IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 29, 2020, accepted March 10, 2020, date of publication March 20, 2020, date of current version March 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982187

Adaptive Medical Image Deep Color Perception Algorithm

XIANHUA ZENG[1], SHIYUE TONG[1], YUZHE LU[1], LIMING XU[1], AND ZHIWEI HUANG[2]

[1]Chongqing Key Laboratory of Image Cognition, College of Computer Science and Technology, Chongqing University of Posts and Telecommunication, Chongqing 400065, China

[2]Department of Computing, Southwest Medical University, Luzhou 646000, China

Corresponding author: Xianhua Zeng (zengxh@cqupt.edu.cn)

This work was supported in part by the Chongqing Research Program of Basic Research and Frontier Technology under Grant cstc2019jcyj-zdxmX0011, in part by the National Natural Science Foundation of China (NSFC) under Grant 61672120, in part by the Doctoral Innovative Talents Project of Chongqing University of Posts and Telecommunications under Grant BYJS201812, and in part by the Doctoral Research Innovation Project of Chongqing under Grant CYB19173.

ABSTRACT The existing medical imaging technologies have little consideration on color information, thus most of medical images are gray. Classical hand-craft features-based methods have obtained unsatisfactory results in colorizing medical images. Moreover, these methods ignore the deep feature of medical images that represent pathology and color information. In this paper, we propose a novel method that iteratively colorizes grayscale medical images under preserving content in fine-tuned deep neural network. To the best of our knowledge, there is no currently work that attempts to colorize the medical image by using deep neural network. Specifically, we propose Y-loss which is defined as nonlinear combination of ℓ_1 and ℓ_2 norm to preserve content invariance between target and colorized medical image. Then, adaptive reference image search algorithm is introduced to code reference and target medical image with D-hash and search reference image in hash code automatically, which free the manual selection of the reference image. Extensive experiment results show that the proposed method can generate higher quality colored medical image than recent state-of-the-art methods, and can be approved by the doctor. The objective evaluation (PSNR and SSIM) outperform an average increment 24% and 47% than baseline method, respectively. Our code is available at: https://github.com/Tongshiyue/Adaptive-medical-image-deep-color-perception-algorithm.

INDEX TERMS Medical images enhance, color images, color transfer, deep neural network, automatic search.

I. INTRODUCTION

Color transfer is a widely used method in grayscale natural images process, which has been proved to be effective in images processing. The color of the reference image is transferred to the target image by choosing a suitable color image as a reference, so the target image has a similar color distribution with the reference one. There are many end-to-end color transfer methods based on deep learning, but most of them are limited to transfer natural image.

In the field of biomedical, there will yield many medical images every day with the popularization of medical diagnostic equipment such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging (UI) and other diagnostic. These medical images can help doctors diagnose and analyze disease, and decide whether need surgery. Medical image processing algorithms have attracted more and more scholars [1]. Among them, medical image coloring is a common medical image processing direction. The color medical image has broad application prospects in the fields of image reconstruction, simulation, segmentation [2], virtual biopsy, virtual surgery and clinical image diagnosis reference, which can enhance the sense of visual reality and assist diagnosis. For example, Virtual Endoscopy(VE) which based on medical imaging and scientific visualization is to provide trusty means for medical diagnoses by noninvasive demonstration anatomical structures and human's lesions. The color medical image can simulate the visual efferent of optical endoscope, and increase the intuitive feeling of virtual observation, which is conducive to the correct judgment of observation [3]. It has been proved by clinical practice practice that pseudo-color images can highlight the details of organs and

The associate editor coordinating the review of this manuscript and approving it for publication was Marcin Woźniak.

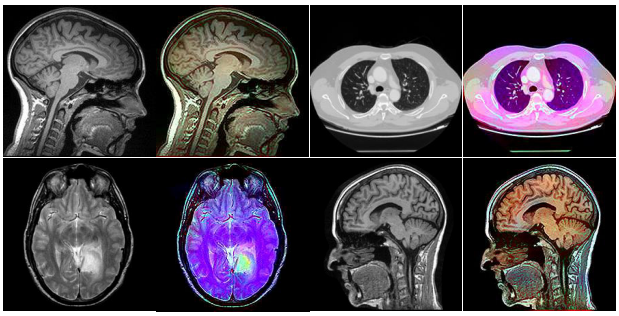


FIGURE 1. The compare of grayscale and color medical image.

tissues better comparing with gray images [4], which can help doctors make correct judgment and avoid misjudgment [5], and also help for medical image segmentation [2], such as the color medical image in Fig. 1 can easily distinguish to the chest and lungs, glioma center, brain and bone. Further, the commons prefer to read color image than grayscale image. The color medical image can help patients observe and understand medical images, which can increase effective communication between doctors and patients, improve their relationship to reduce unnecessary conflicts between them.

A. MOTIVATION AND CURRENTLY THREE PROBLEMS

Traditional colorization methods of medical image use hand-craft features to construct a color mapping. These features include sobel, local binary pattern(LBP), color wavelet, statistic of pixel value, HOG etc. The hand-craft features prefer to expression the properties of the image, e.g., shape or color of the image. The deep features, extracted from deep neural network, can effectively extract semantic content from the medical image. These deep features consist of high-level and low-level features. High-level feature contains global context-aware information, which are more flexible than hand-craft features information [6]. And low-level features contain the spatial pixel details [7]. So deep features can both construct the content and color representations of the medical image. Nevertheless, due to the lack of paired training data, the end-to-end study of medical imaging colorization algorithms based on deep learning has not progressed. Therefore, it is still worthwhile to explore how to apply deep color transfer algorithm to the medical image field.

By analyzing the current research results of medical image coloring, there are three problems need to be solved to achieve accurate and satisfactory color medical images. Therefore, our work can be motivated from this currently three problems.

1. How to reduce the difference between human preference and medical images. **Human preference** refers that people prefer to read color image, but almost all the medical images are grayscale. Color medical images can reduce doctor's visual fatigue, and visualize the tissue area clearly [8].

2. How to make full use of low- and high-level feature during medical image colorization. At present, most of the medical image coloring only utilize hand-craft features, which may lose some important information [6]. To the best of our knowledge, there is currently no medical colorizing algorithm has considered deep feature.

3. How to preserve content invariance between original and transferred medical image. Content invariance ensures that colored medical images still have diagnostic significance. In the process of coloring, the pixels value will be changed, which destroys the content details of the medical image so that yield meaningless colored medical images.

B. OUR CONTRIBUTIONS

From a practical perspective, our approach is an effective algorithm for medical image colorization suitable for medical diagnosis and cleverly avoids the lack of paired training data problem. For medical diagnosis, an intact content of colored result is very important and helpful for doctors to diagnose the illness and make the right therapeutic schemes. To preserve the content invariance, we combine hand-craft and deep features in the process of colorization. And we compare our method with other methods only with hand-craft features, our method take full advantage of the learning capability of deep CNN to extract deep features, which can avoid the paucity of representativeness of color and content of images. The colored medical image can satisfy the need of human preference, reduce visual fatigue of the doctor and promote communication between doctors and patients at the same time. The main contributions of this paper are summarized as follows:

1. Medical images are colored by using content and color representations extracted from deep neural network. These representations can both preserve content of grayscale medical image and extract color of the reference image. This colored medical image has more semantic information than colored ones by traditional methods. This is the first approach to color medical using deep feature of image extracted from deep neural network.

2. The Y-loss function which is defined as nonlinear combination of ℓ_1 and ℓ_2 norm is reported to preserve details of physiological tissues invariance between target and colorized medical image in the colorization process. It is the guarantee for doctors to diagnosis by colored medical image.

3. The reference image is automatically retrieve by adaptive method which codes reference and target medical image with D-hash. This method avoid doctor operation and achieves fully automatic coloring.

The organization of this paper is organized as follows. Section II introduces related work. Section III introduces our proposed method, including content preservation (Section III-A), local feature mapping (Section III-B), adaptive retrieval strategy(Section III-C), other loss functions (Section III-D) and algorithm of overall framework (Section III-E). Section IV introduce the experiments and results, include implementation details (Section IV-A), parameters (Section IV-B), and results and comparison (Section IV-C). Section V gives an analysis about our method. lastly, Section VI summarized our method.

II. RELATED WORK

The existing methods of image colorization can be divided into two classes: one is the hand-craft-features-based colorization method, the other is deep-features-based method. An overview of topic can be summarized as followed.

A. HAND-CRAFT-FEATURES-BASED COLORIZED METHOD

The core of hand-craft-features-based colorization method is the gray-to-color mapping constructed by the hand-craft features of image, such as pixel value, frequency and so on. Then, color the gray-scale image by this mapping relationship.

There are proposed tremendous representative works. Ning *et al.* [9] propose a colorization method. Firstly, this method combines Sobel operator with label watershed segmentation algorithm to highlight the tissues contours to obtain better pseudo-color processing results. Secondly, the Fourier transform-based frequency domain pseudo-color colorization technique is used to color the processed data. Zeng *et al.* [8] propose an image coloring method based on multi-feature fusion. Firstly, this method uses the Sobel operator to find the gradient features of medical images in four directions, and combine the luminance features to obtain a multi-feature vector for each pixel. Then, based on this vector, the K-nearest neighbor map is constructed to find the representative seed pixels, and construct the hierarchical structure. After the dimensional reduction of the topmost representative pixel, the low-dimensional representation of all pixels can be obtained by interpolation from the top to bottom. Lastly, the whole image have color information by the similarity relationship between pixels with color information and pixels without color. On this basis, Zeng *et al.* [10] then propose a method combining hierarchical density peak clustering, which uses local density and neighborhood relationship of pixels to select regional representative points. Finally the color is diffused to the whole image through the colored representative points. The research on medical image colorization has been uninterrupted, indicating that this field still has high research value.

In addition, there are many methods to color grayscale images combined with the color transfer. Welsh *et al.* [11] propose a color transfer approach. This method obtain the best match between the reference and grayscale image by the luminance value and neighborhood statistics of the pixel. Xiao and Ma [12] propose a method of color transfer in any color space. The pixel of the target image is transformed to adapt the reference image by the rotation matrix that obtained by the covariance matrix of the pixels. Huo *et al.* [13] introduce a colorization method combined with clustering. It uses the subtractive clustering to initialize the initial value of FCM, then uses FCM clustering to complete the segmentation of the source and the target image. Find the match between the two images, then transfer color. He *et al.* [14] propose a colorization method based on different emotions. This method uses target scheme, which select directly or put the selected closet scheme by main color in the reference image, to color the target image. Although these methods achieve colorization of medical images, but all have the same limitation that only the hand-craft features of the image are used to color without considering the deep features of images.

B. DEEP LEARNING-BASED METHOD

Recently, with the development of deep learning, convolutional neural network has shown a great potential in computer vision and image processing. Naturally, some incorporate CNN to automatically color images. Many automatic coloring methods are based on neural networks trained by a large number of data sets for special scenes. Iizuka *et al.* [15] train a coloring neural network by using different data sets including indoor and outdoor scenes to achieve fully automatic coloring without human intervention. The coloring network consists of four sub-networks, which combines local and global features, to make gray natural images have real scene colors. Su *et al.* [16] combine the U and V channel of an image, and obtained the value by pre-trained two neural networks respectively, with the brightness value Y to generate color image. However, it is very difficult to acquire enough color medical image datasets to train neural network. Training a colored neural network is not suitable for medical image. Therefore, we will not build and train a coloring neural networks.

Very recently, the method combined with deep learning and color transferring have proposed. Some works show pre-trained networks can be used to generate image that the user expected by extracting deep features to construct and optimize the feature loss functions. In feature inversion [17], feature visualization [18], [19], image stylization [20]–[22] use this strategy [23]. This method does not require a large amount of training data to train a deep network. Considering the application scenario of this paper, it is decided to adopt the same image-based iterative method to color medical images. Gatys *et al*. [21] propose a style transfer method that combines the content of the target image with the style of the reference image by minimizing the feature reconstruction loss. The feature reconstruction loss is constructed by using the features on the different layers, which includes the content loss (1) and the style loss (2). The content loss keeps content of the output image the same as the target image. The style loss (2) preserves style of the reference image. But the result of this method similar to abstract art between two realistic images. Subsequently, Li and Wand [24] propose an approach, which combines the markov random field model (MRF) with the pre-trained deep convolutional neural network, and enhances the spatial relationship between pixels. Therefore, texture distortion can be improved to some extent. But for the realistic image colorization, the better output image is produced because of the similar feature maps in high layer when two images with the similar appearance and size.

In order to solve the problem of detail distortion, Luan *et al.* [25] propose an image photorealism regularization term based on local affine of small regions in the color

space. And they introduce a style loss function combined with semantic segmentation. The regularization (4) can constrain the transformation of edges by calculating the matting Laplace matrix of the target image, so the output image can retain the real details of the target image. Adding semantic segmentation can effectively avoid the content-mismatch problem and improve the color result (3). The total structural loss used by Luan *et al.* [25] is (5). The regularization term can not make human eye to find the loss of its details, but the details still have some losses which seen by the objective evaluation index (SSIM) (in Table 1). In addition, when the target image has N pixels, the matrix has an N×N size which becomes larger as the image becomes larger. Undoubtedly, calculation of this matrix increases the consumption of computer memory and running time. Through experiments, it is found that the convergence of the regularization loss is slow, and the loss of the image field is not considered. Therefore, this paper optimizes it to suitable for medical image coloring.

$$\mathcal{L}_c = \frac{1}{2N_l D_l} \sum_{ij} (F_l[O] - F_l[I])^2_{ij}, \tag{1}$$

$$\mathcal{L}_{style} = \frac{1}{2N_l^2} \sum_{l=0}^{L} w_l \sum_{i,j} (G[O]^l - G[S]^l)_{ij}{}^2, \tag{2}$$

$$\begin{aligned} \mathcal{L}^l_{S^+} &= \sum_{c=1}^{C} \frac{1}{2N_{l,c}{}^2} \sum_{i,j} (G_{l,c}[O] - G_{l,c}[S])^2_{ij}, \\ F_{l,c}[O] &= F_l[O] M_{l,c}[I], \\ F_{l,c}[S] &= F_l[S] M_{l,c}[S], \end{aligned} \tag{3}$$

$$\mathcal{L}_{\mathcal{M}} = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O], \tag{4}$$

$$\mathcal{L}_{total} = \sum_{l=1}^{L} \alpha_l \mathcal{L}^l_c + \Gamma \sum_{l=1}^{L} \beta_l \mathcal{L}^l_{S^+} + \lambda \mathcal{L}_m, \tag{5}$$

where L represents the number of convolutional layers. l represents the lth layer with the N_l fitters each with the feature maps of size D_l. $F_l[\cdot] \in \mathbb{R}^{N_l \times D_l}$ is the feature map in layer l. i, j denote the position i and j. Gram matrix is calculated by $G_l[\cdot] = F_l[\cdot]F_l^T[\cdot] \in \mathbb{R}^{N_l \times N_l}$ that is the inner product between vectorized feature maps in layer l. C is the number of channels of the segmentation mask, $M_{l,c}[\cdot]$ is the c channel of segmentation mask which adapts the size of lth layer's feature map by downsampling. $\mathcal{M}_I$ is Matting Laplacian Matrix with $N \times N$ size that only related to the target image with N pixels. $V_c[O]$ is the vector($N \times 1$) after vectorizing the output image in the c channel. α_l and β_l are the weights that control the effect of each layer, Γ is the weight that balance content and style, the weight λ controls the regularization.

In the medical field, CNN is widely used to deal with medical problem [26]. Liu *et al.* [27] introduce a method to automatically segment brainstem gliomas by CNN model. Their method handles great difficulties in brainstem tumor segmentation by combining two types of multiscale feature-fusion network structures in CNN model. Phan *et al.* [28] introduce a joint classification and prediction CNN framework to automatically classify sleep stage. They use multi-task softmax layer and multitask loss function to train network as multi-task framework. As can be seen from above, deep neural network can promote biomedical research. In this paper, we focus on a deep color medical image model to explore the automatical colorization by using deep neural network.

There are many methods in the field of image matching, such as technologies based on local descriptors [29], hashing algorithms. Hash learning technology uses binary coding to represent data information [30]. Hash technology can be applied to image retrieval tasks to represent images as a series of binary codes [31]. Binary coding is used to implement image retrieval, so the retrieval speed is fast and the storage space is small. Our method use hash technology to retrieve reference images.

III. METHOD

In this section, we present our proposed framework in details. The framework of our model consists of three parts: adaptive retrieval, colorization and imaging module as shown in Fig. 2. Adaptive retrieval strategy can retrieve a reference image as the input of the colorization module. Colorization module can generate a colored medical image as the input of the imaging module. Imaging module can restore colored image to size of original image and enhance its color. In the following, we first describe the details of the vital innovative part, which is the Y-loss $\mathcal{L}_Y$ and Local feature mapping $\mathcal{L}_{color\_swap}$ applied in colorization module. Secondly, we introduce the adaptive retrieval strategy. Lastly, we show the other loss functions and algorithm of our method. The total loss function is as follows:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_c + \lambda \mathcal{L}_Y + \Gamma \sum_{l=1}^{L} \beta_l \mathcal{L}^l_{color^+} + \gamma \mathcal{L}_{color\_swap}, \tag{6}$$

where $\mathcal{L}_c(1)$ and $\mathcal{L}_Y(7)$ are used to preserve content of medical image. $\mathcal{L}^l_{color^+}(11)$ and $\mathcal{L}_{color\_swap}(9)$ are used to color medical image. $\alpha, \lambda, \Gamma, \beta, \gamma$ are the weight of these losses. L is $\{1, 2, 3, 4, 5\}$.

A. CONTENT PRESERVATION

In order to ensure the colored medical image still able to be used for medical diagnosis, the key point is that the content and details of the colored medical image is consistent with grayscale one. Therefore, we constrain the content generation by the content loss in the deep feature field, and Y-loss in hand-craft feature field. As we know, the YUV color space solves the compatibility problem during the transition period between color TV and black-and-white TV. When Chrominance(UV) is removed, the remaining luminance(Y) signal saves all the data recorded by black and white video. So the luminance information of Y channel represents the content of the image in the YUV color space. The value of the image in the Y channel can be used to construct Y-loss to keep the structure of the output image consistent with the target one

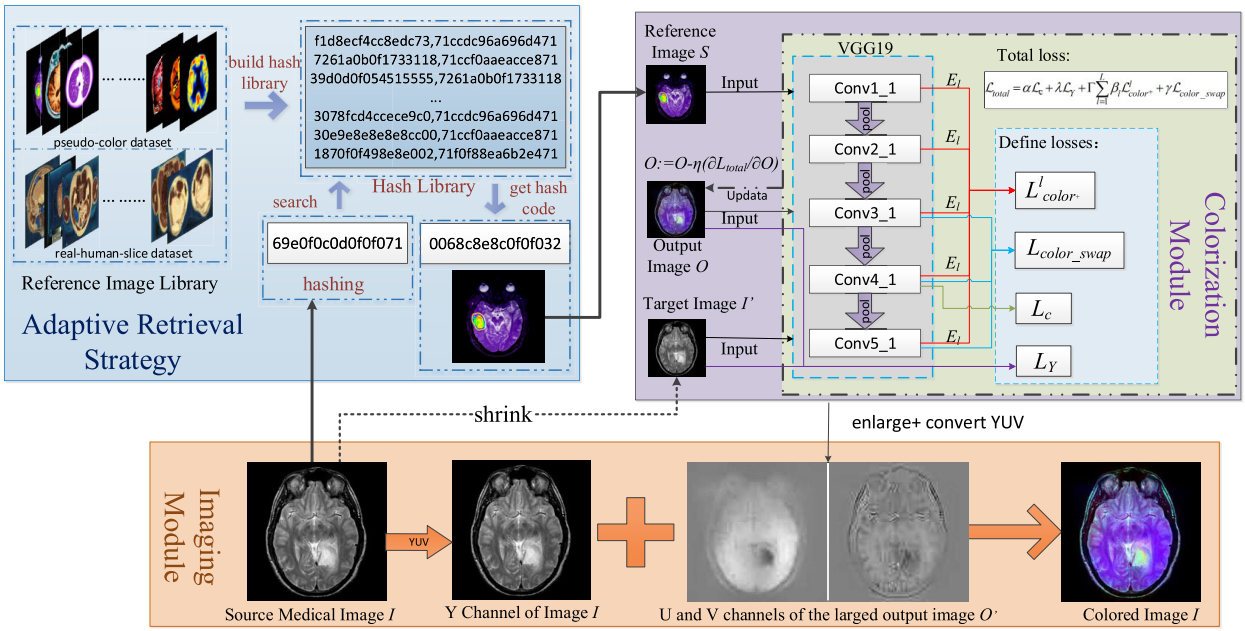


FIGURE 2. The overall framework of our method. Adaptive retrieval strategy is used to search the reference image in image library. To accelerate calculation, the source and reference medical image are uniformly resized in 224×224. Colorization module is adopted to color the target image *I'*, where Y-loss and content loss are introduced to ensure texture structure of target image *I'*. Besides, we use additional image color loss to preserve color invariance between generated and reference images, and local feature mapping loss to maintain the spatial relationship among pixels. Note that all the losses in colorization module are calculated in fine-tuned VGG19 network, which can be clearly observed in Fig. 1. Imaging module is used to enlarge the size of output image *O* by colorization module to be same with source medical image *I*. In order to preserve the texture structure consistency between enlarged output image *O'* and source medical image *I*, we integrate the U&V values of output image *O'* and Y channels of source medical image *I* to form the desired color image.

during the iteratively updating process. We achieve the above goal by minimizing the following loss function:

$$\mathcal{L}_Y = \frac{1}{2D}\left(\alpha_Y \|I_Y - O_Y\|_1 + (1-\alpha_Y)\|I_Y - O_Y\|_2^2\right), \quad (7)$$

where I_Y, O_Y are value of Y channel of image I and O in YUV color space, respectively. D is the number of pixels of I_Y and O_Y, α_Y is the weight to control the rate of ℓ_1 and ℓ_2 norm. $\mathcal{L}_Y$ is ℓ_1 norm when $\alpha_Y = 1$, which can get much sparse solution and automatically select important features, and speed up image update. ℓ_2, when $\alpha_Y = 0$, can make image smoother and suppress noise of the generated image. The user can adjust the value of α_Y based on the image. The derivative of this loss with respect to the pixel values of the O_Y equals

$$\frac{\partial \mathcal{L}_Y}{\partial O_Y} = \frac{1}{D}\left(\alpha_Y C - (1-\alpha_Y)(I_Y - O_Y)\right), \quad (8)$$

where C is a constant.

B. LOCAL FEATURE MAPPING

In the pixel-level model, the colored result of each pixel is independent of other pixels. When coloring images, it is quite difficult to evaluate the semantic relationship between adjacent pixels [24]. The use of the Gram matrix in the color loss only maintains the global consistency of the image in each layer of the segmentation mask, but can not maintain its spatial relationship. Therefore, to enhance the spatial relationship between the pixels during coloring process, we propose local feature mapping.

Firstly the deep features of the 3th, 4th and 5th layers of the target and the reference image are extracted, i.e., F_I^L, F_S^L, $\{L = 3, 4, 5\}$; Feature maps of the target and the reference image on each layer are divided into $3 \times 3 \times n$ size, n is the number of feature map channels. For each patch on the feature maps of the target image, the most matched feature patch can be found by convoluting with the feature patches of the reference image as convolution kernels. $F_i^{IS^l}(I, S)$ is the nearest neighbor matching patch for patch i, instead of $F_i^l(I)$. The replaced feature image of the target image is expressed as $F^{IS^l}(I, S)$.

In the process of iteratively updating the initial image, we hope the local features of the output image are consistent with replaced features as much as possible, so the loss function is defined as the following formula:

$$\mathcal{L}_{color\_swap} = \frac{1}{2N_l D_l}\sum_{l=3}^{5}(F^l[O] - F^{IS^l}[I, S])_{ij}^2, \quad (9)$$

where $F^l[O]$ is the activation of the ith filter at position j in layer ℓ of the output image, $F^{IS^l}[I, S]$ is the replaced feature maps in layer ℓ, N_l represents fitters with the feature maps of size D_l in layer ℓ. The derivative of this loss with respect to

the activations in layer l equals:

$$\frac{\partial \mathcal{L}_{color\_swap}}{\partial F_{ij}^{l}} = \begin{cases} \frac{1}{N_l D_l} \sum_{l=3}^{5} (F^{l}[O] - F^{IS^{l}}[I,S])_{ij}, & if\ F_{ij}^{l}[O] > 0 \\ 0, & otherwise \end{cases} \tag{10}$$

The gradient of $\mathcal{L}_{color\_swap}$ with respect to the pixel values of the output image O can be computed using standard error back-propagation.

C. ADAPTIVE RETRIEVAL STRATEGY

The adaptive retrieval strategy allows the coloring system to automatically retrieval an image as the reference image, thus eliminates human manipulation. We build two reference data sets (the pseudo-color dataset and real-human-slice dataset) for two different research purposes. One of the purpose is to highlight the different organs. The other is to restore the physical color of the organs as much as possible. According to different purposes, users can choose artificially. We use hash retrieval method to retrieve the reference image based on the target image. The specific operation process is as follows, firstly, we get hash library Φ_{H_s} by using difference hashing(D-hash) [32] in one reference image dataset. D-hash is mainly divided into 5 steps. The first step is to reduce the picture to a size of 9×8, a total of 72 pixels. The second step the image is converted to 64-level grayscale. The third step the difference is calculated. The 8 different values of each row of the matrix are obtained by subtract two adjacent elements(left element minus right element). The next step process the binary difference values of the each row. The last step, combine the 64 results to get a hash value. Secondly, we get a hash code H_I of the target image in the same way. Finally, a reference image is retrieved by XOR operation between H_I and all hash codes of Φ_{H_s}. The whole process of adaptive retrieval strategy display in the upper left part of the Fig. 2.

D. COLOR LOSS AND CONTENT LOSS

The color loss can be denoted as:

$$\mathcal{L}_{color^+}^{l} = \sum_{c=1}^{C} \frac{1}{2N_{l,c}^2} \sum_{i,j} (G_{l,c}[O] - G_{l,c}[S])_{ij}^2, \tag{11}$$

which builds upon (3) to maintain the global consistency of color between the output and reference images. $G_{l,c}[\cdot]$ and $N_{l,c}^2$ is defined in Section II-B. This content loss equals (1), which maintains the global consistency of content between the output and target images.

The gradients of two losses with respect to the activations in the layer l are following:

$$\frac{\partial \mathcal{L}_c}{\partial F_{ij}^{l}} = \begin{cases} \frac{1}{N_l D_l} (F^{l}[O] - F^{l}[I])_{ij}, & if\ F_{ij}^{l}[O] > 0 \\ 0, & otherwise \end{cases} \tag{12}$$

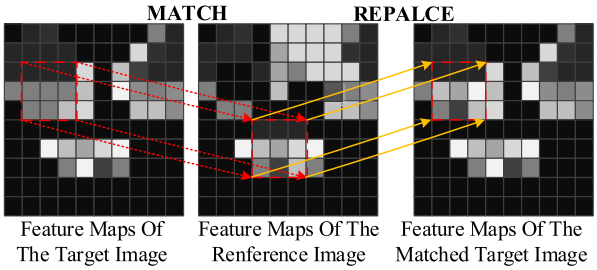


FIGURE 3. The process of local feature mapping. For the each patch of feature maps of the target image, the matched patch is found in feature maps of the reference image, then replace the patch of feature maps of the target image with the matched patch.

$$\frac{\partial \mathcal{L}_{color^+}^{l}}{\partial F_{ij}^{l}} = \begin{cases} \frac{2}{N_{l,c}^2} \sum_{c=1}^{C} \left(M_{l,c}[I] \left(F^{l} M_{l,c}[I] \right)^{T} \left(G_{l,c}[O] - G_{l,c}[S] \right) \right)_{ij}, \\ \quad if\ F_{ij}^{l}[O] > 0 \\ 0, \quad otherwise \end{cases} \tag{13}$$

The standard error back-propagation can be used to readily compute the gradients of $\mathcal{L}_c$ and $\mathcal{L}_{color^+}^{l}$ with respect to the pixel value of output image O.

E. ALGORITHM OVERVIEW

In this section, we give an overview of our algorithm. Firstly, we compress the target image to a fixed size to reduce the resource consumption, because the time and hardware consumption of the algorithm will increase with the increase of the image size. Then, we always resize the reference image to the same size as the target image before computing its feature representations since compare the image information at the same scales. Finally, we enlarge the colored image by interpolation to the original target image size, and then combine the UV channel value of the enlarged image with the Y channel value of the original target image in the YUV color space to obtain a clear colored target image. The algorithm of the whole coloring process is as follows Algorithm 1. The gradient of the total loss with respect to the output image O equals

$$\frac{\partial \mathcal{L}_{total}}{\partial O} = \alpha \frac{\partial \mathcal{L}_c}{\partial F_{ij}^{l}} \cdot \frac{\partial F_{ij}^{l}}{\partial O} + \lambda \frac{\partial \mathcal{L}_Y}{\partial O_Y} + \Gamma \sum_{l=1}^{L} \beta_l \frac{\partial \mathcal{L}_{color^+}^{l}}{\partial F_{ij}^{l}} \cdot \frac{\partial F_{ij}^{l}}{\partial O} + \gamma \frac{\partial \mathcal{L}_{color\_swap}}{\partial F_{ij}^{l}} \cdot \frac{\partial F_{ij}^{l}}{\partial O} \tag{14}$$

The update formula for the output image is: $O := O - \eta \frac{\partial \mathcal{L}_{total}}{\partial O}$.

IV. EXPERIMENTS AND RESULTS

A. IMPLEMENTATION DETAILS AND DATA SETS

In this paper, VGG19 is fine-tuned in medical data sets to adapt to the research of this paper. Some numerical optimization strategy can use the gradient with respect to the pixel

Algorithm 1 Coloring and Imaging

Input:

target medical image I that need to color, and fine-tuned VGG19.

Output:

colored medical image O with the content structure of the target medical image I

1: The reference medical image $S \leftarrow$ adaptive search strategy.
2: Calculate I' and S'. Resize medical image I and S to the fixed size.
3: Calculate I_{seg} and $S_{seg} \leftarrow K\_means(I', S')$, where $K\_means()$ is the segmentation algorithm by using k-means clustering method.
4: Extract feature maps $F^L_{I'}, F^L_{S'}, F^L_{I_{seg}}, F^L_{S_{seg}} \leftarrow VGG(I', S', I_{seg}, S_{seg})$, where $VGG()$ is the VGG19 network.$L = \{1, 2, 3, 4, 5\}$
5: **if** O does not exist **then**
6: Random initialization to generate a noisy image O
7: **end if**
8: **for** $t = 1$ to $maxiter$ **do**
9: Feature maps $F^L_O \leftarrow VGG(O)$
10: Calculate the gradient of whole loss function $\mathcal{L}_{total}$ with respect to the O using Equ 14.
11: Update the output medical image O by $O^{(t)} \leftarrow O^{(t-1)} - \eta \frac{\partial \mathcal{L}_{total}}{\partial O}$, where use the gradient-base optimization to minimize $\mathcal{L}_{total}$.
12: **end for**
13: $O' \leftarrow resize(O)$, where the size of medical image O' is the same as the size of I. $resize()$ is the resize function.
14: $O \leftarrow O'_{UV} + I_Y$, where O'_{UV} is the U and V channels of O' in YUV color space, and I_Y is the Y channel of I in YUV color space.

values $\frac{\partial \mathcal{L}_{total}}{\partial O}$ as input. Here we use L-BFGS-B [33]. The optimization of this paper is initialized with the output of the Style algorithm [21] replaced the style loss by augmented style loss [25]. The result initialized by which has a better optimization work than using (6) directly. For natural images, the region relationship between images can be matched by semantic labels, but it is not applicable for medical images without labels. So we use the similarity region matching map obtained by clustering the target image as the image segmentation mask to guide to color the target image.

For the data sets of our paper, a part of the MRI images come from ADNI dataset, the other come from Glioma dataset. The ADNI dataset is the important component of the comprehensive dataset collected in The Alzheimers Disease Neuroimaging Initiative (ADNI) whose primary purpose is informing the design of therapeutic trials in Alzheimers disease(available at http://adni.loni.usc.edu). Glioma dataset comes from Harvard University Public Dataset (available at http://www.med.harvard.edu/AANLIB/home.htm). Ultrasound images come from a 3-A major Hospital in Chongqing, China, collected from 2014 to 2015, include intrauterine early pregnancy, gallbladder, liver, kidney and so on. CT images come from the dataset of the first Tianchi medical AI competition:intelligent diagnosis of pulmonary nodules. The contest dataset provides thousands of low-dose CT (MHD) images of the lungs of high-risk patients, each contains a series of axial sections of the thoracic cavity (available at https://tianchi.aliyun.com/competition/entrance/231601/information). The real-human-slice dataset is constructed by selecting the color cryosections image. The color cryosections images come from the Visible Human Project of U.S.National Library of medicine. The VHP provides a public-domain library of cross-sectional cryosection, CT, and MRI images obtained from one male cadaver and one female cadaver. The Visible Man data set was publicly released in 1994 and the Visible Woman in 1995(available at https://www.nlm.nih.gov/research/visible/photos.html). The pseudo-color dataset is constructed by selecting the pseudo-color image from the Internet. https://www.vcg.com/creative, this network provides free figures for everyone.

TABLE 1. The objective evaluation of different γ.

γ	1	10	10^2	10^3	10^4
PSNR	30.48±2.43	**29.86 ± 1.53**	29.48 ± 1.62	17.57 ± 0.96	16.15 ± 1.05
SSIM	0.93 ± 0.03	**0.89 ± 0.03**	0.87 ± 0.05	0.35 ± 0.06	0.28 ± 0.05
FSIM	0.97 ± 0.01	**0.96 ± 0.01**	0.97 ± 0.01	0.72 ± 0.04	0.67 ± 0.05

B. PARAMETERS

This section describes the parameters of our experiments. Since pseudo color is richer in color than physical color, the experimental effect can be seen more clearly, so we use the reference image from the pseudo-color dataset to experiment. In this paper, we fix the weight α of the content loss to 1, and the weight Γ of the color loss to 10^2, refer to [21] and [25]. Likewise, the weight β_l of the contribution of the color loss in each layer is fixed to 1, refer to [21] and [25]. To illustrate the effect of γ, we fixe λ to 10^4, and produce results using our approach with different value of γ on the data set. This data set is made up of three modality medical image data sets:MRI,CT and Ultrasound data sets, detailed by Section IV-A. Fig. 4 shows the effect of different value of γ on the result. Table 1 shows the average value and the standard deviation of PSNR, SSIM and FSIM on this data set. We can see from Table 1, $\gamma = 1, 10, 10^2$ can get higher objective evaluation value than other γ value. But when $\gamma = 1$, the same organs has different colors, e.g. the telenecephalon. Combine with visual perception of result, $\gamma = 10$ is better, because results are more clearly, and can keep spatial relationship between the pixels.

To illustrate the effect of λ, we produce colored result by setting different λ value when $\gamma = 0$. The results are shown in Fig. 5. Through experiments, it is found that the feature local mapping loss $\mathcal{L}_{color\_swap}$ can affect the effect of $\mathcal{L}_Y$. There is a certain proportional relationship between weights of λ and γ. So we explore the optimal rate between λ and γ.

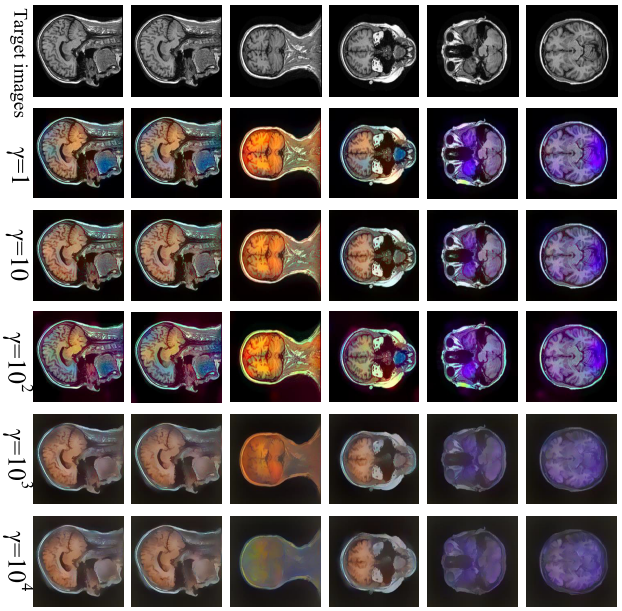


FIGURE 4. Different results about γ ($\alpha_Y = 0$). Different values of γ has different results, when γ is too large, the image blurs. The objective evaluation (Table 1) also shows the content will loss the more when the value is higher. The Feature Local Mapping loss can affect its content structure because of the feature of the output image combined with one of the reference image, therefore, the higher the weight, the lower the SSIM and PSNR. By comprehensively considering the experimental results and the objective evaluation, $\gamma = 10$ is the best parameter. When $\gamma = 10$, the output image have high the objective evaluation while maintain the local relationship between the pixels, since the same organ keeps the same color.

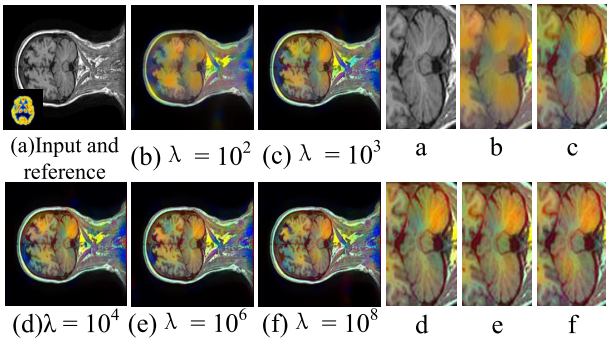


FIGURE 5. Different results about λ. A too small λ value cannot prevent details, the content of the result is not clear, look detail in (b,c). As λ increases, the sharpness of the image increases gradually (a-f show detail of results), in other words, $\mathcal{L}_Y$ can maintain the texture of the target image.

The Fig. 6 shows the different results with the different rates (defined by λ/γ), the objective values are calculated on 100 images, the parameter $\alpha_Y = 0$. Table 2 shows the values of the PSNR,SSIM and FSIM. When rate is too small, 10^{-1}, SSIM(the second column) is the lowest, it means content of the output image is not preserved. As increase with rate, the SSIM and FSIM increase. However, when rate is changed from 10^4 to 10^5, mean values of the three indicators slightly change, but the standard deviation only changes 0.2, which means the result are more sensitive to image quality and

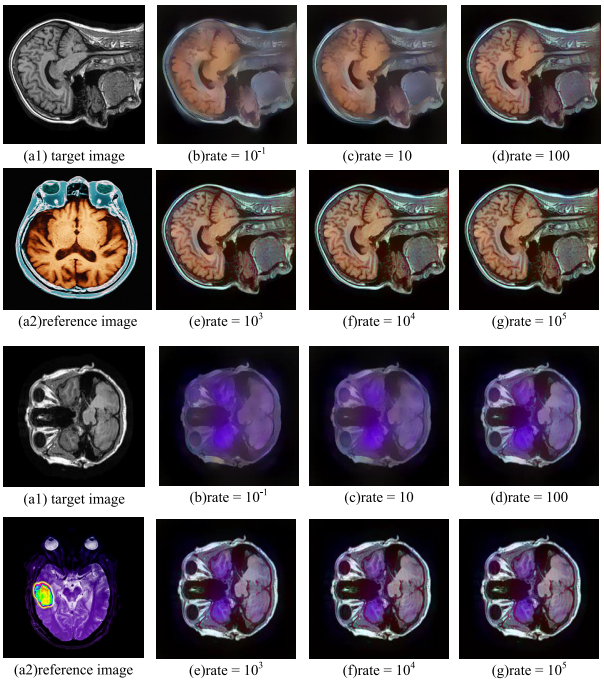


FIGURE 6. Results of The Different Rate (λ/γ). We produce results using our method with different rate. A too small rate make the blur result and thus the result cannot be used to diagnose disease look in ((b),(c)). When rate is more than 10^3, the result is clear look in ((e)-(g)).

TABLE 2. The objective evaluation of different rate.

Rate	-1	10	10^2	10^3	10^4	10^5
PSNR	17.61 ± 1.46	18.71 ± 1.27	23.73 ± 1.51	29.70 ± 1.41	$\mathbf{31.69 \pm 1.94}$	31.78 ± 2.19
SSIM	0.45 ± 0.09	0.49 ± 0.07	0.70 ± 0.06	0.89 ± 0.03	$\mathbf{0.94 \pm 0.01}$	0.95 ± 0.01
FSIM	0.70 ± 0.05	0.73 ± 0.04	0.86 ± 0.04	0.96 ± 0.01	$\mathbf{0.97 \pm 0.01}$	0.98 ± 0.01

more unsteadiness. Therefore, we comprehensively consider objective evaluation value and steadiness, rate $= 10^4$ is the best parameter.

In order to more intuitively demonstrate the effect of different α_Y, we show different experiment results in Fig. 7. We use three medical images data sets obtained from three different imaging principles including CT, MR and Ultrasound. Each of data sets randomly selects 100 images for experiment. We use the same framework only difference in $\mathcal{L}_Y$ to generate result. We calculate the objective evaluation (PSNR and SSIM) of the result when the Y-loss with $\alpha_Y = 1$ and $\alpha_Y = 0$. Fig. 8 show the PSNR and SSIM of different kinds of images, respectively.

In Fig. 7, $L1$ represents ℓ_1 norm, similarly, $L2$ represents ℓ_2 norm. To compare the effective of the norm, we pay more attention to SSIM of the image. As shown in Fig. 8, $\mathcal{L}_Y$ constructed by ℓ_2 can get better SSIM in CT and MRI datasets, because the SSIM of more than half of results obtained by ℓ_2 is higher than results obtained by ℓ_1 in boxplot. From the difference between the minimum and maximum value we can see that, Measures of Dispersion of the results obtained by ℓ_2 is lower than the results obtained by ℓ_1. However, in the Ultrasound dataset, the minimum of the results obtained by ℓ_1 is higher than the minimum obtained by ℓ_2, which because

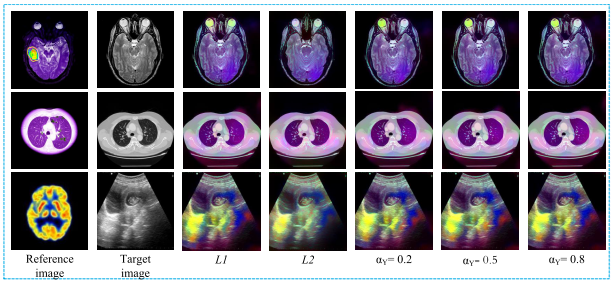
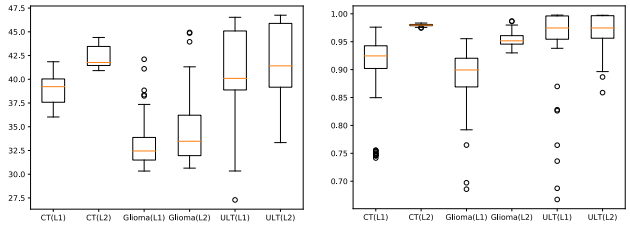


FIGURE 7. Results of The Different α_Y. *L1*(the third column) is ℓ_1 when $\alpha_Y = 1$, *L2*(the fourth column) is ℓ_2 when $\alpha_Y = 0$. The fifth, sixth, and seventh columns show the results of different α_Y value.



(a) PSNR Values (b) SSIM Values

FIGURE 8. The Objective Evaluations on Three Datasets. Both PSNR and SSIM are calculated from the images generated by the Y-loss with $\alpha_Y = 1$ and $\alpha_Y = 0$.

of the sparse selection nature of ℓ_1. Moreover, the Ultrasound image have serious noise due to its imaging principle, the effect of this noise on the colored result is reduced.

C. RESULTS AND COMPARISON

To assess the results comprehensively, the subjective and objective evaluation are conducted on colorized images. The former includes common image transformation metric, the latter focuses on chief doctor diagnosis.

We prove the effectiveness of our method by comparing with the following algorithms. The dataset of our experiment is constructed by MRI images, CT images and ultrasound images, respectively. In order to show the two different purposes of our algorithm at the same time, we use the image of the pseudo-color dataset as the reference image of CT and ultrasound images to highlight the organ, e.g the lung is dyed purple, and use the image of real-human-slice dataset as the reference image of MRI images to restore real tissue color.

In order to compare the experimental results of various color transfer methods, we fixe reference images to maintain the same conditions except for colorization methods. Then we analyze the results by combining the objective evaluation (shown in Table 3) with visual effect (shown in Fig. 9). We compare our method with TCGI [11], ICTEDE [14] and CTCCS [12] across a series of CT, MRI and ultrasound medical images, these method is based on hand-craft feature. This three methods transfer color from the reference image to the target image by the hand-craft feature that include mean, covariance, luminance and pixel value. We can see that the result of ICTEDE [14] have not the color of the reference

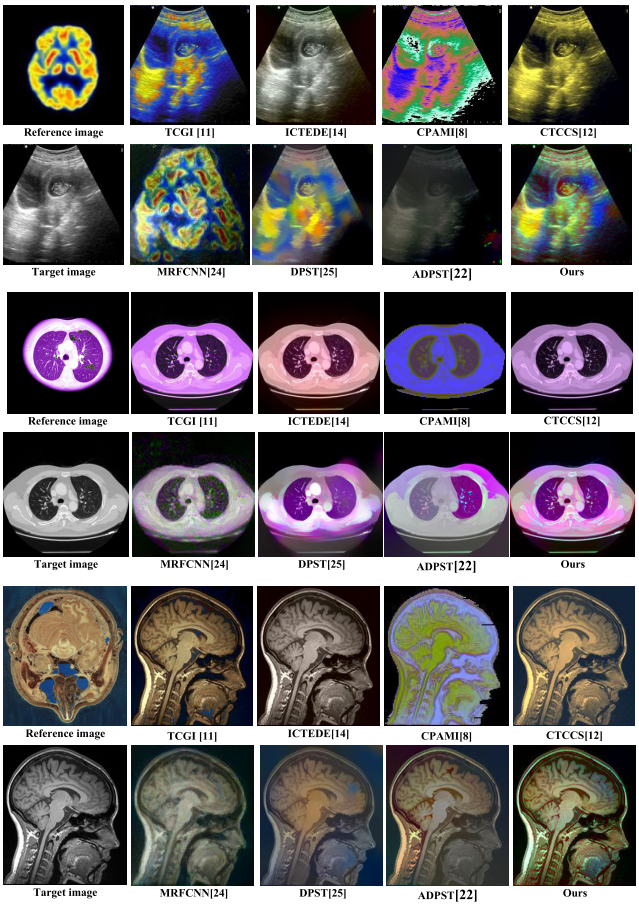


FIGURE 9. Comparison of our method against seven methods. TCGI [11], ICTEDE [14], CPAMI [8], and CTCCS [12] color medical images based on hand-craft features. MRFCNN [24], DPST [25], ADPST [22] are recent color transfer methods based on deep features.

image, because the method of ICTEDE finds the main color of the reference image and further select target scheme to change the pixels of grayscale image to close the theme by linear transformation. Therefore, this method globally change the color of the target image, the result depends on target scheme. The CTCCS [12] is same as ICTEDE in global linear transformation. So, the results of two methods only have less colors, not transfer the whole color of the reference image to the target image. TCGI [11] is the local match method that can color different colors. The visual effect of result based on the luminance and neighborhood statistics of reference image, in the other word, the results of TCGI have not sematic information, e.g. the result should not have the color of water. Our approach adds the segmentation of the reference and target images, so the result have the sematic information.

We also compare with CPAMI [8]. The results of CPAMI are drastically different from ours because CPAMI focuses on regional significance ignore texture details. Therefore, the results of CPAMI have lower PSNR and SSIM, but higher entropy as shown in Table 3. Our result focuses on the detail of content, the entropy is closer to entropy of the target image(the values of the brackets in Table 3).

TABLE 3. The objective evaluation of results.

Methods	TCGI[11]	ICTEDE[14]	CTCCS[12]	CPAMI[8]	MRFCNN[24]	ADPST [22]	DPST[25] $baseline-method$	our approach
MRI_PSNR	30.470	**37.459**	20.686	12.389	20.484	20.249	18.337	34.086
CT_PSNR	33.545	**38.142**	23.357	12.836	21.357	18.681	19.195	35.051
ULT_PSNR	25.997	34.274	25.239	11.074	15.931	12.700	20.611	**41.797**
MRI_SSIM	0.931	0.925	0.775	0.163	0.475	0.710	0.637	**0.964**
CT_SSIM	**0.955**	0.863	0.881	0.577	0.433	0.745	0.714	0.939
ULT_SSIM	0.935	0.808	0.824	0.240	0.218	0.427	0.691	**0.984**
MRI_FSIM	0.946	**0.985**	0.934	0.606	0.736	0.832	0.800	0.942
CT_FSIM	0.925	0.939	**0.944**	0.712	0.744	0.881	0.812	0.924
ULT_FSIM	0.907	**0.986**	0.909	0.651	0.648	0.683	0.769	0.911
MRI_Outrval	89.691	91.913	79.861	58.445	51.738	47.514	42.831	**101.445**
CT_Outrval	**69.088**	62.727	62.689	33.777	56.524	40.162	43.907	64.733
ULT_Outrval	79.270	61.956	64.747	**134.618**	94.195	16.563	40.171	58.780
MRI_Entropy	7.280	6.859	6.990	**5.316**	7.288	6.994	6.758	7.2126(7.168)
CT_Entropy	5.871	6.402	5.318	**3.348**	7.088	5.820	6.664	6.007(5.704)
ULT_Entropy	6.778	7.094	6.238	**4.946**	7.616	5.474	6.754	7.040(6.696)

We compare with DPST [25], MRFCNN [24] and ADPST [22] that image-based iterative method by deep learning, which transfer the color to the target image. The result of MRFCNN [24] has unclear texture, the detail of the colored brain MRI image is blurred such as ectocinerea. ADPST [22] transfer color based on segmentation image obtained by PSPNet, which is trained on ADE20K dataset. Colored lung CT obtained by ADPST has uneven color and the ultrasound has little color on it. Seen from the SSIM of penultimate column in the Table 3, the approach of DPST [25] is blurrier than our result. We calculate average growth rate between the result of DPST and our method. The average growth rate of SSIM is about 0.292, and rate of PSNR is about 0.472. We show more results in Fig. 12. The tissue structure of the colored medical image is more salient than the grayscale one, e.g., The last column in Fig. 12 is the colored chest CT images, the lung in the image is highlighted by the different color, so doctor can easier to observe pulmonary.

D. QUALITATIVE EVALUATION CRITERION

In this subsection, we give the qualitative evaluation criterion of our approach. We display the target image and color results to three doctors (radiologists) who are from the class-A major hospital in Chongqing, China. The two of them are the director doctor and medical students of the Department of Radiology, respectively. Another comes from the Department of Ultrasound. The three doctors give us a report about the results obtained by our algorithm, including false color and physical color results. This report includes three scores, the one is whether the colored results can influence diagnostic, the one is whether histological structure of organs is clear, and the last one judge whether the coloring effect is good or not, which is only for the real physical color result (The level is 1 to 5, 5 represents the best level, 1 represents the worst level). The scores are shown in Table 4.

TABLE 4. The scores of qualitative evaluation criterion.

Score	$score\_one$	$score\_two$	$score\_three$
MRI	4.88	4.88	4.60
CT	4.93	4.87	4.70
Ult	3.32	4.90	4.75

For the score_one in Table 4, it can be seen that the coloring results of MRI and CT can be used as the diagnosis basis for the doctor, which also reflects that the doctors agree with this method. But for the ultrasound results, the doctors give a lower score, because colorization make the information inconspicuous for the original color ultrasound image. The score_two indicates that the doctor could not detect a detailed change in the coloring results. Obviously, score_two explains the proposed method possessed sufficient feasibility. For the score_three, it can be seen that the values are between 4 and 5, which indicates that the doctor is still satisfied with the results with true physical color.

V. DISCUSSION

We first discuss the convergence rate of losses. The convergence rate of losses is plaint in Fig. 10 under the same conditions. The total loss consists of only two losses in the process of colorization (written in parentheses in the top right of Fig. 10), which avoid the result to be influenced by other loss functions. We calculate the value of $\mathcal{L}_c$ (L_content in Fig. 10) by using the total loss that consists of the $\mathcal{L}_c$ loss and $\mathcal{L}_{color^+}$ loss. The value of $\mathcal{L}_{\mathcal{M}}$ (L_m in Fig. 10) and Y-loss are calculated by combining $\mathcal{L}_{color^+}$ and $\mathcal{L}_{\mathcal{M}}$ or Y-loss as the total loss. Y-loss contains $\alpha_Y = 0$ (Yloss(MSE) in Fig. 10) and $\alpha_Y = 1$(Yloss(L1) in Fig. 10), $\mathcal{L}_{\mathcal{M}}$ is defined in (4). In order to see its convergence, the value of each loss is kept in same range, which does not affect its convergence.

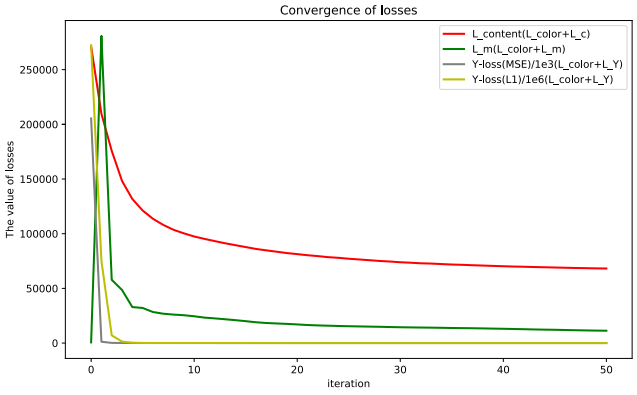


FIGURE 10. Losses Convergence rate. It shows convergence rate of the four losses in the process of colorization, which are $\mathcal{L}_c$ (L_content), $\mathcal{L}_{\mathcal{M}}$ (L_m), $\mathcal{L}_Y$ ((Yloss(MSE)), when $\alpha_Y = 0$, and (Yloss(L1), when $\alpha_Y = 1$)).

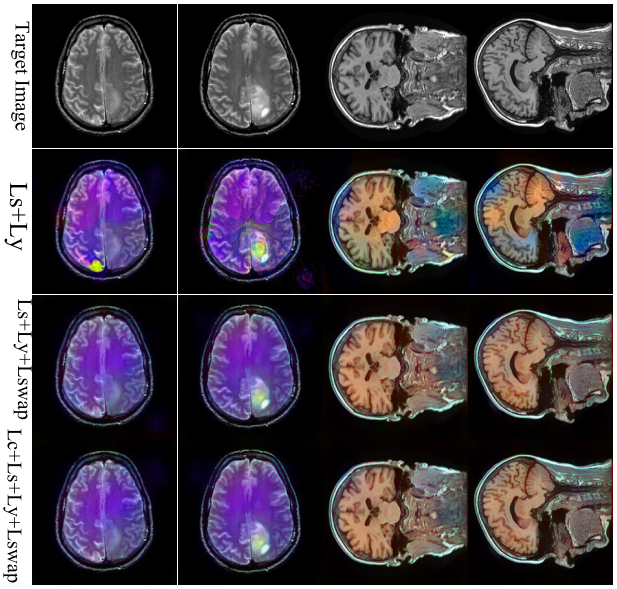


FIGURE 11. Results about different losses combination. It can be seen the first column of the result shown only the style loss combine with Y-loss to be used generate output image, and the second column show the result of adding local feature map to ensure its spatial relationship and reduce the uneven color distribution. The third column show the result of adding the content loss to constrain content changes from deep feature level.

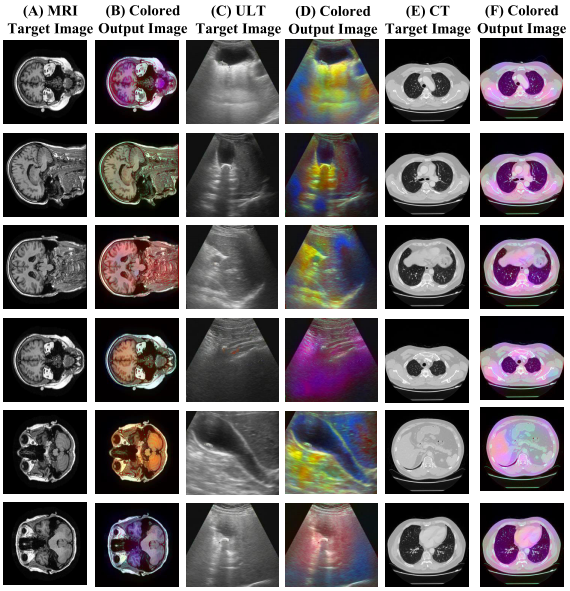


FIGURE 12. Our Result. The (a),(c),(e) columns are the target images from three different medical diagnostic equipments. The reference images come from public images. The (b),(d),(f) columns are the colored medical images using our method.

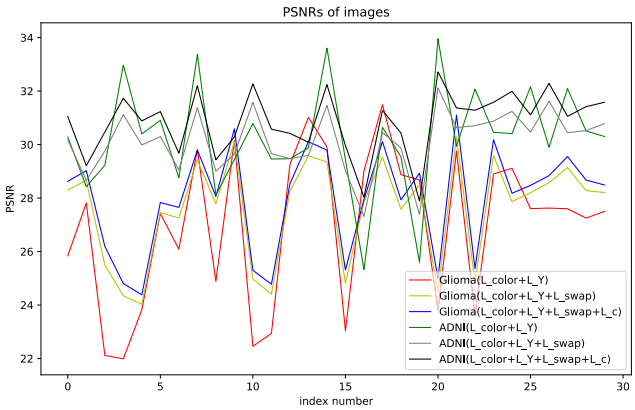


FIGURE 13. PSNR value. It shows that PSNRs of 60 images with different loss combinations. There are 60 images from two different data sets, each have 30 images, we calculate PSNR of each image to draw this figure. Contrast the green and gray lines (or red and yellow lines) can be found, $\mathcal{L}_{color\_swap}$ can change content of the output image. Contrast the gray and black lines (or yellow and blue lines) can be found, $\mathcal{L}_c$ can enhance content invariance.

We iterate 500 times in each experiment, and draw a point every 10 times. It can be seen that the convergence speed of the Y-loss is faster than other losses.In other word, Y-loss can quickly construct the content of the output image.

Secondly, we discuss the rationality of the total loss function in this paper, we analyze the rationality by loss ablation experiment. In order to see the effect clearly, we only show the output images obtained from the colorization module without entering the imaging module to magnification optimization. We randomly select 30 images from ADNI and Glioma databases to color respectively, and then calculate their objective evaluation severally. The Fig. 13 and Fig. 14 show the PSNR and SSIM in two datasets each with 30 images.

In this two data sets, the result obtained by $\mathcal{L}^l_{color+}$ with $\mathcal{L}_Y$ has bad visual effect, which has different colors in one brain tissue. In order to maintain the same color in one structure, the total loss adds $\mathcal{L}_{color\_swap}$. The results of the second and third row in Fig.11 show that the great effect of $\mathcal{L}_{color\_swap}$. But the content could be changed, because the feature maps of the output image is similar to the changed feature maps obtained in local feature mapping. In order to alleviate this change, we add $\mathcal{L}_c$ to preserve the content of image in deep feature domain. In Fig. 13 and Fig. 14, the objective results with added $\mathcal{L}_c$ loss is higher than without it (compare black line with gray line, blue line with yellow line). In other word,

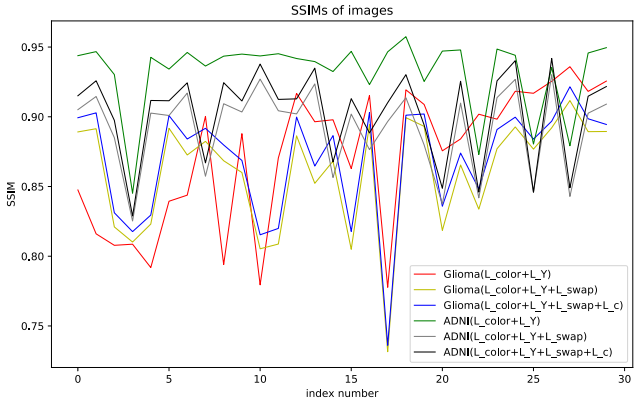


FIGURE 14. SSIM value. It shows that SSIMs of 60 images with different loss combinations. There are 60 images from two different data sets, each have 30 images, we calculate SSIM of each image to draw this figure. This objective evaluation (PSNR shown in Fig. 13 and SSIM) shows the $\mathcal{L}_{color\_swap}$ can influence the content of output images because of the changed feature maps in the process of local feature mapping. And the $\mathcal{L}_c$ can preserve the content which is analyzed from the yellow and blue lines or gray and black lines.

$\mathcal{L}_c$ can preserve content of the output image. Therefore, our total loss consist of four losses, which are $\mathcal{L}_c$, $\mathcal{L}_{color\_swap}$, $\mathcal{L}^l_{color^+}$, and $\mathcal{L}_Y$.

Thirdly, we introduce the runtime of our method. Our method achieve $O(n)$ time complexity. The runtime is related to image size and number of iterations. Most of runtime is spent in the colorization stage, in this stage, we initialize the output image by Section IV-A. We consider the visual effect, iteration times is set to 1000 in our experiment. One color medical image with fixed size (224×224) is generated in approximately 6 minutes on an Nvidia 1050 GPU.

Fourthly, we introduce how to remove noise. We actually use two points to suppress the generation of noise during the experiment. First, we use total variation(TV) loss to the process of generating color medical images to suppress noise and the weight of TV loss is 10^{-3}. The second point is that joint bilateral filter is added in the post-processing stage. We refer to the approach of [25]. The optimal parameters for each locally affine model can be obtained by fitting the local regression of each patch after the optimization iterations. Then we use a joint-bilateral filter guided by the input image to smooth out the noise across the coefficients of the locally affine matrices and reconstruct the output image free of noise. The details of these two points are shown in the code we provided.

VI. CONCLUSION

In this paper, a novel colorization method of medical images has been proposed by combining deep learning with color transferring. The extensive results in multi-modal medical images, such as CT, MRI, and UIT, demonstrate the effectiveness of our method. We use Y-loss builded by hand-craft feature and content loss builded by deep feature to preserve content invariance of the target images. In multi-modal medical images data sets, compared with the methods using deep and hand-craft features respectively, our method works better. Then, we add the automatic retrieval algorithm to avoid to manually retrieve the reference images. Through the selection of different reference image datasets, our algorithm can achieve the purpose of obtaining false color or real physical colors to meet the needs of different people. Our result enhances the image texture feature information and improves the visual effect for clinicians, so our algorithm can apply to medical diagnostic equipment for getting colored medical image in theory. Therefore, the future work will focus on cooperation with doctors or medical college to diagnosis and study diseases. We hope our algorithm will be used in clinical practice and education to improve relationships between hospitals and patients and promote medical development.

REFERENCES

[1] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognit.*, vol. 92, pp. 64–81, Aug. 2019.

[2] M. U. G. Khan, Y. Gotoh, and N. Nida, "Medical image colorization for better visualization and segmentation," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, 2017, pp. 571–580.

[3] T. H. Khan, S. K. Mohammed, M. S. Imtiaz, and K. A. Wahid, "Efficient color reproduction algorithm for endoscopic images based on dynamic color map," *J. Med. Biol. Eng.*, vol. 36, no. 2, pp. 226–235, Apr. 2016.

[4] C. Lan, P. Lan, and Y. Cao, "Pseudo-color processing of medical images," *Chin. J. Stereol. Image Anal.*, vol. 7, no. 3, pp. 166–167, Sep. 2002.

[5] N. A. Semary, "A proposed HSV-based pseudo coloring scheme for enhancing medical image," in *Proc. Comput. Sci. Inf. Technol.*, Feb. 2018, pp. 251–262.

[6] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.

[7] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3085–3094. [Online]. Available: https://arxiv.org/abs/1903.00179

[8] X. Zeng, A. Chen, and M. Zhou, "Color perception algorithm of medical images using density peak based hierarchical clustering," *Biomed. Signal Process. Control*, vol. 48, pp. 69–79, Feb. 2019, doi: 10.1016/j.bspc.2018.09.013.

[9] C. Ning, "Medical image pseudo-color processing based on object enhancement," *J. Changchun Univ. Sci. Technol. (Natural Sci. Ed.)*, vol. 35, no. 4, pp. 185–187, Apr. 2012.

[10] X. Zeng, A. Chen, and S. He, "Color perception algorithm of medical images using multi-features fusion," *J. Comput.-Aided Des. Comput. Graph.*, vol. 30, no. 3, pp. 375–384, Mar. 2018.

[11] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 277–280, Jul. 2002.

[12] X. Xiao and L. Ma, "Color transfer in correlated color space," in *Proc. ACM Int. Conf. Virtual Reality Continuum Appl. (VRCIA)*. Hong Kong: Chinese Univ. of Hong Kong, 2006, pp. 305–309.

[13] X. Huo, J. Tan, and R. Wang, "Color transfer based on combining subtractive clustering with FCM clustering," in *Proc. 10th IEEE Int. Conf. Comput.-Aided Design Comput. Graph.*, Beijing, China, 2007, pp. 461–464.

[14] L. He, H. Qi, and R. Zaretzki, "Image color transfer to evoke different emotions based on color combinations," *Signal, Image Video Process.*, vol. 9, no. 8, pp. 1965–1973, Nov. 2015.

[15] 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, pp. 1–11, Jul. 2016, doi: 10.1145/2897824.2925974.

[16] Z. Su, X. Liang, J. Guo, C. Gao, and X. Luo, "An edge-refined vectorized deep colorization model for grayscale-to-color images," *Neurocomputing*, vol. 311, pp. 305–315, Oct. 2018.

[17] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5188–5196.
[18] S. Karen, V. Andrea, and Z. Andrew, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2013, pp. 1–8, [Online]. Available: https://arxiv.org/abs/1312.6034
[19] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," in *Proc. ICML Deep Learn. Workshop*, Lille, France, 2015, pp. 5188–5196.
[20] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, 2015, pp. 262–270.
[21] 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)*, Las Vegas, NV, USA, Jun. 2016, pp. 2414–2423.
[22] S. Penhouët and P. Sanzenbacher, "Automated deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jan. 2019, pp. 1–14. [Online]. Available: https://arxiv.org/abs/1901.03915v1
[23] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 694–711.
[24] C. Li and M. Wand, "Combining Markov random fields and convolutional neural networks for image synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2479–2486.
[25] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6997–7005.
[26] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
[27] J. Liu, F. Chen, C. Pan, M. Zhu, X. Zhang, L. Zhang, and H. Liao, "A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1943–1952, Sep. 2018.
[28] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, May 2019, doi: 10.1109/TBME.2018.2872652.
[29] B. Zhou, X.-M. Duan, W. Wei, D.-J. Ye, M. Wozniak, and R. Damasevicius, "An adaptive local descriptor embedding Zernike moments for image matching," *IEEE Access*, vol. 7, pp. 183971–183984, 2019.
[30] E. Yang, C. Deng, C. Li, W. Liu, J. Li, and D. Tao, "Shared predictive cross-modal deep quantization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5292–5303, Nov. 2018.
[31] C. Deng, E. Yang, T. Liu, J. Li, W. Liu, and D. Tao, "Unsupervised semantic-preserving adversarial hashing for image search," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4032–4044, Aug. 2019, doi: 10.1109/TIP.2019.2903661.
[32] M. Fei, J. Li, L. Shao, Z. Ju, and G. Ouyang, "Robust visual tracking based on improved perceptual hashing for robot vision," in *Proc. Int. Conf. Intell. Robot. Appl. (ICIRA)*, Portsmouth, U.K., 2015, pp. 331–340.
[33] C. Y. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Soft.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.

XIANHUA ZENG was born in 1973. He received the Ph.D. degree in computer software and theory from Beijing Jiaotong University, in 2009. He was a Visiting Scholar with the University of Technology, Sydney, from August 2013 to August 2014. He is currently a Professor at the Chongqing Key Laboratory of Computational Intelligence, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China. His main research interests include medical image processing, machine learning, and data mining.

SHIYUE TONG was born in Sichuan, China, in 1995. She received the B.S. degree in computer science and technology from Southwest Petroleum University, China, in 2017. She is currently pursuing the M.S. degree with the Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications. Her research interests are in machine learning, color transfer, and medical image processing.

YUZHE LU was born in Hubei, China, in 1996. He received the B.S. degree in computer science and technology from the Chongqing University of Posts and Telecommunications, China, in 2018, where he is currently pursuing the M.S. degree with the Department of Computer Science and Technology. His research interests are in machine learning, style transfer, and image processing.

LIMING XU received the B.S. and M.S. degrees in computer science and technology from China West Normal University, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interests are in machine learning, generative adversarial nets, and medical image processing.

ZHIWEI HUANG received the bachelor's degree from the Department of Mechanical Engineering, North University of China, Taiyuan, China, in July 2001, the M.S. degree from the School of Electronics and Mechanical Engineering and the School of Manufacturing Science and Engineering, Sichuan University, in July 2006. He is currently pursuing the Ph.D. degree with the Chongqing University of Posts and Telecommunications. He is currently a Vice Professor of Southwest Medical University. His research focuses on image processing analysis by means of artificial intelligence.