The Journal of the Society for Art and Science, Vol. 22, No. 4, pp. 14:1-14:9 (2023)

A Cross-domain Color Mapping from Exemplar Anime Image Colorization Networks

Jincheng Peng[1)] (Student Member) **Guoyue Chen**[2)] **Kazuki Saruta**[2)] **Yuki Terata**[2)]

1) Graduate School of Akita Prefectural University

2) Department of Information and Computer Science, Akita Prefectural University

{d23s002, chen, saruta, terata} at akita-pu.ac.jp

Abstract

Image colorization is a hot topic in the research of image processing and generation techniques. It plays an important role in restoring old photos, colorization medical images, and colorization anime manuscripts. In recent years, with the rapid development of deep learning, some colorization methods based on deep learning have been proposed, and satisfactory colorization results have been achieved. However, these colorization methods generally use unguided feature learning to achieve color prediction, it cannot predict, select, and intervene in image colors based on the subjective ideas of the creator. At the same time, the generated colorization results may have unreasonable phenomena such as color concentration, saturation and brightness distribution. Therefore, we propose a cross-domain color mapping from exemplar anime image colorization method, which is used for coloring a given anime image with a reference style. This method involves a supervised colorization network, which is divided into two structures: a domain semantic matching sub-network and a colorization sub-network. The domain semantic matching sub-network maps the pixels of the input grayscale image domain and the reference color image domain to an intermediate domain by calculating the similarity of the correlation matrix, thereby establishing a corresponding semantic relation and obtaining a coarse colorization map. We also employed an end-to-end CNN encoder and decoder network to further extract the feature information of the matched image colors, and used the Lab color space to select, propagate, and predict the color distribution features of the image. Experimental results have demonstrated that the anime coloring images generated by our network have high restoration similarity compared to the original reference images, reasonable coloration in the coloring area, and this colorization method solves the problem of prior color style transfer in anime images.

1. Introduction

With the development of the animation, gaming, and multimedia industries, users are increasingly demanding high-definition, high-quality image experiences. However, most anime images are produced in grayscale, and readers can only understand the meaning and information expressed by the anime images through a dozen or so gray categories. In contrast, colored images, with their highly distinctive features, allow users to have more imagination and rich visual experiences of what the creator wants to express. Therefore, in the process of anime image creation, anime coloring is one of the most important steps. However, due to its complex and labor-intensive process, traditional manual anime coloring is no longer suitable for the current big data information era. Therefore, new research methods are needed in the anime industry to improve coloring efficiency and achieve automated coloring processes.

In recent years, with the development and application of multimedia and deep learning network technology in the field of computer vision, the mainstream method of image colorization has shifted from traditional machine learning to using the powerful parameter learning ability of deep neural networks to learn how to select, propagate, and predict color distributions from large-scale data. Deep learning image colorization is generally divided into two modes: guided colorization and unguided colorization. Unguided colorization typically involves the underlay network synthesizing a large amount of extracted edge, shape, and color information data to make reasonable choices, propagate and predict color distribution for the generated image. However, this colorization method is an ill-posed problem, meaning the network has a large degree of randomness in predicting color distribution, resulting in areas of unrealistic coloring and artifacts, which negatively affects the overall visual quality. Additionally, the generated color style is determined by the network learning and cannot be well adapted to user understanding and artistic creativity. Guided colorization refers to coloring a specified area in a target grayscale image with user-provided color prompts. One type of guided colorization research is based on exemplar colorization, where a reference exemplar is used to transform the grayscale image domain X to a reference image domain Y. The content of the original image $\alpha(a, X)$ is extracted from its domain properties and its semantic relevance to the properties of the reference image domain Y is identified, allowing for reasonable color prediction and distribution in the original image based on the properties of the reference image domain Y, resulting in a new image $\beta(a, Y)$ with the reference image color style attributes.

The main focus of this study is to establish a cross-domain color correspondence relationship between the grayscale domain image X and the reference domain image Y before coloring the image. To address this issue, this paper proposes A Cross-domain Color Mapping method for anime images exemplar colorization. The colorization network consists of two structures: the domain matching sub-network and the colorization sub-network. The domain semantic matching module refers to the design idea of CoCosnet [1] semantic matching. First, two pyramid network models are used to extract feature maps between different domains from the image to be colored and the reference image. Then, the domain semantic matching sub-network maps the images of the two domains to an intermediate domain. By calculating the pairwise similarity between pixel positions in the two domains using a correlation matrix, a correspondence relationship is established between the input image A domain and the reference instance image B domain, and a rough chromaticity map is obtained. Meanwhile, we design an end-to-end CNN encoder and decoder network to learn and extract the color features of the input image of the domain semantic matching module. The network selects, propagates, and predicts the color distribution of the ab channel by skipping layers to the Lab color space image feature. Finally, by combining the original L channel, the transformation from the feature domain to the image domain is completed. Through experiments, we compare the network with previous relevant networks and prove that the colorization network method produces images with better restoration accuracy and coloring effects that are closer to the reference image. Utilizing this method for interactive coloring enables the restoration of older black and white anime images, avoiding the cumbersome manual coloring process. Additionally, it addresses the limitation of conventional coloring networks in generating creative coloring, allowing users with limited artistic knowledge to effortlessly create vibrant colored anime images. Therefore, it holds significant practical applications and commercial value. In the future, there is an opportunity to apply this model to line art anime coloring and conduct further in-depth research.

2. Related Work

Among the non-guided colorization studies, Lizuka et al [2] converted the image colorization problem into an image classification problem using a two-channel network that combines information about local features and global prior information in the image. Automatic colorization of gray scale images of any size. Larsson et al [3] used VGG networks as semantic analysis and retrieval of localized information integrated into a colorization system. Have the coloring system predict the color histogram of each image location to predict the color distribution of each pixel. While these networks can roughly colorize grayscale images based on color distribution, using convolutional neural networks alone to extract color features can lose semantic information of the domain, resulting in misclassification and unsatisfactory visual experience for viewers in terms of coloring effect.

In recent years, with the rapid development of generative adversarial networks consisting of generators and discriminators, a conditional generative adversarial network-based colorization method was proposed in reference [4]. This method predicts the color distribution of each domain by using the discriminator to predict the loss between generated images and real images. Deshpande et al. [5] embedded VAE-encoded color modeling into the Gaussian distribution of grayscale images and sampled different colors from the Gaussian distribution. However, this method has the drawback of excessive randomness due to the addition of strong noise in each layer, which leads to uncontrollable coloring results and affects the quality of image generation. Although these colorings achieve satisfactory results, they are unable to predict and control the color distribution according to the subjective will of the creator since they are unguided colorings.

In guided coloring-related research, Welsh et al. [6] proposed a general technique for coloring grayscale images by transferring colors between source color images and target grayscale images based on matching brightness and texture information between images. Levin et al. [7] further proposed a colorization algorithm based on color tags for grayscale images, which solves the sparse Markov random field for color propagation. Fields with similar grayscale values and adjacent brightness values have similar colors. However, this method cannot guarantee color space continuity after image coloring.

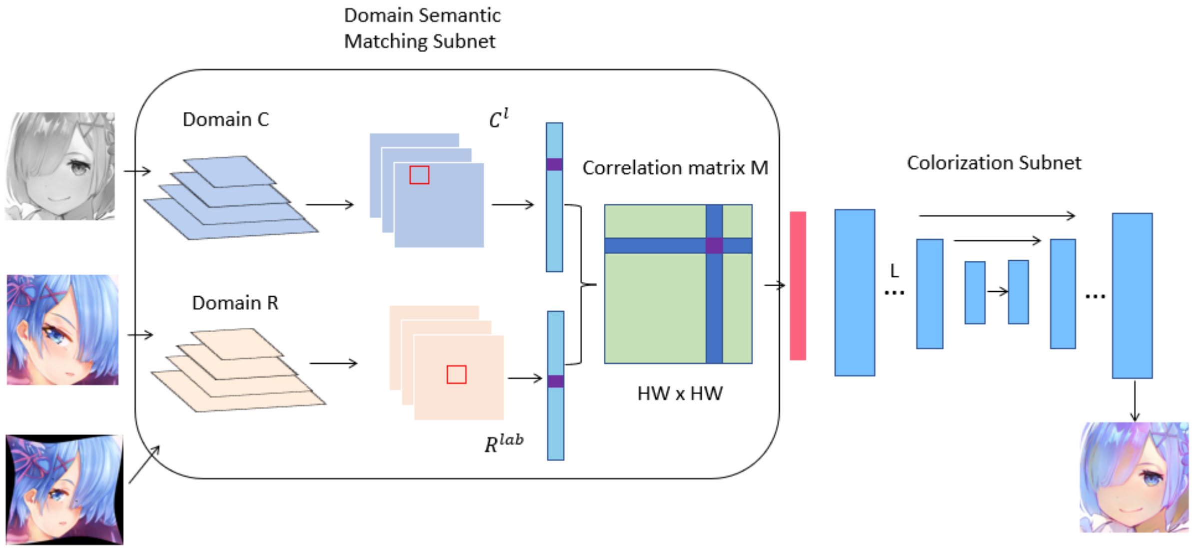


Figure 1. The Overall Structure of The Colorization

Adjacent regions with similar colors may be assigned different colors due to different brightness, resulting in a confusing coloring effect. He et al. [8] proposed an image colorization method that extracts and aligns the features of target and reference images' luminance channels and calculates the bidirectional similarity between the two types of images using discrete cosine distance d. In anime coloring research, Hensman et al. [9] used conditional generative adversarial networks to colorize grayscale anime images, which does not require user interaction to refine coloring results. However, this method only learns the relationship between grayscale and color images, resulting in poor anime coloring effects. In addition, Liao et al. [10] used the PatchMatch algorithm to match high-dimensional features extracted from reference and target images, achieving style transfer between target and reference images. This method can be used for line drawing colorization, but the effect is poor for images with different global line structures, and it changes the texture style of the target image.

3. Methods

3.1 Network structure

We will provide a detailed description of a cross-domain color mapping from exemplar anime image colorization networks. As shown in Figure 1, given a grayscale anime image C and a reference example image R from our dataset, we use the Lab color model for color prediction. First, we convert the RGB color model to the Lab color model to obtain the L channel target grayscale image and a color reference image containing the L channel and ab channels. Simultaneously, we apply thin plate spline (TPS) transformation [11] to generate an enhanced self-reference image $R^{\widetilde{lab}}$. Using these two images C^l and R^{lab}, along with the enhanced data image $R^{\widetilde{lab}}$ as inputs. Sharing the input domain space and the reference domain space as the semantic match domain, the domain matching module calculates the similarity between the target image domain A and the reference image domain B, resulting in a bidirectional mapping function $sim_{C\leftrightarrow R}$. Using the colorization sub-network to output predicted chromaticity values in the Lab color space based on the mapping $sim_{C\leftrightarrow R}$ relationship. $P^{ab} \in R^{H\times W\times 2}$ Finally, obtaining P^{lab} through $C^l = P^l$ to complete the transfer with the reference image color style.

3.2 Domain semantic matching sub-network

The solid box in Figure 1 represents the structure of the entire domain semantic matching sub-network. We use a network structure similar to that of the CoCosnet, and to extract semantic connections, we use a VGG19 pre-trained classification network model for the backbone network for cross-domain feature extraction. Additionally, considering the different layer features between the grayscale image domain and the reference example image domain, we use a pyramid-structured multi-scale input layer network as the feature extractor. The purpose of this design is to provide the cross-domain network with local and global feature information of different scales for fusion, reduce the loss of color information, and greatly improve the overall performance of the network, further learning higher quality feature representations. Then, we use the global matching structure proposed in reference [12] to match the features of the two domains and establish the correspondence between them. The grayscale image domain and the reference example image domain are located in the same shared S domain. Meanwhile, we further match the pairwise similarity measurement between pixel positions in the two domains through the correlation matrix $M \epsilon R^{HW\times HW}$, as shown in the following formula:

$$M(i,j)=\frac{\overline{C^l}(u)^T\,\overline{R^{lab}}(u)}{\left\|\overline{C^l}(u)\right\|_2\left\|\overline{R^{lab}}(u)\right\|_2} \tag{1}$$

Where $\overline{C^l}(u) = C^l(u) - \mu C^l(u)$ and $\overline{R^{lab}}(u) = R^{lab}(u) - \mu R^{lab}(u)$, μ is its mean based on the calculated mean and variance, a distorted sample is obtained. Then, it is combined with thin-plate spline (TPS) transformation to generate an enhanced distorted sample y_B By selecting the most correlated pixel in y_B and taking a weighted average, an approximation of the color sampling in y_B is obtained [13].

$$f_{mat}(u)=\sum_v soft\max(\alpha M(u,v))\cdot y_B(v) \tag{2}$$

We calculate the pairwise similarity of the feature space between the C^l and R^{lab} domains through the domain sharing module and establish the feature matching f_{mat} from image C^l to the R^{lab} domain. The domain semantic matching module generates a coarse chromaticity map by sharing the features of the two domains and establishing color mapping $sim_{C\leftrightarrow R}$ through matching.

3.3 Colorization sub-network

Since these domain semantic matches are not very precise in all regions of the target image, we build a colorization sub-network on top of the domain semantic matching module to correct the color distribution of the coarse chromaticity map produced by the module. The colorization sub-network, shown on the right in Figure 1, adopts a U-shaped encoder-decoder structure [14], where the left part is the encoder consisting of L downsampling modules, each of which consists of a 4x4 convolution layer, a normalization layer, and a LeakyRelu layer. The downsampling convolution module is responsible for extracting the structural and semantic information, as well as the color texture information of the image at different levels. The right part is the decoder consisting of L upsampling modules, each of which corresponds to a 4x4 deconvolution layer, a normalization layer, and a LeakyRelu layer. Each layer of the decoder can predict the color space for calculating the loss function, and the downsampling module is responsible for combining the features extracted by the encoder and restoring the image information and precision through upsampling. Each encoding region is connected to its own decoding region block through skip layers, connecting the features outputted by each downsampling layer to the corresponding upsampling layer. The purpose is to directly pass shallow information to the same level of deconvolution layers to form richer features, improve the details of image color generation, use the brightness channel of the coarse chromaticity map generated by the domain semantic matching module as input, output the predicted color space ab, and finally obtain P^{lab} through $C^l = P^l$ to generate a new image with the style of the reference. The Colorization sub-network can produce a colorization map with overall reasonable colors.

3.4 Loss function

Perceptual loss function

In order to obtain more reasonable perceptual outputs, we use perceptual loss [15] to predict the semantic difference between the output grayscale image domain and the reference example. The feature maps of the VGG19 network pre-trained on ImageNet are used to calculate the identity mapping loss. The definition of the formula is as follows：

$$L_{perc}=\left\|\phi_l(\overline{y})-\phi_l(y)\right\|_2^2 \tag{3}$$

Where specifically refers to a certain convolutional layer in VGG. In the experiments, this paper selects $relu1\_2$, $relu2\_2$, $relu3\_3$ and $relu4\_3$ are input as the perceptual loss computed in this paper.

Smooth L1 loss function

Smooth L1 loss, also known as Huber loss, is a type of loss function commonly used in colorization tasks. It is a combination of the mean absolute error (MAE) loss and the mean squared error (MSE) loss. In this paper, smooth L1 loss is used instead of MSE loss directly in order to obtain a more robust solution that is less sensitive to outliers in the colorization problem. The following formula (Formula 4) can be used to calculate smooth L1 loss.

$$Loss_{reg}=HuberLoss(\overline{Y},Y)+\sum_{i=1}^{H*W}\frac{1}{2}(y_i-\overline{y_i})^2\Big|_{|y_i-\overline{y_i}|\le 1} +(\left|y_i-\overline{y_i}\right|-\frac{1}{2})\Big|_{|y_i-\overline{y_i}|\ge 1} \tag{4}$$

3.5 Evaluation method

Due to the uncertainty of the colorization task, the general mainstream evaluation method uses the peak signal-to-noise ratio and structural similarity metrics in the image restoration task to evaluate the quality of the images generated by the coloring algorithm. peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible signal power and the destructive noise power which affects its accuracy. The maximum signal-to-noise ratio is often expressed in logarithmic decibels. A measurement that measures the quality of an image. PSNR was defined as Mean Square Error (MSE). For the generated anime images and the reference images, If one is an approximation of another's noise, the PSNR between them is defined as：

$$PSNR=10\log_{10}(\frac{MAX^2}{MSE}) \tag{5}$$

structural similarity [16] (SSIM) is often used as an indicator to assess image quality, generally measuring the similarity of two images in terms of contrast, brightness and resulting information. In this paper, structural similarity is used to compare the similarity between the generated anime images and the reference images. and the larger the structural similarity value, the closer the two images are, the better the learning effect. structural similarity can be defined as :

$$SSIM(X,Y) = \frac{(2\mu_X \mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (6)$$

where μ_X is the mean of image X, μ_y is the mean of image Y, σ_X^2 is the covariance of X, σ_Y^2 and similarly is the covariance of Y, σ_{XY} representing the covariance of X and Y.c_1 and c_2 are constants that maintain stability.

4. Experiments

4.1 Evaluation method

All training and testing experiments were conducted on both a server and Google Colab. The server used an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz processor and two NVIDIA GeForce GTX 1080Ti GPUs with a total of 24GB of VRAM. Google Colab allocated 12GB RAM and a GPU for the experiments. The proposed network architecture was implemented using the PyTorch framework and optimized using the Adam optimizer with an initial learning rate of lr =0.0001. The loss functions used were the perceptual loss function and the Smooth L1 loss function mentioned earlier. using waifu2x and then cropped to an appropriate size. We used the grayscale images from this dataset as input grayscale domain images for our network, and established source-reference pairs that were semantically similar to the grayscale domain information of the reference images. However, due to the uniqueness of the coloring in the reference examples, it was not possible to find enough source-reference pairs for training. Therefore, during the network training process, we used a thin-plate spline (TPS) geometric distortion transformation to augment the dataset. We generated reference information from the original images through geometric distortion, simulating semantic similarity to the reference images and providing complete color information for the target images. Finally, we randomly selected 25 images from the test set for evaluation.

4.2 Experimental process

We used the Re:Zero Rem Anime Faces For GAN Training dataset [17] downloaded from the Kaggle website, as well as 100 randomly selected cartoon anime images from the internet, as the test experimental dataset. The Re:Zero dataset contains 725 images of Rem anime faces, which were captured from Pixiv and used to train StyleGAN2. The images were cropped to their original size of 512x512 using a custom-trained YOLOv5 model. Images smaller than 512x512 were enlarged using waifu2x and then cropped to an appropriate size. We used the grayscale images from this dataset as input grayscale domain images for our network, and established source-reference pairs that were semantically similar to the grayscale domain information of the reference images. However, due to the uniqueness of the coloring in the reference examples, it was not possible to find enough source-reference pairs for training. Therefore, during the network training process, we used a thin-plate spline (TPS) geometric distortion transformation as shown in Figure 2 to augment the dataset. We generated reference information from the original images through geometric distortion, simulating semantic similarity to the reference images and providing complete color information for the target images. Finally, we randomly selected 25 images from the test set for evaluation.

a. Origin reference image b. TPS enhancement image

Figure 2. Origin Reference Image and TPS Enhancement Image

4.3 Comparisons with state-of-the-arts

Table 1 shows the evaluation results of PSNR and SSIM for several commonly used example colorization networks. It can be seen from the table that the cross-domain color mapping reference example network model has the highest average PSNR and SSIM scores compared to other colorization networks. Figure 3, the third column, shows the coloring effect of Welsh's traditional machine learning method. It can be clearly seen from the figure that the accuracy of color prediction distribution of the network model is low, the coloring effect is relatively simple, and it cannot achieve diversified colorization. The coloring effect and color prediction are still inferior to commercial anime. The last three methods are all based on deep learning. Although Liao's method achieved multi-color distribution transfer for reference images, it is essentially an image translation method. When there is a large difference in content between the two images, it may lead to inaccurate similarity calculation and poor coloring in some details. Moreover, the texture style of the target image is changed, inheriting some of the texture feature styles of the reference example, and cannot fully retain all the texture information of the target image. He's coloring method calculates the similarity of the correlation matrix between the target image and the reference image and predicts multiple color distributions based on the similarity combined with a coloring subnetwork. Although it fully retains the target image's texture details, it slightly surpasses Liao's method in semantic similarity.

Figure 3. Different Colorization Method Coloring Effect Comparison

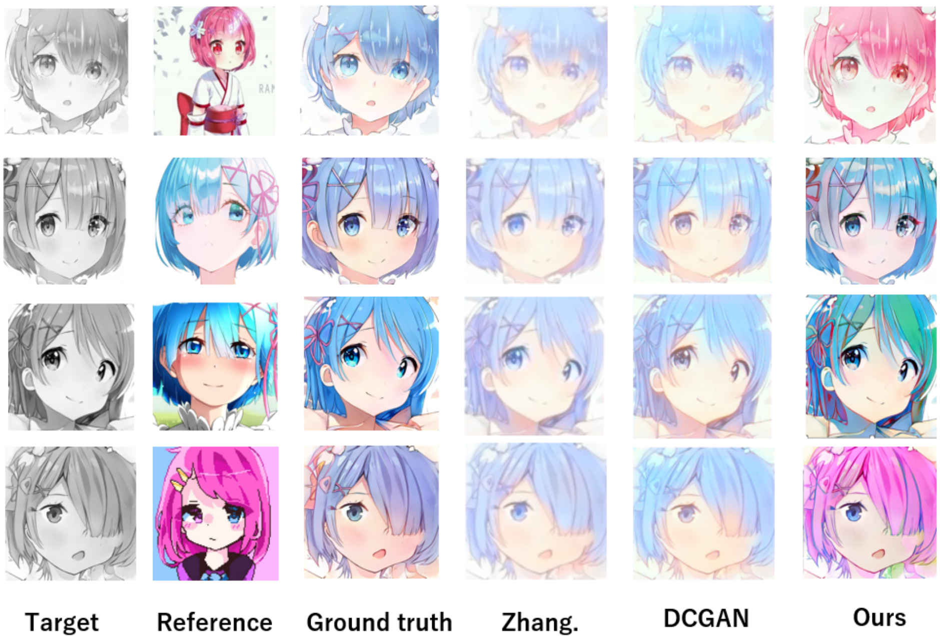


Figure 4. Different Colorization Method Coloring Effect Comparison

However, in the color prediction of the red hair in Figure 5, the first image in the 5th row, there was a misjudgment, resulting in large areas of mixed colors being preserved in the originally classified red hair. Compared with previous reference-based methods, our method achieves the best fidelity of the target image's texture style, transfers the colors of the reference image's corresponding semantic regions through domain matching, and fills the color distribution of unmatched areas between the reference and target images with TPS data augmentation, avoiding the impact of incorrect color distribution information caused by mismatching, further enhancing the coloring result. The colored images are more harmonious, coordinated, and closer to real natural anime images, providing a better visual effect. We also compared the coloring results of other non-reference-based methods and displayed several representative results. All comparison results were obtained using publicly available code. We present qualitative comparison results in Figure 4, which show that Zhang's coloring method [18] inherits the purple hair color of Rem in the dataset, resulting in a coloring effect that is basically consistent with the original image, but with some errors in color distribution prediction.

Table 1. Different Exemplar Colorization Methods PSNR and SSIM On The Dataset

Rem Dataset		
Methods	PSNR	SSIM
Welsh	18.49	0.821
Liao	20.56	0.804
He	22.32	0.916
Our	**23.61**	**0.934**

Although DCGAN's coloring method has been optimized to make the colors in the anime character area more vivid and clear, these methods only predict similar color distributions for the original images in the dataset and cannot transfer colors from reference examples to target images, while our method has a higher degree of color restoration for the reference examples. Furthermore, we also verified the model's ability to colorize based on different specified styles. As shown in Figure 5, we colored grayscale images according to their style using four different reference images. It can be observed from the figure that our coloring method closely aligns with the styles of the reference images.

Figure 5. The Coloring Images Generated Based On Different Reference Style Images

To demonstrate the importance of TPS data augmentation in the domain semantic matching module, we conducted a comparative experiment. The first column shows the distorted images resulting from the reference image and the image after passing through the domain semantic matching module. The second column shows the colorized image results after passing through the coloring module. The left image shows the results without using TPS data augmentation, while the right image shows the results after using TPS data augmentation. By comparing the results in Figure 6, it can be seen that the coloring effect produced by TPS data augmentation does not have mixed colors in Rem's hair, and the coloring effect is better restored to the reference image.

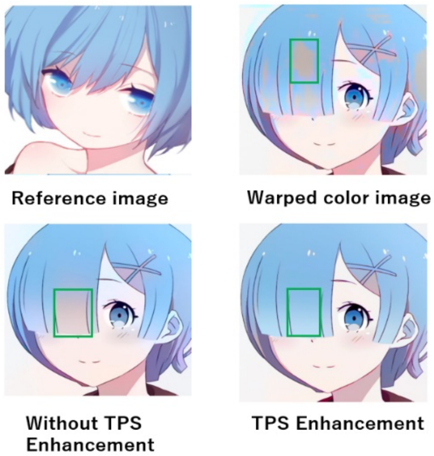


Figure 6. Comparison of The Coloring Effect of TPS Data Enhancement

4. Conclusions

In this paper, we propose a new anime image coloring mode, called reference example anime image coloring mode, using deep learning. We adopt a cross-domain color mapping approach for reference example anime image coloring, where a supervised coloring model is trained to achieve basic consistency in color style between the target image and reference image. Additionally, we incorporate thin-plate spline (TPS) transformation as a data augmentation technique. The coloring network consists of two main structures, domain matching module and coloring module. The domain matching module establishes a relationship between feature maps from two domains, by mapping the feature spaces to a certain color feature space and obtaining a mapping TT between the two domains. The coloring module then transforms the mapping TT from feature domain to image domain. This coloring method can obtain reasonable color transfer results from the reference domain to the target domain, given a similar semantic reference image. We train and test our model on the Rem dataset, and compare it with other reference-based coloring methods using metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) on the validation set. The experimental results show that the generated images are clear, natural, and closer to the original reference anime images. There is still room for improvement in our coloring method, as we have not applied it to anime line coloring. Future work can focus on further research in this area.

References

[1] P. Zhang, B. Zhang, D. Chen, et al, Cross-domain correspondence learning for exemplar-based image translation Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.5143-5153, 2020.

[2] S. LIZUKA, E. SIMOSERRA, 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 Transactions on Graphics (ToG), Vol. 35, No.3, pp.1-11, 2016.

[3] G. Larsson, M. Maire, G. Shakhnarovich, Learning representations for automatic colorization European conference on computer vision. Springer, Cham, pp.577-593, 2016.

[4] K. Nazeri, E. Ng, and M. Ebrahimi, Image colorization using generative adversarial networks, International conference on articulated motion and deformable objects, 2018.

[5] A. Deshpande, J. Lu, M. C. Yeh, et al, Learning diverse image colorization Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.6837-6845, 2017.

[6] T. Welsh, M. Ashikhmin, K. Mueller, Transferring color to greyscale images, in ACM Transactions on Graphics (TOG), Vol. 21, pp. 277–280, 2002.

[7] A. Levin, D. Lischinski, and Y. Weiss, Colorization using optim ization, in ACM transactions on graphics (TOG), Vol. 23, pp. 689–694, 2004.

[8] M. He, D. Chen, J. Liao, P. V. Sander, & L. Yuan, Deep exemplar-based colorization, ACM Transactions on Graphics (TOG), Vol. 37, No.4, pp.1-16, 2018.

[9] P. Hensman, K. Aizawa, cGAN-based manga colorization using a single training image, 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, pp.72-77, 2017.

[10] J. Liao, Y. Yao, L. Yuan, G. Hua, & S. B. Kang, Visual attribute transfer through deep image analogy, arXiv preprint arXiv:1705.01088, 2017.

[11] J. Lee, E. Kim, Y. Lee, D. Kim, J. Chang, J. Choo, Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence, In: CVPR 2020. IEEE Computer Society, pp. 5800–5809, 2020.

[12] X. Wang, R. Girshick, A. Gupta, & K. He, Non-local neural networks, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7794-7803, 2018.

[13] B. Zhang, M. He, J. Liao, et al, Deep exemplar-based video colorization, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8052-8061, 2019.

[14] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation International Conference on Medical image computing and computer-assisted intervention, pp.234-241, 2015.

[15] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, pp.694-711, 2016.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, et al, Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing, Vol 13(4), pp. 600-612, 2004.

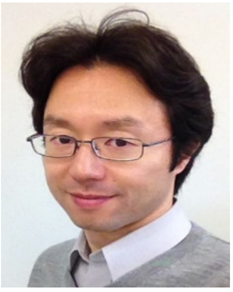
[17]Re:Zero Rem Anime Faces For GAN Training. https://www.kaggle.com/datasets/andy8744/rezero-rem-anime-faces-for-gan-training, 2023.

[18] R. Zhang, P. Isola, A. A. Efros, Colorful image colorization, European conference on computer vision, pp.649-666, 2016.

Jincheng Peng received the ME degree from Chengdu University Of Technology in 2020. From 2021, he has been staying at PhD. course in Systems Science and Technology at Akita Prefectural University Graduate School. His research interests include medical image segmentation, medical image processing, and deep learning.

Guoyue Chen received his ME and PhD degrees from Tohoku University, Japan in 1993 and 1996, respectively. Then, he was an assistant professor in Tohoku University, Japan. He is currently a professor in Akita Prefecture University, Japan. His research interests include digital signal processing and its applications to active noise control, image processing and medical image processing.

Kazuki Saruta received his BE and ME degrees from Akita University, Japan, in 1991 and 1993, and his PhD degree from Tohoku University, Japan, in 1996. Then, he was a lecturer in Yamagata University. He is currently a professor in Akita Prefecture University, Japan. He has been engaged in research on pattern recognition and image processing.



Yuki Terata received his BE degree from Akita University, Japan, in 1998, and the ME and PhD degrees from the same university in 2000 and 2008. He is an associate professor in Akita Prefecture University, Japan. He is interested in VR/AR for cognitive psychology of elderly people.