

Received July 24, 2019, accepted August 9, 2019, date of publication August 19, 2019, date of current version August 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936258

Hybrid Scheme of Image's Regional Colorization Using Mask R-CNN and Poisson Editing

WUJIAN YE¹, HAOWEN CHEN¹, ZIWEN ZHANG¹, YIJUN LIU¹,
SHAOWEI WENG¹, AND CHIN-CHEN CHANG²

¹School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China

²Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan

Corresponding author: Yijun Liu (yjliu2002@163.com)

This work was supported in part by the Department of Science and Technology at Guangdong Province under Grants 2018B030338001 and 2018B010107003, in part by the Peng Cheng Laboratory, in part by the Guangdong Education Department, and in part by the Guangdong University of Technology under Grant 220413548.

ABSTRACT Image colorization is a creative process of reasonably adding colors on gray-scale images to generate well-pleasing colorized images. The existing colorization methods normally require user-supplied hints of color points and doodles, or handpicked color reference images for transferring colors, or diverse color images for predicting the colorized results; but the final colorized results generated by most of them may seem unnatural as a consequence of the unprofessional users' skills, inaccurately color transferring, or limited scale of color image collection. To overcome these limitations, a hybrid scheme consisting of two modules is proposed for images' region colorization by combining semantic segmentation and seamless fusion techniques in this paper. In the first module, the masks and category of input image's regions and background are derived from a Mask R-CNN model, and the corresponding reference images of each region are selected from a pre-classified color image database. In the second module, the background and various regions of an image are colorized by a U-Net model and a VGG model respectively. Then, the Poisson editing technique is applied for fusing all the colorized results to generate the final whole colorized image. The experiments show that our scheme can not only flexibly select the appropriate reference images for different regions of the image according their semantic information, but also effectively merge colorized results to generate a plausible colorful image. By reasonably combining different CNN-based models according to their superiority, our scheme avoids the limitation of failures caused by single-step methods, and achieves richer artistic visual effect compared with other existing methods.

INDEX TERMS Hybrid scheme, image's region colorization, deep learning, Mask R-CNN, seamless Poisson fusion.

I. INTRODUCTION

Image colorization is an artistic creation process of colorizing images, like gray-scale photos, skeletons, monochrome picture, ink paintings and so on, to generate some perceptually meaningful and visually appealing colorful images [1], [2]. As early as 1970, Wilson Markle has firstly introduced the concept of colorization, which means adding colors to the black-and-white movies and television [3]. Nowadays, the demands for colorization have been rapid increasing continuously in the fields of art-fonts rendering, graphic and animation design, movie production and other artistic works. The problem of image colorization is ill-conditioned and

inherently ambiguous since many different colors can be assigned to the same gray pixels of an input image according to human imagination and aesthetic standards [1]. Thus, there is no unique correct colorization solution.

Early colorization requires completely hand-design, which is an intensive, expensive and time-consuming process. Thus, many computer-aided methods have emerged to relieve time-consuming user intervention during the coloring process, including user-hints-based, transfer-based, auto-prediction, and hybrid methods [2]. Although these existing methods achieve great results, they still have some limitations. The user-hints based method needs intensive manual work and professional skills for providing good scribbles; the transfer-based method relies on a careful selection of colorful reference image, but it is hard to find a suitable

The associate editor coordinating the review of this article and approving it for publication was Kok Lim Alvin Yau.

one; the auto-prediction method can predict the colors by learning a large-scale of images, but it cannot predict the colors beyond the training database. The hybrid methods combining various single-step techniques can obtain better results, but their results are not always good enough due to the direct combination of different single methods without considering the semantic information of different main regions in an image, resulting in poor visual effect in some cases. Therefore, an improved hybrid scheme is needed to avoid the defects of a single-step method.

Combining with image semantic segmentation and fusion techniques, we propose a hybrid scheme for images' region colorization to enable users to individualize the colors of target image's regions, the final generated image can present more diversified artistic colorful effects.

The proposed scheme mainly consists of region segmentation module and colorization module. In the first module, the masks and categories of background and each region in an input gray-scale image are derived from a Mask R-CNN model, which is a well-trained CNN network; then, the appropriate reference images of every region are selected manually or randomly from a pre-classified image database according to the category information of each region. In the second module, the background is automatically colorized by a U-Net model; the different target regions are colorized with suitable reference images by a VGG model. Finally, the colorized background and regions are fused by Poisson seamless fusion to generate the final satisfactory colorful image.

With the similar reference images well selected from the pre-classified database, our scheme can make different target regions have their own natural semantic colors, rather than having the same color from a single reference image. By reasonably integrating two different CNN-based colorization models according to their superiority, the background and regions of an image are effectively colorized respectively, avoiding the easily failure of applying a single-step method. The experiments show that our scheme can achieve better visual effect compared with other existing schemes.

The remaining paper is organized as follows. In Section 2, the related works about techniques of image colorization are introduced. In Section 3, an improved hybrid scheme is proposed for image's region colorization. In Section 4, the performance evaluations of proposed scheme are presented. Finally, Section 5 makes a conclusion and gives the future works.

II. RELATED WORKS

Traditional hand-colorization is quite useful for drawing pictures, editing images, or making movies, but it is very time-consuming, labor intensive and expensive [4]. Thus, many computer-aided methods have merged for efficient colorization, which can be categorized into different types based on different criteria.

According to the level of user participation, the existing colorization methods are divided into automatic, semi-automatic, and user-guided colorization methods [4].

The automatic methods can colorize monochrome images by training a CNN network with large-scale image datasets. The semi-automatic methods mainly transfer color patterns from some reference images to the monochrome image. The user-guided methods directly give colors to the corresponding regions of image with user decision.

According to the source of color information, there are user hints-based, transfer-based, auto-prediction and hybrid colorization methods. They are introduced as follows in details.

A. USER HINTS-BASED METHODS

Levin et al. firstly introduce this method in 2004 [5], which focuses on propagating the user hints, such as color points or strokes [1]. These kinds of methods work on a common hypothesis, that adjacent pixels with similar luminance tend to have similar colors and can be optimized using normalized cuts [6]. To achieve better performance, the colorized results can be further interactively refined through users' additional scribbles [7]. Huang et al. exploit adaptive edge extraction to accelerate the process of optimization and bleeding effects for colorization [8]. Yatziv and Sapiro present the idea of color blending with chrominance value of pixels [9]. Luan et al. combine both the neighboring pixels with similar intensity and remote pixels with similar textures for enhancing the visual performance in the case of sparse strokes provided by users [10]. Xu et al. decide the color of each pixel by seeking out the most confident color of strokes with the probability distribution [11]. These methods have great interactivity, but they entirely rely on users intervention for choosing the position and range of colors and strokes, and heavily require repetition tests to obtain an acceptable colorized results, and the processing time is still too long [4], [12].

B. TRANSFER-BASED METHOD

To reduce the user efforts, transfer-based methods complete the colorization task by using the colors of some selected reference image(s) similar to the grayscale image [12]. The mapping between input images and reference images is established automatically by some local descriptors, or is specified manually with human intervention. Welsh et al. firstly transfer colors by matching global color statistics. This method yields unsatisfactory results in many cases since it ignores spatial pixel information [1], [13]. For more accurate color transfer, different correspondence techniques are considered, including segmented region level, super-pixel level, and pixel level [1], [14]–[17].

Chia et al. search a suitable reference image from Internet conveniently for colorization. By using a user-provided semantic text label and segmentation cues of main foreground objects in the scene, they can download many different images from photo sharing websites, and filter them to select some appropriate reference images that are reliable for transferring color to the given gray-scale image [18]. Morimoto et al. propose an automatic method using multiple images collected from the Websites, which can generate various and natural colorized images by using the

gray-scale images information of the scene structure [19]. Two recent works utilize deep features extracted from a pre-trained VGG network for color transferring [20], [21]. He *et al.* train a CNN model for directly mapping an input gray-scale image into an output colorized image with a given reference image. They present an image retrieval algorithm to automatically recommend references by considering both semantic and luminance information, which can reduce manual effort in selecting the references [1]. Luan *et al.* propose a VGG-based photographic style transfer model, which can faithfully transfer the colors of reference image to an input image while keeping the textures and content of input image the same. This method can effectively restrain distortion and generate photorealistic colorful images in a great variety of scenarios [2].

However, the transfer-based methods require the chosen reference images having the same information of scene semantics with the input gray-scale images, which is also a time-consuming task and is still an obstacle for users [12], [22]. Actually, it is hardly to seek out an approximate reference image that is similar to original gray-scale image, leading to poor colorized results.

C. AUTO-PREDICTION METHOD

The auto-prediction methods rely entirely on learning to produce the colorization results from large-scale color images. Most color prediction models are built from large-scale image collections without user intervention [1]. Deshpande *et al.* define colorization as a linear system and learn its parameters [23]. Cheng *et al.* concatenate several pre-defined image features as the input of a three-layer fully connected neural network [24]. Larsson *et al.* design a deep network model for predicting per-pixel color histograms, which can be the intermediate output of the network for automatically generating a color image, or further being manipulating prior to image formation [7]. Other end-to-end prediction methods [12], [25], [26] build different CNN-based models with corresponding loss functions to extract features and predict the color results automatically. Iizuka *et al.* propose a deep network containing a fusion layer which can fully fuse both global and local information of image features and is trained in an end-to-end fashion with L2 loss to achieve realistic colorization [12]. Zhang *et al.* train a feed-forward CNN with a well-chosen objective function (classification loss) for considering the multi-modal colorization [25]. Isola *et al.* present a U-Net based GAN network with L1+GAN loss for solving the image-to-image translation, which can effectively synthesize photos from label maps, reconstruct objects from edge maps, and colorize images [26]. However, most auto-prediction methods only generate a single plausible result for each input; even colorization is essentially an ill-posed problem of multi-modal uncertainty [1]. And the visual effects produced by them rely on the diverse and scale of training images in database, and it cannot learn the other mapping between luminance and colors beyond the database.

D. HYBRID METHODS

The hybrid methods combine all or parts of above single-step techniques to achieve well-pleasing colorized results [1]. Zhang *et al.* [27] and Sangkloy *et al.* [28] propose hybrid frameworks that inherit controllability and the robustness from user hints-based method and auto-prediction method respectively. Zhang *et al.* present a U-Net-based colorization framework the adopts user- provided color points and a gray-scale image as input, which can reduce users interventions and improve the performance of colorization [27]. Sangkloy *et al.* present a deep adversarial architecture for image synthesis to generate realistic objects like cars, bedrooms, or faces, working on the setting conditions of sketched boundaries and sparse color strokes [28]. While Cheng *et al.* apply a joint bilateral filtering in post-processing step, and further develop an adaptive image clustering technique to incorporate the global image information, which are used to train the three-layer CNN colorization model [24]. Numerous experiments demonstrate that this method outperforms in terms of quality and speed, but the color appears to be lower saturation, with some monotonous feeling. Though these hybrid methods combining the advantages of various single-step methods can improve the visual effects of colorized results obviously, they still have some inherent defects of the single-step methods. Specially, they tend to colorize the whole image, lacking of regional semantic consideration, whose performance is still not ideal.

E. ANALYSIS AND PROBLEMS

As discussed in the above review, the single-step colorization methods are hard to achieve efficient effects, and the existing hybrid methods also have some limitations. In recent years, deep learning techniques have achieved great breakthrough in the fields of image classification, semantic segmentation, style transfer, speech recognition and so on.

And the existing colorization methods successfully combine with deep learning techniques by constructing a variety of CNN-based models. However, Due to the differences of the network architectures and objective loss function, the function and performance of different CNN models are also different. That is, CNN algorithms are not universal algorithms, and a single CNN model is difficult to meet all the requirements of colorization task.

Thus, it needs to comprehensively consider the advantages and disadvantages of different colorization techniques, to reasonably combine and optimize different existing CNN-based models, and to propose an automatic hybrid scheme for image's region colorization, which can overcome the limitations of existing methods, and find more appropriate reference images for the main regions of the input gray-scale image, so as to make the colorized effects more natural, rich and diverse.

III. PROPOSED SCHEME

As shown in Fig. 1, an improved hybrid scheme is proposed for auto/semi-auto image's region colorization, consisting of

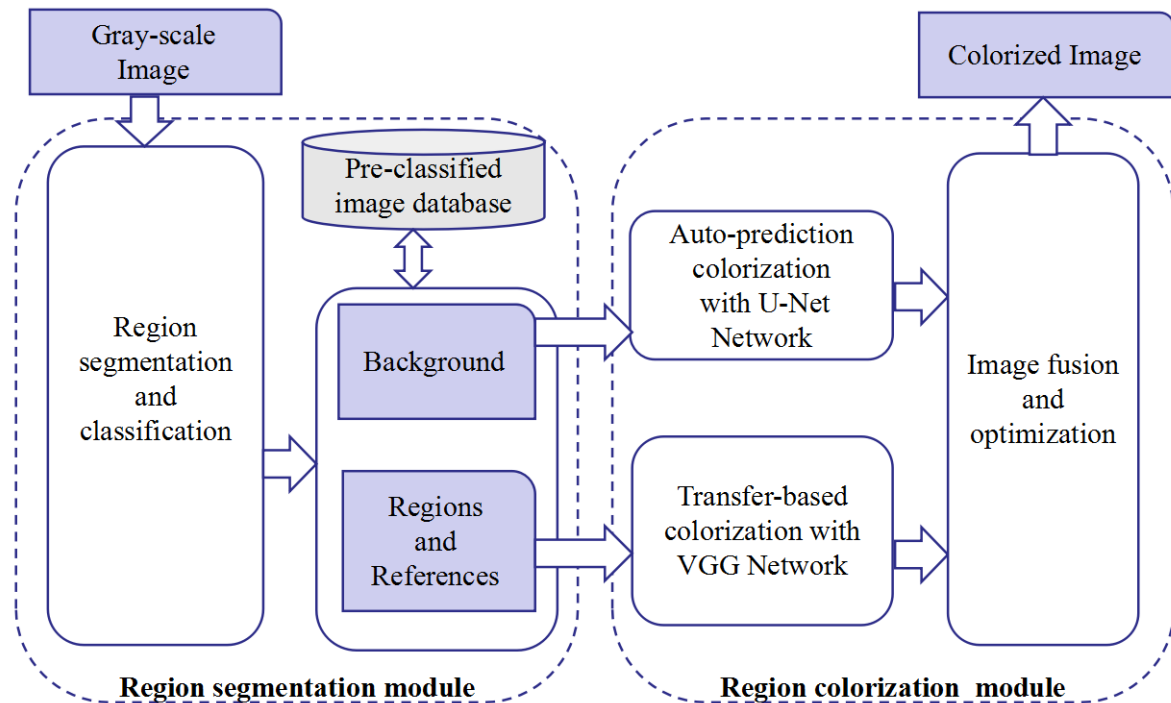


FIGURE 1. Proposed hybrid scheme of image's regional colorization.

segmentation module and colorization module. In the segmentation module, the background and main regions of input gray-scale image are segmented by Mask R-CNN model, and their category information is also attached. According to the category of the region, the appropriate reference image is manually or randomly selected from the pre-classified color image database. In the colorization module, the image background is automatically colorized based on the U-Net base auto-prediction model; and the main regions of images are colorized using VGG-based transfer model with the corresponding references selected in the first module. Finally, an image fusion algorithm is used to fuse and optimize the colorized background and regions to generate a reasonable and satisfactory colorful image. The details of our scheme are introduced in the following subsections.

A. REGION SEGMENTATION MODULE

Object recognition aims at detecting and classifying target objects in images by using the theories and methods of image processing and pattern recognition, and semantic segmentation is an extension of object recognition, which is a process of classifying and separating pixels detected in images belonging to the same class of objects. It is a key technology in content-based multimedia applications. The instance segmentation requires finer segmentation of similar objects on the basis of semantic segmentation. In this paper, we choose instance segmentation technique to identify the regions of an image and their corresponding categories.

1) REGION SEGMENTATION AND CLASSIFICATION

Mask R-CNN is an instance segmentation framework proposed by He et al in 2016 [29]. It is an end-to-end neural network, which can accomplish different tasks by adding different branches, such as target classification, object detection, semantic segmentation, posture recognition, and so on. This framework is highly efficient, flexible and easy to be expanded.

As shown in figure 2, Mask R-CNN is comprised of three basic parts [30], and an extended filtering part added by us to filter masks of main regions. The first part is a backbone network ResNet-FPN, which acts as a feature extractor combining the convolutional residual neural network ResNet and the feature pyramid network (FPN). The forward propagation of the images' feature maps through the backbone network serves as the input for the next part. The second part is a lightweight neural network - region-proposal network (RPN) which scans the feature maps of image with a sliding window and looks for the area proposals with the target. The third part includes a ROI (region of interests) classifier, a boundary regression box and a segmentation mask branch. The ROI classifier generated by RPN gives classification of the region target in each ROI as a specific category (person, car, chair, etc.), and it also gives the classification of image background. The border box fine tunes the position and size of the border to encapsulate the target region. The mask branch is a convolutional network. The positive region selected by the ROI classifier is taken as the input, and the mask of each target can be generated.

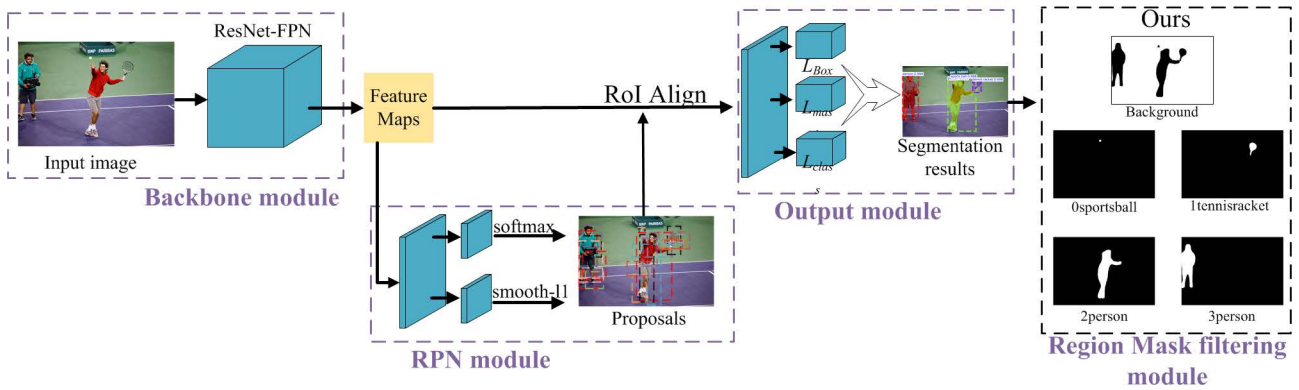


FIGURE 2. The architecture of modified MASK R-CNN.

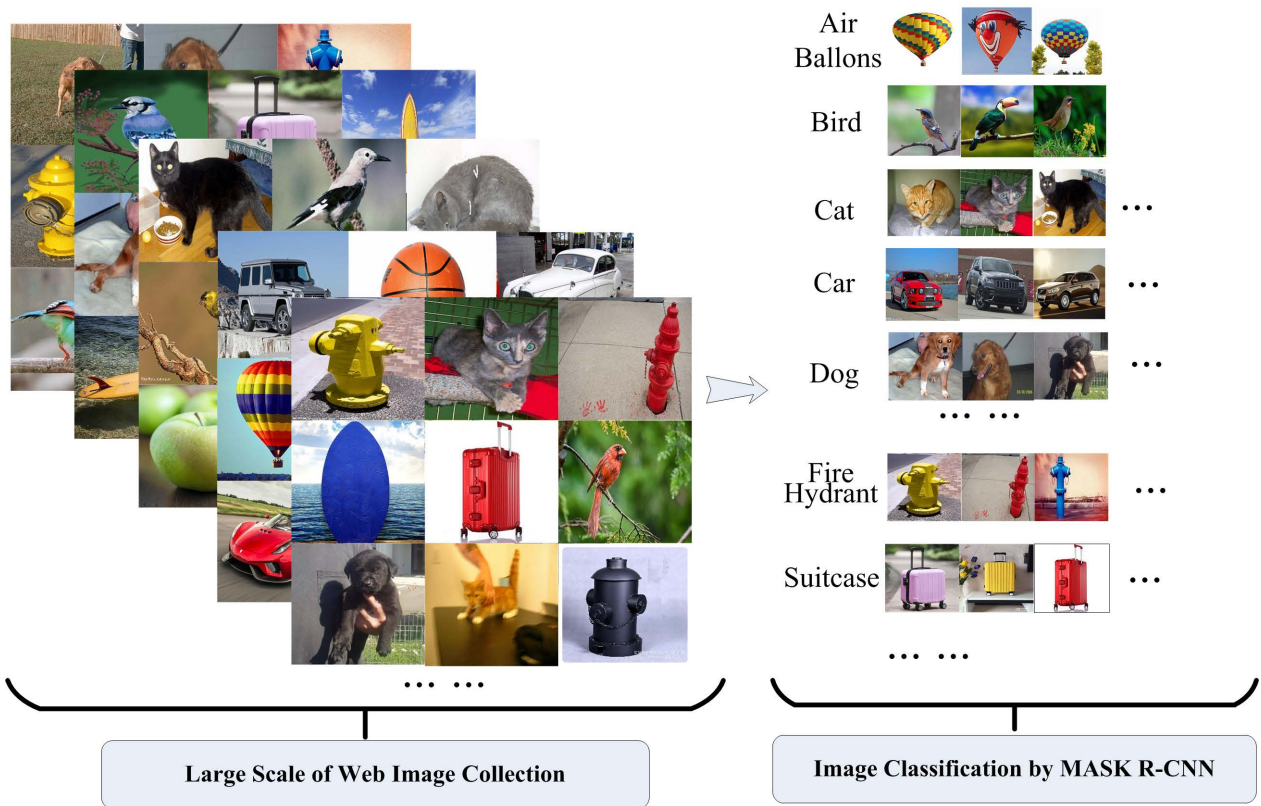


FIGURE 3. Pre-classified image database.

There are many regional targets in an image, but only the mask and class of the main targets and background are needed. Thus, the region mask filtering part is designed to filter the masks of main regions with corresponding rules, and the expected outputs are shown as follows.

- a) Masks of each main instance region;
- b) $N+1$ classes, 0 to N stands for the category of each instance region and $N+1$ refers to the background.

And the proposed filter rules are as follows.

- a) Sort the masks according to the category and area of the target regions;

- b) Draw up the threshold value for screening the mask image. Only when the area of the region is larger than the threshold value, its corresponding masks can be output, otherwise it is a part of the background.

2) BUILDING THE PRE-CLASSIFIED IMAGE DATABASE

The pre-classified image database is built for selecting reference image for each region. This database contains large-scale Web images obtained by the crawler software. The class labels of these images are needed to mark based on MASK R-CNN, then to be stored into this database.

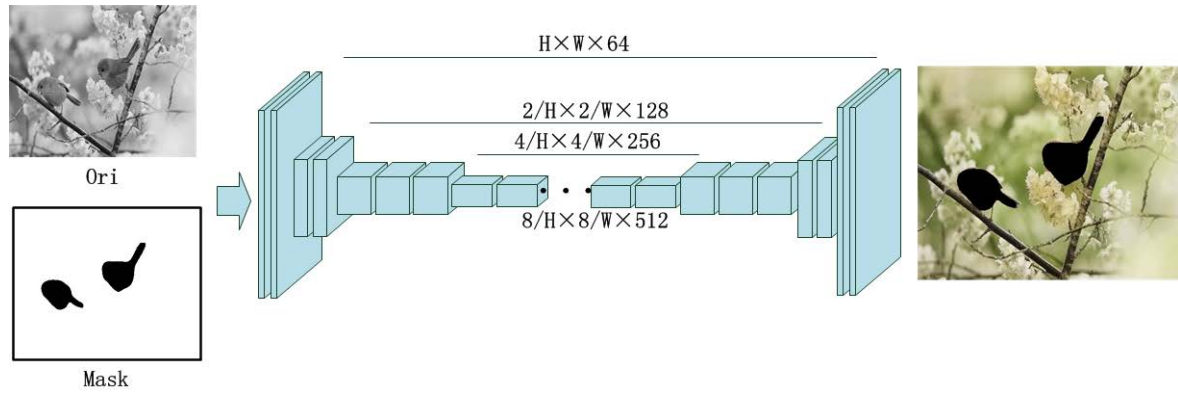


FIGURE 4. The U-Net architecture for auto-predicting the background colors.

According to the category of a region, the appropriate reference image is selected manually or randomly in the corresponding labeled image subsets of the database, which can not only make different regions have their own reference image, but also make the image color conform to the actual color effect.

B. REGION COLORIZATION MODULE

In this module, two different methods are applied for coloring the segmented background and main regions of an input gray-scale image, which are introduced below.

1) AUTO-PREDICTION MODEL WITH U-NET

In this paper, the U-Net used by Zhang *et al.* [27] is applied for coloring the background since it performs very well in many conditional generation tasks with the skip connections, which provide easy access to important low-level information in later layers.

As shown in Fig. 4, the U-Net network consists of 10 convolutional blocks, called conv1 to conv10. In the blocks of conv1 to conv4, the feature tensors are gradually halved in space and the feature dimension is doubled; each block contains 2 to 3 conv-relu pairs. In the conv7 to conv10, the spatial resolution is recovered, and feature dimensions are halved. In the blocks of conv5 to conv6, the dilated convolutions are used to replace the halving of spatial resolution. The symmetric shortcut connections are added to assist the network recover spatial information, making it easy to access the important low-level information from later layers [27]. The BatchNorm layers are added after every convolutional block, which helps with training.

The U-Net network is trained to learn the pixel mapping between grayscale image and color image, parameterized by θ , with minimizing the objective function (1) to obtain the optimized parameter θ^* . D represents a dataset of grayscale images, user inputs, and desired output colorization. X is the input of original grayscale image, along with an input user tensor U ; M is the background mask; Y is the output of the colorized background image Y ; the loss function L represents how close the final colorized output generated by network is

to the ground truth.

$$\theta^* = \arg \min_{\theta} E_{X,U,Y \rightarrow D} [L(F(X, M, U, \theta), Y)] \quad (1)$$

2) TRANSFER-BASED MODEL WITH VGG

For adding colors to specific regions in the image and maintaining their texture information, the color transfer model proposed by Luan *et al.* [2] is applied. This model can only transfer the color information of the reference image to a corresponding content image, by using Matting Laplacian to suppress the distortion of image texture during the style transferring process.

This model is carried out in YIQ color space, Y indicates luminance channel, I and Q mean color information. At first, the regions of image and their corresponding masks are input into the pre-trained VGG19, then the network is optimized by minimizing the total loss L_{total} , which is calculated by equation (2). Finally, a final transferred color image is obtained from this model until exceeding a setting threshold value set by user. In the following equation, L_c is the content loss derived from conv4_2 layer; and L_s is the style loss derived from conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1 layers of VGG19; L_m is the added photorealism regularization loss, used to constraint the texture changes of content images; the α , β and γ are the weighted factors set by the user, for tuning the final transfer effects flexibly.

$$L_{total} = \alpha L_s + \beta L_c + \gamma L_m \quad (2)$$

3) FUSION OF IMAGE'S REGIONS

The image fusion and optimization sub-module seamlessly fuses the colorized background and each region to generate the final colorized image. If all the colorized results are directly integrated, the final effect will be abrupt and unnatural because the cutting edges of each regions are obviously different from the tone, there will be obvious boundaries among the final colorized regions and background.

Poisson editing [31] is a fusion optimization method that can better integrate the target images into the designated background image naturally. In this method, a Poisson equation is solved according to the image boundary conditions

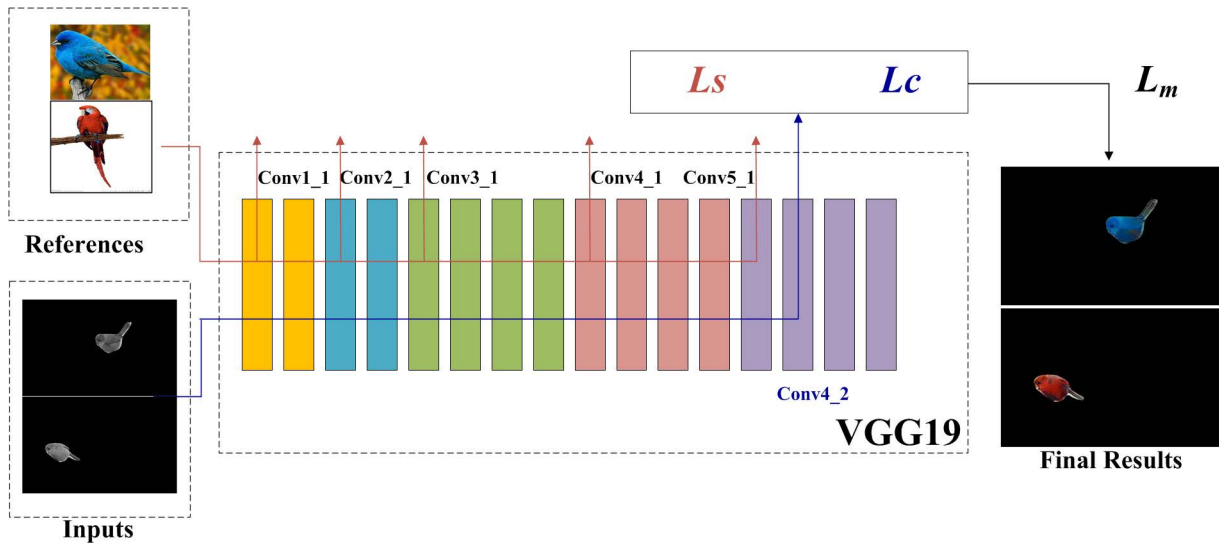


FIGURE 5. Transfer-based colorization.

specified by the user. In addition, this method can also realize some functions such as texture flattening, local illumination changes, local color changes, and so on. Therefore, we apply Poisson editing to tackle the colorized regions and background, then to obtain the final colorized image with natural edge transition and coordinated regional colors. The main steps of the algorithm are as follows.

a) Calculate the gradient field of background image and N target region images respectively;

b) Operate the image dilation on the masks of N regions with corresponding category information. The specific process for each mask is as follow: Scan each pixel of the mask image with the structural element, which is a 3×3 matrix, and its default values of all elements are 1. Then do “and” operation between the structural element and its covered part of mask image; if all are 0, the pixel of the mask image is 0, otherwise it is 1. The result is to make the mask image expand one circle.

c) The N dilated masks are used to overlay the gradient field of the target region images onto the gradient field of the source background image.

d) The gradient value of each pixel point is obtained, that is, the gradient field of the image to be reconstructed, and then the partial derivative of the gradient is obtained to calculate the divergence.

e) Construct the Poisson equation (3), where b represents the divergence obtained in the third step above, A is the coefficient matrix, x is solved to obtain the R, G and B values of each pixel of the fused colorized image.

$$Ax = b \quad (3)$$

IV. EVALUATION

A. DATASETS AND PLATFORM

In this paper, the training of instance segmentation and image colorization models are based on MS-COCO [32],

and ImageNet [33] datasets respectively. Due to the fact that few real images can be used to evaluate the colorization performance of different schemes, the input grayscale images used for training the colorization model are derived from the corresponding color images, which are also the expected output and ground truth. The experiments carry on the platform of CentOS7.0, with the deep learning supported software of OpenCV and TensorFlow, and hardware of Nvidia P40 GPU and dual Xeon E5-6280 CPU.

B. EVALUATING THE GRAYSCALE IMAGE SEGMENTATION BY MASK R-CNN

To evaluate the segmentation effect of Mask R-CNN model on grayscale images, we have trained two different models, called color model and gray model respectively through color image dataset and gray image dataset. As Fig. 6 shows, we compare their performance on segmenting the color image and gray image with the same contents.

The experiment indicates that both models can effectively mark the main target regions of the input images. The color model can mask more regions from the color image than from the grayscale image, and especially including the small regions and targets in complex background. The gray model can focus on learning more information from gray images, so it performs better on capturing the details of gray image, and achieves better segmentation effects. Therefore, the gray model is selected in this paper.

C. ANALYZING THE VISUAL EFFECTS OF THE PROPOSED SCHEME

This subsection analyzes the performance of our proposed schemes. There are four different schemes according to the input and output processing in our paper. The scheme 1 works with a single reference image, and non-edge optimization; the scheme 2 works with a single




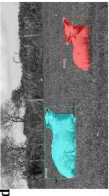

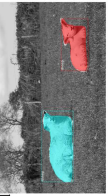



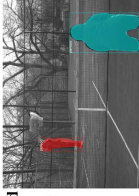
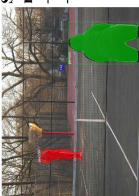
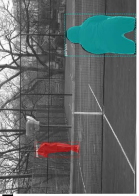


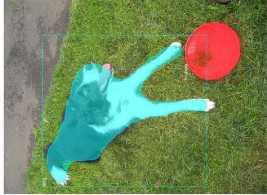
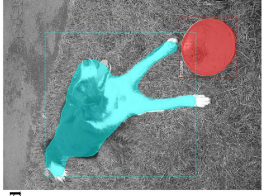
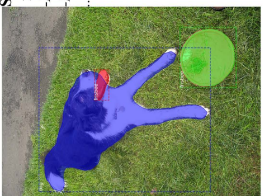
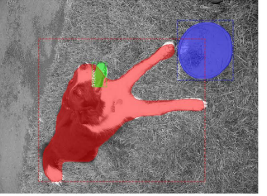



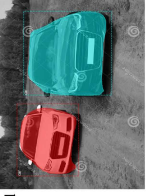




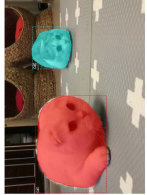
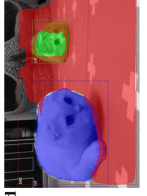
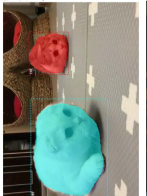
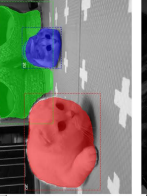









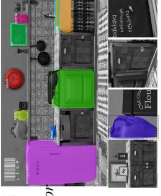


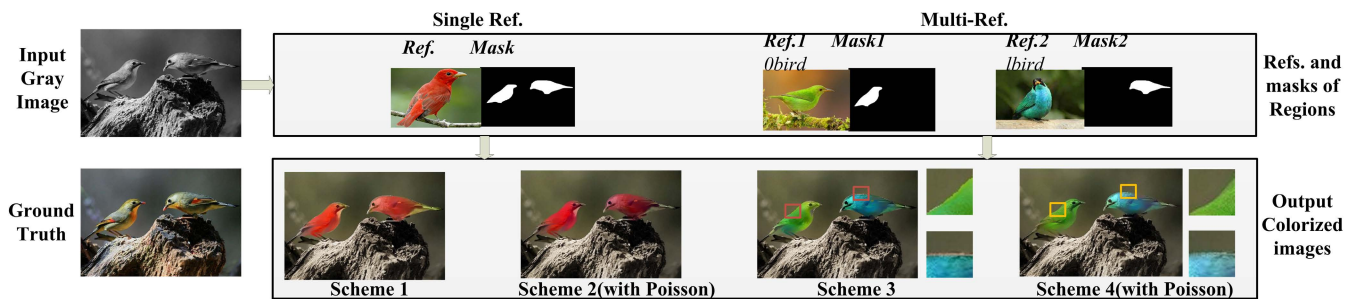
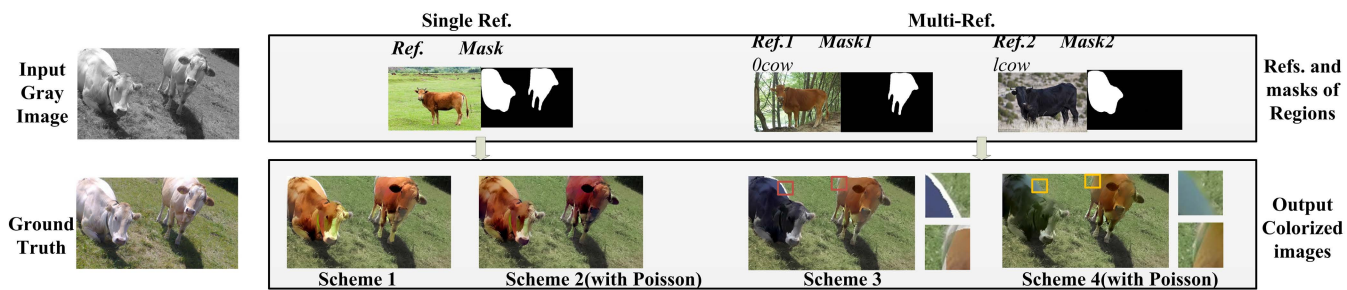
		Color-images based segmentation model		Gray-images based segmentation model	
	Inputs	Outputs on color images	Outputs on gray images	Outputs on color images	Outputs on gray images
1	 	 Segmented regions: 3 -Sheep x2 -Sports ball	 Segmented regions: 2 -Sheep x2	 Segmented regions: 2 -Sheep x2	 Segmented regions: 2 -Sheep x2
2	 	 Segmented regions: 5 -Person x2 -Tennis racket -Car	 Segmented regions: 2 -Person x2	 Segmented regions: 3 -Person x2 -Car	 Segmented regions: 2 -Person x2
3	 	 Segmented regions: 2 -Dog -Frisbee	 Segmented regions: 2 -Dog -Frisbee	 Segmented regions: 3 -Dog -Frisbee ...	 Segmented regions: 3 -Dog -Frisbee ...
4	 	 Segmented regions: 2 -Car x2	 Segmented regions: 2 -Car x2	 Segmented regions: 2 -Car x2	 Segmented regions: 3 -Car x2 -Frisbee
5	 	 Segmented regions: 2 -Cat x2 -Bed	 Segmented regions: 3 -Cat x2 -Bed	 Segmented regions: 2 -Cat x2	 Segmented regions: 3 -Cat x2 -Chair
6	 	 Segmented regions: 14 -Donut x14	 Segmented regions: 13 -Donut x13	 Segmented regions: 11 -Donut x10 -Dining table	 Segmented regions: 13 -Donut x13
7	 	 Segmented regions: 16 -Refrigerator -Oven -Sink	 Segmented regions: 10 -Refrigerator -Oven -Sink	 Segmented regions: 13 -Refrigerator -Oven -Clock	 Segmented regions: 15 -Refrigerator -Oven -Sink

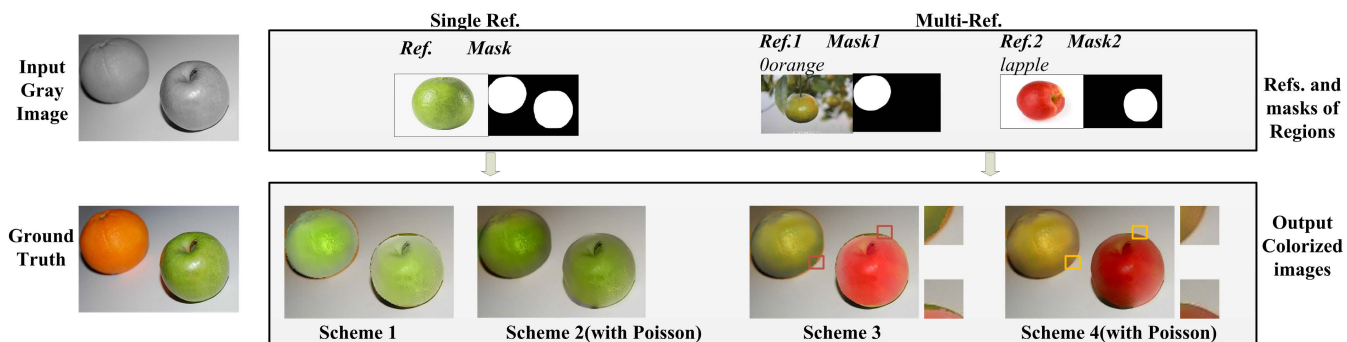
FIGURE 6. The segmentation effects of Mask R-CNN on color and grayscale images.



(a) Birds image colorization by different schemes



(b) Cows image colorization by different schemes



(c) Fruits image colorization by different schemes

FIGURE 7. The visual effect analysis of proposed scheme in different situations.

reference, and edge optimization; the scheme3 works with multi-reference images, and non-edge optimization; the scheme 4 works with multi-reference images, and edge optimization.

As shown in Fig. 7, the main region masks obtained by Mask R-CNN are selected for auxiliary colorization according to the filtering principle. The scheme 1 and scheme 2 apply the same coloring treatment to all target regions based on a selected reference image; on the other hand, the scheme 3 and scheme 4 can colorize the different regions of the gray-scale image with corresponding reference images, which are selected from the constructed large-scale database according to the semantic information of each regions; and the colors of final generated images are richer and the subjects are more various.

The scheme 3 has richer colors than scheme 1, but neither of them is optimized for region edges, resulting in a sharp edge gap, with an unnatural region transition. Like the

scheme 2, the scheme 4 applies the Poisson editing to optimize not only the edges of each region, but also the brightness and color of the region, so that all the colored regions can naturally be blended with the colored background, and the final colorful image has more harmonize and richer colors. Therefore, the scheme 4 is chosen for colorization in this paper.

D. PERFORMANCE COMPARISON WITH EXISTING METHODS

This subsection compares the performance of proposed scheme with other existing methods. As shown in Fig. 8, the auto-prediction methods can automatically predict the overall color of the gray-scale images, but its color is limited by the training datasets. Specially, the visual effect is plain in color and luminance.

The transfer-based method can colorize gray-scale images according to a single reference image, but not

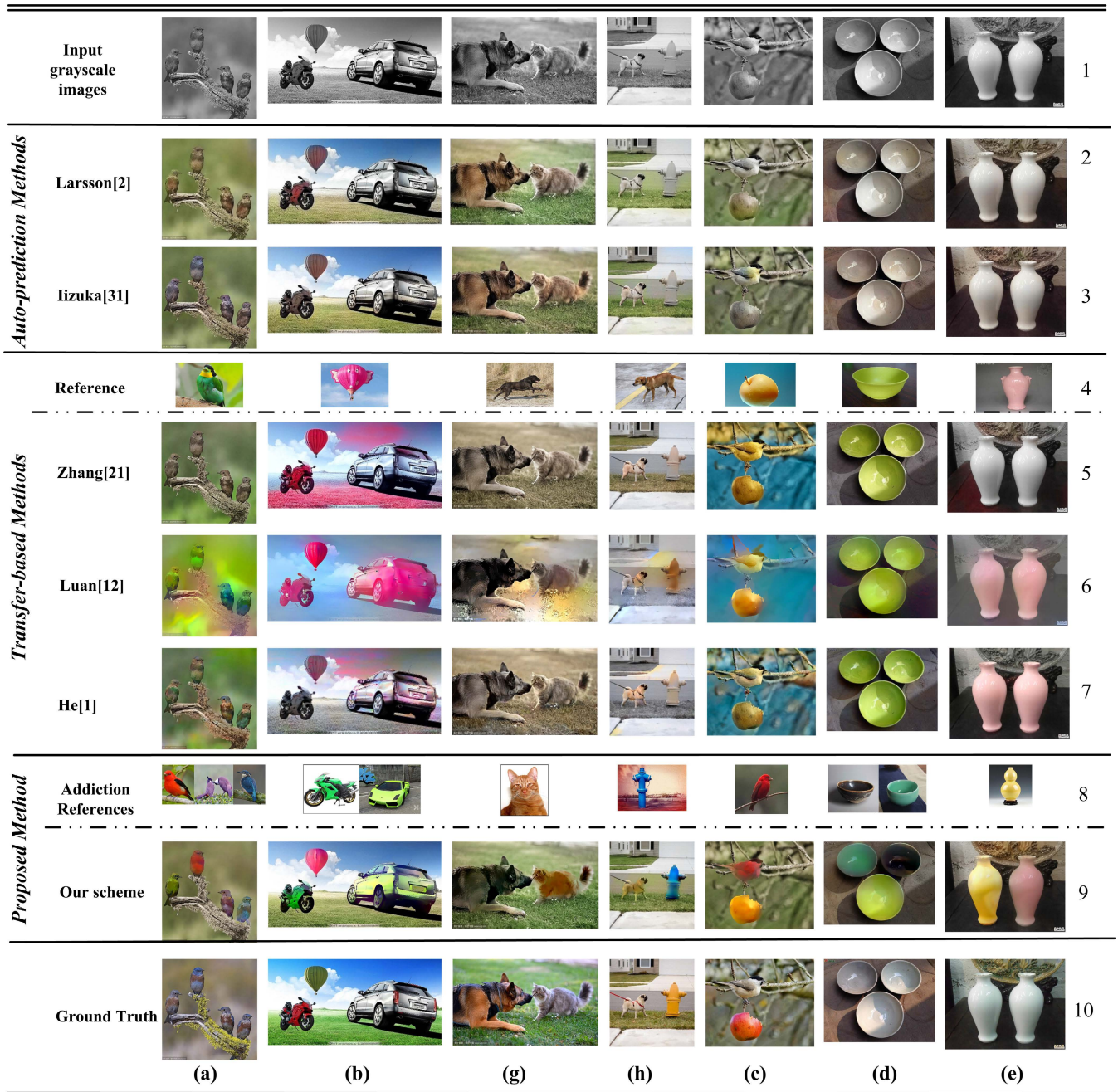


FIGURE 8. The performance evaluation of different existing schemes (The rows 1 and 10 are input and Ground Truth; the reference image in the fourth row is used by the schemes of paper [1] and [2]. The row 8 is the additional reference images added by our scheme on the basis of the previous one to achieve the colorizing different regions).

considering the specific semantics of images regions. Thus, the new added colors of image's main regions show similar tones, and some of them overflow the region's boundary.

Combining the advantages of auto-prediction method and transfer-based method, our scheme can automatically or manually select the corresponding reference images from the large-scale image database for each region according to the specific semantic category of different regions. For more complex and diverse backgrounds, auto-prediction method

TABLE 1. Average ranking of subjective effects.

Method	Classification categories	Average rankings
[7]	Auto-prediction	4.28
[12]		3.57
[2]		5.28
[1]	Transfer-based	2.26
[27]		4
Our	Hybrid	1.29

is used for color prediction. The visual effects colorized by our scheme can achieve stable quality, rich colors and natural fusion of images regions.

TABLE 2. Comparison of different existing methods.

Scheme	Classification	Network architecture	Datasets	Interactivity	Large datasets	Region coloring	Multi refs.
[7]	Auto-prediction	- VGG+hypercolumn descriptor	ImageNet; c10k	x	✓	x	x
[12]		- Low-Level Features network; - Global Features Network; - Mid-Level Features Network; - Colorization Network	Places	x	✓	x	x
[2]	Transfer-based	- VGG+Photorealism regularization	-	✓	x	x	x
[1]		- Similarity sub-net(VGG); - Colorization sub-net(U-Net);	ImageNet	✓	x	x	x
[27]		- Main colorization network layers (U-Net); - Local Hints Network; - Global Hints Network	ImageNet	✓	✓	✓	x
Our	Hybrid	- Mask R-CNN; Poisson editing; - Net for Background; - VGG for regions	CoCo ImageNet;	✓	✓	✓	✓

Colorizing grayscale image is one of artistic creation process, there is no unique ground truth since it relies on human imagination. Different people may have different or even opposite views on the same colorized results. Therefore, how to evaluate the visual results of the coloring algorithm is still an important problem. In general, qualitative evaluation (QA) is mainly used in evaluating the effect of colorization. QA requires participants to evaluate the results of different algorithms in the form of questionnaire, which depends on the observation of participants. However, the assessment results may vary according to the participants' attributes (such as age and occupation). Although there is a certain degree of uncertainty in the QA method, this method can at least provide some information about people's preference for color combination.

In this paper, questionnaire survey is conducted by different kinds of people through the Internet. Thirty participants take part in our questionnaire, who are mainly college students and teachers. The contents of the questionnaire are to sort the coloring effects of every input grayscale images in Fig. 8.

The subjective average ranking of each method's performance is finally calculated by ranking the different coloring effects of the same image. As can be seen from Table 1, our proposed scheme achieves better evaluation in human subjective vision compared with the other existing methods.

The table 2 compares different existing methods in details. Most of them are built using the public ImageNet datasets, and relying on deep convolutional network such as VGG to extract image features. The auto-predication methods can work automatically without user interaction, and they need to learn the color mapping between gray-scale images and color images, so they require a large scale of datasets. The methods of [1], [2], [7] also require large image database since they need to select specific suitable reference images for the input images. Specially, our scheme can generate the colorful image by colorizing different main regions of an image with multiple reference images.

V. CONCLUSION

The task of image colorization is to reasonably add colors to grayscale/monochrome images, photos, videos and movies, then to obtain richer and more meaningful artistic works. Generally, this task does not have a unique solution since it involves assigning RGB pixel values to an image whose elements are characterized by luminance only, and often requires much manual adjustment to achieve lifelike quality.

The computer-assisted methods have emerged to make colorization task more easily and faster, but with some common drawbacks. These methods normally require a color scribble on the input image, a special selection of reference images, or a diverse image collection. But colorized results may appear inauthentic and unnatural as a result of unprofessional user skills, inaccurate transfer of colors, or limited scale of image collections. Existing hybrid methods combining different single-step methods also have some inherent defects since they ignore the regional semantic information of input gray-scale images. It is necessary to propose an improved hybrid method for image's region colorization.

To overcome the above limitations, a hybrid scheme is proposed for image's region colorization in this paper, which enables users to colorize different target regions of the grayscale image using different reference images according to regions category information, while avoids the defects of a single-step methods, the layout of image colors can be realized reasonably. The proposed scheme is mainly composed of segmentation module and colorization module. In the first module, the background and target regions are segmented from the input grayscale image using the Mask R-CNN model. According to the categories of main regions, the appropriate reference images corresponding to each region are selected manually or randomly from the pre-classified database. In the second module, the image background is automatically colorized using the U-Net model; for different target regions, color transferring is realized using the VGG model. Finally, the colorized background and regions are fused and optimized by Poisson editing techniques to generate a satisfactory colorful image.

By making full use of the segmentation and category information derived from the Mask R-CNN based segmentation model, our scheme not only builds a pre-classified database, but also chooses the appropriate reference for each region from the database according to the category semantics of the target region, which makes all the regions of image have their own appropriate references, avoiding to face the situation that a single suitable reference is needed to be selected for colorizing all the regions of an input grayscale image. By reasonably integrating transfer-based method and auto-prediction method, which are built based on different CNN architectures and purposes, different colors are effectively added to the corresponding background and regions respectively. Our scheme not only overcomes the limitations of single-step methods, but also achieves high controllability and well robustness. Finally, the seamless fusion algorithm is applied to optimize the edge of every colorized part in an image. Therefore, the final generated color image can present more diversified artistic effects visually.

In this paper, our scheme still has some limitations. Firstly, the reference selection algorithm is still needed to be further improved to search more suitable reference images for each region. Secondly, the existing transfer-based methods still have some shortcomings, such as texture fuzzy, inconsistency of color distribution with the reference image, which need to be further optimized. And the proposed scheme cannot work in real-time since the processing time of Mask R-CNN and VGG is high relatively. Third, our scheme can also be extended to colorize each region of image with several different reference images, which needs to study the automatic algorithm for selecting and sorting reference images, and the image color fusion and transfer algorithm. Finally, how to combine different methods reasonably is still needed to be discussed in the future.

REFERENCES

- [1] M. He, D. Chen, J. Liao, P. V. Sander, and L. Luan, "Deep exemplar-based colorization," *ACM Trans. Graph.*, vol. 37, no. 4, 2018, Art. no. 47.
- [2] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," 2017, *arXiv:1703.07511*. [Online]. Available: <https://arxiv.org/abs/1703.07511>
- [3] W. Markle, "SMPTE periodical—The development and application of colorization," *SMPTE J.*, vol. 93, no. 7, pp. 632–635, Jul. 1984.
- [4] S. Li, Q. Liu, and H. Yuan, "Overview of scribbled-based colorization," *Art Des. Rev.*, vol. 6, no. 4, pp. 169–184, 2018.
- [5] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.
- [6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [7] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 9908, 2016, pp. 577–593.
- [8] Y. Qu, T.-T. Wong, and P.-A. Heng, "Manga colorization," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1214–1220, 2006.
- [9] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1120–1129, May 2006.
- [10] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y. Xu, and H. Shum, "Natural image colorization," in *Proc. 18th Eurograph. Conf. Rendering Techn.*, 2007, pp. 309–320.
- [11] L. Xu, Q. Yan, and J. Jia, "A sparse control model for image and video editing," *ACM Trans. Graph.*, vol. 32, no. 6, 2013, Art. no. 197.
- [12] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, 2016, Art. no. 110.
- [13] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 277–280, Jul. 2002.
- [14] R. Irony, D. Cohen-Or, and D. Lischinski, "Colorization by example," in *Proc. 16th Eurograph. Conf. Rendering Techn.*, 2005, pp. 201–210.
- [15] G. Charpiat, M. Hofmann, and B. Schölkopf, "Automatic image colorization via multimodal predictions," in *Proc. 10th Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 126–139.
- [16] R. Gupta, A. Y.-S. Chia, D. Rajan, and D. Rajan, "Image colorization using similar images," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 369–378.
- [17] A. Bugeau, V.-T. Ta, and N. Papadakis, "Variational exemplar-based image colorization," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 298–307, Jan. 2014.
- [18] A. Y.-S. Chia, S. Zhuo, R. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin, "Semantic colorization with Internet images," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. no. 156.
- [19] Y. Morimoto, Y. Taguchi, and T. Naemura, "Automatic colorization of grayscale images using multiple images on the Web," in *Proc. ACM SIGGRAPH, Talks*, 2009, Art. no. 59.
- [20] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. Kang, "Visual attribute transfer through deep image analogy," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–15, 2017.
- [21] M. He, J. Liao, L. Yuan, and V. P. Sander, "Progressive color transfer with dense semantic correspondence," 2017, *arXiv:1710.00756*. [Online]. Available: <https://arxiv.org/abs/1710.00756>
- [22] C.-M. Wang and Y.-H. Huang, "A novel automatic color transfer algorithm between images," *J. Chin. Inst. Eng.*, vol. 29, no. 6, pp. 1051–1060, 2006.
- [23] A. Deshpande, J. Rock, and D. Forsyth, "Learning large-scale automatic image colorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 567–575.
- [24] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," 2016, *arXiv:1605.00075v1*. [Online]. Available: <https://arxiv.org/abs/1605.00075v1>
- [25] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. ECCV*, 2016, pp. 649–666.
- [26] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, Jul. 2017, pp. 1125–1134.
- [27] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, 2017.
- [28] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. CVPR*, Jul. 2017, pp. 5400–5409.
- [29] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–9, Jun. 2018.
- [30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2015, pp. 91–99.
- [31] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, Jul. 2003.
- [32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.
- [33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.



WUJIAN YE received the B.S. degree in computer science and technology from the School of Computers, Guangdong University of Technology, Guangzhou, China, in 2010, and the M.S. and Ph.D. degrees in computer science from Dankook University, South Korea, in 2012 and 2015, respectively.

Since 2016, he has been a Lecturer with the School of Information Engineering, Guangdong University of Technology. His research interests include deep learning and computer vision, machine learning application, computer networks and security analysis, and voice recognition.



YIJUN LIU received the B.S. degree from Beijing Normal University, in 1999, the M.Sc. degree from the Guangdong University of Technology, Guangzhou, China, in 2002, and the M.Phil. and Ph.D. degrees from the University of Manchester, U.K., in 2003 and 2005, respectively, all in computer science.

He is currently a Full Professor with the School of Information Engineering, Guangdong University of Technology. His research interests include neuromorphic computing, deep learning, computer architecture, and GPS/Beidou Navigation.



HAOWEN CHEN received the B.S. degree in ship electronic and electrical engineering from the Guangzhou Maritime Institute, China, in 2017. He is currently pursuing the master's degree with the Guangdong University of Technology. His research interests include deep learning and their applications in images generation.



SHAOWEI WENG received the B.S. degree from North China Electric Power University, and the Ph.D. degree from Beijing Jiaotong University, in 2009. From October 2016 to October 2017, she was as a Visiting Scholar with the New Jersey Institute of Technology, USA. She is currently an Associate Professor with the School of Information Engineering, Guangdong University of Technology. Her research interests include image processing, data hiding and digital watermarking, pattern recognition, and computer vision.



CHIN-CHEN CHANG received the Ph.D. degree in computer science from National Tsing-Hua University, in 1982.

He was a Visiting Scholar/Researcher with Tokyo University and Kyoto University. Since 1982, he has been an Associate Professor with National Chiao-Tung University, a Professor with National Chung-Hsing University, the Chair and a Professor with the Computer Science Department, National Chung-Cheng University, the Director of the Automation Research Center, the Dean of the College of Engineering, Provost, and the Acting President of National Chung-Cheng University. He is currently the Chair Professor with Feng-Chia University, an Honorary Professor with National Chung-Cheng University, and holds a joint appointment with National Chiao-Tung University. He served as the Director of Advisory Office, Ministry of Education, Taiwan. He was involved in many different topics in information security, cryptography, and multimedia image processing. He has published several hundreds of papers in international conferences and journals and over 30 books. He was cited over 27 668 times and has an h-factor of 80 according to Google Scholar. Several well-known concepts and algorithms were adopted in textbooks.



ZIWEN ZHANG received the B.S. degree from the Department of Surveying and Mapping Engineering, East China Jiaotong University, in 2007, and the M.S. and Ph.D. degrees in geodesy and surveying engineering from Liaoning Project Technology University, China, in 2013 and 2017, respectively.

He is currently a Postdoctoral Researcher with the Guangdong University of Technology. His research interests include geotechnical application of differential interferometry for spaceborne radar and deep learning.

...