Saliency Guided Deep Neural Network for Color Transfer With Light Optimization

Yuming Fang, *Senior Member, IEEE*, Pengwei Yuan, Chenlei Lv, *Member, IEEE*, Chen Peng, Jiebin Yan, and Weisi Lin, *Fellow, IEEE*

***Abstract*— Color transfer aims to change the color information of the target image according to the reference one. Many studies propose color transfer methods by analysis of color distribution and semantic relevance, which do not take the perceptual characteristics for visual quality into consideration. In this study, we propose a novel color transfer method based on the saliency information with brightness optimization. First, a saliency detection module is designed to separate the foreground regions from the background regions for images. Then a dual-branch module is introduced to implement color transfer for images. Finally, a brightness optimization operation is designed during the fusion of foreground and background regions for color transfer. Experimental results show that the proposed method can implement the color transfer for images while keeping the color consistency well. Compared with other existing studies, the proposed method can obtain significant performance improvement. The source code and pre-trained models are available at https://github.com/PlanktonQAQ/SCTNet.**

***Index Terms*— Color transfer, saliency detection, light optimization.**

I. Introduction

LIMITED by the performance of camera devices and complex environmental conditions, the visual quality of a raw image cannot be guaranteed to satisfy the requirements in practice. To obtain the images with good quality, a series of image enhancement techniques have been proposed to improve the image properties, including contrast, color, resolution, clarity, *etc*. As one type of image enhancement, color transfer aims to improve the visual quality of image color. It establishes color mapping function between different images and transfer the color of the reference image with good quality to the source one. It is widely used in various applications including computational/digital imaging systems, medical testing, cultural relics restoration, *etc*.

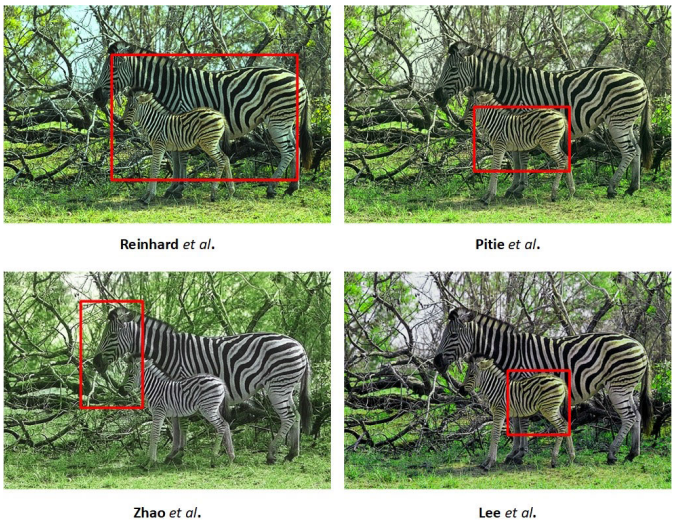

Fig. 1. Instances of classical color transfer methods. The regions with mixed colors (labeled by red boxes) reduce the saliency-based subjective visual sensitivity. The methods are: Reinhard et al. [1], Pitie et al. [2], Zhao et al. [3], and Lee et al. [4].

Following the pioneer work by Reinhard et al. [1], many researchers attempt to establish global color mapping function between different images to implement color transfer. The basic assumption is that the color between images can be transferred by the correspondence of global color distributions. Once the correspondence is established, the color transfer can be implemented by the related global color mapping function. The solution reflects the accurate color distribution properties in global view and keeps the color consistency. However, the local color distribution properties and the semantic correspondence are ignored. Thus, it limits the flexibility and precision of the color transfer in related tasks.

To improve the performance, many semantic correspondence based methods are proposed to implement more accurate color transfer [5], [6]. They attempt to search the semantic correspondence areas between the reference and the source images, which help establish the color mapping function to balance different factors, including global color distribution, semantic correspondence, and color consistency. Benefited from the semantic correspondence, the scheme can achieve

Manuscript received 9 January 2023; revised 29 August 2023 and 28 December 2023; accepted 11 March 2024. Date of publication 12 April 2024; date of current version 16 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62132006 and Grant 62311530101; and in part by the Natural Science Foundation of Jiangxi Province of China under Grant 20223AEI91002, Grant 20232BAB202001, and Grant 20224BAB212012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giulia Fracastoro. *(Corresponding author: Chenlei Lv.)*

Yuming Fang, Pengwei Yuan, Chen Peng, and Jiebin Yan are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032, China (e-mail: fa0001ng@e.ntu.edu.sg; 1317091982@qq.com; ispengchen@outlook.com; jiebinyan@foxmail.com).

Chenlei Lv is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: chenleilv@mail.bnu.edu.cn).

Weisi Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

Digital Object Identifier 10.1109/TIP.2024.3381833

more accurate and flexible color transfer. The performance of these methods depends on the accuarcy of semantic correspondence. Once the semantic regions of images are affected by some factors such as abnormal exposure, various views, noise, and motion-based blur, the quality of the color transfer is reduced with higher probability. With the development of deep learning techniques, some recent methods use deep neural network to improve the semantic region detection in the color transfer. These works can achieve more accurate semantic regions while keeping better robustness to transitional schemes. As we know, color information is sensitive to the human visual system and the perceptual mechanism should be taken into consideration in color transfer, which is ignored by previous works. The semantic regions with mixed colors are difficult to be distinguished and this reduces the saliency-based subjective visual sensitivity (some instances are shown in Fig. 1).

In this paper, we propose a saliency-guided color transfer deep network with brightness optimization (SCTNet). It is constructed by three parts: separation module, color transfer module, color fusion. The separation module is designed to highlight foreground objects in images. The color transfer module, i.e., the core part of the proposed model, adopts a dual-branch color transfer to focus on the salient regions to ensure the accuracy of color transfer, therefore improving the flexibility and accuracy of the model. The color fusion module is an auxiliary module for better performing color fusion and brightness optimization on the transferred images and improving the visual quality. The contributions of the proposed method are summarized below.

- A separation module is proposed based on U2 saliency detection model and Yolact model with multi-scale anchor method for salient foreground objects from the input image, to significantly improve the segmentation accuracy of large objects and the effect of color transfer.
- A color transfer module is presented by a double branch structure, which transfers the color information of the foreground and background regions, respectively. It can effectively improve the accuracy and flexibility of the color transfer model while enhancing the contrast of the salient area.
- A color fusion method is designed to aggregate foreground and background color transfer results with lighting optimization. It generates more accurate color transfer result.

The pipeline of our method is shown in Fig. 2. The rest of the paper is organized as follows. In Sec. II, we discuss the related works about color transfer. In Sec. III-V, we introduce the implementation details of the separation module, color transfer module and color fusion. We show the performance of proposed method in Sec. VI and Sec. VII provides the conclusion.

II. Related Work

According to the recent survey [7], color transfer methods can be divided into several categories, including statistical color transfer, semantic correspondence-based color transfer, and deep feature-based color transfer. In following parts, we discuss such categories and saliency detection methods.

A. Statistical Color Transfer

According to the scope of color mapping, the statistical color transfer methods can be divided into two categories: global color transfer and local color transfer. The global color transfer methods [1], [8], [9] utilize the global color statistics to implement color transfer between images. Reinhard et al. [1] proposed the pioneer work based on $l\alpha\beta$-based global color transfer scheme. Pitie et al. [8] mapped the N-dimensional distribution of the reference image to the target image by applying a designed novel continuous transformation method. Pitie et al. [10] constructed a one-to-one color mapping that transfers the palette of an example target picture to the original picture. Pouli and Reinhard [11] introduced different scale histograms to consider coarse and fine features separately for color transfer. Wang et al. [12] learned implicit mappings with enough examples for global color transfer. Abadpour and Kasaei [13], and Xiao and Ma [14] used principal component analysis (PCA) to compute the decorrelated color space of the input image. Chang et al. [15], [16] classified colors into perceptual-based categories, and then perform color transfer based on this classification by restricting the resulting colors to similar categories. Su et al. [17] proposed a self-learning filtering scheme and a multi-scale detail manipulation scheme to construct a probabilistic color map between the source image and the reference image. Global color transfer is mainly designed based on the selected statistical information, and it is difficult to avoid the following two problems: 1) the global transfer method is prone to local color confusion if the source image or the destination image contains regions of different colors; 2) if the color distributions of the two images are very different, the results tend to appear unnatural and saturated with the $l\alpha\beta$ color space.

In order to overcome the above limitations, many local color transfer algorithms are proposed based on local spatial relationship. For these algorithms, representative colors are first obtained from the reference image to the source image (such as dominant colors [18], [19], [20], [21], feature point colors [22]). Then, the color correspondences between the source and reference images are established. Tai et al. [20] proposed a local color transfer model based on probability segmentation, which segments the source image and the reference image by probability, and then uses the Gaussian mixture model to model the components of different regions respectively. Xia et al. [23] proposed a saliency guided image transfer model, which obtains the color information of the salient regions of the original image and the reference image, and establishes the corresponding color transfer relationship. Wen et al. [24] used sketches in the source and target images to define the corresponding regions. Yoo et al. [25] applied statistical transfer by exploiting their dominant colors to find local region correspondences between two images. Li et al. [26] proposed a local color transfer method to extract salient regions. Xiao and Ma [27] adopted a local color transfer method of histogram matching and color gradient optimization

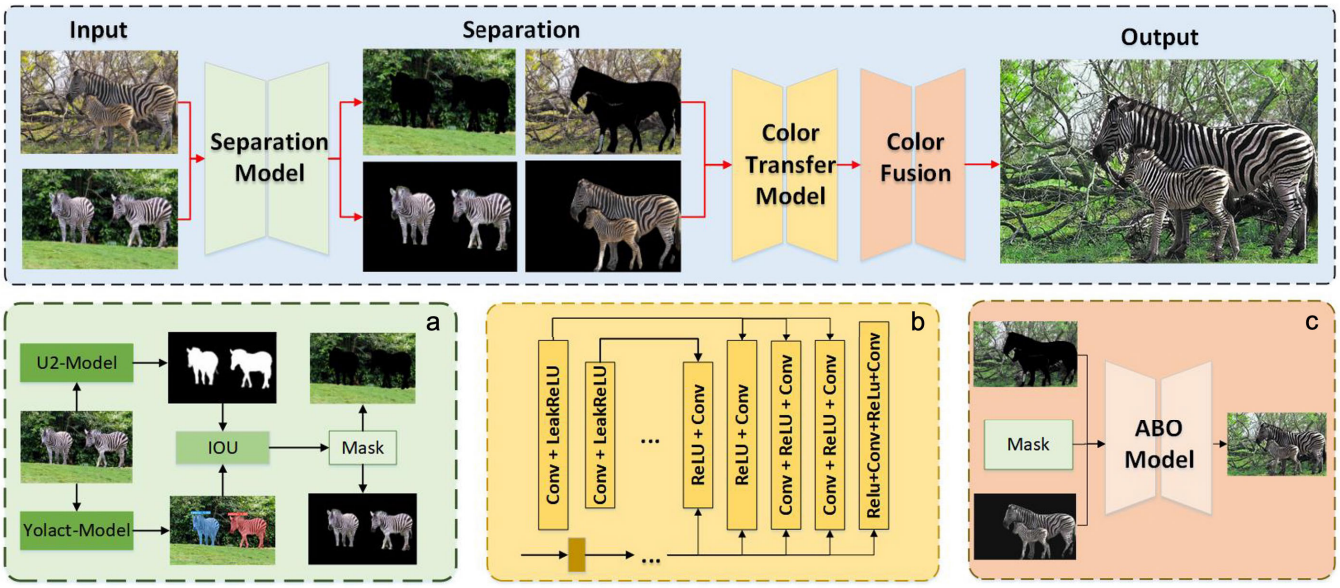

Fig. 2. The pipeline of the proposed method. (a) Separation Module; (b) Color Transfer Module; (c) Color Fusion.

to alleviate the fidelity problem of color transfer in terms of scene details and colors. Although the local color transfer methods can effectively alleviate the limitations of the global mapping scheme, the advantages might be lost if a general color transfer is performed when the reference image and the original image lack an obvious semantic relationship.

B. Semantic Correspondence-Based Color Transfer

In order to improve the performance of color transfer, many researchers have tried to establish the semantic correspondence between the original image and the reference image, thereby improving the accuracy and flexibility of the color transfer process. Based on the color transfer work of dense correspondence, HaCohen et al. [5] proposed to use the improved GPM [28] to obtain semantic corresponding regions, and cluster the consistent matching regions from coarse to fine through aggregating consistent regions to find the dense correspondence of images, then the image color transfer is completed by learning the global color change between corresponding regions. Considering that patch correspondence and dense point correspondence are time-consuming, some researchers propose to use a color transfer algorithm based on sparse correspondence. Park et al. [6] proposed a color transfer method based on SIFT keypoint matching, a color matrix was constructed based on a series of image features to fit the local feature description, then the color parameters were calculated by matrix decomposition to establish the color transfer model.

In addition to dense correspondences and sparse features for semantic correspondence schemes, Tai et al. [20] and Hristova et al. [29] employed the EM algorithm to find corresponding color clusters in images. Arbelot et al. [30] proposed a unified framework that uses edge-aware texture descriptors of the regional covariance of images to establish correspondences and conduct guided color transfer. He et al. [31] used neural representations for more accurate color transfer between objects for semantically meaningful dense correspondences. Gao et al. [32] proposed a multi-label semantic description of wallpaper textures for image style transfer. Liao et al. [33] proposed a new semantic context-aware image style transfer method, which carried out style transfer by implementing semantic context matching. Li et al. [34] proposed a new global and local semantic coloring method. However, the above schemes does not solve the problem of semantic consistency, and in the process of color transfer, the situation of local color confusion still exists. Although these methods address the global color transfer problem when the source and reference images have similar structures, the model robustness still suffers when these two images have completely unrelated content and styles.

C. Deep Feature-Based Color Transfer

The statistical learning and the semantic corresponding color transfer methods still have some problems. The statistical methods generally have the problem of poor robustness and local image color imbalance. Although the improvement scheme based on semantic correspondence alleviates to a certain extent, the problem still exists. Subsequently, many models based on deep learning techniques have been proposed to solve these problems. Due to the properties of convolutional neural networks, such works can achieve more accurate semantic regions while maintaining better robustness to the color transfer process. He et al. [35] proposed to use the reference image to guide the color transfer by combining the similarity network and the colorization network, using the feature vector encoded by the CNN network to establish the semantic correspondence between the original image and the reference image pixels, and successfully transferred to the video color transfer [36]. Liu et al. [37] proposed a color transfer framework including feature network, sentiment classification network, fusion network and colorization network. Liao et al. [38] proposed a depth image analogy technique to establish a dense semantic correspondence between two input images for color

transfer. Song and Liu [39] proposed a new image appearance transfer method that combines depth color and texture features. Gatys et al. [40] extracted leveraging features by a pretrained VGG network [41], and iteratively optimizing the output color transfer images. Xu et al. [42] proposed an end-to-end two-subnet color transfer scheme, where the previous part uses the codec architecture to replace the similar subnets in [35] to establish the semantic correspondence between the original image and the reference image, and uses AdaIN to perform feature matching and blending to speed up color transfer work.

However, it is difficult to obtain a suitable reference image, Zhao et al. [3] proposed a method to randomly intercept patch blocks from the ground truth image as the reference image, and use the color network to extract the color information as the color information in the grayscale image coloring process. Huang et al. [43] proposed an innovative high dynamic range image color transfer generative adversarial network (HDRCT-GAN), which encodes raw images and learns fine features in a self-supervised manner through a generative adversarial network (GAN). Zhao et al. [44] proposed a fully automatic saliency guided coloring and generation adversarial network. Dong et al. [45] proposed a spatially uniform crossed attention (SCCA) block to encourage the slice pixels of different outer polar lines in the gray image to correspond spatially to the pixels of the reference color image. Dou et al. [46] propose a new sketch coloring method, two-color space-guided Generative Adversarial Network (DCSGAN), which takes into account complementary information contained in RGB and HSV color spaces. Besides, there are many different studies on color transfer based on deep learning features, e.g. [47], [48], [49], [50], and [51] etc. However, in these works, salient regions or foreground objects can not be well enhanced, making the color distribution of the corresponding semantic regions indistinguishable, which reduces the saliency-based subjective visual sensitivity.

D. Saliency Detection and Instance Segmentation

Saliency detection plays a crucial role in visual scene analysis as it is capable of identifying and extracting objects that attract human attention. In recent years, research efforts have mainly focused on convolutional neural networks (CNN) using an encoder-decoder architecture. Hou et al. [52] incorporated short connections to fuse global and local information from deep and shallow layers. Liu et al. [53] proposed a pyramid pooling module to capture global semantic information from the encoder and recovers sparse information in the encoder through a feature aggregation module. Qin et al. [54] designed the Residual U (RSU) block to capture more contextual information from different scales and utilizes pooling operations within RSU to increase depth without significantly increasing computational cost. With the introduction of Transformer models to visual tasks, some studies have extended Transformers to the task of saliency detection. Yun and Lin [55] utilized Transformer as the encoder backbone network and developed a branch to learn global context, achieving outstanding performance.

Instance segmentation aims to predict a pixel-wise mask for each object of interest. He et al. [56] introduced Mask R-CNN, an extension of Faster R-CNN that incorporates a dedicated branch to predict object masks alongside the existing bounding box recognition branch. Yolact, proposed by Bolya et al. [57], decomposes instance segmentation into two parallel subtasks, leading to fast segmentation. Chen et al. [58] introduced BlendMask, which utilizes instance-level information, semantic information, and low-level fine-grained information to achieve improved mask prediction. Cheng et al. [59] proposed SparseInst, which employs sparse instance activation maps to emphasize the informative areas of foreground objects, and subsequently accomplishes segmentation through feature aggregation.

Besides the saliency used as the guidance for color transfer, the proposed method introduces and brightness optimization, to change the color distributions for the foreground and background independently, which improves the flexibility and accuracy of the color transfer model. In the following parts, we will introduce the details of the proposed method.

III. Separation Module

In this part, our primary goal is to separate salient object from their backgrounds. To achieve this, we have developed a specialized separation module. The purpose of this module is to effectively extract salient objects while preserving the overall structure of the target. To attain stable and significant foreground object segmentation while minimizing computational costs, we have opted for the efficient saliency extraction model U2Net [54] and the instance segmentation model Yolact [57].

As shown in Fig. 2, an instance of segmentation model is shown for segmenting objects. It extracts salient regions as the input for the subsequent color transfer model, which can be formulated as follows:

$$I_s^f, I_r^f, I_s^b, I_r^b = SegNet(I_s, I_r), \tag{1}$$

where I_s is the source image, I_r is the reference image, I_s^f refers to the foreground area map created by the source image after passing through the separation module, and I_s^b refers to the background area map created by the source image after passing through the separation module. The subscript f indicates foreground, and subscript f indicates background. Take I_s as an example, I will proceed with a more meticulous exposition of this module. And the processing procedure of I_r is the same as I_s

$$M_i^s, S^s = Yolact(I_s),\ U2Net(I_s), \tag{2}$$

$$M_s = Max((S^s \cap M_i^s)/(S^s \cup M_i^s)), \tag{3}$$

where M_i^s is the output of the instance segmentation model for the source image, and $i \in \{0, \ldots, 4\}$ indicating the selection of the top 5 instance results with the highest confidence scores, and M_i^s contains five different instance object masks of the source image. Where S^s is the result of the Salience extraction model, it only contains the salient object masks of the source image. Finally, we opt for the instance segmentation object that has the greatest IoU (Intersection over Union) with the

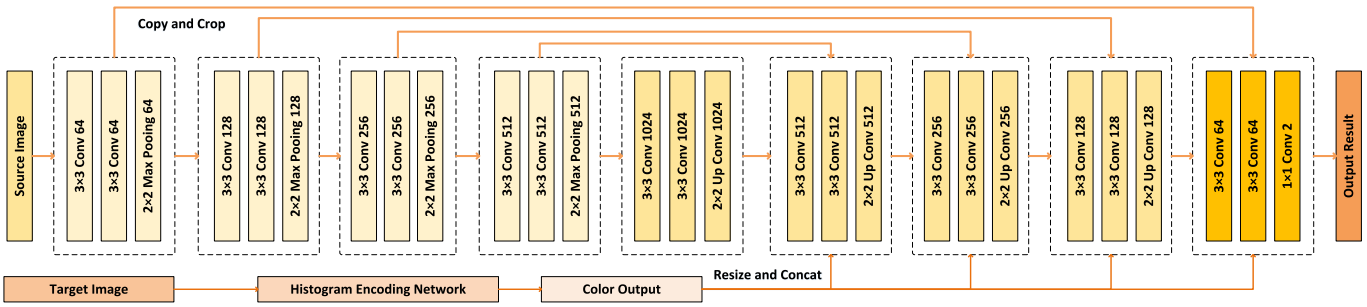

Fig. 3. The main network structure of the dual-branch color transfer, in which the color transfer model parameters are shared, and the corresponding color transfer is performed on the segmented foreground image and the corresponding color transfer on the background image.

salient object as the final mask M_s for the source image

$$\begin{cases} I_s^f = I_s * M_s \\ I_s^b = I_s - I_s^f, \end{cases} \tag{4}$$

which is represented as a boolean matrix, and leveraging the broadcasting mechanism, we perform a matrix multiplication between the mask M_s and the input image I_s to obtain the corresponding saliency object I_s^f. Similarly, subtracting the saliency image I_s^f from the input image yields the background image I_s^b.

As we know, it is a fact that color mixing is prone to occur in some regions between objects. Without elaborative processing of the color transfer of objects and scenes in different semantic backgrounds, the consistency of brightness and chroma would be too high, and the contrast of the generated images is reduced, which reduces the overall visual quality of generated image and destroy their texture information. To effectively alleviate the above problems, the proposed method proposes a separation module which consists of an improved real-time instance segmentation model and a saliency detection model. It is used to separate salient regions in the original image and the reference image, and as the pre-processing part of the entire SCTNet color transfer architecture.

In order to better ensure the texture of the generated image and highlight the color transfer effect between the foreground and the background, we need to accurately obtain the semantically corresponding foreground regions with larger scales in the original image and the reference image. For the above considerations, we believe that the accuracy of large image objects in the instance segmentation process should be improved, and we have made the following improvements. As far as we know, the setting of anchor size and aspect ratio has a great impact on the performance of instance target segmentation methods. Intuitively, the size and aspect ratio of the target varies greatly between tasks. In terms of length and width, setting an anchor point similar to the target is more conducive to the segmentation of the target. In order to improve the effect of segmentation of large objects in images, we design a multi-scale anchor (MSA) module to get more reasonable anchors. The size of each layer is changed to three ratios of the original size $\left\{1x, 1.6^{\frac{1}{3}}x, 2^{\frac{2}{3}}x\right\}$ and we enlarge the anchor size from [24, 48, 96, 192, 384] to [32, 64, 128, 256, 512]. The experimental results prove that they can obtain better performance than the traditional Yolact segmentation model [57].

IV. Color Transfer Module

After obtaining the foreground and background, we aim to achieve end-to-end color transfer through a module. In order to ensure the accuracy and consistency of color transfer, we have specifically designed a dual-branch color transfer module that shares weights between the foreground and background.

As shown in Fig. 3, when building the main network of the color transfer model, a dual-branch architecture is adopted, which includes a salient foreground branch and a background branch. The foreground branch takes the salience target image extracted by the instance segmentation model as the input, then carries on the color transfer to the foreground salience target of the original image and the reference image, and outputs the result of foreground color transfer. The background branch takes the background image of the segmented source image and reference image as the input for color transfer, and outputs the background color transfer results.

$$\begin{cases} I_c^f = ColNet(I_s^f, I_r^f) \\ I_c^b = ColNet(I_s^b, I_r^b), \end{cases} \tag{5}$$

where $I_s^f, I_r^f, I_s^b, I_r^b$ results from our separation model as a color transfer model input volume, after ColNet model (Foreground Background Color Transfer Network) obtains the I_c^f and I_c^b color transfer results.

A. Foreground-Background Color Transfer Network

When building the two-stage color transfer model SCTnet network module (b), a dual-branch architecture is adopted, which includes a salient foreground branch and a background branch. The specific situation is shown in Fig. 3, where the foreground branch uses the salient object extracted by the first-stage sub-network as input, and performs color transfer for the foreground salient object to obtain the foreground mapping results.

The two branches use the same network architecture and implement parameter sharing, including encoder-decoder, refinement and auxiliary modules, where the encoder-decoder uses U-Net as the main color transfer network. In fact, we choose U-Net as the color transfer module, because its effectiveness has been demonstrated in the computer vision

TABLE I
THE IMPACT OF α AND β ON IMAGE QUALITY

	α:10 β:180	α:30 β:160	α:30 β:190	α:50 β:200
SGDnet ↑	3.440	3.441	**3.442**	3.441
NIQE ↓	**3.595**	3.610	3.610	3.621
IL-NIQE ↓	**23.043**	23.068	23.070	23.150

community, and our backbone is adapted from the work by Lee et al. [4].

As shown in Fig. 3, our color transfer model consists of two modules: an encoder module and a U-Net module. Specifically, our network consists of 9 convolutional blocks. Each convolutional block contains 3 Conv-relu units followed by a batch normalization layer, except the last block. The feature maps in the former 4 convolutional blocks are gradually halved in space while doubling the number of feature channels. To aggregate multi-scale contextual information without loss of resolution. In the 4 convolutional blocks of the decoder, the size of the feature maps are gradually doubled in space, while the number of feature channels is halved, and the output features are concatenated with the features of the auxiliary module. All downsampling layers use stride 2 convolution, while all upsampling layers use stride 2 deconvolution. Symmetrical jump connections have been added between the outputs of the 1nd and 9th, 2rd and 8th, 3rd and 6th, 4th and 6th blocks, respectively. Finally, a convolutional layer with kernel size 1 $\times$1 is added after the 9th block, and at the end of each layer of the decoder, we append an auxiliary layer to guide the multi-scale prediction of the color transfer image.

We design a histogram encoding network based on convolutional neural network, which encodes the image histogram into color feature vector. The histogram H of the image I as input to the HEN network. In the Lab color space, we first divide the image I into I^l and I^{ab}, which respectively correspond to the brightness and color information of the image. Second, we compute the histograms $H^l \in \Lambda^p$ of I^l and $H^{ab} \in \Lambda^{q*r}$ of I^{ab}, where p, q, r are the histogram sizes. We tile H^l and concatenate it with H^{q*r} in the channel direction, resulting in $H \in \Lambda^{q*r*(p+1)}$. HEN takes H and generates an encoded histogram $e \in \Lambda^k$, where k is the histogram feature vector size. In our work, we use $q = 64$, $r = 64$, $p = 8$, and $k = 64$. We determined the values of alpha and beta through a series of experiments. The main criterion for selection was evaluating the quality of the results using various Image Quality Assessment (IQA) methods. According to Table I, by comparing a series of different values for alpha and beta, we observe that the image generated has the better quality when alpha is 10 and beta is 180.

Finally, based on the front and back background result images by the dual-branch color transfer model, the foreground information I_f and the background I_b information are weighted and fused on different lab channels to obtain the final color output transfer result I_c.

B. Loss Function

The training process of color transfer deep model is difficult. On the one hand, we expect that color transfer network can not only transfer "correct" colors to "correct" regions based on semantic features, but also refer to meaningful colors of objects when the reference is not available. On the other hand, there is usually no exact Ground Truth to supervise the network to learn the corresponding color transfer process. In order to effectively solve the above two problems, we construct two perceptual loss terms to measure the content difference between the output image I_o and the input source image I_s, and the style difference between I_o and the input reference image I_r. The first loss function includes two items, which are denoted by $\mathcal{L}_{mse}$ and $\mathcal{L}_{vgg}$, respectively, and these two items can be calculated as follows:

$$\mathcal{L}_{mse} = \frac{1}{n}\sum_{i=1}^{n}(I_o - I_s)^2, \tag{6}$$

$$\mathcal{L}_{vgg} = \frac{1}{N_l}\sum_{i=0}^{N_l}\left\|\phi_i\left(I_o\right) - \phi_i\left(I_s\right)\right\|_2, \tag{7}$$

where $\phi_i(\cdot)$ denotes features extracted from the i-th layer in a pretrained VGG19 and N_l is the number of layers. To better transfer the color information, we use the calcStylLoss [23] to measure the color loss between the generated image and the reference image. The specific formula is calculated as follows:

$$\mathcal{L}_{color} = \frac{1}{n}\sum_{i=1}^{n}(gramMat\left(I_r\right) - gramMat\left(I_o\right))^2, \tag{8}$$

By separately computing the inner product of I_r and I_o to obtain their $gramMat$, and then calculating the mean square error between the corresponding $gramMat$ to obtain L_{color}. The Gram matrix is a representation method used to describe the texture features of an image, and it is computed based on the inner product between the feature vectors of the image. In color transfer tasks, by calculating the similarity between the Gram matrices of two images, it is possible to quantify their texture differences and optimize the process of generating images.

In this paper, we address the training problem of a color transfer model in the form of a weighted loss function:

$$l_{total} = \lambda_1 * \mathcal{L}_{vgg} + \lambda_2 * \mathcal{L}_{color} + \lambda_3 * \mathcal{L}_{mse}, \tag{9}$$

where λ_1, λ_2, and λ_3 control the relative importance of the corresponding terms respectively, while λ_1, λ_2 and λ_3 are set to 0.2, 0.00002 and 0.8.

V. COLOR FUSION

In order to better adjust the brightness optimization of the foreground and background images, we design the following image color fusion module with adaptive brightness optimization (ABO) module based on saliency guidance. Based on the foreground and background result images by the dual-branch color transfer model, the foreground information I_f and the background information I_b are weighted and fused on different channels to obtain the final color output transfer result I_o. The specific formula is as follows:

$$I_o = mask * I_f * W_f + (1 - mask) * I_b * W_b, \tag{10}$$

In the above formula, the mask comes from the significance target mask obtained by the separation module. WF and WB are super parameters, which determine the change of brightness. The calculation formula is as follows:

$$W_f = \begin{cases} \frac{L_f}{L_f + L_b} + 1, & L_f <= \alpha \\ 1, & \alpha < L_f < \beta \\ \frac{L_f}{L_f + L_b}, & L_f >= \beta, \end{cases} \tag{11}$$

$$W_b = \begin{cases} \frac{L_b}{L_f + L_b} + 1, & L_b <= \alpha \\ 1, & \alpha < L_b < \beta \\ \frac{L_b}{L_f + L_b}, & L_b >= \beta, \end{cases} \tag{12}$$

where W_f and W_b are computed from the adaptive calculation of the brightness, which are set based on the foreground image and the background image after the color transfer. On this basis, we use α and β to constrain the range of brightness value, which are calculated as:

$$L_f = (0.299 * I_f^r) + (0.587 * I_f^g) + (0.114 * I_f^b), \tag{13}$$

$$L_b = (0.299 * I_b^r) + (0.587 * I_b^g) + (0.114 * I_b^b), \tag{14}$$

where I^r, I^g, I^b are the three channel values of the image; the α and β values are set to 10 and 180 respectively.

VI. Experiments

A. Evaluation Methodology

1) Model and Parameter Settings: We select training data of fixed categories (zebra, airplane, truck, bus, bird) from the COCO dataset [60]. The size of the dataset we used for training consists of 8540 pairs of images, and the test dataset consists of 987 pairs of images. First, we use the self-designed and pre-trained separation model to extract the saliency mask on the training data, and then calculate the foreground and background information of the original image and the target image by using the mask. Second, the weight of the deep color transfer model [4] is used to initialize our two-branch color transfer network. Third, the learning rate is set to e^{-5}, and the network is trained 5 epochs. Throughout the training process, we use ADAM optimizer with $\beta_1 = 0.99$ and $\beta_2 = 0.999$ to update the parameters. All images are resized to 256×256 resolution, and it costs about one day to train the model on a desktop with a single 3090Ti GPU.

2) Compared Methods: We compare the proposed model with five color transfer methods, including GCT [1], IDT [8], MKL [2], C2E [3], DCT [4], WCT [61], PhotoWCT [62] and WCT2 [63]. GCT, IDT and MKL are classical color transfer methods, which rely on global color statistics. C2E, IDT, WCT, PhotoWCT and WCT2 are deep learning-based methods. Among of them, PhotoWCT and WCT2 are the neural network-based method for photorealistic style transfer.

3) Evaluation Metrics: Since the color transfer work lacks Ground-Truth, we use three no-reference metrics, including SGDnet [64], NIQE [65] and IL-NIQE [66]. Among them, the SGD method [64] emphasizes the evaluation of the image quality of the salient area of the image, and the higher the value, the better; IL-NIQE [66] integrates the color statistical features, structural statistical features, multi-scale orientation and frequency statistical features, and it is based on NIQE [65]. At the same time, we use other two quality evaluation indicators, including contrast and entropy, to calculate the quality of generated images.

For visual comparative evaluation, we test the compared models on 987 pairs of images, and show several representative results in Fig. 4. It is more aligned with our expectations when both the source and reference images contain a distinct object. So, we also select images from the COCO dataset, which focus on specific categories such as zebras, airplanes, trucks, buses, and birds, to construct our test dataset. From Fig. 4, we can clearly see that the deep learning based methods can produce perceptually more attractive results than traditional methods. It is not difficult to find that the salient object of other schemes has obvious color mixing with the background color, the color contrast of the overall image is low, and the color of the salient object cannot be well represented. However, our method can better solve the color mixing problem, on the one hand, it has good color similarity with the reference image, and on the other hand, the details of the image can be better preserved.

B. Quality Results

1) Quantatitive Analysis: Table I shows that the comparison results with the above eight methods using the no reference quality evaluation methods. Due to the lack of ground-truth label data for the transferred images, we employ three classical quality assessment methods [64], [65], and [66] to evaluate the quality results of the generated images from multiple perspectives.

As shown in Table II, WCT2 exhibits some advantages in terms of NIQE, while our method still maintains the best performance in terms of the SGDNet and IL-NIQE metric. In the visual comparison in Fig. 4, it can be observed that WCT2 still produces some inaccurate color transfer results for certain objects. In contrast, our model is able to maintain better color consistency in the images and also performs very well in color transfer between salient objects.

To further demonstrate the generalization of our scheme more concretely and accurately, we randomly selected 1000 pairs of image data from the ImageNet dataset and generate related color transfer results. In Fig. 6, we present a visual comparison between our method and other methods.It has been proved that Our method better preserves the color consistency of the results compared to other methods, and it also maintains superior color transfer between salient objects. Furthermore, we conducted a no-reference image quality assessment on the 1000 images generated by the color transfer method. The results are presented in Table III. Based on these metrics, including NIQE [65], IL-NIQE [66], and SGDnet [64], our method achieves the best results among color transfer schemes, which demonstrate its superiority over other methods.

2) Qualitative Analysis: For the purpose of showcasing the effect of our method more intuitively, we select two sets of representative scheme comparison charts to visually demonstrate

Fig. 4. Comparison results of different methods. Regions of mixed colors (labeled by red boxes) reduce saliency-based subjective visual sensitivity; Regions of color error (labeled by yellow boxes) where the color of the reference image cannot be effectively transferred to the original salient region or target; Regions of color cast (labeled by green box). The methods: GCT [1], IDT [8], MKL [2], C2E [3], DCT [4], WCT [61], PhotoWCT [62], WCT2 [63], SVCNet [67], PDNLA [68].

TABLE II
QUANTITATIVE COMPARISON OF COLOR TRANSFER METHODS SCORES ON TEST DATASET OF 987 SETS IMAGES

	GCT	MKL	IDT	C2E	DCT	WCT	PhotoWCT	WCT2	SVCNet	PDNLA	Our	Our-SAM
SGDnet[54] ↑	3.407	3.406	3.429	3.433	3.425	3.164	3.323	3.297	3.414	3.215	**3.440**	**3.440**
NIQE[55] ↓	3.942	3.913	4.519	3.782	3.617	4.794	3.566	**3.278**	3.773	3.674	3.596	3.599
IL-NIQE[56] ↓	24.683	26.669	24.791	23.913	23.888	27.572	24.833	23.102	23.596	24.057	23.048	**23.043**

TABLE III
QUANTITATIVE COMPARISON ON 1000 RANDOM IMAGES FROM IMAGENET DATASET. THE BEST RESULT BOLDFACED

	GCT	IDT	MKL	C2E	DCT	SVCNet	PDNLA	Our
SGDNet[54] ↑	3.353	3.403	3.372	3.423	3.409	3.213	3.091	**3.456**
NIQE[54] ↓	4.515	5.113	4.500	4.367	4.165	4.771	4.196	**4.073**
IL-NIQE[54] ↓	29.433	28.533	31.605	27.102	27.708	28.501	29.571	**26.986**

the characteristics of our method from a visual point of view. From Fig. 5, the source image selected in the figure itself has a certain exposure problem, and other color transfer schemes obviously wrongly transfer the green background color to the salient target giraffe and the sky, resulting in extremely poor overall image contrast. Because our method adopts saliency guidance and simultaneously transfers the bi-branch color by local and background, a color fusion module based on brightness adjustment is also added before image generation, which enables our method to accurately perform effective detection on salient objects and the sky. While reasonably adjusting the brightness of the sky, the proposed method enhances the image contrast and the accuracy of color transfer, and the contrast effect is remarkable.

As can be seen from Fig. 5, our scheme is designed based on saliency guidance, and the core goal is to more accurately and efficiently establish the target semantic correspondence between the original image and the reference image. Among them, the zebra contrast map clearly proves the accuracy of our method in color transfer from the edge details of salient objects. The zebra target edges generated by the transfer of other deep learning methods are mixed with background colors, while our method does not have edge mixed color. In the duck comparison image, the salient target of the reference image is a gray bird. Through the comparison of various color transfer methods, we find that the color transfer effect of the salient object is not significant in other generated images. Relatively speaking, our scheme is designed based on saliency prediction, so that the color of the object is basically consistent with the color of the gray birds in the reference image. Also the color transfer of the salient object between the original image and the reference image is better, which proves that the proposed method can obtain color transfer results accurately and efficiently.

In order to demonstrate the superiority of our method more objectively and accurately, we analyze it from the perspective of contrast and information entropy. Contrast expresses the degree of stretched contrast between the light and dark of the image which can accurately express the clarity of the image quality. We randomly selected 8 images from a test dataset that consisted of 987 pairs of images as the Fig. 7 shown. Subsequently, we computed the respective scores for different

TABLE IV
FOREGROUND AND BACKGROUND COLOR HISTOGRAM COMPARISON VALUE OF BABBITT DISTANCE ON 500 PAIRS IMAGES

	GCT	MKL	IDT	C2E	DCT	OURs
Foreground	0.229	0.224	0.235	0.219	0.216	**0.145**
Background	0.474	0.473	0.469	0.455	0.440	**0.392**

methods. In Fig. 8, our quantitative comparison of contrast scores for eight random images from the test dataset is the best on average. In Fig. 9, our quantitative comparison of contrast scores is the best on average. Through the score comparison, our scheme ranks in the top 2 among multiple scheme comparisons, and the improvement effect was significant. In addition, we noticed that AttentionGAN [70] also focuses on extracting salient foreground for translation tasks, so we compared our approach with it. In Fig. 12, our separation module demonstrates enhanced accuracy in extracting salient objects. And our results are further improved by the adaptive brightness optimization module, which effectively enhances the deficiencies related to lighting. By combining the related histograms in Fig. 13, we can also discover that our results have a clear advantage in brightness.

In Fig. 10, we visually show the contrast results of the color distribution between the foreground images of the generated image and the reference image and between the background images in the form of color histograms. Color histogram is the most basic and common representation of image color features which accurately reflects the composition distribution of colors in the image, and color histogram's advantage is not affected by image translation and rotation. In the case of global color similarity, the difference in the global color distribution of two images is measured by comparing the difference in color histograms. In Fig. 10, we can clearly find that the color distribution curves of the foreground and background images generated by our method are basically consistent with the trend of the color distribution curve of the reference image itself, and they are better than the color curves of other methods. Our saliency-guided color transfer network can better accurately transfer the color of the reference image to the original image. From Table IV, we calculate the histogram comparison values

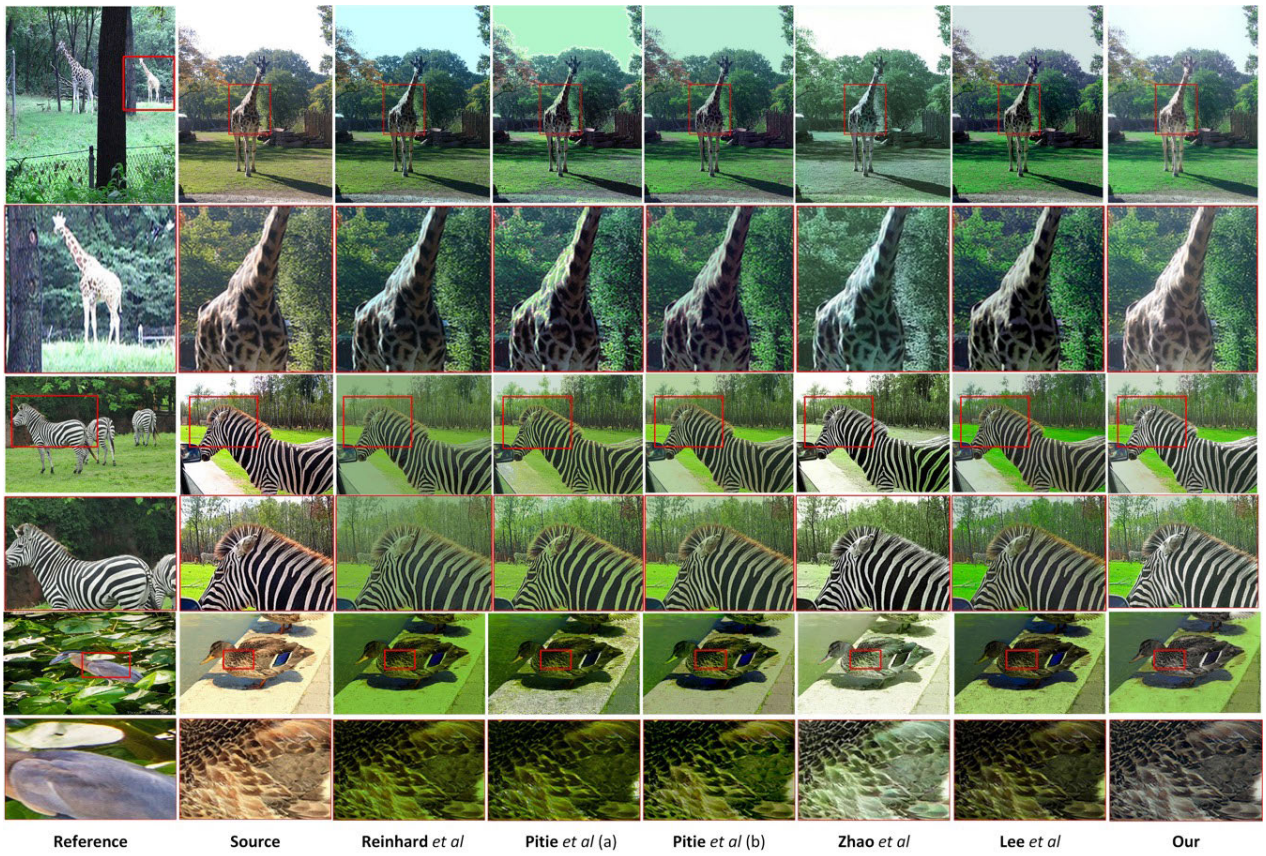

Fig. 5. Visual comparison on the dataset images among state-of-the-art color transfer approaches.

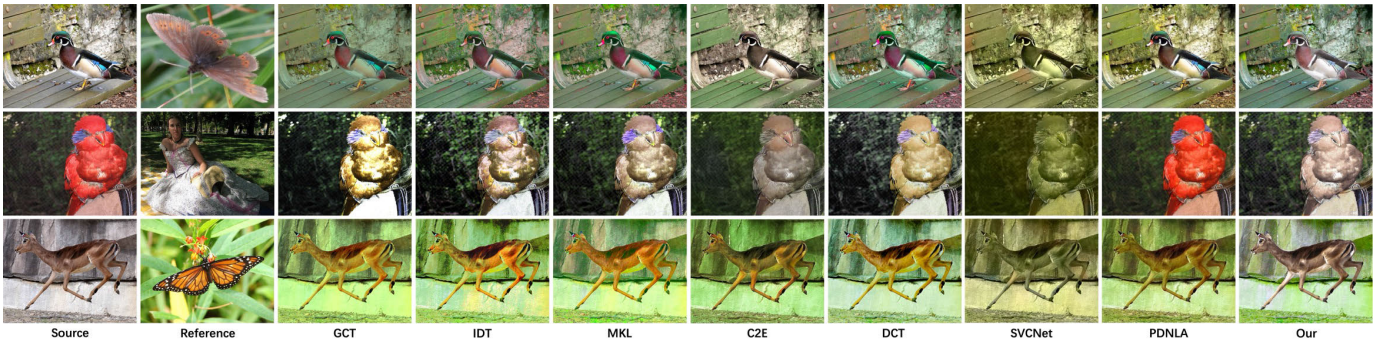

Fig. 6. Comparison of results with different methods on ImageNet Dataset [69]. The methods: GCT [1], IDT [8], MKL [2], C2E [3], DCT [4], SVCNet [67], PDNLA [68].

of Bhattacharyya Distance for the foreground and background of 500 pairs of images. Bhattacharyya Distance measures the similarity of the distribution of two discrete samples. Based on this, we measure the similarity of the color histogram between the generated image and the reference image. The lower the value of Bhattacharyya Distance indicates the higher accuracy of color transfer between the two images.

C. Ablation Study

In order to intuitively show the innovation and superiority of our method, we have designed the following ablation experiments. As shown in Fig. 11, comparing our method with the basic deep color transfer model in the third column, we can intuitively feel that the color of the remarkable region of the transfer result is not coordinated with the reference image color. The fourth column result chart is that our method does not adopt the model renderings of the adaptive brightness optimization module and the upgraded segmentation model. The five column result chart is that our method does not adopt the adaptive brightness module, and the sixth column results chart is the whole. From the renderings of the method, we can clearly see that the color consistency of the front and rear background and the reference image is more consistent, and at the same time, the brightness is more coordinated.

To evaluate the influence of segmentation performance on our model, we substituting the Yolact segmentation model with the high-performing segment anything model (SAM)

Fig. 7. Randomly selected 8 specific source/reference/our images from our dataset.

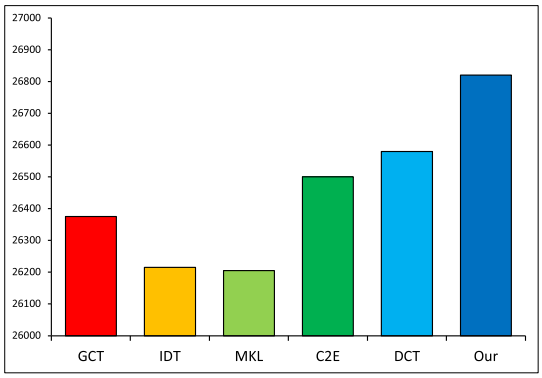

Fig. 8. Contrast Score Quantitative comparison on eight random images from dataset.

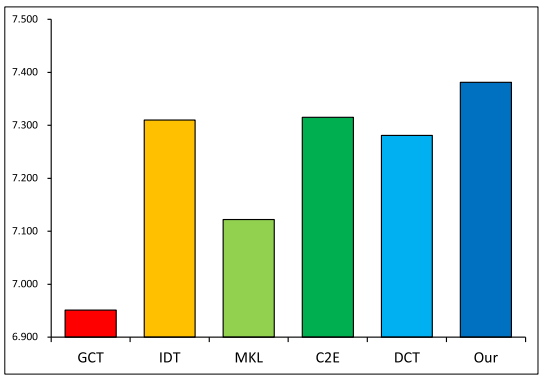

Fig. 9. Information Entropy Score Quantitative comparison on eight random images from dataset.

[71]. SAM has gained significant popularity and achieved outstanding results in various segmentation tasks. We present a comparative analysis of SAM and Yolact for instance segmentation on the COCO2017 dataset in Table V. AP(Average Precision) is a comprehensive evaluation metric that measures the performance of an algorithm on detecting and segmenting object instances across different categories. AP_S, AP_M, and AP_L are variants of AP used specifically to measure the performance of algorithms on small, medium, and large-sized targets. AP_{50} and AP_{75} represent the average precision at overlap thresholds of 0.50 and 0.75 respectively. Higher values indicate superior performance of the model on the respective

TABLE V
INSTANCE SEGMENTATION ON COCO 2017

	AP	AP_S	AP_M	AP_L
Yolact	29.8	9.9	31.3	47.7
SAM	46.5	30.8	51.0	61.7

TABLE VI
THE RESULTS OF MULTI-SCALE ANCHOR MODULE WHICH IMPORVE DIFFERENT ANCHOR CHOICES OF PREDICTION HEAD ON YOLACT SEGMENTATION MODEL ON COCO DATASET

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Base	29.8	48.5	31.2	9.9	31.3	47.7
w/ MSA	30.8	49.5	32.5	9.0	34.0	51.5

task. We conducted experiments using the SAM model based on a test set consisting of 987 pairs of images. The results in Table II indicate that improving the segmentation performance can not lead to a significant enhancement in our overall performance. This can be attributed to the separation module, which is not only essential for instance segmentation but also for extracting salient features.

Although SAM exhibits superior segmentation performance, it encounters limitations such as slow computation speed and large number of model parameters. Consequently, we opt for the comparatively lightweight Yolact as the component for instance segmentation, with the goal of enhancing its performance as much as possible without inflating the parameter count. From the quantitative target indicators in Table VI, the performance of the basic model of Yolact and the performance of the multi-scale anchor model we used. As far as the semantic segmentation of the COCO data set is concerned, the segmentation accuracy of our MSA method has improved significantly on big and medium targets.

In order to further prove the effectiveness and superiority of our method, we design a large data set of 1974 images from the COCO dataset to pair it, resulting in 987 original images and reference images. With this database, we conduct the comparison experiments. In Table VII, IL-NIQE, NIQE and SGDNet scores on the images, these results are randomly

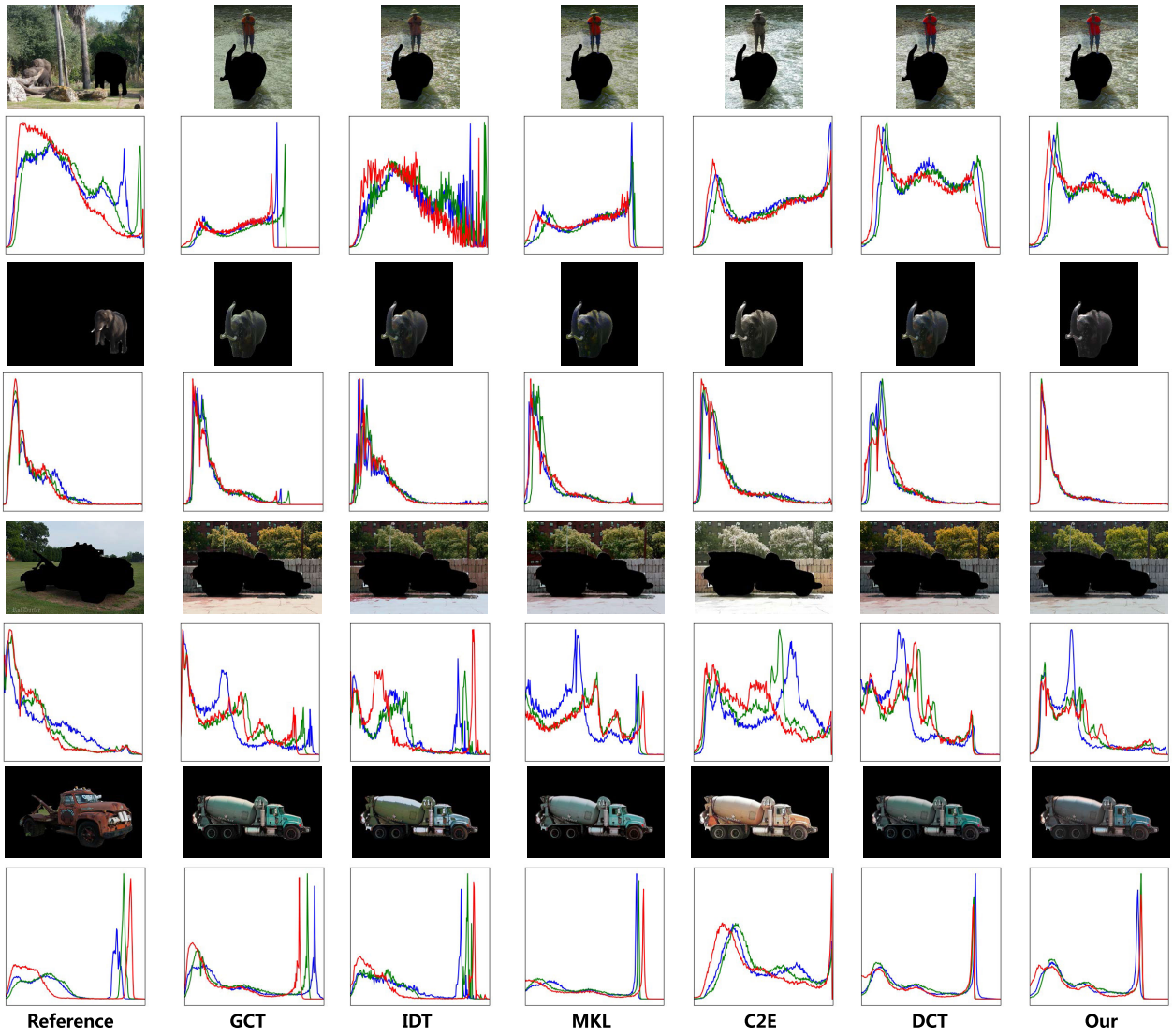

Fig. 10. Foreground and background images color histogram comparison results of different methods. The methods includes GCT: Reinhard et al [1]; IDT: Pitie (a) et al [8]; MKL: Pitie (b) et al [2]; C2E: Zhao et al [3]; DCT: Lee et al [4].

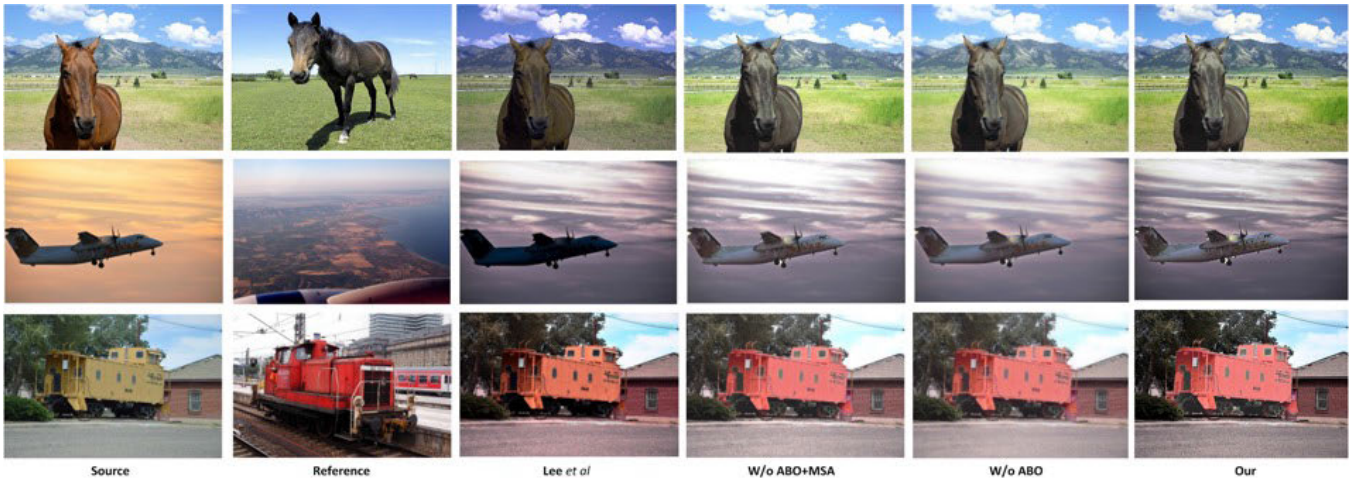

Fig. 11. The results of ablation experiments, each row is a set of pictures, where the third column is the result of the base color transfer model, and the fourth,fifth and sixth columns are the results of our model. The baseline result is produced by Lee et al [4].

selected from the COCO dataset. Based on the experimental results, the proposed method can obtain promising results.

D. Limitation

Since our model is based on deep neural network, it has certain limitations in some specific scenarios. We have provided

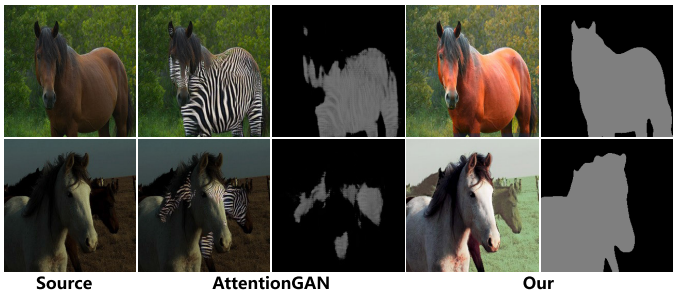

Fig. 12. Visual comparison between different methods.

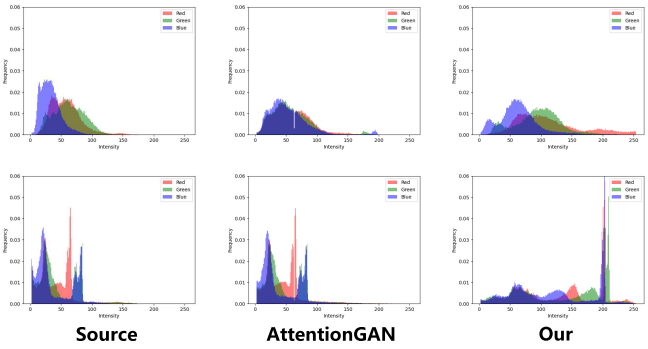

Fig. 13. Color histogram comparison between different methods based on pictures of Fig. 12.

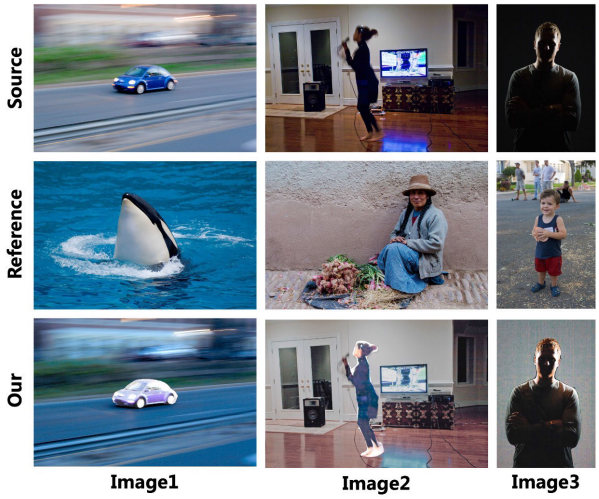

Fig. 14. The limitation of our color transfer model in some certain scenarios.

some examples in Fig. 14. From the first column in Fig. 14, it can be observed that when the blurriness of the image is relatively high, the effectiveness of color transfer is limited while resulting in some inaccurate results. For the second column, it is obviously that when source images are blurry, the salience extraction module fails to separate objects effectively. In such conditions, the proposed color transfer method may produce fragmentation or noticeable boundaries that break the visual consistency. Similarly, in the third column, when the images are affected by poor exposure, the color transfer may generate prominent boundaries, which affects the quality of final result.

TABLE VII

QUANTITATIVE RESULTS IN TERMS OF IL-NIQE, NIQE AND SGDNET SCORE ON 987 PAIRS IMAGES WHICH RANDOMLY SELECTED FROM COCO DATASET, WHERE ABO MODULE, MSA MODULE ARE ABOUT THE PROPOSED ADAPTIVE BRIGHTNESS OPTIMIZATION MODULE AND MULTI-SCALE ANCHOR MODULE

Method	NIQE	ILNIQE	SGDNet
w/o MSA+ABO	3.662	22.429	3.368
w/o MSA	3.605	23.374	3.436
w/o ABO	3.650	**22.339**	3.366
OURs	**3.596**	23.311	**3.438**

VII. CONCLUSION

In this paper, we have proposed a novel two-stage saliency-guided color transfer deep network model, where the improved object segmentation model and adaptive brightness optimization model are introduced. Our model effectively reduces the problem of background color and prospect color mixing in the current mainstream methods. Experiments show that the proposed methods are significantly better than the existing ones. However, our approach relies on the object segmentation model to be used, and in some cases color overflows near the edges. In the future, we will improve the algorithm by considering edge preservation.

REFERENCES

[1] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 34–41, Jan. 2001.

[2] F. Pitie and A. Kokaram, "The linear Monge–Kantorovitch linear colour mapping for example-based colour transfer," in *Proc. IET 4th Eur. Conf. Vis. Media Prod. (CVMP)*, 2007, pp. 1–9.

[3] H. Zhao, W. Wu, Y. Liu, and D. He, "Color2Embed: Fast exemplar-based image colorization using color embeddings," 2021, *arXiv:2106.08017*.

[4] J. Lee, H. Son, G. Lee, J. Lee, S. Cho, and S. Lee, "Deep color transfer using histogram analogy," *Vis. Comput.*, vol. 36, nos. 10–12, pp. 2129–2143, Aug. 2020.

[5] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–10, Jul. 2011.

[6] J. Park, Y.-W. Tai, S. N. Sinha, and I. S. Kweon, "Efficient and robust color consistency for community photo collections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 430–438.

[7] C. Lv, D. Zhang, S. Geng, Z. Wu, and H. Huang, "Color transfer for images: A survey," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 2023, pp. 1–28, Jan. 2023.

[8] F. Pitie, A. C. Kokaram, and R. Dahyot, "N-dimensional probability density function transfer and its application to color transfer," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2005, pp. 1434–1439.

[9] D. Freedman and P. Kisilev, "Object-to-object color transfer: Optimal flows and SMSP transformations," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 287–294.

[10] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," *Comput. Vis. Image Understand.*, vol. 107, nos. 1–2, pp. 123–137, Jul. 2007.

[11] T. Pouli and E. Reinhard, "Progressive color transfer for images of arbitrary dynamic range," *Comput. Graph.*, vol. 35, no. 1, pp. 67–80, Feb. 2011.

[12] B. Wang, Y. Yu, and Y.-Q. Xu, "Example-based image color and tone style enhancement," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–12, Jul. 2011.

[13] A. Abadpour and S. Kasaei, "A fast and efficient fuzzy color transfer method," in *Proc. 4th IEEE Int. Symp. Signal Process. Inf. Technol.*, Feb. 2004, pp. 491–494.

[14] X. Xiao and L. Ma, "Color transfer in correlated color space," in *Proc. ACM Int. Conf. Virtual Reality Continuum Appl.*, Jun. 2006, pp. 305–309.

[15] Y. Chang, S. Saito, K. Uchikawa, and M. Nakajima, "Example-based color stylization of images," *ACM Trans. Appl. Perception*, vol. 2, no. 3, pp. 322–345, Jul. 2005.

[16] Y. Chang, K. Uchikawa, and S. Saito, "Example-based color stylization based on categorical perception," in *Proc. 1st Symp. Appl. perception Graph. Visualizat.*, Aug. 2004, pp. 91–98.

[17] Z. Su, K. Zeng, L. Liu, B. Li, and X. Luo, "Corruptive artifacts suppression for example-based color transfer," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 988–999, Jun. 2014.

[18] Y.-W. Tai, J. Jia, and C.-K. Tang, "Soft color segmentation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1520–1537, Sep. 2007.

[19] H. Chang, O. Fried, Y. Liu, S. DiVerdi, and A. Finkelstein, "Palette-based photo recoloring," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–11, Jul. 2015.

[20] Y.-W. Tai, J. Jia, and C.-K. Tang, "Local color transfer via probabilistic segmentation by expectation-maximization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 747–754.

[21] Q. Zhang, C. Xiao, H. Sun, and F. Tang, "Palette-based image recoloring using color decomposition optimization," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1952–1964, Apr. 2017.

[22] Y. Hwang, J.-Y. Lee, I. S. Kweon, and S. J. Kim, "Color transfer using probabilistic moving least squares," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3342–3349.

[23] J. Xia, "Saliency-guided color transfer between images," in *Proc. Int. Symp. Vis. Comput.*, 2013, pp. 468–475.

[24] C. Wen, C. Hsieh, B. Chen, and M. Ouhyoung, "Example-based multiple local color transfer by strokes," *Comput. Graph. Forum*, vol. 27, no. 7, pp. 1765–1772, Oct. 2008.

[25] J.-D. Yoo, M.-K. Park, J.-H. Cho, and K. H. Lee, "Local color transfer between images using dominant colors," *J. Electron. Imag.*, vol. 22, no. 3, Jul. 2013, Art. no. 033003.

[26] Z. Li, Z. Tan, L. Cao, H. Chen, L. Jiao, and Y. Zhong, "Directive local color transfer based on dynamic look-up table," *Signal Process., Image Commun.*, vol. 79, pp. 1–12, Nov. 2019.

[27] X. Xiao and L. Ma, "Gradient-Preserving color transfer," *Comput. Graph. Forum*, vol. 28, no. 7, pp. 1879–1886, Oct. 2009.

[28] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized patchmatch correspondence algorithm," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 29–43.

[29] H. Hristova, O. Le Meur, R. Cozot, and K. Bouatouch, "Style-aware robust color transfer," in *Proc. Workshop Comput. Aesthetics*, 2015, pp. 67–77.

[30] B. Arbelot, R. Vergne, T. Hurtut, and J. Thollot, "Local texture-based color transfer and colorization," *Comput. Graph.*, vol. 62, pp. 15–27, Feb. 2017.

[31] M. He, J. Liao, D. Chen, L. Yuan, and P. V. Sander, "Progressive color transfer with dense semantic correspondences," *ACM Trans. Graph.*, vol. 38, no. 2, pp. 1–18, Apr. 2019.

[32] Y. Gao et al., "Wallpaper texture generation and style transfer based on multi-label semantics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1552–1563, Mar. 2022.

[33] Y.-S. Liao and C.-R. Huang, "Semantic context-aware image style transfer," *IEEE Trans. Image Process.*, vol. 31, pp. 1911–1923, 2022.

[34] H. Li, B. Sheng, P. Li, R. Ali, and C. L. P. Chen, "Globally and locally semantic colorization via exemplar-based broad-GAN," *IEEE Trans. Image Process.*, vol. 30, pp. 8526–8539, 2021.

[35] M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, "Deep exemplar-based colorization," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–16, Aug. 2018.

[36] B. Zhang et al., "Deep exemplar-based video colorization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8044–8053.

[37] D. Liu, Y. Jiang, M. Pei, and S. Liu, "Emotional image color transfer via deep learning," *Pattern Recognit. Lett.*, vol. 110, pp. 16–22, Jul. 2018.

[38] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang, "Visual attribute transfer through deep image analogy," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–15, Aug. 2017.

[39] Z.-C. Song and S.-G. Liu, "Sufficient image appearance transfer combining color and texture," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 702–711, Apr. 2017.

[40] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[42] Z. Xu, T. Wang, F. Fang, Y. Sheng, and G. Zhang, "Stylization-based architecture for fast deep exemplar colorization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9360–9369.

[43] Y. Huang, S. Qiu, C. Wang, and C. Li, "Learning representations for high-dynamic-range image color transfer in a self-supervised way," *IEEE Trans. Multimedia*, vol. 23, pp. 176–188, 2021.

[44] Y. Zhao, L.-M. Po, K.-W. Cheung, W.-Y. Yu, and Y. A. U. Rehman, "SCGAN: Saliency map-guided colorization with generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 8, pp. 3062–3077, Aug. 2021.

[45] X. Dong, C. Liu, X. Hu, K. Xu, and W. Li, "Spatially consistent transformer for colorization in monochrome-color dual-lens system," *IEEE Trans. Image Process.*, vol. 31, pp. 6747–6760, 2022.

[46] Z. Dou, N. Wang, B. Li, Z. Wang, H. Li, and B. Liu, "Dual color space guided sketch colorization," *IEEE Trans. Image Process.*, vol. 30, pp. 7292–7304, 2021.

[47] X. Zhong, T. Lu, W. Huang, M. Ye, X. Jia, and C.-W. Lin, "Grayscale enhancement colorization network for visible-infrared person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1418–1430, Mar. 2022.

[48] R. Li et al., "SDP-GAN: Saliency detail preservation generative adversarial networks for high perceptual quality style transfer," *IEEE Trans. Image Process.*, vol. 30, pp. 374–385, 2021.

[49] J.-W. Su, H.-K. Chu, and J.-B. Huang, "Instance-aware image colorization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7965–7974.

[50] P. Lu, J. Yu, X. Peng, Z. Zhao, and X. Wang, "Gray2ColorNet: Transfer more colors from reference image," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 3210–3218.

[51] M. M. Ho and J. Zhou, "Deep preset: Blending and retouching photos with color style transfer," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2112–2120.

[52] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3203–3212.

[53] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 3917–3926.

[54] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U^2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.

[55] Y. Ke Yun and W. Lin, "SelfReformer: Self-refined network with transformer for salient object detection," 2022, *arXiv:2205.11283*.

[56] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[57] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9157–9166.

[58] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan, "BlendMask: Top-down meets bottom-up for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 8573–8581.

[59] T. Cheng et al., "Sparse instance activation for real-time instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4433–4442.

[60] T. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[61] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 386–396.

[62] Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, and J. Kautz, "A closed-form solution to photorealistic image stylization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 453–468.

[63] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9036–9045.

[64] S. Yang, Q. Jiang, W. Lin, and Y. Wang, "SGDNet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1383–1391.
[65] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Apr. 2012.
[66] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[67] Y. Zhao et al., "SVCNet: Scribble-based video colorization network with temporal aggregation," *IEEE Trans. Image Process.*, vol. 32, pp. 4443–4458, 2023.
[68] H. Wang, D. Zhai, X. Liu, J. Jiang, and W. Gao, "Unsupervised deep exemplar colorization via pyramid dual non-local attention," *IEEE Trans. Image Process.*, vol. 32, pp. 4114–4127, 2023.
[69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[70] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe, "AttentionGAN: Unpaired image-to-image translation using attention-guided generative adversarial networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1972–1987, Apr. 2023.
[71] X. Zhao et al., "Fast segment anything," 2023, *arXiv:2306.12156*.
[72] A. Abadpour and S. Kasaei, "An efficient PCA-based color transfer method," *J. Vis. Commun. Image Represent.*, vol. 18, no. 1, pp. 15–34, Feb. 2007.
[73] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4990–4998.
[74] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2017.
[75] W. Liu et al., "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[76] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 385–400.
[77] R. Zhang, Z. Tian, C. Shen, M. You, and Y. Yan, "Mask encoding for single shot instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10223–10232.

Yuming Fang (Senior Member, IEEE) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, computer vision, and 3D image/video processing. He serves on the editorial board for IEEE TRANSACTIONS ON MULTIMEDIA and *Signal Processing: Image Communication*.

Pengwei Yuan received the B.M. and M.E. degrees from the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China, in 2020 and 2023, respectively. His research interests include image processing and computer vision.

Chenlei Lv (Member, IEEE) received the Ph.D. degree from the College of Information Science and Technology, Beijing Normal University (BNU). He was a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University (NTU). He is currently an Assistant Professor with the Visual Computing Research Center (VCC), College of Computer Science and Software Engineering (CSSE), Shenzhen University (SZU). His research interests include computer graphics, 3D vision, differential geometry, and machine learning. He has served as the ICME2023 Area Chair and the CVM2023 Session Chair. For more information: https://aliexken.github.io/.

Chen Peng received the B.E. degree from the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China, in 2022, where he is currently pursuing the M.E. degree. His research interests include image processing and computer vision.

Jiebin Yan received the Ph.D. degree from Jiangxi University of Finance and Economics, Nanchang, China. He was a Computer Vision Engineer with the MT Lab, Meitu Inc., and a Research Intern with MOKU Laboratory, Alibaba Group. From 2021 to 2022, he was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a Lecturer with the School of Information Management, Jiangxi University of Finance and Economics. His research interests include visual quality assessment and computer vision.

Weisi Lin (Fellow, IEEE) received the Ph.D. degree from King's College London, London, U.K. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University. His research interests include image processing, perceptual signal modeling, video compression, and multimedia communication, in which he has published over 200 journal articles, over 230 conference papers, filed seven patents, and authored two books. He is a fellow of the IET and an Honorary Fellow of Singapore Institute of Engineering Technologists. He was the Technical Program Chair of the IEEE ICME 2013, PCM 2012, QoMEX 2014, and IEEE VCIP 2017.