Journal of Imaging Science and Technology® 67(5): 1–6, 2023.

Acquisition of Color Reproduction Technique based on Deep Learning Using a Database of Color-converted Images in the Printing Industry

Ikumi Hirose[†] **and Ryosuke Yabe**[†]
Graduate School of Science and Engineering, Chiba University, Chiba, Japan
E-mail: ry0419@chiba-u.jp

Toshiyuki Inoue
Sanko Corporation, Tokyo, Japan

Koushi Hashimoto
Nikko Process Corporation, Tokyo, Japan

Yoshikatsu Arizono
Sanko Corporation, Tokyo, Japan

Kazunori Harada
Nikko Process Corporation, Tokyo, Japan

Vinh-Tiep Nguyen, Thanh Duc Ngo, and Duy-Dinh Le
University of Information Technology, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Norimichi Tsumura[▲]
Graduate School of Engineering, Chiba University, Chiba, Japan

Abstract. *Color-space conversion technology is important to output accurate colors on different devices. In particular, CMYK (Cyan, Magenta, Yellow and Key plate) used by printers has a limited range of representable colors compared with RGB (Red, Green and Blue) used for normal images. This leads to the problem of loss of color information when printing. When an RGB image captured by a camera is printed as is, colors outside the CMYK gamut are degraded, and colors that differ significantly from the actual image may be output. Therefore, printers and other companies manually correct color tones before printing. This process is based on empirical know-how and human sensitivity and has not yet been automated by machines. Therefore, this study aims to automate color correction in color-space conversion from RGB to CMYK. Specifically, we use machine learning, utilising a large color-conversion database owned by printing companies, which has been cultivated through past correction work, to learn the color-correction techniques of skilled workers. This reduces the burden on the part of the work that has been done manually, and leads to increased efficiency. In addition, the machine can compensate for some of the empirical know-how, which is expected to simplify the transfer of skills. Quantitative and qualitative evaluation results show the effectiveness of the proposed method for automatic color correction.* © *2023 Society for Imaging Science and Technology.*

[DOI: 10.2352/J.ImagingSci.Technol.2023.67.5.050402]

1. INTRODUCTION

Color management is an indispensable technology for working with color, as it ensures that colors are maintained as much as possible when images such as photographs are represented on different devices, such as cameras, displays, and printed matter. Profile connection space (PCS) is defined by the CIELAB and CIEXYZ color systems, which are defined by the International Commission on Illumination (CIE). The L*a*b* and XYZ color spaces are called device-independent colors because they reproduce the same color no matter what device is used to output the image as long as the values are identical. On the other hand, RGB and CMYK, which are generally used for color reproduction, are called device-dependent colors because different colors are reproduced on different output devices even if the numerical values are the same [1].

Figure 1 shows the workflow for creating an image for printing [2]. In the printing industry, the first step in printing is photographing and development by a photographer,

[▲] IS&T Member.

[†] The first and second authors were equally contributed to this paper.

Received May 30, 2023; accepted for publication Aug. 5, 2023; published online Sept. 13, 2023. Associate Editor: Marius Pedersen.

1062-3701/2023/67(5)/050402/6/$25.00

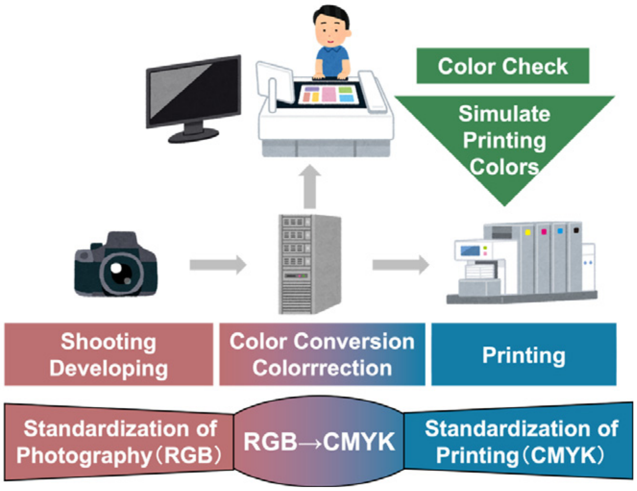

Color Check
Simulate Printing Colors
Shooting Developing
Color Conversion Colorrrection
Printing
Standardization of Photography (RGB)
RGB→CMYK
Standardization of Printing (CMYK)

Figure 1. Printable image creation workflow.

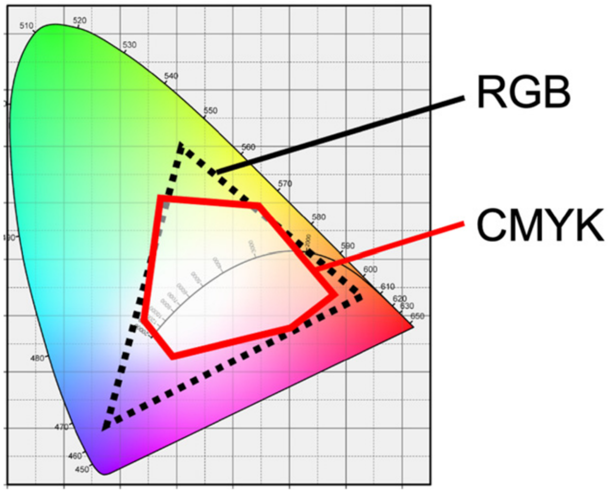


Figure 2. CIE chromaticity diagram.

followed by color conversion and color-tone correction, plate making, and printing, in that order. To construct an optimal workflow, it is necessary to standardize the upstream photography process and the downstream printing process, respectively. The standardization of imaging and printing can be said to be the standardization of colors in RGB and CMYK. However, as shown in Figure 2, the color-reproduction areas of RGB and CMYK are different, making it difficult to standardize color conversion. Furthermore, in such a process, customers often judge the quality of the color based on sensory criteria, and to meet their sensory demands, the color tone must be corrected many times in the plate-making process. Since colors are managed subjectively rather than by objective numerical values, the reproducibility and stability of colors are not maintained.

To promote printing standardization, some printing companies have introduced the Japan Color reference color and certification standards, which can quantitatively express colors, as color management standards [3]. This enables accurate evaluation of printed materials and numerical management that is independent of differences in the skills and experience of printing operators. In some cases, when color conversion is performed, the color is not accurately reproduced, but is built in so that it is the optimal color when printed. For example, skin tones are corrected to be brighter and less muddy than the actual colorimetric values by increasing the brightness and saturation, and the color of cherry blossoms is corrected to be closer to pink rather than white as the human eye perceives it. This is because the color that a person remembers in an image is called a memory color (impression color), which may differ from the actual color captured by a camera or other means. Preferred color reproduction involves human memory, and since humans tend to store colors in their brains as more vivid than they are, there are differences between the remembered and recorded colors [4]. However, the mechanism of desirable color reproduction involves higher-order information processing in the brain, which has not yet been elucidated. In addition to the enormous amount of time required for this kind of correction work, which requires capturing the characteristics of each image one by one, it is also a task filled with empirical know-how, making it difficult to pass on the skills to the next generation.

In this study, therefore, we propose color-correction skills in the printing industry through deep learning. We built an automatic color-correction model by learning from a large database of color-converted images that has been developed by printing companies.

The remainder of this paper is structured as follows, Section 2 explains related works, which are similar to generative adversarial network (GAN). Section 3 explains our approach, which consists of dataset construction, and how to evaluate learning models. The results of the quantitative and qualitative evaluation are discussed in Section 4. Section 5 presents the conclusions and future works.

2. RELATED WORK

In this section, we introduce some research about GANs [5], one of which is used in the proposed approach. Research on GANs has been active, and there are many derivatives of the original GAN. Among them, those related to the original research on which this method is based are mentioned below.

Conditional GAN (CGAN), a network proposed by Mirza et al. in 2014 [6] that follows the basic structure of GANs using generators and discriminators but is extended to perform learning conditionally by providing additional information. The basic structure of GANs using generators and discriminators is retained. While ordinary GANs could only use noise vectors as input to the generator, CGAN has been improved to allow the input of condition data corresponding to the condition vectors to the discriminator.

Pix2pix is an image transformation method based on CGAN, proposed by Isola et al. as an image-generation algorithm [7]. Therefore, it is not possible to control the generated data by specific conditions such as categories.

Pix2pix is a type of CGAN in which the generator takes the input image x as input and generates an image that is close to the correct image y. The discriminator, on the other hand, identifies whether the input is a pair of the generator image $G(x)$ and the input image x or a pair of the correct image y and the input image x.

The generator employs a U-Net-based network [8], which consists of encoder path and decoder paths. In addition, skip connections are introduced between each of them to transfer feature maps at each layer in the encoder path to the same depth layer in the decoder path. The feature map at each layer in the encoder path is transmitted to a layer of the same depth in the decoder path. This enables learning without losing detailed information in the image. The encoder passes extract local features of the input image while down-sampling by convolution, and the decoder pass reconstructs the image based on the feature map while up-sampling by inverse convolution.

PatchGAN architecture is used as the discriminator [9]. PatchGAN divides the entire image into patches of fixed size and determines the truth or falsity of each patch. Finally, the output of all blocks is averaged, and the average value is used as the final output. This provides validity for detailed parts of the image. The network is a fully convolutional network, which consists of only convolutional layers, where the input is the entire image and the output is a feature map consisting of the results of the patch judgments. Pix2pix is an effective network for image transformation and has been studied for various images. It can be directly applied to real image transformation from labeling information, colorization of black-and-white images, land use analysis, and detection of geospatial elements.

There are some approaches for image-to-image translation, the most famous being pix2pix, cycleGAN [10] and adversarial inverse graphics networks [11]. The image sizes in adversarial inverse graphics networks are 128 × 128, while the sizes of the other two are 256 × 256. In addition, pix2pix is better at directly converting images [12]. These are the reasons we used pix2pix in this approach.

Not only pix2pix, but a lot of studies have also contributed to image transformation. Luan et al. proposed a method for style transfer of images and allowed input images to be transferred to the same style as the target images [13]. Also, Huang and Belongie used adaptive instance normalization to realize style transfer in real time, which enables more flexible style transfer by linking the feature map of input images to style images [14].

3. PROPOSED METHOD

3.1 *Dataset*

We created a dataset that we used for learning, shown in Figure 3. The model pix2pix, which is the basis for learning, requires pairs of condition images and images as a dataset, and learns the correspondence between them. In this study, RGB images captured by a camera were used as the condition images. The RGB image was converted to CMYK, and the image that had been further corrected manually by an expert was taken as the correct image. These images were paired to create a dataset for learning.

Because printing companies have large databases of converted images that they have developed over the years, we used images provided by the printing companies rather than having experts perform new corrections for the dataset. This enables the construction of a model that acquires the expertise of experts from printing companies through deep learning.

Out of these large databases of converted images, 7379 pairs of images randomly selected by the experts were used as the big dataset. The big dataset contained images of various categories, such as people, landscapes, and product shots, but the image categories and colors in the images were set so that there was no bias. On the other hand, a small dataset for each subject was also created by extracting only images of the same subject from the big dataset, and the accuracy of each dataset was compared. For this big dataset, we manually classified four types of subjects—plants, landscapes, people, and objects—and used them as small datasets for training. All images in the datasets were resized to 256 × 256 pixels and used for training.

3.2 *Network Architecture*

The learning model in this study was built based on pix2pix, which is widely used for image-to-image conversion problems [7]. The architecture of the proposed method is shown in Figure 4. The main flow of training consists of the following three steps.

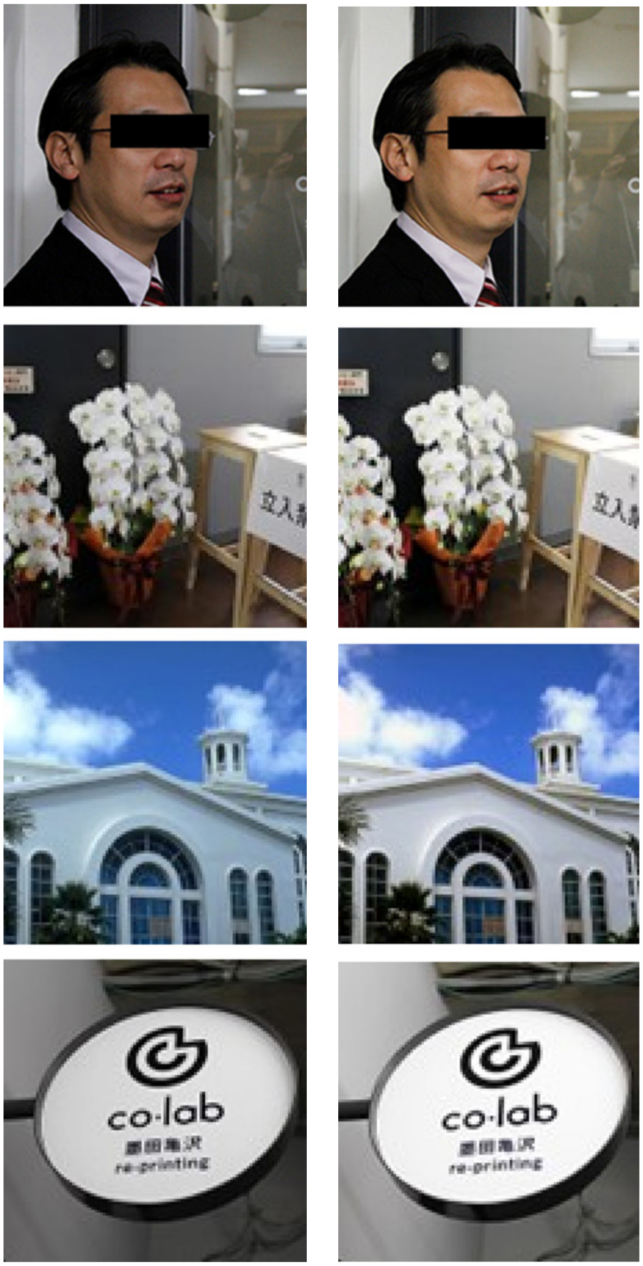
(1) Using the generator, the original image x in the datasets is transformed so that it is close to the correct image, ground-truth, which is the target of the transformation.

(2) Discriminate between the image $G(x)$ generated by the generator and the correct image y using the discriminator.

(3) The loss function is calculated from the image identification results by the discriminator, and the weights of the generator and the discriminator are updated and optimized accordingly.

As in the conventional pix2pix, U-Net was used for the generator and PatchGAN was used for the discriminator.

In this study, learning was performed using three loss functions. In addition to the L1 loss and the adversarial loss used in the conventional pix2pix method, we used the newly introduced structural similarity (SSIM) loss. SSIM is a measure of image similarity that is widely used to evaluate image-processing algorithms and as a loss function in many image-processing applications [15]. The image transformation in this study involved human perceptual color reproduction, and we believe that the introduction of SSIM loss, in addition to L1 loss expressed as a simple absolute error of pixel values, may improve accuracy over conventional pix2pix. We define conventional pix2pix as the conventional method and pix2pix with SSIM loss as the proposed method. Equation (1) is SSIM loss, where a is the



(a) Original

(b) Ground-truth

Figure 3. Examples from big dataset.

weight.

$$\mathcal{L}_{\mathrm{ssim}}(G) = \mathbb{E}[\log(1 + (a(1 - SSIM))^2)]. \qquad (1)$$

3.3 *Learning Parameters*

We set the learning rate to 0.001 for both the generator and discriminator, referring to the conventional pix2pix; the hyperparameters of each loss function to 0.1 for the L1 and adversarial losses, and to 1.0 for the SSIM loss. The number of epochs was set to 200 when the loss was sufficiently lowered by the learning curve.

The number of images in each dataset is shown in Table I. All image sizes were 256 × 256 pixels.

Table I. The number of images in each dataset.

	All	Flower	Scene	Human	Object
Train	6641	171	1056	612	544
Validate	738	21	119	72	54
Test	1000	190	87	71	30

Table II. Classification performance of photo → labels for different methods on Cityscapes [10].

	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [16]	0.45	0.11	0.08
BiGAN/ALI [15, 17]	0.41	0.13	0.07
SimGAN [18]	0.47	0.11	0.07
CycleGAN [10]	0.58	0.22	0.16
Pix2pix	0.85	0.40	0.32

4. ACCURACY VERIFICATION

4.1 *Comparison with Other Methods*

First, we discuss the accuracies of image transformation for pix2pix and the other methods. Table II shows the performance of the photo → labels task on the Cityscapes [10]. In this case, pix2pix outperforms the other methods.

4.2 *Quantitative Evaluation based on Indicators*

The learning accuracy was verified using three quantitative evaluation indices; PSNR (peak signal-to-noise ratio) [19], SSIM [20, 21], and LPIPS (Learned Perceptual Image Patch Similarity) [22].

As shown in Figure 5, the proposed method had the best evaluation values for all quantitative evaluation indices. Comparing the quantitative evaluation results of each image, it can be said that the proposed learning model achieves more accurate image transformation than the conventional pix2pix.

4.3 *Qualitative Evaluation by Experts*

We conducted a qualitative evaluation by experts because the correct images in the dataset were created based on human perceptual judgment, and evaluation using a quantitative index alone was not sufficient. From all the training results, two images with high and two images with low quantitative evaluation values were selected for each dataset; a total of 16 images were evaluated. Color correction is a technique that involves a great deal of empirical know-how, and it is difficult for a novice to judge which images are reproduced correctly and which are not by comparing the RGB images—the correct images and the resulting images. For this reason, three experts conducted the evaluation. During the qualitative evaluation, the RGB image, the correct image, and the training result image were presented while the results of the quantitative evaluation were hidden. The experts were asked to evaluate the images by focusing on

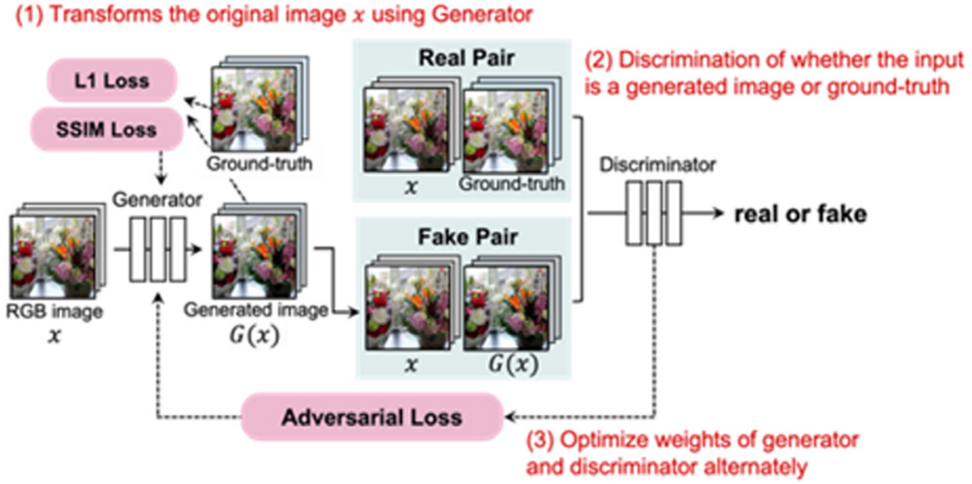

(1) Transforms the original image x using Generator
L1 Loss
SSIM Loss
Ground-truth
Generator
RGB image
x
Generated image
G(x)
Real Pair
x
Ground-truth
Fake Pair
x
G(x)
(2) Discrimination of whether the input is a generated image or ground-truth
Discriminator
real or fake
Adversarial Loss
(3) Optimize weights of generator and discriminator alternately

Figure 4. Architecture of proposed method.

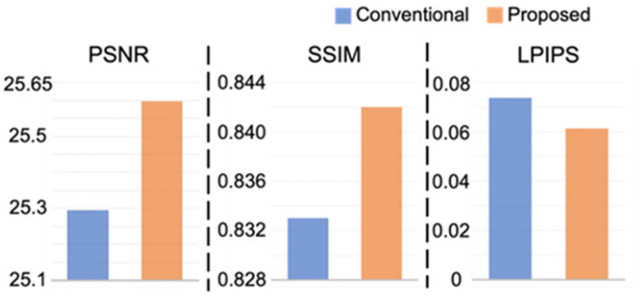


Figure 5. Quantitative evaluation result. PSNR (peak signal-to-noise ratio), SSIM (structural similarity), LPIPS (Learned Perceptual Image Patch Similarity).

"what kind of correction was performed in the case of manual work" and "which parts were realized by machine learning and which parts were not". Based on the results of the qualitative evaluation, we discuss the challenges of the proposed method.

4.4 *Consideration*

We discuss the training results for each dataset.

In the case of the plant images, the results of the test data were significantly less accurate than the training results of the other three small datasets. One of the reasons was the color bias between the training and test data. The color of the plants in the training dataset was limited to red, blue, yellow, etc., which prevented the learning of the color changes and thus did not yield good results. Therefore, we believe that the accuracy of the small dataset can be improved if the data is carefully examined and augmented while reducing the color bias.

The results for the human images showed that many of the images were not corrected well for brightness. In the case of human images, the brightness of the face was dark or white, depending on the shooting environment, thus many images required brightness correction. To improve the training accuracy for such images, it was necessary to classify the images into those that are brightened overall, those that are darkened overall, and those that are corrected in the middle when creating the training data set. In addition, when a person is recognized in an image, it would be effective to automatically convert the skin color of the person to the appropriate color using a target value.

The results for object images were the best. The object images in this study were mainly product images used in advertisements and catalogs. Many of them have simple structures, such as a single background color, few uneven colors due to lighting, and a single object in the image. Therefore, the amount of manual correction was relatively small, and the image transformation by learning may have produced better results compared to other datasets. However, in practical applications, the proposed method may be applied to multiple images on the same page of a brochure or catalog to perform color correction. In such a case, the color-correction results for the background color must all match. Currently, unevenness in background color occurs in different images. Therefore, in the future, it is necessary to compare the results for images with the same background color to verify not only whether there is no unevenness in color, but whether the brightness and tones of the images are equal across multiple images.

Furthermore, two of the quantitative evaluation indices used in this study, PSNR and SSIM, were calculated based on image luminance values. Therefore, it is possible that they did not accurately evaluate the color component, which should be considered important for this study. Therefore, we believe that more accurate quantitative evaluation results will be obtained in the future by using image evaluation indices specialized for color components, such as hue, saturation, and value.

5. CONCLUSION AND FUTURE WORKS

In this study, we proposed a method based on pix2pix to learn and automate techniques for correcting color changes caused by the conversion from RGB to CMYK using a color-transformation image database. The effectiveness of automatic color correction was demonstrated by the proposed method.

Future works include improving the accuracy of local features in the image and extending the image size. One problem with the current learning results is that the correction tends to be applied in the same direction and with the same strength to the entire image. Manual correction, on the other hand, considers various objects and backgrounds in the image and applies the appropriate correction for each of them. We believe that image transformation using deep learning, by introducing object recognition features and recognizing each object and color in the image, and deepening learning tailored to each of them, will make it possible to correct images according to local features of the image.

In addition, the proposed method can only process images of size 256 × 256 pixels. However, considering practical situations, image sizes vary, and we consider it a prerequisite to be applicable to larger images. Pix2pixHD [23], an extension of pix2pix, achieves image transformation for high-resolution images of 2048 × 1024 pixels. We would like to consider applying such a model to improve the size of images that can be learned.

Furthermore, the images in all datasets in this study were compressed to 256 × 256 pixels, resulting in low-resolution and blurred images. This caused some of the images to be corrupted during image processing owing to the reduced resolution. Therefore, it is necessary to construct a model that can be applied to high-resolution images, as described above, or to use networks that restore low-resolution images to high-resolution, such as SRCNN [24] and GAN for super-resolution [25], in response to the results of this study.

ACKNOWLEDGMENT

This work was supported by Sumida Ward Office, Japan.

In addition, we would like to thank Editage (www.editage.jp) for English language editing.

REFERENCES

[1] T. Fujiwara, "Difference in color among rendering intents for the profile conversion," Bulletin **62**, 1–9 (2018).

[2] T. Sugiyama and F. Nakasai, "Color management and standardization in printing factory," J. Printing Sci. Technol. **48**, 117–123 (2011).

[3] M. Asano, "Color management and standardization in the pressroom," J. Printing Sci. Technol. **48**, 124–129 (2011).

[4] L. Ghent, "Recognition by children of realistic figures presented in various orientations," Can. J. Psychol. **14**, 249–256 (1960).

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *NIPS* (Curran Associates, Red Hook, NY, 2014).

[6] M. Mirza and S. Osindero, "Conditional generative adversarial nets," CoRR, abs/1411.1784, (2014).

[7] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CVPR* (IEEE, Piscataway, NJ, 2017), pp. 5967–5976, 3, 4.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015*, edited by N. Navab, J. Hornegger, W. Wells, and A. Frangi, Lecture Notes in Computer Science (Springer, Cham, 2015), Vol. 9351.

[9] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Computer Vision–ECCV 2016. ECCV 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling, Lecture Notes in Computer Science (Springer, Cham, 2016), Vol. 9907.

[10] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc. 2017 IEEE Int'l. Conf. on Computer Vision* (IEEE, Piscataway, NJ, 2017), pp. 2223–2232.

[11] H. F. Tung, A. W. Harley, W. Seto, and K. Fragkiadaki, "Adversarial inverse graphics networks: learning 2D-to-3D lifting and image-to-image translation from unpaired supervision," *The IEEE Int'l. Conf. Computer Vision* (IEEE, Piscataway, NJ, 2017).

[12] E. Lin, "Comparative analysis of Pix2Pix and cycleGAN for image-to-image translation," *2023 Int'l. Conf. on Computers, Machine Learning and Artificial Intelligence* (CMLAI, San Francisco, USA, 2023).

[13] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2017).

[14] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," *2017 IEEE Int'l. Conf. on Computer Vision (ICCV)* (IEEE, Piscataway, NJ, 2017).

[15] V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro, and A. Courville, "Adversarially learned inference," *5th Int'l. Conf. on Learning Representations, ICLR 2017, Toulon, France, April 24–26* (2017), Conference Track Proceedings. OpenReview.net (2017).

[16] M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," *NIPS* (Curran Associates, Red Hook, NY, 2016).

[17] J. Donahue, P. Krahenbuhl, and T. Darrel, "Adversarial feature learning," *Int'l. Conf. on Learning Representations* (2017).

[18] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2017).

[19] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, Inform. Fusion **55**, 220–244 (2020).

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13** (2004).

[21] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *NIPS* (Curran Associates, Red Hook, NY, 2016).

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2018), pp. 586–595.

[23] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2018), pp. 8798–8807.

[24] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *TPAMI* (IEEE, Piscataway, NJ, 2015).

[25] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic signal image super-resolution using a generative adversarial network," *CVPR* (IEEE, Piscataway, NJ, 2017).