Unsupervised Image-to-Image Translation from MRI-based Simulated Images to Realistic Images Reflecting Specific Color Characteristics

Naoya Wada* and Masaya Kobayashi*
KYOCERA Corporation, 3-7-1 Minatomirai Nishi-ku, Yokohama, Japan

Keywords: Generative Adversarial Networks (Gans), Image-to-Image Translation, Domain Adaptation, Unsupervised Learning, MRI.

Abstract: In this paper, a new domain adaptation technique is presented for image-to-image translation into the real-world color domain. Although CycleGAN has become a standard technique for image translation without pairing images to train the network, it is not able to adapt the domain of the generated image to small domains such as color and illumination. Other techniques require large datasets for training. In our technique, two source images are introduced: one for image translation and another for color adaptation. Color adaptation is realized by introducing color histograms to the two generators in CycleGAN and estimating losses for color. Experiments using simulated images based on the OsteoArthritis Initiative MRI dataset show promising results in terms of color difference and image comparisons.

1 INTRODUCTION

Image synthesis can now achieve sufficient quality for practical use thanks to the progress of generative adversarial networks (GANs) (Goodfellow et al., 2014). GANs have also been used in various tasks, including image-to-image translation (Zhu et al., 2017; Isola et al., 2017), image interpolation (Yu et al., 2018), and data synthesis other than images (Yu et al., 2017). While GANs now have a wide range of applications, image synthesis to realize specific characteristics is one of the main tasks for practical use, and the demand for image synthesis has grown as the quality of the generated images has increased. One example is the task of generating images that reflect the specific color or other characteristics of a person.

This paper presents a new image-to-image translation method that reflects specific color characteristics, with the aim of generating a realistic image reflecting a person's characterics from simulated images based on MRI data. In conventional image translation methods, there are three main considerations. The first is how to capture specific characteristics of a domain (Karras et al., 2019). The second is the necessary amount of training data (Liu et al., 2019). The third is the labeling cost for supervised learning, which is required for several methods that involve pairing images across the domains (Isola et al., 2017). To address these issues, we propose a CycleGAN-based network model to achieve unsupervised learning that requires a smaller amount of data. However, the original CycleGAN architecture cannot accept the characteristics of a specific target as input and is unable to adapt the generated image accordingly. Therefore, we import the color characteristics as color histograms to the middle layer between the encoder–decoder structures of the two generators in the cyclic architecture of CycleGAN. Then, color loss is independently estimated and combined with other losses such as cycle consistency loss and adversarial loss. We also performed experiments to demonstrate image-to-image translation across two domains. One domain consisted of 3D simulated knee images obtained from MRI data provided by the OsteoArthritis Initiative (OAI). The other domain consisted of real-world knee images. Image-to-image translation from MRI to realistic knee images would make it easier for non-healthcare professionals to understand MRI images.

* https://www.kyocera.co.jp

Wada, N. and Kobayashi, M.
Unsupervised Image-to-Image Translation from MRI-based Simulated Images to Realistic Images Reflecting Specific Color Characteristics.
DOI: 10.5220/0010916300003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 2: BIOIMAGING*, pages 184-189
ISBN: 978-989-758-552-4; ISSN: 2184-4305

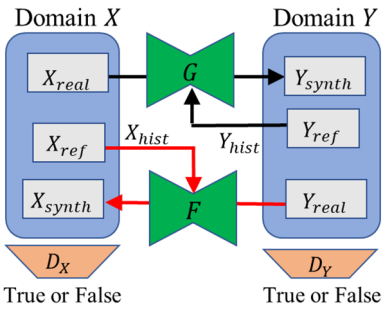

Figure 1: Schematic of the image-to-image translation network.

It could also support communication between patients and healthcare professionals by aiding in explanations of MRI images from other patients.

In the remainder of this paper, we discuss related work in Section 2, introduce our proposed method in Section 3, present our experimental methods and results in Section 4, and give our conclusions in Section 5.

2 RELATED WORK

GANs have been used previously for image-to-image domain adaptation. Pix2pix (Isola et al., 2017) is an early, well-known method for image-to-image translation that uses conditional GANs to learn paired images. However, Pix2pix uses supervised learning and a large number of paired images, so some labeling cost is required. CycleGAN (Zhu et al., 2017) is a method for image-to-image translation without learning paired images. The model has a cycle architecture that consists of two generators and two discriminators and cycle consistency loss for unsupervised learning. CycleGAN requires a relatively small number of unpaired images for training. However, it is difficult to adapt the generated image to specific characteristics (color, illumination, etc.) in a dataset. Recently, StarGAN-v2 (Choi et al., 2018; Choi et al., 2020) has been proposed for image-to-image translation across domains and can reflect styles by obtaining style codes from datasets automatically. However, it requires a large dataset for training, and it is difficult to specify a specific style in a non-biased dataset because style codes are automatically obtained by capturing bias in a training dataset. DeepHist (Avi-Aharon et al., 2020) can reflect specific color characteristics in generated images by using a differentiable network with kernel density estimation of color histograms from the generated image. However, paired images are required for training.

3 PROPOSED METHOD

Here we propose a network model for image-to-image translation that is based on the cyclic architecture of CycleGAN and is able to reflect target color characteristics. For training, our method requires only a relatively small amount of data and does not require pairing images. Furthermore, we introduced an architecture to accept color histograms for the target domain. This section describes the details of this architecture and the estimation of losses during training.

3.1 Overview of the Network

The network consists of two generators and two discriminators as shown in Figure 1. X and Y denote the domains in image-to-image translation. G and F denote the generator from X to Y and Y to X respectively. Each of them generates a synthetic image (X_{synth} and Y_{synth}) in the target domain from a real image (X_{real} and Y_{real}) in the source domain. D_X and D_Y denote the discriminators and X_{hist} and Y_{hist} denote the color histograms for X and Y, respectively. These histograms are input into F and G, respectively, so that the generated synthetic images reflect the color characteristics obtained from reference images (X_{ref} and Y_{ref}) in the respective target domains. Spectral normalization (Miyato et al., 2018) is adopted for both generators and discriminators to stabilize training of this network.

3.2 Importing Color Characteristics

The architecture adopted for G and F is shown in Figure 2. The color distribution is imported with reference to previous methods. First, an RGB histogram is obtained from an image in the target domain. Histograms for each color are concatenated and imported to the middle layer between the encoder and the decoder of the generator. The purpose of importing the histograms is to import color information after spatial features have been convoluted. A translated image is output from the decoder. To evaluate the color of the output image, L2 loss between histograms of the source and output images is obtained. The histograms of the output image are obtained by kernel density estimation because it enables backpropagation and updating of the network. Kernel density estimation is done using the following probability density function:

$$f_I(g) = \frac{1}{NB}\sum_{x\in\Omega} \mathrm{K}\left(\frac{I(x)-g}{B}\right), \quad (1)$$

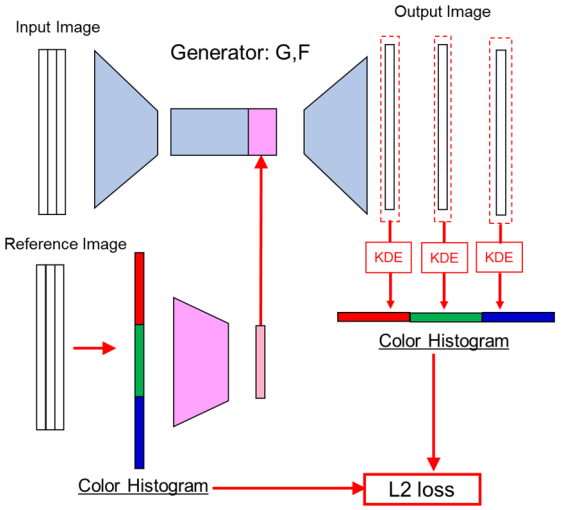

Figure 2: Architecture of the generators (KDE: kernel density estimation).

Here, x denotes the pixel position, $g \in [-1, 1]$ denotes the pixel value, B denotes the bandwidth, $N = |\Omega|$ denotes the number of pixels, $I(x) \in [-1, 1]$ denotes the luminance, and $\mathrm{K}(\cdot)$ denotes the kernel function, which is defined as

$$\mathrm{K}(z) = \frac{d}{dz}\sigma(z) = \sigma(z)\sigma(-z). \quad (2)$$

Here, $K(z)$ is the derivative of the sigmoid function $\sigma(z)$. Using $f_I(g)$, each pixel in an image is assigned to a histogram bin according to

$$P_I(k) = \int_{\mu_k - \frac{L}{2}}^{\mu_k + \frac{L}{2}} f_I(g)dg\,, \quad (3)$$

where $L = \frac{2}{K}$ denotes the bin width and $\mu_k = -1 + L\left(k + \frac{1}{2}\right)$ denotes the center of the k-th bin when normalized pixel values $[-1, 1]$ are divided into K bins $\{B_k\}_0^{K-1}$. Equation (3) can be developed by using (2) to give

$$P_I(k) = \frac{1}{N}\sum_{x \in \Omega} \Pi_k(I(x)) \quad (4)$$

where $\Pi_k(z)$ is defined as

$$\Pi_k(z) \triangleq \sigma\left(\frac{I(x) - \mu_k + \frac{L}{2}}{B}\right) - \sigma\left(\frac{I(x) - \mu_k - \frac{L}{2}}{B}\right). \quad (5)$$

Thus, $P_I(k)$ gives the value of the k-th bin of the color histograms.

3.3 Loss Function

The loss function for the entire network is defined as

$$\mathcal{L} = \lambda_{\mathrm{GAN}}\mathcal{L}_{\mathrm{GAN}} + \lambda_{\mathrm{CYC}}\mathcal{L}_{\mathrm{CYC}} + \lambda_{\mathrm{IDT}}\mathcal{L}_{\mathrm{IDT}} + \lambda_{\mathrm{HIST}}\mathcal{L}_{\mathrm{HIST}}, \quad (6)$$

where $\mathcal{L}_{\mathrm{GAN}}$ denotes the adversarial loss, $\mathcal{L}_{\mathrm{CYC}}$ denotes the cycle consistency loss, $\mathcal{L}_{\mathrm{IDT}}$ denotes the identity mapping loss, and $\mathcal{L}_{\mathrm{HIST}}$ denotes the color loss. $\mathcal{L}_{\mathrm{HIST}}$ is defined by using L2 losses as

$$\mathcal{L}_{\mathrm{HIST}} = \left|\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{R}} - \boldsymbol{h}_{\mathrm{REF}}^{\mathrm{R}}\right|_2 + \left|\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{G}} - \boldsymbol{h}_{\mathrm{REF}}^{\mathrm{G}}\right|_2 + \left|\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{B}} - \boldsymbol{h}_{\mathrm{REF}}^{\mathrm{B}}\right|_2, \quad (7)$$

where $\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{R}}$, $\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{G}}$, and $\boldsymbol{h}_{\mathrm{OUT}}^{\mathrm{B}}$ denote the histograms of each RBG color estimated from the generated image and $\boldsymbol{h}_{\mathrm{REF}}^{\mathrm{R}}$, $\boldsymbol{h}_{\mathrm{REF}}^{\mathrm{G}}$ and $\boldsymbol{h}_{\mathrm{REF}}^{\mathrm{B}}$ denote the histograms of each color estimated from the imported reference images.

4 EXPERIMENTS

Image-to-image translation experiments were performed to evaluate the proposed method. The task is to translate images across two domains.

4.1 Dataset

One domain consisted of 3D simulated knee images generated from OAI MRI data. MRI data were converted to 3D structure and a front-view image was obtained as a 3D simulated knee image by using the 3D medical imaging software InVesalius 3.1. The other domain consisted of real-world knee images (the color source) obtained from several datasets including the Fashion Product Images Dataset provided by Kaggle. The resolution of both input and output images was 256 × 256 pixels. The training dataset consists of 477 3D simulated images and 438 real-world images because we sought to consider the situation with a relatively small training dataset. For learning, the network was trained for 400 epochs. The test dataset consists of 28 images from each domain, and image translation was performed from the MRI-based 3D simulation domain to the real-world domain to reflect the color characteristics of an input real-world knee image. As a result, 784 images were generated for each input combination comprising a 3D simulated image and a real-world image (28 × 28 images) in round-robin manner.

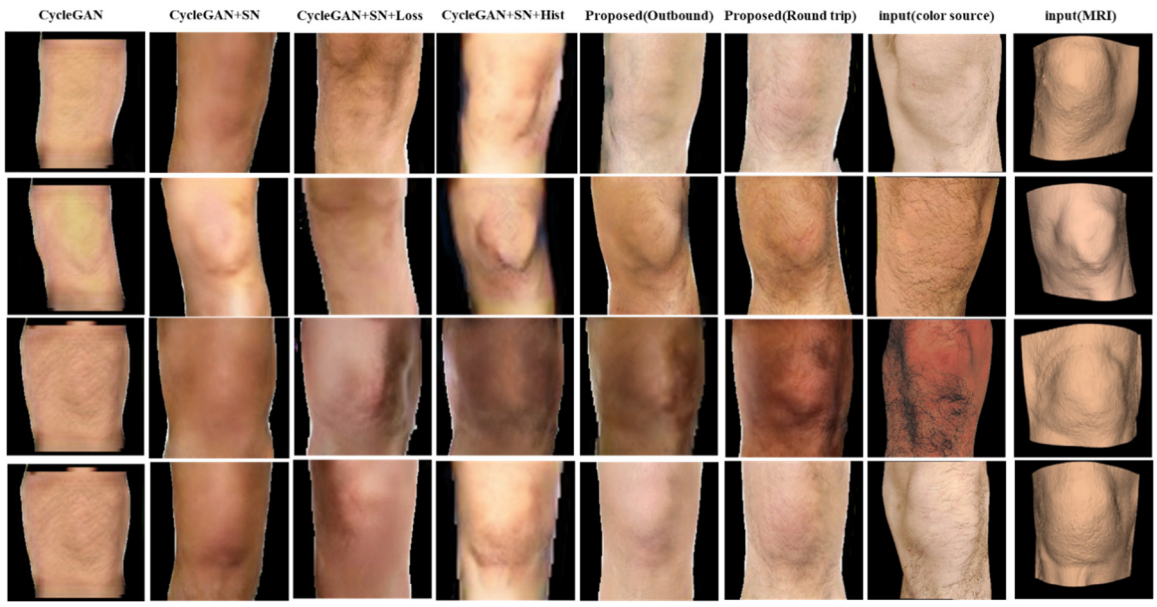

Figure 3: Comparison of images generated by the six methods shown in Table 1.

4.2 Evaluation

To evaluate how well the color characteristics in the generated output images reflected those in the input source images, their mean colors were compared. The color difference ΔE^*_{ab} was calculated using CIE76 to compare the mean colors in the L*a*b* color space:

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}. \quad (8)$$

We compared the color difference among conventional CycleGAN, the proposed method, and some incomplete methods based on the proposed method but with the omission of some of its techniques. These incomplete methods are intended to evaluate the effect of each technique. The evaluated methods are summarized in Table 1.

4.3 Results

Figure 3 shows representative examples of images generated by the evaluated methods for five combinations of source MRI-simulation and real-world images, and their color differences are summarized in Table 1. Note that Methods 1-3 do not have an architecture to accept input color histograms; thus, unspecified color characteristics extracted from whole dataset during training are reflected in the generated images and cause the differences from the color source images. The images generated by methods 1 and 2 all have the same color characteristics. The result for method 3 shows that it is not very effective on its own. On the other hand, methods 4-6 which accept input color histograms improve the color differences. The effect of using color histograms as input and evaluating losses for color can be clearly seen. Especially, the result for method 6 achieving $\Delta E^*_{ab} < 5.0$ is promising because this threshold is same as the color tolerance of printed solids defined in ISO 12647-2. The improvement in

Table 1: Summary of methods and color differences.

No.	Method	Description	Color Diff.
1	CycleGAN	Original CycleGAN	14.1
2	CycleGAN +SN	Spectral normalization is applied to method 1.	13.7
3	CycleGAN +SN+Loss	Loss function considering colors (6) is applied to method 2.	13.0
4	CycleGAN +SN+Hist	Color histograms are applied to method 2.	8.9
5	Proposed (outbound only)	Loss function (6) and color histograms are applied to generator G (MRI to real-world) only.	5.0
6	Proposed (round trip)	Proposed method (the method applied to G in method 5 is also applied to F)	4.6

color difference between methods 5 and 6 seems not large. However, the images generated by method 6 show clear differences in brightness within each image, whereas the images generated by method 5 are slightly blurry. More examples of images generated by method 6 are shown in Figure 4, which gives an overview of the relationships between the source 3D simulated images and the source colors. We can see that the generated images change according to the source model and color source.

4.4 Discussion

In this section, we discuss the importance on incorporating color histograms and the loss function considering color loss into CycleGAN. Method 5 incorporated loss function (6) and color histograms into only generator G (for translating MRI-simulation images into realistic images), where the aim is to reflect the color characteristics of an input image. On the other hand, method 6 incorporated them into not only G but also F, where the color characteristics of the MRI-simulation images are reflected. The surfaces of MRI-simulation images are colorized by the medical imaging software with a certain fixed color and brightness as shown in Figure 3 because the source MRI data does not include the surface colors of the scanned individual. Therefore, color histograms input to the generator F mainly reflect the contrast of brightness rather than the color of the person's skin. The detailed characteristics of the images generated by method 6 are attributable to the architecture of the proposed method. In other words, incorporating color histograms into CycleGAN was effective not only for reflecting color characteristics of the source image, but also for better representing contrasts, unevenness, and other characteristics.

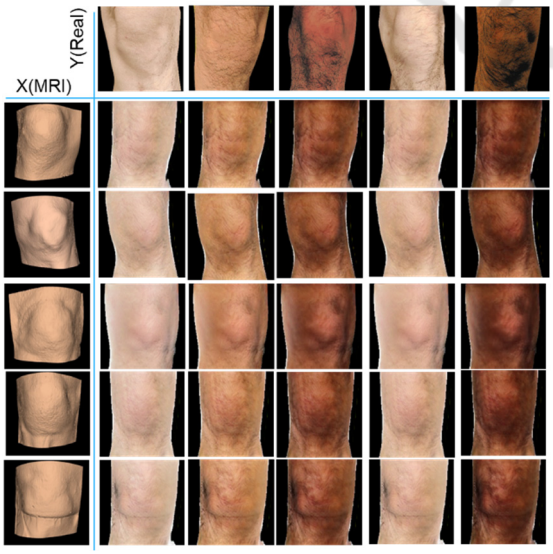

Figure 4: Examples of generated images for 25 combinations of source 3D simulated images and source real-world images. The first column and first row show the respective source images.

5 SUMMARY

In this paper, a new technique for image-to-image translation reflecting specific color characteristics was presented. That technique was intended to translate MRI-based 3D simulated images to realistic images reflecting the characteristics of a specific person's appearance. In our image-to-image translation network, which was based on CycleGAN, color histograms of an input image were concatenated to the input vector as reference color characteristics and the generated image was evaluated with a loss function considering color loss. The experimental results showed that the presented technique was effective not only for reflecting the color characteristics of the input source image but also for better representing contrasts, unevenness, and other characteristics.

Topics for future work include realizing more useful image-to-image translation with representation of various lighting environments. In the 3D simulated image domain, it is easy to obtain images under many different lighting conditions. If this can be reflected in the generated image along with color characteristics, then image translation to the real-world image domain would be more flexible and useful.

ACKNOWLEDGEMENTS

Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive (NDA). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide resource to quicken the pace of biomarker identification, scientific investigation and OA drug development. Dataset identifier: 2343.

REFERENCES

Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014). Generative Adversarial Nets. In *Neural Information Processing Systems*.

Zhu J., Park T., Isola P., Efros A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *International Conference on Computer Vision (ICCV)*.

Isola P., Zhu J., Zhou T., Efros A. (2017). Image-to-image Translation with Conditional Adversarial Networks. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yu J., Lin Z., Yang J., Shen X., Lu X., Huang S. T. (2018). Generative Image Inpainting with Contextual Attention. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yu, L.; Zhang, W.; Wang, J.; and Yu, Y. (2017). SeqGAN: Sequence generative adversarial nets with policy gradient. In *Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence*.

Karras T., Laine S., Aila T. (2019) A Style-based Generator Architecture for Generative Adversarial Networks. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liu M., Huang X., Mallya A., Karras T., Aila T., Lehtinen J., Kautz J. (2019) Few-Shot Unsupervised Image-to-Image Translation. In *International Conference on Computer Vision (ICCV)*.

Choi Y., Choi M., Kim M., Ha J., Kim S., Choo J. (2018) Stargan: Unified Generative Adversarial Networks for Multi-domain Image-to-image Translation. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Choi Y., Uh Y., Yoo J., Ha J. (2020) Stargan v2: Diverse Image Synthesis for Multiple Domains. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8188–8197.

Avi-Aharon M., Arbelle A., Raviv R. T. (2020). Deephist: Differentiable Joint and Color Histogram Layers for Image-to-image Translation. In *arXiv* preprint arXiv:2005.03995.

Miyato T., Kataoka T., Koyama M., Yoshida Y. (2018) Spectral Normalization for Generative Adversarial Networks. In *International Conference on Learning Representations (ICLR)*.