The Visual Computer
https://doi.org/10.1007/s00371-023-02962-3

ORIGINAL ARTICLE

AddCR: a data-driven cartoon remastering

Yinghua Liu[2] · Chengze Li[1] · Xueting Liu[2] · Huisi Wu[2] · Zhenkun Wen[2]

Accepted: 9 June 2023

Abstract
Old cartoon classics have the lasting power to strike the resonance and fantasies of audiences today. However, cartoon animations from earlier years suffered from noise, low resolution, and dull lackluster color due to the improper storage environment of the film materials and limitations in the manufacturing process. In this work, we propose a deep learning-based cartoon remastering application that investigates and integrates noise removal, super-resolution, and color enhancement to improve the presentation of old cartoon animations. We employ multi-task learning methods in the denoising part and color enhancement part individually to guide the model to focus on the structure lines so that the generated image retains the sharpness and color of the structure lines. We evaluate existing super-resolution methods for cartoon inputs and find the best one that can guarantee the sharpness of the structure lines and maintain the texture of images. Moreover, we propose a reference-free color enhancement method that leverages a pre-trained classifier for old and new cartoons to guide color mapping.

Keywords Cartoon remastering · Color enhancement · Denoising · Deep learning · Multi-task learning

1 Introduction

Many people find enjoyment in watching cartoons; however, the video quality of older cartoons often fails to meet modern expectations. To address this issue, media publishers and conservator-restorers dedicate their efforts to remastering these vintage animations with the goal of delivering a superior visual experience. Cartoon remastering typically involves three key processes: denoising, super-resolution, and color enhancement. These techniques transform antiquated, noisy, low-resolution, and faded animations into high-definition, vibrant works of art. Given the time-consuming and labor-intensive nature of manual remastering, there is growing interest in leveraging learning-based solutions to accelerate the process. By employing advanced algorithms and techniques, these methods have the potential to not only accelerate the remastering of cartoons but also maintain, or even improve, the quality of the final output.

Figure 1 presents an example of cartoon remastering, comparing the original animation (a) with a professionally remastered version (e). Close-ups of the yellow and green boxes in (a) reveal substantial noise in the old cartoon, which may be further exacerbated by compression artifacts introduced during network transmission. The original cartoon has a relatively low resolution of 640 × 480 pixels, whereas the manually remastered version boasts a significantly higher resolution of 1920 × 1080 pixels. Consequently, the application of super-resolution techniques is crucial for enhancing the visual quality of the old animation. Moreover, the color palette of the original cartoon appears less vivid and bright compared to its remastered counterpart. According to Fig. 1, the sky in the remastered version has been manually enhanced to appear more vibrant. Consequently, color enhancement is another critical aspect of cartoon remastering. Although previous research has predominantly addressed denoising and super-resolution, the challenge of color enhancement in old cartoons remains relatively unexplored. Two primary reasons contribute to this gap: the lack of frame-by-frame paired old and new cartoons for guiding color enhancement through color distribution mapping, and the limited consideration given to structure lines in existing methods. Some existing

✉ Chengze Li
czli@cihe.edu.hk

✉ Huisi Wu
hswu@szu.edu.cn

[1] School of Computing and Information Science, Caritas Institute of Higher Education, Hong Kong, China

[2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

Published online: 23 July 2023

methods, such as Waifu2x [25], account for structure lines in cartoons but do not fully exploit the characteristics of structure lines and color regions.

We examined relevant works and classified them into two categories based on their reference requirements. Semi-interactive methods [15, 33] can restore and colorize old films but require manual input of suitable keyframes. We attempted to use old cartoon colors as color references for colorization, but the results were unsatisfactory, as the enhanced colors remained unimproved and inconsistencies in hue were apparent in areas like the man's hair and eyebrows. A complete redrawing of the keyframes with more modern-ish colors seems also not to be an option, as it requires tedious work in manual drawing and color editing. In contrast, some approaches do not require reference images. For example, Old Photo Restoration [34] leverages the similarity between the latent spaces of synthetic degraded images and ground-truth old cartoon images to generalize well on ground-truth old cartoon images. However, this method fails to consider the characteristics of cartoons, leading to issues such as inaccurately colored eyes.

Other techniques, like DeOldify [12], rely solely on input grayscale image values, resulting in abrupt pixel value changes in the same color region of the output image. Furthermore, several related methods focus on addressing a single type of degradation, such as super-resolution [14, 35] or colorization [8, 37], whereas real-world cartoon images often suffer from a combination of multiple degradations.

In this paper, we present an integrated solution for cartoon remastering that combines denoising, super-resolution, and color enhancement in a comprehensive, data-driven pipeline. For the denoising component, we employ a multi-task learning method, incorporating a line extraction task to guide the model in accounting for structure lines. We then evaluate various single-image super-resolution techniques and identify the most effective approach for cartoon inputs. Subsequently, we introduce an unsupervised color enhancement model featuring a color distribution classifier. This model utilizes a pre-trained classifier for new and old cartoons as a mapping function to enhance the color of old cartoons without requiring additional reference color information. To preserve sharp structure lines and prevent outliers, we propose a line-preservation cosine loss mixed with a novel line extraction task in the color enhancement phase. Our color enhancement model is lightweight and efficient. As demonstrated in our qualitative and quantitative evaluation, the resulting images are high-definition, vibrantly colored, and exhibit clear structure lines, showcasing the effectiveness of our proposed solution.

The main contributions of our method are as follows.

- We make the first attempt in an integrated solution for cartoon remastering.
- We leverage multi-task learning in both denoising and color enhancement to guide the model to focus on line extraction. Among them, we propose a color enhancement network that learns with the pre-trained classifier to enhance the color of cartoons without additional reference images or color information and suggest a line-preservation cosine loss to prevent odd pixel values of the structure line.

2 Related work

Cartoon remastering is to input a noisy, low-resolution, dull and lackluster cartoon and output high-definition, color-enhanced ones after processing. As mentioned above, the existing remastering works do not work for cartoon remastering. Therefore, we not only discuss the methods in color

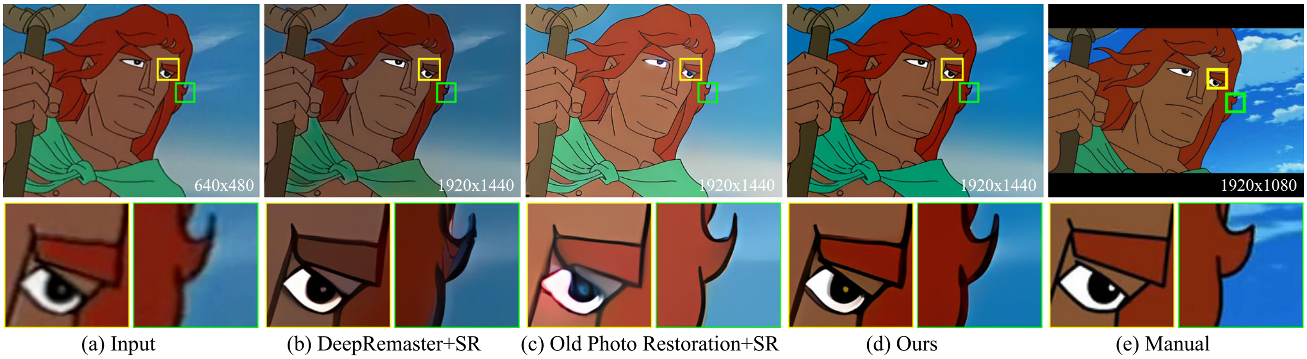


Fig. 1 Cartoon remastering. **a** old cartoon. **b**, **c** and **d** the blows-up of a frame processed by DeepRemaster [15], Old Photo Restoration [34], and the proposed integrated solution, **e** the manually remastered frame. We padded the manually remastered image with the black pixel to make sure it was the same height as the displayed image. For methods [15, 34], and ours, we employ the same super-resolution method [35] to adhere to the single variable principle. The corresponding blown-ups are shown in the below yellow box and the green box. The resolutions are in the lower right of the images

remastering but also the following three parts, color enhancement, denoising, and super-resolution.

2.1 Cartoon remastering

The existing methods related to remastering are mainly DeepRemaster [15] and Old Photo Restoration [34]. Among them, the remastering method for old films [15] can reduce some noise, but it requires appropriately selected keyframes to provide color references. Old Photo Restoration [34] utilizes latent space of both synthetic and ground-truth deteriorated images to restore old photos. There is no need to provide reference images. However, it generates colorful structure lines because it ignores the characteristics of the cartoon. The initial finding is that the existing remastering methods are unsuitable for cartoon remastering. Therefore, we present an integrated solution for cartoon remastering that combines denoising, super-resolution, and color enhancement in a comprehensive, data-driven pipeline.

2.2 Color enhancement

There are many color transfer methods in traditional algorithms and deep learning. For example, the traditional method [27] changes the color of source images according to the mean and standard deviation of the reference image. The deep learning methods [3, 6, 9–11, 21, 23, 29] use the style transfer processing method, that is, its color converts to the reference image while retaining its semantic information. The methods [1, 20] use histogram to guide the color transfer of semantic information. However, the selection of reference images requires manual screening, which is undoubtedly inconvenient for cartoon remastering work. Actually, the laborious process of choosing reference images can be avoided by using the input image as the reference image for colorization. However, such methods tend to generate results with uneven color regions. Additionally, by blending the reference color information into the features of the input image, it is challenging to maintain the sharpness of the image. So we exclude methods that need given reference images. StarEnhancer [30] needs no reference images. It embeds the penultimate layer of the style classifier, then takes it as a latent code to guide the color curve prediction for downsampled images, and finally applies the color map to the full-scale image. Since old cartoons lack remastered cartoons with consistent content, methods that require paired datasets are also not suitable for color enhancement in cartoon remastering. Low-light image enhancement method [24] is trained without paired dataset. However, we should implement the color distribution mapping from the old to the new cartoon, not just meeting the normal exposure conditions. The method [4] fuses local and global information to guide the input image for color enhancement. But this method fails to take the structure lines into account and lacks judgment as to whether the generated image conforms to the color distribution of new cartoons. CycleGAN [42] is based on GAN [5], adding a cycle loss to improve the quality of the generated images in both ways. This method also does not take into account the specificity of structure lines. Moreover, its discriminator's full power for guiding color enhancement mapping remains to be improved. We split the techniques related to retaining lines or improving line quality into two categories, transform and restoration. The former will change the shape of the structure line, which is conflict with our goal that retaining the content of the image. The latter includes denoising, super-resolution, and other methods proposed for manga and animation. However, we have not found such models tailored for structure lines yet. We think using the extracted structure line straightly may be the better choice. To fully use the discriminator's guiding role in the generator, we exploit the fixed discriminator by transfer learning [32].

2.3 Denosing

The traditional joint bilateral filtering denoising method based on [31] weighs the denoising effect and blurring degree by setting the range and difference of adjacent reference pixels. As we all know, it is difficult to choose a suitable value to avoid the phenomenon of image gelatinization caused by mutual reference between structural lines and general color regions. Recently, to improve the representation learning capabilities of the algorithm, many researchers focused on deep learning methods, especially those based on the UNet structure. For example, the denoising capability of Waifu2x [25], an application designed for anime, is somewhat restricted. DIDN [36] leverages the advantage of the deep network by iterative downsampling and upsampling processing to improve the denoising ability. However, this iterative method also increases the amount of computation in the network. NAFNet [2] is a lightweight and nonlinear network. However, the methods mentioned above fail to harness the full power of the convolutional neural network (CNN) by treating the structure lines as the general color regions. Thanks to the attention mechanism of the Transformer, SwinIR [22] achieves a better denoising effect. However, it takes a lot of time and does not ensure that the structure lines are all identical colors. To ensure the clarity and uniform color of the structure lines, we add a line extraction task in the denoising part.

2.4 Super-resolution

Traditional bicubic interpolation obtains the pixel value of the point by the weighted average of the nearest 16 sampling points on the rectangular grid. Since this method gives the weighting coefficient by a fixed formula, it cannot predict the upsampled image well. We further studied some classic and excellent algorithms based on deep learning and found that the algorithms based on CNN generally use the minimum mean square error, which makes the generated images lack high-frequency information and over-smooth textures. For example, UNet for super-resolution uses classic up- and downsampling structures and skip-connection to perform better in image restoration. Waifu2x [25], a classical super-resolution application, is proposed to improve the quality of cartoon images. The large-scale image restoration network [40] combines short and long skip connections and does not have downsampling modules. RealCUGAN [14] is trained with massive data and has a strong generalization. As we all know, the quality of cartoon datasets affects the restoring performance in a supervised training manner. Affected by network transmission, it is difficult to obtain high-quality cartoon pictures or cartoon animations. Hence, the limitations brought by the training mechanism also have an impact on how effective CNN-based methods perform. Under the adverse condition of the limitation of the quality of images, unsupervised methods have become the model of choice in image restoration. Among them, RealESRGAN [35] is an improvement on GAN [19] and has the advantage of introducing perceptual loss [16] to improve the fidelity of the super-resolution image texture. Moreover, it enhances the network's representation learning capability by introducing the high-order simulation degradation modeling process and generating data in a dynamic form. We empirically demonstrate that employing RealESRGAN [35] to improve the quality of old cartoons is the best choice.

3 Method

3.1 Overview

Our remastering process is composed of three successive stages: a multi-task UNet denoising model, the RealESRGAN super-resolution model [35], and an unsupervised color enhancement model with a color distribution classifier, as depicted in Fig. 2. Color distribution refers to the frequency and distribution of different colors. Existing denoising methods, while effective at noise reduction, often struggle to maintain uniform colors in structure lines. To address this, we employ a multi-task learning approach based on the lightweight UNet structure [28], which guides the model to consider structure lines during denoising. After evaluating various techniques, we opted to use RealESRGAN for super-resolution, as the extracted structure lines from the higher-resolution images are utilized in the subsequent color enhancement stage. Crucially, we have developed a color enhancement model that maps the color distribution of cartoons from old to new without requiring reference images. By pre-training the classifier with ground-truth old and new cartoon frames, the propagated gradients regularize the Residual Color Enhancement Module, allowing it to derive an appropriate mapping function for enhancing the color distribution of old cartoons during the training phase. This semi-supervised approach eliminates the need for additional reference color distributions throughout the pipeline, as the necessary information is already learned during classifier pre-training. Moreover, the color distribution classifier is not required during the inference phase.

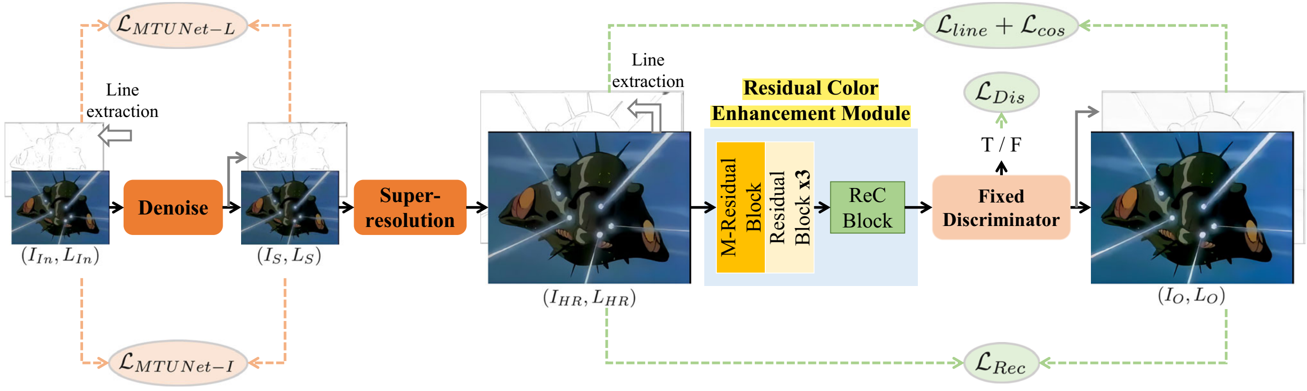


Fig. 2 Method overview. We put the noisy, low-resolution, and dull and lackluster images I_{In} into the multi-task UNet denoising model, the super-resolution model, and color enhancement model (consists of the Residual Color Enhancement Module and the Fixed Discriminator) orderly, then correspondingly get the cleared images I_S and the structure lines L_S, the high-resolution images I_{HR}, the final outputs I_O in color-enhanced version and the structure lines L_O. L_{In} and L_{HR} are the structure lines of the target images at the denoising part and the high-resolution images, respectively

We will further elaborate on the denoising and color enhancement components in Sects. 3.2 and 3.4, respectively.

3.2 Denoising

As previously mentioned, existing denoising methods often fail to account for the structure lines in cartoons, leading to issues such as colored structure lines due to neighboring color blocks or blackened color blocks influenced by reference structure lines. To address this problem, we propose the multi-task UNet denoising model (MTUNet), which incorporates a line extraction task to guide the network to consider structure lines. We select UNet [28] as our backbone network because its downsampling structure facilitates the extraction of high-level semantic information, while skip connections minimize information loss during downsampling. Building upon the UNet network, we adopt multi-task learning [41], with one task focusing on image restoration and the other on line extraction. By having these two tasks independently learn across three convolution layers, our network effectively reduces artifacts while preserving the original color of the structure lines.

3.2.1 Training

To achieve denoising, our proposed MTUNet model necessitates the use of a paired dataset. To create the noisy version of cartoon frames, we degrade the JPEG quality of the source image by setting the compression level to a random number between 30 and 70, and then save the resulting image in JPG format as MTUNet's input. This compression level range should sufficiently cover the noise levels found in old cartoons. We employ the line extraction method [13] to obtain the structure line of Input L_{In}. And we employ the Adam optimizer [17] to train our networks with a learning rate of 10^{-4} and no decay. The patch size is set to 512 following a random crop operation, and the batch size is 8.

3.2.2 Loss

In order to strike a balance between the denoising effect and the preservation of structure lines in the generated image, we set the weight ratio of the image restoration task to the line extraction task as 1:1 during the training phase. Compared to other weight configurations, our approach is better suited for denoising tasks that aim to maintain the clarity and color of structure lines. We will define this specific loss as follows.

$$\mathcal{L}_{MTUNet} = \mathcal{MSE}(I_S, I_{gt}) + \mathcal{MSE}(L_S, L_{In}) \quad (1)$$

$\mathcal{L}_{MTUNet}$ refers to the total loss objective of the denoising module, $\mathcal{MSE}$ denotes the mean square error, and I_{gt} denotes the target images of denoising. L_S and L_{In} denote the predicted and target structure lines.

3.3 Super-resolution

In order to select an appropriate super-resolution method to enhance the resolution of old cartoons, it is important to be familiar with a variety of super-resolution approaches. After reviewing the literature, we ultimately chose RealESRGAN [35] to improve image quality, as it not only enhances image resolution but also preserves textural information and predicts clear structure lines. Besides the primary task of super-resolution, we consider the clarity of structure lines to be a crucial criterion, as the extracted structure lines from the higher-resolution images will be utilized in the color enhancement stage to further supervise and maintain the integrity of these lines. To maintain the performance of pre-trained models such as Waifu2x [25], RealCUGAN [14], and RealESRGAN [35], we directly test them. For other super-resolution methods like UNet [28] for super-resolution and RCAN [40] designed for natural images, we establish uniform training configurations as follows. We continue to use the Adam optimizer. The learning rate is initially set at 10^{-4} and decays at milestones 500 and 1000 with a multiplicative factor of 0.5. After a random crop, the patch size is 256, and we set the batch size to 64.

3.4 Color enhancement

The task of color enhancement is considered challenging because of the lack of paired supervision between old and remastered cartoons for model training. Some existing methods such as [3, 9–11, 20, 23, 26, 27, 29] considered using reference frames as an additional source of color regularization. However, we argue that selecting reference images can be time-consuming and may result in color inconsistencies between frames or clips. Therefore, we propose a color enhancement model that operates *without any references*. Unlike previous methods, which necessitated supplying the model with a global set of color-specific reference data, our approach employs a color mapping function. Specifically, the gradient of the fix-weight pre-trained color distribution classifier (which works similarly as a non-trainable GAN discriminator) is propagated to the Residual Color Enhancement Module, further regularizing the output color to closely resemble that of recent cartoons. As illustrated in Fig. 3, our method achieves pixel-level color mapping. It is important to note that the corresponding color mappings for different input images will vary.

We designed the Residual Color Enhancement Module with one M-Residual Block and three Residual Blocks. The M-Residual Block is based on the Residual Block and incorporates an additional convolution in the shortcut connection,

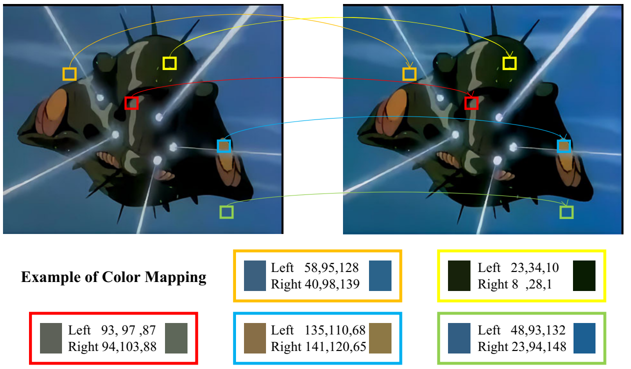


Fig. 3 Pixel-level color mapping. The left and right labels in the box are the values in the RGB space of the input image and the output image of the color enhancement model. We display a colormap of the median value for each image box

allowing the model to fuse information from both branches. Since convolutional layers tend to treat structure lines as general color regions, leading to unclear structure lines in the generated images, we also added a line extraction task by introducing a secondary path of convolutional layers to predict the structural lines of the input frame. The pre-trained color distribution classifier is fine-tuned based on the ResNet-101 classification model [7], achieved by replacing its final fully connected layer with a sequence of a fully connected layer, ReLU, Dropout, and another fully connected layer. The classification model is binary, to predict if the input image is with the color of old-style cartoons. By leveraging the discriminative power of high-level features, we successfully adapted ResNet-101 [7] to classify the color of old and new cartoons.

3.4.1 Training

In the model training process, we first focus on the color distribution classifier. We extracted frames from 17 old and 25 new cartoons at a frame rate of 0.1, removed blank frames and those with little semantic information manually, and ultimately obtained 10,000 new and old cartoon frames each. We set the labels for new and old cartoons as 0 and 1, dividing the total of 20,000 images into a training dataset and a validation dataset with a 9:1 ratio. We trained the color distribution classifier first and then the color enhancement module. The optimizer for both models is Adam. The learning rate for the color distribution classifier is set to 10^{-4} with a weight decay of 10^{-4}. The learning rate for the color enhancement model is 10^{-3} and decays after each epoch with a multiplicative factor of 0.8. Furthermore, the depth of the middle feature representations is 16. During the training process of the unsupervised color enhancement, we encountered two major challenges. The first challenge involves the conflict between color mapping the input image (which aims to *increase* the distance between input and generated images) and maintaining faithful to the input frame (which aims to *narrow* the distance between input and generated images). The second challenge is the difficulty in propagating the gradients of the pre-trained color distribution classifier to the Residual Color Enhancement module.

To address the first challenge, we employ perceptual loss [16] instead of mean squared error to constrain the generated and input images, mitigating the conflict between content preservation and color distribution mapping. Additionally, we added a line extraction task using mean squared error as the loss function to ensure that the generated images have clear structural lines. However, due to the conflict, the structure lines often exhibit abrupt pixel values. We propose using a line-preservation cosine loss to prevent unusual pixel values in the structure lines. For the second challenge, we achieve training balance by reweighting the objectives of image-to-image color enhancement and color distribution classifier objectives to maximize the likelihood of color distribution. Based on experiments, we found that a learning weight ratio of the color distribution classifier and Residual Color Enhancement module of 0.02:1 achieves the balance of training.

3.4.2 Loss

For color enhancement, we apply four loss functions. One is the reconstruction loss function $\mathcal{L}_{Rec}$, one is the discrimination loss function $\mathcal{L}_{Dis}$, one is the line extraction loss function $\mathcal{L}_{line}$ and the last one is line-preservation cosine loss $\mathcal{L}_{cos}$. Among them, the $\mathcal{L}_{Rec}$ mainly constrains the feature gap between the predicted image $Output$ and the input image $Input$ to ensure that the prediction has the same feature content as the input image. The formula for $\mathcal{L}_{Rec}$ is as follows:

$$\mathcal{L}_{Rec}{}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \left\| \phi_j(\hat{y}) - \phi_j(y) \right\|_2^2 \tag{2}$$

$\hat{y}$ represents the predicted image $Output$, y represents the input image $Input$, ϕ represents the neural network VGG19 for extracting features, j represents the number of layers of the VGG19, and we calculate the *ReLU_2_1*, *ReLU_3_1* and *ReLU_4_1* layers here. We mainly use $\mathcal{L}_{Dis}$ to reduce the difference between the predicted label and the target label corresponding to $Input$ so that the color of the generated image is more in line with the color distribution of the new cartoon. The formula is as follows:

$$\begin{aligned} \mathcal{L}_{Dis} &= \frac{1}{N} \sum_i \mathcal{L}_i \\ &= \frac{1}{N} \sum_i - \left[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \right] \end{aligned} \tag{3}$$

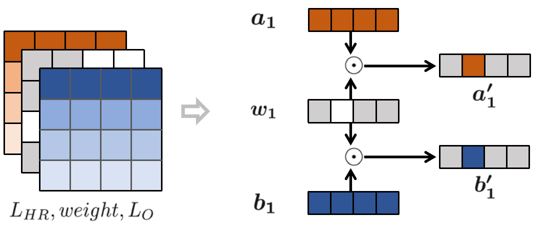


Fig. 4 Reweight of line-preservation cosine loss. $\boldsymbol{a_1}$, $\boldsymbol{w_1}$ and $\boldsymbol{b_1}$ are the first row of high-resolution image L_{HR}, $weight$ and final output L_O, respectively. $\boldsymbol{a_1^{'}}$ and $\boldsymbol{b_1^{'}}$ are the vectors after reweighting

N represents the total number of training dataset, $\mathcal{L}_i$ represents the cross-entropy loss of the i-th image, y represents 1 (the label of the old cartoon in the color distribution classifier), $1 - y$ represents 0 (the label of the new cartoon in the color distribution classifier), p and $1 - p$, respectively, represent the prediction probability of the old and new cartoon images. When $1 - p$ is greater, the probability that the color distribution representing the predicted image belongs to the color distribution of the new cartoon is greater, which means that the predicted label is closer to the target label.

As mentioned in Sect. 3.4.1, we set $\mathcal{L}_{line}$ to ensure the clear lines of the generated images I_O while avoiding outliers due to conflicts between preserving the image content and changing the color. The loss function of $\mathcal{L}_{line}$ is as follows,

$$\mathcal{L}_{line} = \mathcal{MSE}(L_O, L_{HR}) \qquad (4)$$

We introduced additional **line-preservation cosine loss** $\mathcal{L}_{cos}$ because the images generated without line-preservation cosine loss frequently display local swollen structure lines that affect the visual experience. To calculate the line-preservation cosine loss, we first adjust the weights for structure line of high-resolution image L_{HR} and the structure line of final output L_O by $weight = 1 - L_{HR}$ shown as Fig. 4. Then compute the cosine similarity [18] by row and take the mean. Last, we get the line-preservation cosine loss. The formula of the line-preservation cosine loss L_{cos} is as follows:

$$\mathcal{L}_{cos} = 1 - \frac{1}{N} \times \sum_{i}^{N} \frac{\boldsymbol{a_i}' \cdot \boldsymbol{b_i}'}{\|\boldsymbol{a_i}'\| \, \|\boldsymbol{b_i}'\|} \qquad (5)$$

$$\boldsymbol{a}_i^{'}, \boldsymbol{b}_i^{'} = \boldsymbol{a_i} \odot \boldsymbol{w_i}, \boldsymbol{b_i} \odot \boldsymbol{w_i} \qquad (6)$$

N is the number of rows of the matrix, that is, the height of the image. $\boldsymbol{a}_i^{'}$ and $\boldsymbol{b}_i^{'}$ are the vectors reweighted from $\boldsymbol{a_i}$ and $\boldsymbol{b_i}$ of L_2 and L_O by multiplying $\boldsymbol{w_i}$ of $weight$, as the formula 6 shown. The total formula of the loss function in the color enhancement part is as follows:

$$\mathcal{L}_{UCEC} = \lambda_{CLS}\mathcal{L}_{Dis} + \lambda_{RCE}(\mathcal{L}_{Rec} + \mathcal{L}_{line} + \mathcal{L}_{cos}) \qquad (7)$$

λ_{CLS} and λ_{RCE} represent the learning weights of the color distribution classifier and Residual Color Enhancement module. Their values are 0.2 and 1, respectively.

4 Experiments and results

4.1 Denoising

In this section, we first contrast our method with the state-of-the-art denoising methods to evaluate the output quality as well as the preservation of structure lines while denoising.

Visual comparisons In contrast to other current denoising techniques, our enhanced approach effectively prevents the coloration of structure lines due to reference from nearby color block pixels and efficiently eliminates noise from cartoon images, as demonstrated in Fig. 5. From the comparison, it may be concluded from comparing Waifu2x [2, 25] with ours that the denoising part is essential for the cartoon remastering application. Additionally, it is demonstrated from DIDN [22, 36] and our comparison that multi-task learning is effective for improving the network.

Ablations We also carried out ablations to demonstrate the necessity of denoising in cartoon remastering and how the line extraction work helps to guarantee the consistency of the color of the structure lines. The results are shown in Fig. 6.

4.2 Super-resolution

We thoroughly compare the existing single-image super-resolution methods [14, 25, 28, 40] to demonstrate that RealESRGAN [35] is the most effective for enhancing the resolution of cartoons. Figure 7 demonstrates that the structure lines generated by [35] are the clearest, and the optimized structure lines can effectively supervise the line extraction task in the color enhancement stage.

4.3 Color enhancement

We also compare our color enhance model for cartoon remastering purposes, against both the unpaired CycleGAN [42] which does not require a reference, reference-based [37] and low-light enhancement method [24]. As the reference-based method [37] is hard to find the modern-ish look counterparts for each frame for color enhancement, we have to use the input frame as the reference.

Visual comparisons Fig. 8 proves the low-light image enhancement method SCI [24] only focuses on the illumination channel and has nearly no difference from the input. The second and fifth column shows that our method generates images with more vibrant colors and clearer structure lines compared to CycleGAN [42]. Furthermore, DeepEx-

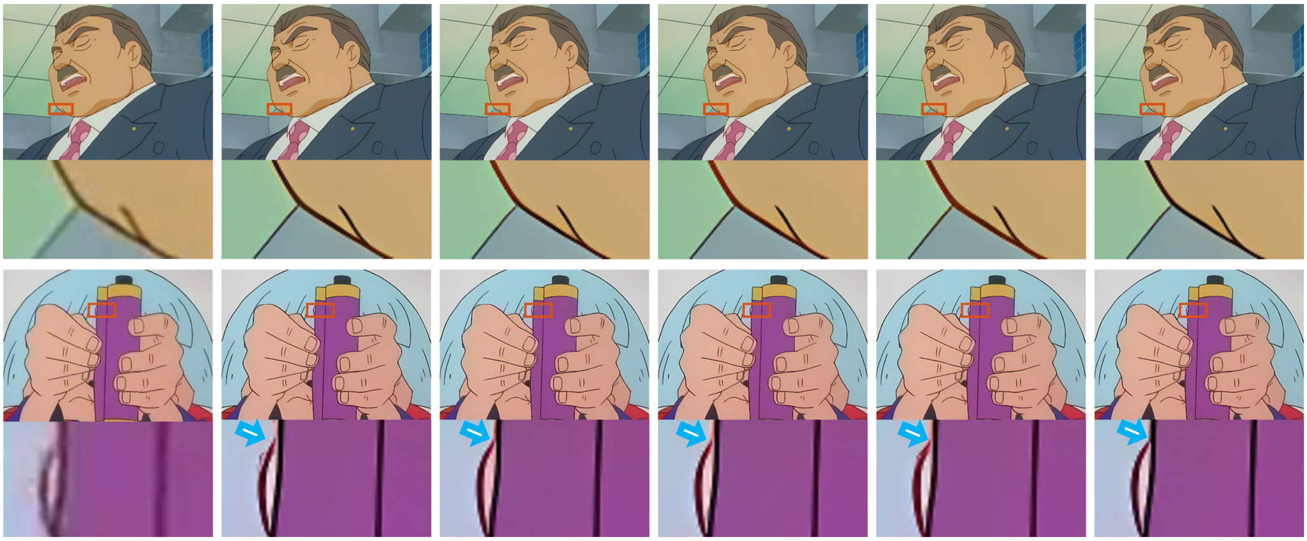


Fig. 5 Comparison results of denoising. The first column shows the inputs after being interpolated with bicubic interpolation. The second to sixth columns are the results generated by methods Waifu2x [25], DIDN [36], SwinIR [22], NAFNet [2], and Ours, respectively. The even rows show the blown-up. To better show the necessity of the denoising part, the example images are all processed by the super-resolution method [35]

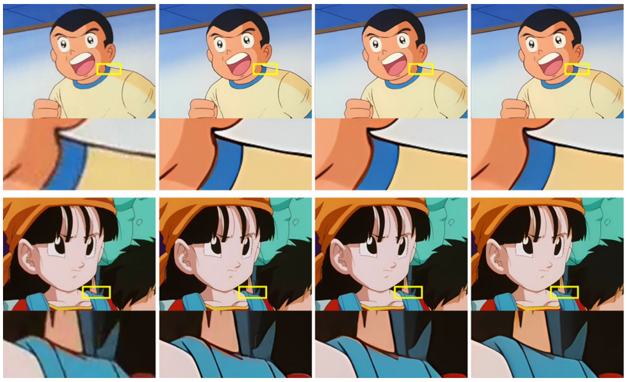
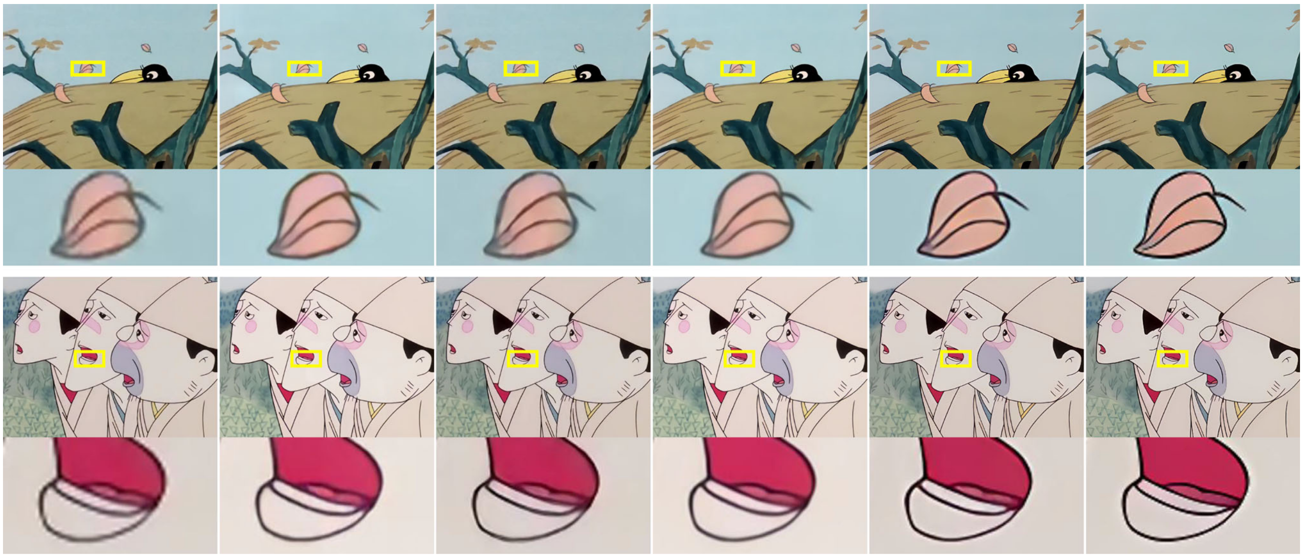


Fig. 6 Ablation for denoising. **a** Input after being interpolated with bicubic interpolation. **b**, **c**, and **d** are increased resolution by the method [35], but **b** the results without denoising, **c** the results of our method without the line extraction task and **d** Ours. The even rows show the corresponding blown-up of images on the odd row

emplar [37] chooses feature information from the reference image that most closely resembles the input image for coloring, but this often leads to uneven color regions due to incorrect references and loss of details. As seen in the second row and third column, the mouse exhibits inconsistent coloration on its belly and loses the shading indicative of lying down. As shown in Table 1, our method is also more efficient than global color mapping techniques [37, 42], nearly 10 times faster. Though our method completed slower than the low-light image enhancement method SCI [24], it has better performance on color enhancement due to the sufficient space to improve the color distribution.

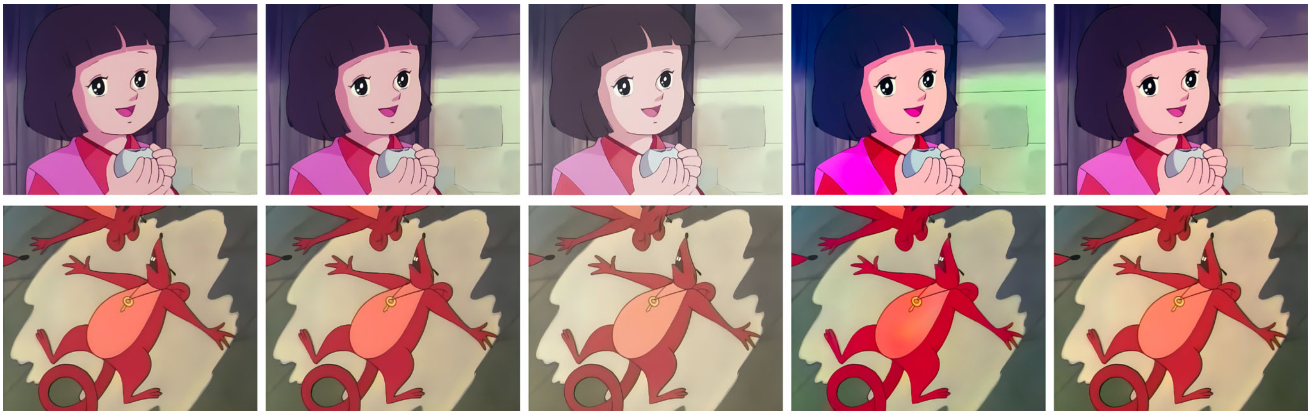
Quantitative evaluation In the color enhancement evaluation, we carried out two distinct types of assessments. The first assessment focused on image quality, while the second was a subjective evaluation based on human perception. We decided not to compare with SCI [24] due to its results showing no difference from the inputs. For evaluating the quality of the color enhancement results, a pixel-wise estimation is not suitable as the pixel values of the enhanced image can significantly differ from those of the input image. Consequently, we utilized the Learned Perceptual Image Patch Similarity (LPIPS) [39] and the Feature Similarity Index Measure (FSIM) [38] to assess the effectiveness of two unsupervised methods, as well as our proposed method. The results are displayed in Table 2 that represents the comparison of these methods. For subjective studies, we collected feedback from 20 participants who were asked to determine which method in the video clip appeared the closest to a new cartoon in terms of color richness, clarity, and temporal smoothness. We presented them with a total of 18 video clips. As shown in Table 3 summarizing the user study, the majority of the participants found that our proposed color enhancement method produced cartoons that more closely resembled new cartoons. Furthermore, we gathered feedback from 28 viewers who provided their overall impressions for each video clip. Our color enhancement method outperformed the others in 18 different cartoons, indicating its superior performance. Note that we decided not to compare them with [37] because it usually produces apparent artifacts.

Ablations We performed ablations in accordance with the single variable principle to show the necessity of each loss



(a) Bicubic (b) UNet (c) Waifu2x (d) RCAN (e) RealCUGAN **(f) RealESRGAN**

Fig. 7 Results of super-resolution. **a**, **b**, **c**, **d**, **e**, and **f** represent the images obtained by upsampling the input image through bicubic interpolation, UNet [28], Waifu2x [25], RCAN [40], RealCUGAN [14], and RealESRGAN [35], respectively



(a) Input (b) SCI (c) CycleGAN (d) DeepExemplar **(e) Ours**

Fig. 8 Comparisons of color enhancement part. **a** is the input, while **b**, **c**, **d** and **e** display the results generated by the methods SCI [24], CycleGAN [42], DeepExemplar [37], and ours, respectively. According to the principle of a single variable, we use only the input images as the reference images for [37]

Table 1 Elapsed time

Method	SCI [24]	CycleGAN [42]	DeepExemplar [37]	Ours
Time (ms) (↓)	4.1	103.0	135.1	19.0

Time (ms) means average milliseconds spent on NVIDIA GeForce RTX 3090 per 1280 × 960-pixel cartoon image. ↓ denotes lower is better

Table 2 LPIPS and FSIM among different methods of color enhancement

Method	CycleGAN [42]	DeepExemplar [37]	Ours
LPIPS(↓)	0.0705 ±0.0315	0.0702 ±0.0151	**0.0165** **±0.0050**
FSIM(↑)	0.9727 ±0.0115	0.9732 ±0.0115	**0.9907** **±0.0026**

All results are listed as mean±std across testing frames. Best results are denoted in bold. ↓ denotes lower is better, while ↑ indicates the opposite

function. As illustrated in Fig. 9, replacing the perceptual loss with the mean square error results in a distorted generated image. This distortion occurs due to the inconsistencies between the two objectives: preserving the image's content and altering the image's color. The former requires a smaller gap between the input and output images, while the latter requires a larger. The sequence from (c) to (d) to (e) reveals that the line extraction task can mitigate the conflict caused by

Table 3 User study of color enhancement. The figures in the table denote the corresponding proportion

Method	Input	CycleGAN [42]	Ours
Richness of color	0.2730	0.1455	**0.5641**
	±0.1261	±0.0793	**±0.0998**
Clearness	0.1663	0.2459	**0.6047**
	±0.0781	±0.1037	**±0.0883**
Temporal smooth	0.2189	0.1971	**0.5829**
	±0.1126	±0.0928	**±0.1324**
Overall consideration	0.2710	0.2122	**0.5159**
	±0.1104	±0.1143	**±0.1023**

All results are listed as mean ± std in all the testing videos. Best results are denoted in bold

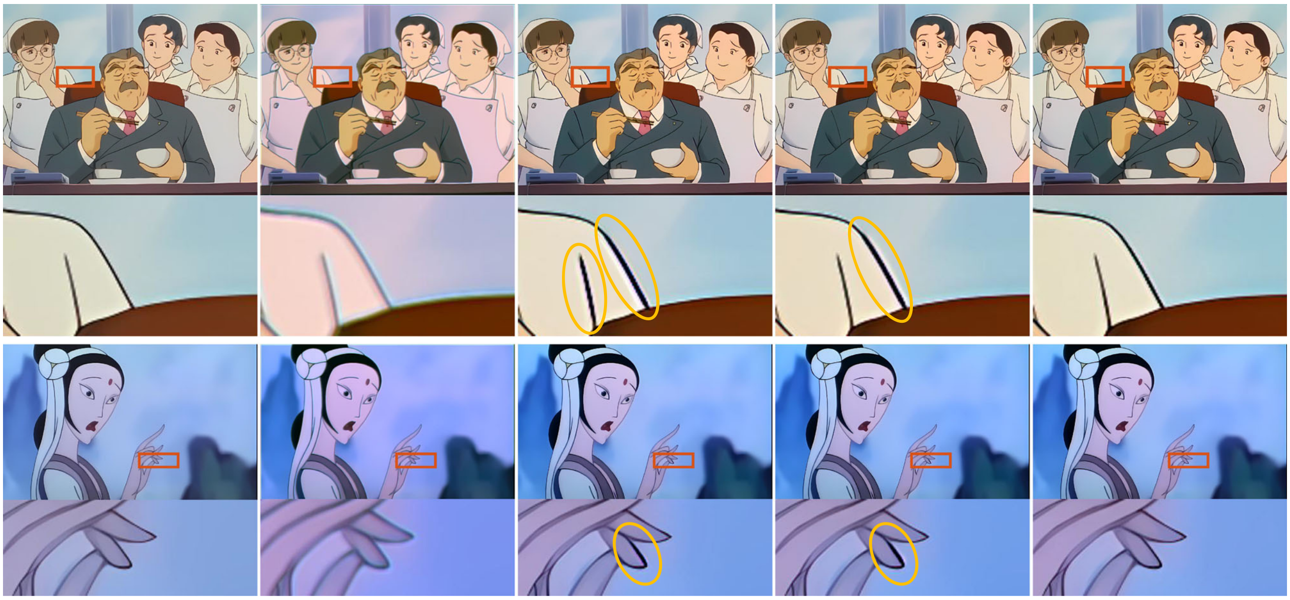
the unusual value of the structure line, highlighted in yellow, in the color enhancement task. However, there is still room for improvement. Furthermore, our suggested line-preservation cosine loss effectively addresses this issue, providing a more consistent solution.

4.4 Cartoon remastering

There are currently existing remastering methods, such as [15] and [34], which are highly relevant to our work. Consequently, we attempted to apply these two methods to the cartoon remastering process and compare them to our pipeline. We use the same approach to upsample the results of the two methods since they do not involve the super-resolution process.

Visual comparisons As in Fig. 1, we observe that Deepremaster [15] does not enhance the image's color because it takes the input images as references. It may be concluded from comparing to Deepremaster that our integrated solution has a better performance in artifact removal and color enhancement. By comparing to Old Photo Restoration [34], we concluded that ours can better retain structure lines. The hue of the man's eyebrows is uneven due to the mutual reference between structure lines and color regions.

Ablations In order to prove the necessity of denoising first, then super-resolution, and finally color enhancement, we compare with the alternatives. If denoising is performed beforehand, the resulting image is likely to enhance the original unwanted noise, negatively impacting the visual impression of the image. This can be seen in the wall of (d), (e), and (f) in rows 2 and 4 of Fig. 10. (b), (c), and (d) demonstrate that when color enhancement is placed before super-resolution, the resulting lines change color due to the reference of general color regions.



(a) Input (b) MSE loss (c) w/o line loss (d) w/o cosine loss **(e) Ours**

Fig. 9 Ablations of color enhancement. **a** the input image; **b** results by replacing the perceptual loss in the reconstruction objective $\mathcal{L}_{Rec}$ with the MSE loss, **c** results by removing the line extraction task from our method, **d** results by extracting structure lines from inputs by [13] to supervise the line extraction task without $\mathcal{L}_{cos}$, **e** adding the $\mathcal{L}_{cos}$ based on **d**

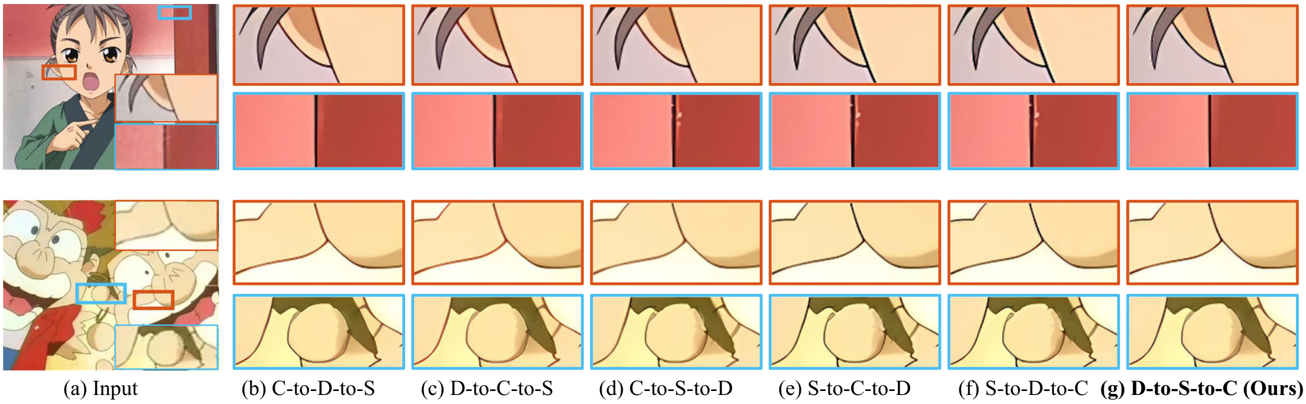


Fig. 10 Ablations of the whole cartoon remastering application. D, S, and C represent denoise, super-resolution, and color enhancement, respectively. (b) C-to-S-to-D means color enhancement first, then super-resolution, denoising at last, and the same applies to other methods

5 Conclusion

In this paper, we propose an application for cartoon remastering, which includes denoising, super-resolution, and color enhancement. In the denoising part, we add a line extraction task to guide the model to focus on the structure line. In the super-resolution part, we choose RealESRGAN [35] to enhance the solution of images after comparing the classic and excellent models. In the color enhancement, we propose a color enhancement model for old cartoons, which in the context of unsupervised. We empirically demonstrate that the proposed application is efficient and effective. Our method is also considerably restricted by the super-resolution method, such as the occasional hollow line, when using super-resolution to generate optimized structure lines.

Funding This work was supported in part by grants from the National Natural Science Foundation of China (Nos. 61973221, 62002232 and 62273241), the Natural Science Foundation of Guangdong Province, China (No. 2019A1515011165), the Major Project of the New Generation of Artificial Intelligence (No. 2018AAA0102900), the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No.UGC/FDS11/E02/21).

Data availibility The data are not publicly available due to the containing information that could compromise the privacy of research participants.

Declarations

Conflict of interest The authors have no conflicts of interest/competing interests to declare that are relevant to the content of this article.

References

1. Afifi, M., Brubaker, M.A., Brown, M.S.: Histogan: Controlling colors of gan-generated and real images via color histograms. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7941–7950 (2021)
2. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. arXiv preprint arXiv:2204.04676 (2022)
3. Chen, S.Y., Zhang, J.Q., Gao, L., He, Y., Xia, S., Shi, M., Zhang, F.L.: Active colorization for cartoon line drawings. IEEE Trans. Vis. Comput. Graph. **28**(2), 1198–1208 (2020)
4. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM Trans. Graph. (TOG) **36**(4), 1–12 (2017)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
6. Gu, C., Lu, X., Zhang, C.: Continuous color transfer. arXiv preprint arXiv:2008.13626 (2020)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
8. He, M., Chen, D., Liao, J., Sander, P.V., Yuan, L.: Deep exemplar-based colorization. ACM Trans. Graph. (TOG) **37**(4), 1–16 (2018)
9. He, M., Liao, J., Chen, D., Yuan, L., Sander, P.V.: Progressive color transfer with dense semantic correspondences. ACM Trans. Graph. (TOG) **38**(2), 1–18 (2019)
10. Ho, M.M., Zhou, J.: Deep preset: Blending and retouching photos with color style transfer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2113–2121 (2021)
11. Hong, K., Jeon, S., Yang, H., Fu, J., Byun, H.: Domain-aware universal style transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14609–14617 (2021)
12. Higumax, I.: Github. [EB/OL]. https://github.com/Dakini/AnimeColorDeOldify Accessed November 2, 2021 (2021)
13. Higumax, I.: Github. [EB/OL]. https://github.com/higumax/sketchKeras-pytorch Accessed August 25, 2020 (2020)
14. Nihui, I.: Real-cugan. [EB/OL]. https://github.com/nihui/realcugan-ncnn-vulkan Accessed July 28, 2022 (2022)
15. Iizuka, S., Simo-Serra, E.: Deepremaster: temporal source-reference attention networks for comprehensive video enhancement. ACM Trans. Graph. (TOG) **38**(6), 1–13 (2019)
16. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14, pp. 694–711. Springer (2016)
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
18. Lahitani, A.R., Permanasari, A.E., Setiawan, N.A.: Cosine similarity to determine similarity measure: Study case in online essay

assessment. In: 2016 4th International Conference on Cyber and IT Service Management, pp. 1–6. IEEE (2016)

19. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 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, pp. 4681–4690 (2017)
20. Lee, J., Son, H., Lee, G., Lee, J., Cho, S., Lee, S.: Deep color transfer using histogram analogy. Vis. Comput. **36**(10), 2129–2143 (2020)
21. Li, Y., Liu, M.Y., Li, X., Yang, M.H., Kautz, J.: A closed-form solution to photorealistic image stylization. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 453–468 (2018)
22. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844 (2021)
23. Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep photo style transfer. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4990–4998 (2017)
24. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5637–5646 (2022)
25. Nagadomi, V.I.: waifu2x. [EB/OL]. https://github.com/nagadomi/waifu2x Accessed October 11, 2015 (2015)
26. Pitie, F., Kokaram, A.C., Dahyot, R.: N-dimensional probability density function transfer and its application to color transfer. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, pp. 1434–1439. IEEE (2005)
27. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2001)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)
29. Shi, M., Zhang, J.Q., Chen, S.Y., Gao, L., Lai, Y.K., Zhang, F.L.: Deep line art video colorization with a few references. arXiv preprint arXiv:2003.10685 (2020)
30. Song, Y., Qian, H., Du, X.: Starenhancer: Learning real-time and style-aware image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4126–4135 (2021)
31. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271), pp. 839–846. IEEE (1998)
32. Torrey, L., Shavlik, J.: Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms Methods and Techniques, pp. 242–264. IGI global, Pennsylvania (2010)
33. Wan, Z., Zhang, B., Chen, D., Liao, J.: Bringing old films back to life. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 17694–17703 (2022)
34. Wan, Z., Zhang, B., Chen, D., Zhang, P., Chen, D., Liao, J., Wen, F.: Bringing old photos back to life. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2747–2757 (2020)
35. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1905–1914 (2021)
36. Yu, S., Park, B., Jeong, J.: Deep iterative down-up CNN for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0 (2019)
37. Zhang, B., He, M., Liao, J., Sander, P.V., Yuan, L., Bermak, A., Chen, D.: Deep exemplar-based video colorization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8052–8061 (2019)
38. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011)
39. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric (2018)
40. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV), pp. 286–301 (2018)
41. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. **34**(12), 5586–5609 (2021)
42. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yinghua Liu received her B.Sc. degree from Anhui University of Science and Technology in 2019. She is currently a graduate student in the College of Computer Science and Software Engineering, Shenzhen University. Her research interests include computer vision and computer graphics.

Chengze Li received their B.Eng. degree from University of Science and Technology of China in 2013, and Ph.D. degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2020. He is currently an Assistant Professor in the School of Computing and Information Sciences, Caritas Institute of Higher Education. His research interests include 2D nonphotorealistic media analysis and processing, computational photography, and computer graphics.

Xueting Liu received her BE degree in Computer Science and Technology from Tsinghua University and Ph.D. degree in Computer Science from The Chinese University of Hong Kong in 2009 and 2014, respectively. Her research interests include computer animation, computer graphics, computer vision, and deep learning.

Huisi Wu received his B.E. and M.E. degrees both in Computer Science from the Xi'an Jiaotong University (XJTU) in 2004 and 2007, respectively. He obtained his Ph.D. degree in Computer Science from The Chinese University of Hong Kong (CUHK) in 2011. He is currently a Professor in the College of Computer Science and Software Engineering, Shenzhen University.

Zhenkun Wen received his M.Sc. degree in Science and Technology from Tsinghua University in 1999. Since 1987, he has been engaged in computing research and teaching in Shenzhen University. He is currently a professor of computing and software, and director of the Science and Technology Department of Shenzhen University. His research interests are in video tampering detection and location, video information security, and information management system design and implementation.