownet Spynet. The disadvantage of these optical flow methods is low accuracy, so directly applying them to video restoration can easily lead to restoration failure. Meanwhile, due to the fact that most previous methods estimate optical flow and perform restoration on the original image, they can only handle low resolution videos and are difficult to effectively integrate with Transformer methods. Therefore, this paper proposes an optical flow propagation method based on global matching and Transformer encoder to effectively improve the accuracy of optical flow restoration [15].

Assuming that the two consecutive image frames of the input video sequence are A_1 and A_2, this paper first uses convolutional layers to downsample by 1/8 and extract initial features, and then uses Transformer blocks to extract contextual features F_1 and F_2 with long-term dependency information. Due to the high dimensionality of image features, this article adopts the Transformer based on hybrid pooling attention proposed in the previous section to reduce computational complexity and memory usage. Then, by constructing a 4D cost volume on features F_1 and F_2, and calculating global matching information based on the cost volume, output the rough flow f_1 between A_1 and A_2, and optimize it through the RAFT model. The optimization process is shown in Fig. 1.

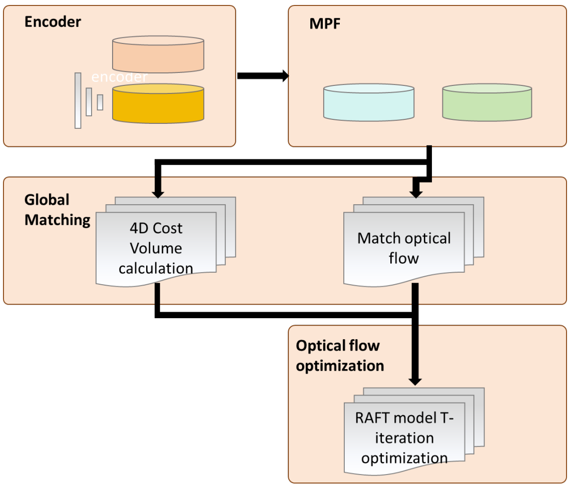


Fig. 1. Optimize the flowchart

Build image pixel vector:

$$C\left(i, j, c_a, c_b\right) = F_1\left(i, j\right) \cdot F_2\left(c_a, c_b\right) \tag{4}$$

Then, this article uses a dual operator to convert the cost quantity into matching reliability, and generates matching optical flow based on the obtained matching reliability:

$$M_i\left(i, j\right) = \arg\max\, p_c\left(i, j, c_a, c_b\right) \tag{5}$$

After obtaining the matching optical flow, it is used as the initial optical flow of the RAFT model, and RAFT is further trained as the optimizer to obtain the final refined optical flow. After obtaining the complete optical flow, the forward and backward optical flows are mapped to the input feature map through optical flow warp operation, thus completing the feature propagation of the optical flow. Taking the feature of frame i as an example, through the optical flow $F_{i \to i+1}$ from frame i to frame i +1, the region in frame i +1 can be backpropagated to the corresponding missing region in frame i +1, thus completing the update of the feature of frame G. This update process is implemented through the backpropagation function, and the optical flow propagation module is shown in Fig. 2.

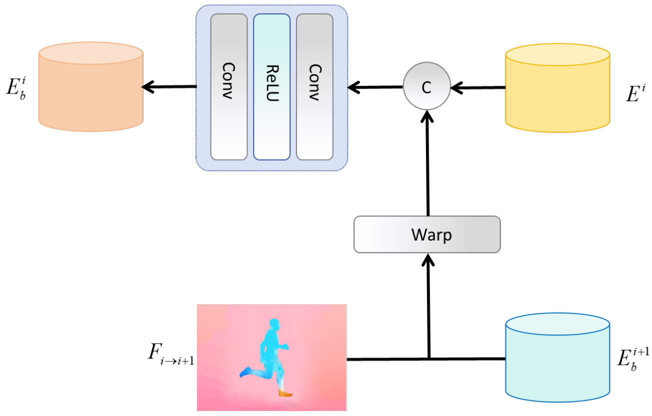


Fig. 2. Schematic diagram of optical flow propagation module

Therefore, the video restoration process in this article consists of two stages. The first stage is the pre restoration stage, which uses the optical flow propagation module to perform optical flow restoration at the feature level. At the same time, the output restored optical flow features can provide clear local spatiotemporal correlations for the next stage. Then, using the hybrid pooling based Transformer model [16] from the previous section as the content repair module for the second stage, modeling the temporal correlation of global long-distance frames to complete the repair of missing video content in the region. By dividing the modeling into two stages of near distance spatiotemporal modeling and long-distance spatiotemporal modeling, the local spatiotemporal characteristics of the optical flow method and the global characteristics of the attention mechanism can be fully utilized. Decomposing the task into two modules also helps reduce computational costs, making the model more economical and efficient while maintaining the overall temporal consistency of the model.

For a video sequence with missing areas, it is first fed into an optical flow estimation model based on global matching and Transformer encoder to calculate the optical flow of downsampled features. Then, an encoder is used to obtain initial features with the same downsampling ratio as the optical flow, and the features are grouped according to the distance from the center frame of the sequence. Adjacent frames are grouped together, while distant frames are grouped together. Next, the adjacent frame features and optical flow are fed into the bidirectional optical flow propagation module at the feature level to obtain the repaired optical flow and features, completing the first stage of repair. In the second stage, the repair features obtained in the first stage are fused with the initial features through a convolutional layer and fed into a content repair module composed of stacked hybrid pooling Transformer blocks of the same specifications. This module models the global spatiotemporal context and performs further repairs. Finally, the repaired features are fed into the decoder for upsampling to obtain the repaired video sequence. Dividing the repair process into two stages: on one hand, it clarifies the division of labor among each repair module, avoiding errors caused by the optical flow model during the initial training stage that may affect the training of subsequent modules; on the other hand, it can further save computing resources, ensuring that the model can be trained and deployed in a more economical way. The specific process is shown in Fig. 3.

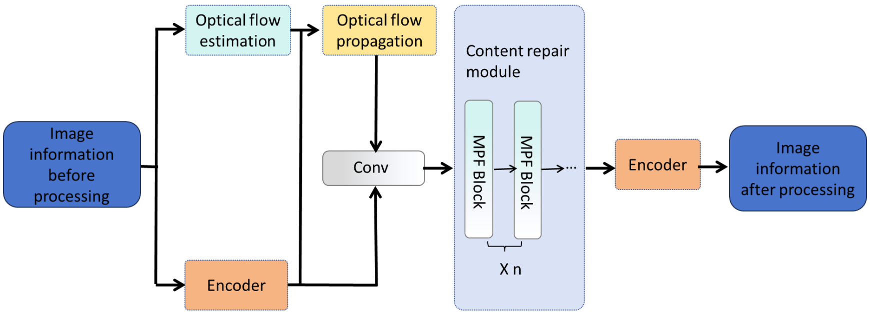


Fig. 3. Video repair process

3.3 Application of Deep Learning in Video Color Enhancement

Firstly, based on the characteristics of the video image, the original image is downsampled and adaptively partitioned to obtain a sparse image, in order to reduce the amount of image data [17]. Then, the MBLLEN network [18] is applied to enhance and combine the sparse image in true color to obtain an enhanced sparse image. Finally, a mixed histogram matching pair is used to map the color of the original image, and the color enhanced sparse image is upsampled and restored. Combined with the restored image, the mixed histogram matching image is color compensated to obtain a high spatial resolution video image that conforms to human vision and color consistency. The process of color enhancement method is shown in Fig. 4.

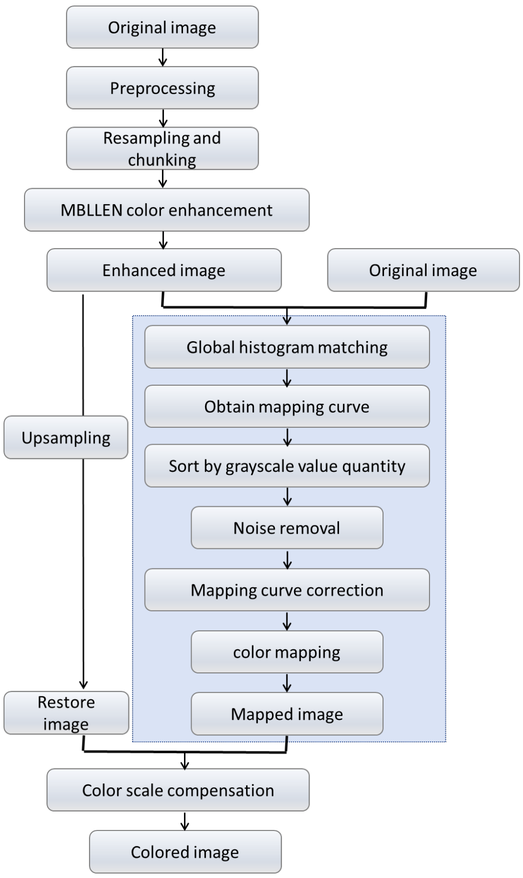


Fig. 4. Flowchart of color enhancement methods

Due to the large amount of short video data, color enhancement processing of raw data is not only time-consuming and labor-intensive, but also requires high computer performance. Before processing, it is necessary to perform a certain proportion of downsampling and thinning on the video impact, and adaptively divide the image into blocks according to the resampling ratio. Then, the determination of the downsampling ratio of the image is based on the principle that the smallest interested ground unit in the processed image exists and can be distinguished, and the number of blocks is determined according to the downsampling ratio. The schematic diagram of the sampling principle is shown in Fig. 5.

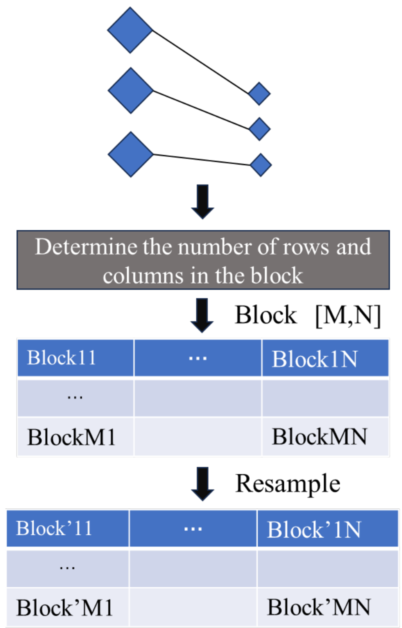


Fig. 5. Schematic diagram of sampling principle

Then, the MBLLEN deep learning model is used to intelligently enhance the colors in the video, as shown in Fig. 6.

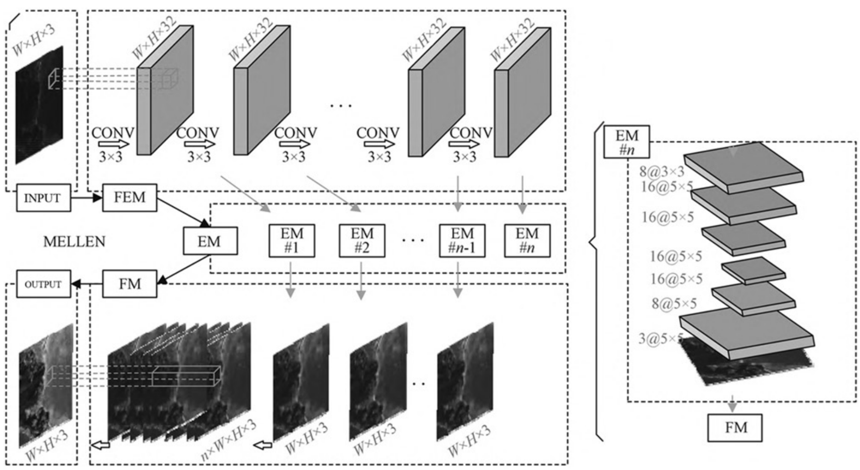


Fig. 6. MBLLEN network structure

The MBLLEN deep learning model (Fig. 4) mainly consists of three parts: feature extraction module (FEM), enhancement module (EM), and fusion module (FM), FEM utilizes different convolutional layers to extract rich features, enhances features through different subnets in EM, and fuses multiple branches for FM output, which can simultaneously solve the problems of image denoising and low light enhancement.

The method of using mixed histogram matching combined with color level compensation for color mapping can effectively compensate for the color loss caused by histogram matching, and achieve high reproduction of color enhancement in high-resolution images through sparse image. The specific steps are as follows:

1) Calculate the histogram of the original image and the histogram of the thinned image after true color enhancement, and calculate their respective cumulative histograms. For any gray level of the reference image, determine a corresponding gray level in the original image to obtain a mapping curve that matches the histogram:

$$M\left[i_t\right] = i_s, F_s\left[i_s\right] \le F_t\left[i_t\right] \le F_s\left[i_s + 1\right] \quad \textbf{(6)}$$

In the formula, $F[x]$ represents histogram data and i represents grayscale.

2) Reorder the grayscale values of the original image histogram according to the corresponding number of grayscale values, so that they meet the following criteria:

$$H\left[i_0\right] \le H\left[i_1\right] \le \cdots \le H\left[i_{255}\right] \quad \textbf{(7)}$$

$H[x]$ represents the grayscale value function.

3) For the sorted grayscale values, filter out the corresponding grayscale values with fewer pixels than the total pixel ratio. Then, deduce the original image grayscale values at the corresponding positions as histogram noise pixels and remove the noise pixels to obtain the mapping curve. The curve representation method is as follows:

$$Q(i) = \left\{ i_k \middle| H\left[i\right] \le T \right\} \quad \textbf{(8)}$$

T represents the pixel ratio.

4) There are some vacant areas, and for the vacant intervals generated by denoising the mapping curve, an optimized mapping curve is obtained by first connecting straight lines and then smoothing the curve. The original image and the reference image are color mapped using the mapping curve to obtain the mapped image;

5) Upsampling the reference image to obtain a restored reference image, calculating the mapping between the restored reference image and the mapped image on the wavelength band, and then performing color compensation to compensate for the color level loss after matching the mixed histogram, resulting in a high-resolution video image with color enhancement [19].

3.4 Artificial Intelligence Video Color Grading Method

Color is an essential element in videos, and the adjustment of color tones in conventional videos often relies on manual labor, which consumes a lot of time and effort, and cannot guarantee satisfactory results. Therefore, this article uses reinforcement learning methods to change image color information to meet the color requirements of different films. Meanwhile, color is a visual expression of emotions in videos. For example, black tones have inclusiveness and invasiveness, giving people a sense of mystery and nobility, while blue tones have flexibility and rationality, giving people a sense of melancholy and detachment from the mundane. Therefore, in the post-production process of film and television, how to accurately identify video emotions and automatically match video tones and emotions is the focus of this study.

Based on the K-Nearest Neighbor regression model [20], a color transfer algorithm is used to fuse the local color information of the image with the Reinhard algorithm without requiring a large amount of sample data, achieving a combination of local and global information for color transfer. For the target pixel in the video, K sample points closest to the point are selected, and then assuming there is a set containing m sample points, the expression is as follows:

$$S_{Pixel} = \left\{ \left(p_{x1}, p_{y1}\right), \left(p_{x2}, p_{y2}\right), \cdots, \left(p_{xm}, p_{ym}\right) \right\} \quad \textbf{(9)}$$

In the formula, $p_{xi,i\in(1,m)}$ is a point in the M -dimensional Euclidean space, and $p_{yi,i\in(1,m)}$ is the value corresponding to each sample point. K sample coordinates closest to the new sample point can be selected, and their y values can be weighted and averaged to obtain the y value of the new sample point, which is expressed as:

$$p_y = \sum_{(p_{xi},p_{yi})\in S_K(p_x)} \lambda_i p_{yi} \tag{10}$$

In the formula, $S_K(p_x)$ represents the set of K points closest to the sample point, and λ_i represents the weights corresponding to each of the K sample points. In channel L of the *Lab* color space, two KNN models are established from channel L to channels a and b using 3×3 rectangular neighborhood blocks, namely the L values corresponding to pixel points and a total of 9L values in their *N8* neighborhood. The color transfer process is as follows:

1) Convert the source image S_{pixel} and the target image T_{pixel} from the RGB color space to the *Lab* color space using color space transformation matrices;

2) Establish separate datasets for source image S_{Pixel} and target image T_{Pixel} . The training set for the model consists of a 3×3 rectangular neighborhood block in channel L of target image T, as well as the values of channel a and channel b corresponding to the center point of the rectangular neighborhood block. The prediction set for the model consists of a 3×3 rectangular block in channel L of source image S_{Pixel} .

3) For the dataset obtained from target image T_{Pixel} , train two KNN regression models, one for channel a and the other for channel b, using 3×3 rectangular blocks in color channel L;

4) Using the two models obtained above, predict the values of channel a and channel b corresponding to channel L of source image S_{Pixel} , respectively.

5) Convert the mapped result image from the *Lab* color space back to the *RGB* space, achieving color transfer from the target image T_{Pixel} to the source image S_{Pixel} .

The advantage of color transfer algorithm is that it utilizes the K nearest 3×3 rectangular blocks in the pixel block for learning, and uses local features of the target image for transfer. And this algorithm, like the Reinhard color transfer algorithm, allows for different sizes between the source image and the target image. However, at the same time, due to the fact that the K nearest neighbor sample points found by the KNN algorithm may be far apart in the image, the color differences between adjacent pixels in the final result image may vary, resulting in unnatural color transitions. To address the aforementioned issues with the algorithm, this paper employs the Reinhard algorithm that combines global statistical features and the KNN algorithm that utilizes local pixel features.

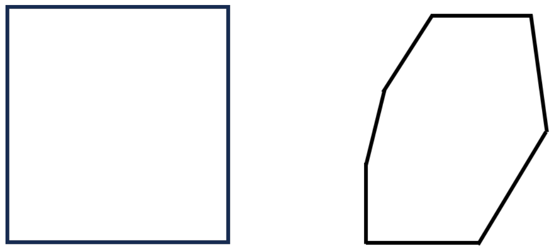
The Reinhard algorithm [21] improvement calculates the standard deviation and mean of the three channels of the target image and the source image separately in the *Lab* space, and calculates the transformed mean. Using X to represent a certain color channel in *Lab* space, and XT_{Pixel} to represent the value of a pixel point in the target area in the X channel of *Lab* space, the resulting new image has the mean and standard deviation information of the reference image, achieving the goal of color similarity with the reference image.

For special cases where the reference image standard deviation is the same as the source image standard deviation, color transfer that fully preserves details is necessary. For color transfer that balances both color and detail, the Reinhard algorithm only considers the overall color of the image. For images with complex color combinations and detailed information, the effect of the Reinhard algorithm is not obvious. By analyzing the principle of the Reinhard algorithm, it can be seen that the Reinhard algorithm focuses on the overall color tone transfer, but ignores the correlation between pixels in small areas of the image. Therefore, it is very likely to modify the detail effect of the texture information of the image. In fact, the color values of pixels in two-dimensional space exist in a certain distribution, and the color value of a certain pixel is only related to a certain neighborhood around it and is not affected by color values outside the neighborhood. Therefore, after converting the *RGB* of the source image and reference image to *Lab*, the three component values of *Lab* are still affected by the component values in the neighborhood themselves. Therefore, based on the Reinhard algorithm process, this improved algorithm completes the Reinhard algorithm improvement process by transforming color channels, calculating standard deviation and mean, and introducing a new local standard deviation ratio factor, adding local reference image and source image information.

4 Verification of Post-production Effects in Film and Television

Research on a film and television post-production image segmentation method based on interactive object segmentation algorithm. To verify the practical application performance of the proposed method, a daily shot video is used as a film and television post-production video as the research object, and the method proposed in this paper is used for image segmentation processing.

ROI region extraction is performed within the selected video set by randomly selecting an image frame and using the proposed method to select the ROI region within that image. The results are shown in Fig. 7(a). According to the analysis of Fig. 7(b), the proposed method can effectively determine the ROI region in the video image.



(a) Initial contour of image (b) Processed contour

Fig. 7. Initial contour and converged target contour

The experiment on target segmentation performance is divided into two groups. The first group analyzes the ability of target segmentation, and the second group analyzes the convergence speed. In order to effectively improve the problem of re initialization, a regularization term is introduced into the CV model during the experiment. Analyzing Fig. 7(a), it can be concluded that when using the proposed method for object segmentation, due to the introduction of interactive information and the use of prior graphics to drive the evolution of the level set function, the image contour closest to the prior graphics is obtained, which improves the accuracy of object segmentation. Analyzing Fig. 7(b), it can be concluded that due to the introduction of interactive information in the proposed method and the influence of prior graph guidance, the convergence efficiency of the level set function is significantly improved. Fig. 7(b) shows that the proposed method has reached a convergence state under the condition of the 40th iteration, while the basic CV model is still in a state of large segmentation error under the condition of the 40th iteration. This indicates that using the proposed method for post-production image segmentation can greatly improve the efficiency of the image segmentation process.

At the same time, in order to verify the effectiveness of video restoration, video restoration experiments were conducted. In terms of datasets, this article uses the most widely used video restoration datasets currently available: YouTube VOS and DAVIS. YouTube VOS contains 3471 training videos, 474 testing videos, and 508 validation videos, with an average frame length of 150 frames. The DAVIS dataset contains 150 video clips characterized by foreground motion and camera movement. This article uses 50 of them as the test set to maintain consistency with the method to be compared. All models and variants were trained on the YouTube VOS dataset and tested and evaluated using the validation sets of YouTube VOS and DAVIS. During the training process, randomly generated masks are used to mimic missing areas in the video, and randomly shaped masks are used to evaluate repair performance. Meanwhile, the original high-density labeled object mask provided by the dataset is used to evaluate object removal performance. It can be observed that after introducing the bidirectional optical flow propagation module (FP), the model not only recovers to the repair level of the original method, but also slightly surpasses it in some aspects. This progress is attributed to the unique properties of the two-stage training model, which optimizes in two independent stages to improve the quality of video restoration. The comparison between traditional video restoration methods and the video restoration method proposed in this article in terms of video restoration speed, namely the frames per second (FPS) of processed images, is shown in Fig. 8.

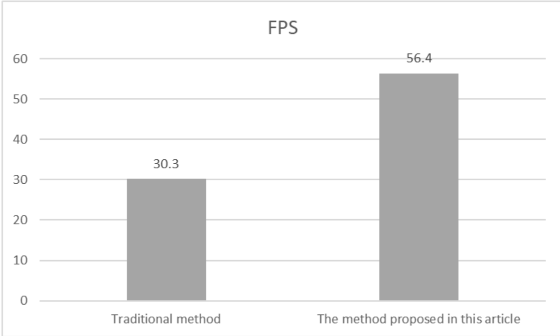


Fig. 8. Comparison of improved video enhancement model effects

While maintaining the same repair effect as the original method, the improved method significantly improves running speed and reduces computational resource requirements. This result means that the improved network structure not only performs excellently in terms of performance, but also is more efficient in practical applications, which is particularly important for resource constrained environments.

Based on the YouTube VOS dataset, MBLLEN is trained by simulating images synthesized under low light conditions as labeled images. The original image and the synthesized low light image form a data pair, and the simulated low light conditions are as follows: Gamma correction and Poisson noise with a peak value of 200 are used for end-to-end training of the network to obtain the trained model. Import the resampled image into the trained MBLLEN model to obtain the true color enhanced result, and perform geographic information recovery on the output result. The processing procedure is shown in Fig. 9.

Fig. 9. Comparison of the effects before and after video enhancement

For video color intelligent color matching, the experiment used the color transfer algorithm proposed in this article to perform common color transfer based on image color emotion on the image. Choose an image of green grass as the source image, and the storyline of the video at this time is post-war calmness. The video should express a certain sadness and confusion, and the color transfer algorithm of the K-Nearest Neighbor regression model improved by the Reinhard algorithm was used to approach the overall color tone of the target image more closely. Fig. 10 shows the comparison of the images before and after color matching, and Fig. 11 shows the comparison of the Colorfulness (CS) and Structural Similarity (SSIM) index between the target image and the source image when the improved algorithm deals with color tone problems.

This section has completed the post-production image extraction method of interactive object segmentation algorithm for film and television, and verified the effectiveness of this method from real videos and theoretical aspects. At the same time, for video restoration and enhancement, the proposed method can compensate for the color level loss after mixed histogram matching in the video, obtain high-resolution video images with color enhancement, and finally achieve the effect of automatically adjusting the color tone of the video based on the movie theme emotion in color adjustment.

Fig. 10. Comparison of video color adjustment results before and after algorithm improvement

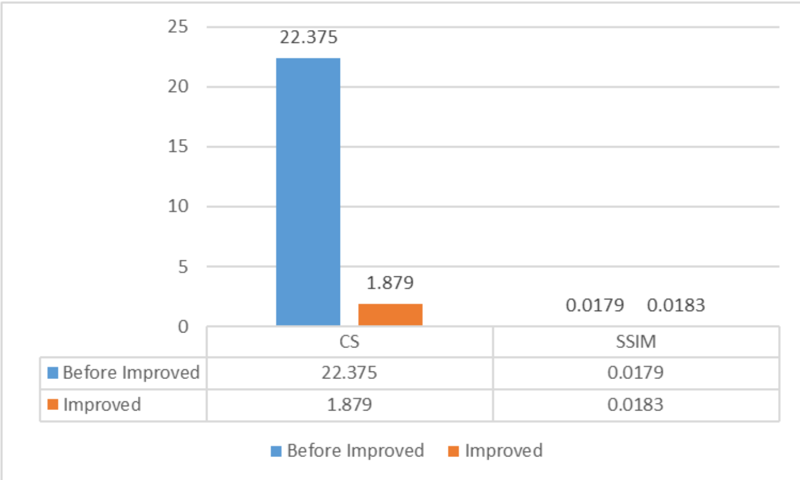


Fig. 11. comparison of the CS and (SSIM) index between the target image and the source image

5 Conclusion

This article focuses on the research direction of video post-production based on deep learning. Video post-production technology has shown significant potential in the fields of video processing and computer vision, aiming to repair damaged parts in video data in an efficient and spatiotemporal consistent manner to achieve high-quality visual experience, while enhancing video color effects and locating areas of interest in the video. The core of this technology lies in simulating the cognitive process of the human visual system to achieve accurate restoration of damaged video content. Video post-production plays an important role in various practical scenarios, including but not limited to film post-production, cultural heritage preservation, and facial feature restoration. At present, video post-production methods based on deep learning suffer from issues such as blurry details, difficulty in maintaining good temporal coherence, and high training and inference costs that make it difficult to put into practical applications. In response to the problem of insufficient utilization of temporal spatial contextual information in current methods, this paper proposes a novel end-to-end network architecture based on spatiotemporal cross window Transformer. Improved window attention is used to enhance the local spatiotemporal modeling ability of Transformer and reduce computational complexity, effectively improving video restoration performance and mitigating temporal artifacts caused by blurry details and poor temporal coherence.

Currently, the accuracy and efficiency of video restoration methods are steadily improving, but there are still many problems and challenges. In the future, this article will continue to conduct research on video restoration methods from the following aspects:

1) The research in this article is mainly based on the current mainstream public datasets. In the future, we will attempt to shift our research direction towards video restoration of real-world data, collecting data in actual and specific video scenes to test the performance of the model in real-world environments.

2) The ultimate goal of video post-production tasks is to produce repair results that are difficult for the human eye to detect defects, and the current methods still have difficulty achieving the level of accuracy required to "confuse the real with the fake". In the future, we will attempt to further improve the video restoration model and conduct research on video restoration of missing areas in various complex scenes. At the same time, further compressing the model size, reducing the number of parameters, and improving the efficiency of video restoration.

3) In the future, we will explore how to combine the video repair model and system proposed in this paper with Internet applications and websites, build a convenient and fast online video repair website, and realize the practical application of more portable video repair technology.

References

[1] X.-J. Zhao, Research on Post-Production and Editing of TV News Programs from the Perspective of New Media, Video Engineering 46(2)(2022) 123-125.

[2] A.-C. Yi, J.-J. You, Research on post production image cutting method for film and television based on interactive object segmentation algorithm, Electronic Design Engineering 32(11)(2024) 188-191+195.

[3] K. Hu, P. Xie, Research on intelligent filmmaking technology based on deep learning, Advanced Motion Picture Technology (3)(2024) 12-19.

[4] Y. Hong, Y.-Y. He, Research on Film Technology and Film Color Design Based on AIGC Enablement, Color (9)(2023) 44-46.

[5] M.-Y. Liu, W.-R. Zhao, A research on the workflow of creating movie visual effects based on the interaction engine technology, Advanced Motion Picture Technology (8)(2023) 34-40+33.

[6] Z.-X. Xie, Research on the Characteristics of Architectural Landscape Pattern Based on Multi Adversarial Information Generation Network, Journal of Jiamusi University (Natural Science Edition) 42(5)(2024) 132-135.

[7] Y.-J. Pan, Research on Application of Video Repair and UHD Reproduction Technology Based on Intelligent Algorithm, Radio & TV Broadcast Engineering 50(6)(2023) 28-32.

[8] Y. Zhao, H.-F. Xia, G.-B. Liu, Application and Challenge of AI in Automated Television Program Editing and Broadcasting, Video Engineering 48(6)(2024) 102-104+108.

[9] J.-H. Du, Application of Computer Network Technology in Post-Production of Radio and Television, Audio Engineering 47(5)(2023) 71-73+78.

[10] Z.-Y. Yin, C. Yin, F.-Q. Zhang, G.-Y. Xu, F.-L. Xu, Landmark Detection Method That Combines Adaptive Thresholds with Dynamic ROI, Journal of Chinese Computer Systems 45(2)(2024) 345-350.

[11] F.-J. Feng, Y. Yang, M. Tan, H.-S. Gou, Y.-H. Liang, L. Wang, An Alpha Matting Algorithm Based on Micro-scale Searching for High-Resolution Images, Pattern Recognition and Artificial Intelligence 36(6)(2023) 530-543.

[12] H.-M. Chen, A Review of Semantic Image Segmentation Based on Deep Learning in Image Processing, Science & Technology Information 22(6)(2024) 10-13.

[13] A.-D. Yang, X.-D. Xie, N.-M. Luo, J. Zhang, N. Jiang, S.-T. Wang, Implicit Heaviside filter with high continuity based on suitably graded THB splines, Frontiers of Mechanical Engineering 17(1)(2022) 120-148.

[14] J. Chen, K.-X. Wang, Y.-T. Zuo, Q. Lin, H.-Q. Zeng, Video Inpainting Based on Deep Learning: An Overview, Journal of Signal Processing 40(6)(2024) 1171-1184.

[15] B. Liao, W. Wu, Video shadow removal method using region matching guided by illumination transfer, Journal of Computer Applications 39(2)(2019) 556-563.

[16] M.-Y. Li, Y. Fu, Joint self-attention Transformer for multispectral and hyperspectral image fusion, Journal of Image and Graphics 28(12)(2023) 3922-3934.

[17] B.-R. Zhao, S.-W. Niu, L.-Y. Wang, X.-T. Yang, H.-B. Jiao, Z.-K. Wang, An intelligent color enhancement method for high-resolution remote sensing images of the coastal zone of an island, Remote Sensing for Natural Resources 36(2)(2024) 70-79.

[18] L.-G. Wang, J.-J. Tong, J.-L. Chen, Y.-C. Liu, Research on Endoscopic Image Enhancement Algorithm Based on Improved MBLLEN Network, Software Engineering 27(8)(2024) 12-15.

[19] J.-C. Zhou, X.-J. Wei, J.-Y. Shi, Underwater Image Enhancement Algorithm Based on Blue-green Channel Color Compensation, Journal of Electronics & Information Technology 44(8)(2022) 2932-2939.

[20] Z.-F. Wu, M.-M. Wang, T. Lan, A.-Y. Zhang, GHM-FKNN: a generalized Heronian mean based fuzzy k-nearest neighbor classifier for the stock trend prediction, High Technology Letters 29(2)(2023) 122-129.

[21] Y.-M. Yu, D. Jin, Q. Wang, Q. Zhang, X. Chen, X.-J. Wang, Color Image Segmentation Algorithm Based on Lab Sub Channel Histogram and Its Application, Imaging Science and Photochemistry 37(1)(2019) 18-32.