Learning From Paired and Unpaired Data: Alternately Trained CycleGAN for Near Infrared Image Colorization

Zaifeng Yang and Zhenghua Chen
Institute of High Performance Computing, A*STAR, Singapore
Institute for Infocomm Research, A*STAR, Singapore

***Abstract*—This paper presents a novel near infrared (NIR) image colorization approach for the Grand Challenge held by 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP). A Cycle-Consistent Generative Adversarial Network (CycleGAN) with cross-scale dense connections is developed to learn the color translation from the NIR domain to the RGB domain based on both paired and unpaired data. Due to the limited number of paired NIR-RGB images, data augmentation via cropping, scaling, contrast and mirroring operations have been adopted to increase the variations of the NIR domain. An alternating training strategy has been designed, such that CycleGAN can efficiently and alternately learn the explicit pixel-level mappings from the paired NIR-RGB data, as well as the implicit domain mappings from the unpaired ones. Based on the validation data, we have evaluated our method and compared it with state-of-the-art method in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and angular error (AE). The experimental results validate the proposed colorization framework.**

***Index Terms*—VCIP challenge 2020, NIR image colorization, CycleGAN, UNet, paired/unpaired images, data augmentation.**

I. Introduction

NEAR Infrared (NIR) is invisible to human eyes with wavelength from 780 nm to 1000 nm. It has been widely used in imaging under low-light conditions such as night time video surveillance, object detection and remote sensing [1], [2]. The NIR images are all monochromatic without color information like grayscale images. The absence of color makes people difficult to perceive the scene in the real world. Recently, learning-based methods for colorization have been developed to restore the color information from monochromatic images. Although both grayscale and NIR images are monochromatic with only one channel and share some similarities in colorization techniques, the NIR image colorization is much more challenging. The intensity of NIR images depends on the absorption and reflection of the materials in addition to the color itself. Besides, the NIR images, unlike grayscale images, cannot be directly converted from the RGB ones. NIR image colorization based on paired NIR-RGB data have been developed in the past few years [3]-[6]. In [4], an additional edgenet was introduced to improve the edges and textures of the colorized NIR images. Suarez *et al.* [5], [6] proposed a triplet loss to improve the colorization performance. However, these approaches are based on paired data only. The public available paired data are limited and it is labour intensive and time consuming to generate these aligned data. To explore more on colorization using unpaired data, CycleGAN [7] has been applied to learn the mapping from the NIR domain to the RGB domain [8], [9]. In [8], CycleGAN with Unet was utilized to learn NIR colorization based on unpaired data. In [9], asymmetric CycleGAN was proposed to simplify the projection from RGB to NIR domains, and a ResNet was embedded in the innermost layer to enlarge the capacity for high-level information.

In this paper, a novel NIR image colorization approach is developed for the Grand Challenge held by 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP). The challenge provides 372 paired NIR-RGB images and additional 1020 RGB images without the corresponding NIR images. To learn the NIR colorization from both available paired and unpaired data, a CycleGAN with cross-scale dense connections is designed to learn the color translation from the NIR domain to the RGB domain. An alternating training strategy is developed so that CycleGAN can efficiently and alternately learn the explicit pixel-level mappings from the paired NIR-RGB data and the implicit domain mappings from the unpaired ones. The rest of the paper is organized as follows: In Section II, the training scheme and the proposed CycleGAN architecture including the generator, discriminator, loss functions and data augmentation via cropping, scaling, contrast and mirroring operations will be described. In Section III, we will compare with the state-of-the-art method in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and angular error(AE) on the validation data.

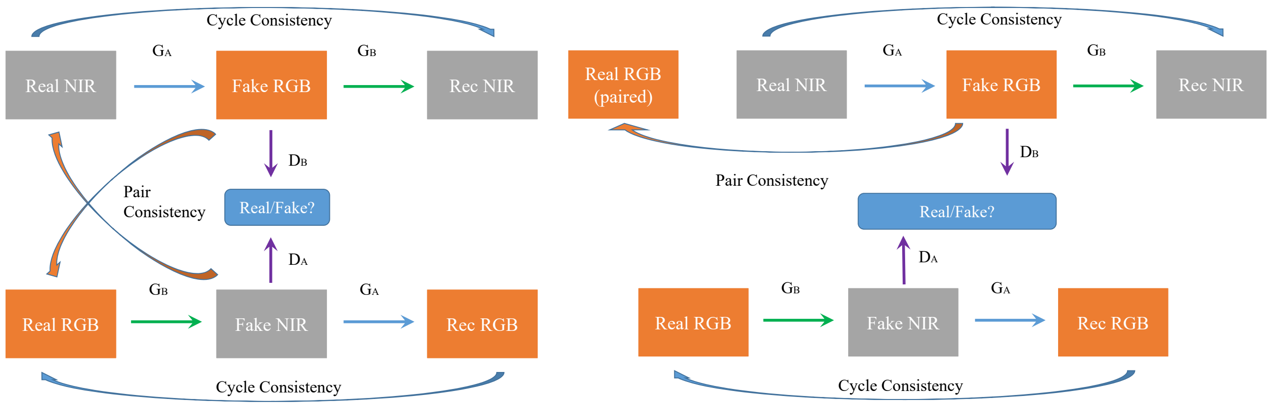
II. Proposed Approach

The proposed CycleGAN built upon UNet256 with cross-scale dense connections is able to increase the learning capacity for high-level semantic information. Considering the provided paired and unpaired data of the challenge, a three-phase training scheme with alternating training strategy is

Zaifeng Yang is with the Department of Electrical and Photonics, Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore.

Zhenghua Chen is with the Department of Machine Intellection, Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore.

Our team name for the VCIP 2020 Grand Challenges on NIR Image Colorization is A*StarTrek.



(a) Paired Data (b) Unpaired Data

Fig. 1: Illustration of the proposed CycleGAN architecture for NIR colorization with (a) paired data and (b) unpaired data.

proposed and data augmentation is applied to introduce more variations on the limited NIR-RGB paired data.

A. *CycleGAN Architectures for Paired and Unpaired Data*

As shown in Fig. 1, different from the conventional CycleGAN in [8], the proposed CycleGAN learns colorization from both paired and unpaired data. in Fig. 1a, when it is trained by the paired data, pair consistency losses are imposed on the generated RGB/NIR and their original input images. For the unpaired data shown in Fig. 1b, the pair consistency loss can be only applied to the generated RGB images and their corresponding ground truth. The cycle consistency losses will be employed in both paired and unpaired data for the input images and the reconstructed images. The generator A (G_A), the generator B (G_B), the discriminator A (D_A) and the discriminator B (D_B) in Fig. 1 are optimized with special training schemes. It will be introduced with details in Section II-D.

B. *Generator and Discriminator*

We use UNet256 with cross-scale dense connections in both generators (i.e., G_A and G_B) to enlarge its learning capacity instead of using additional embedded ResNet [9]. The idea is borrowed from [11], but we only use a few connections on the decoder side because this part is the key steps to restore the color information. The structure of the generators using UNet256 with cross-scale dense connections is illustrate in Fig. 2. For each encoder, it includes the convolution layers with the kernel size of 4×4 and the stride size of 2, followed by instance normalization [10] (exclude encoder x1) and a leaky ReLU (0.2) activation function. For each decoder, the kernel and stride sizes of the de-convolution layers are the same as the ones in the encoder, except that ReLU activation function is used. As illustrated in Fig. 2, for G_A learning colorization from NIR $\rightarrow$ RGB, the input image is the NIR image with one channel, while for G_B learning from NIR $\rightarrow$ RGB, the input image is the RGB image with three channels. On the decoder side, y4 is upscaled using bilinear

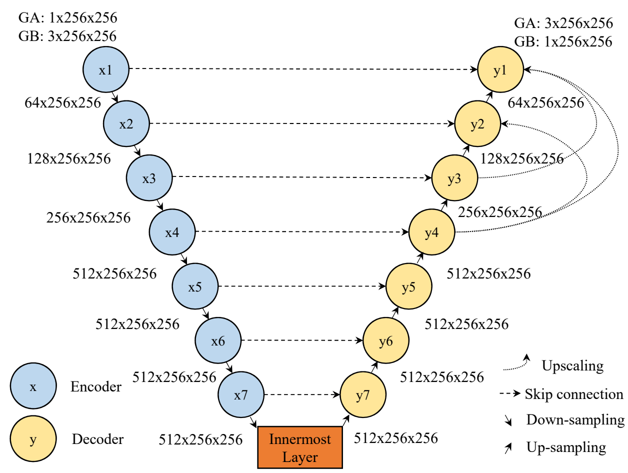


Fig. 2: Illustration of the structure of the generators G_A and G_B using UNet256 with cross-scale dense connections.

interpolation by 4 and 8 times for the connections with y2 and y1, respectively. y3 is upscaled by 4 times and connected to y1 together with the upscaling from y4. For the discriminators DA and DB, conventional PatchGAN is utilized.

C. *Loss Functions*

For the paired and unpaired data, the total losses ($\mathcal{L}_{\text{total\_paired}}$ and $\mathcal{L}_{\text{total\_unpaired}}$) are different by combining GAN loss, cycle consistent loss and pair consistent loss. The detailed losses are given in the following paragraphs.

1) GAN Loss: Taking G_A : NIR to RGB as an example, the GAN loss can be expressed as

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G_A, D_B, A, B) &= \mathbb{E}_{B\sim p_{data}(B)}[\log D_B(B)] \\ &+ \mathbb{E}_{A\sim p_{data}(A)}[\log(1 - D_B(G_A(A)))]. \quad (1)\end{aligned}$$

where $\mathbb{E}_{A\sim p_{data}(A)}$ and $\mathbb{E}_{B\sim p_{data}(B)}$ are the distributions of NIR images and generated fake RGB images, respectively, and $\log(1 - D_B(G_A(A)))$ and $\log D_B(B)$ are the probabilities of the discriminators for real and fake data.

2) Cycle Consistent Loss: Cycle consistent loss [7] is applied for reconstructed and input images in the both data. G_A and G_B are the generators to map from NIR to RGB domains and from RGB to NIR domains, respectively. The cycle consistent loss for both data can be expressed as

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G_A, G_B) = \mathbb{E}_{A\sim p_{data}(A)}[||G_B(G_A(A)) - A||_1] \\ + \mathbb{E}_{B\sim p_{data}(B)}[||G_A(G_B(B)) - B||_1]. \end{aligned} \quad (2)$$

3) Pair Consistent Loss: Different pair consistency losses are imposed on paired data and unpaired data for handling the provided mixed data. For the pair consistent losses on the paired data, they can be calculated as

$$\begin{aligned}\mathcal{L}_{\text{pairA}}(G_A, A, B) = \mathbb{E}_{A\sim p_{paired\_data}(A)}[||(G_A(A)) - B||_1] \\ \mathcal{L}_{\text{pairB}}(G_B, A, B) = \mathbb{E}_{B\sim p_{paired\_data}(B)}[||A - (G_B(B))||_1]. \end{aligned} \quad (3)$$

For the pair consistent loss for the unpaired data, it can be calculated as

$$\begin{aligned}\mathcal{L}'_{\text{pairA}}(G_A, A, B_{paired}) =& \mathbb{E}_{A\sim p_{unpaired-data}(A)} \\ & [||(G_A(A)) - B_{paired}||_1]. \end{aligned} \quad (4)$$

4) Total Loss: The total losses for both data can be expressed as follows:

$$\begin{aligned}\mathcal{L}_{\text{total\_paired}} = \mathcal{L}_{\text{GAN}} + \lambda_1\mathcal{L}_{\text{cyc}} + \lambda_2\mathcal{L}_{\text{pairA}} + \lambda_2\mathcal{L}_{\text{pairB}} \\ \mathcal{L}_{\text{total\_unpaired}} = \mathcal{L}_{\text{GAN}} + \lambda_1\mathcal{L}_{\text{cyc}} + \lambda_2\mathcal{L}'_{\text{pairA}}. \end{aligned} \quad (5)$$

where λ_1 and λ_2 indicate the contributions of the cycle consistent loss and the pair consistent loss, respectively.

D. Training Scheme

Unlike the conventional CycleGAN in [8], the proposed CycleGAN treats the paired and unpaired data using different schemes. The whole training procedure can be divided into three phases. In the first phase, only the paired images are employed for training. In this period, the CycleGAN learns using the paired data efficiently because the pair consistency losses can lead it to convergence faster. In the second phase, the unpaired and paired data will be used in an alternating manner. This scheme can restrict the learning from generating unrealistic colorized images because it will learn from the paired data with ground truth for every other epoch. In the third phase, the learning rate will be decayed to zero gradually. The CycleGAN will learn the colorization with smaller learning rates to improve the performance delicately.

E. Data Augmentation

In order to make full advantages of the available data, the 372 NIR and their corresponding paired RGB images will be augmented based on random size cropping, scaling, contrast adjustment and mirroring as follows.

- **Contrast adjustment**: For each image loaded to the input of the network, 80% of its contrast will be randomly adjusted from 0.5 to 1.5 times of the original image.
- **Cropping and scaling**: After contrast adjustment, the NIR images will be randomly cropped to a size of 200×200 and resized to 256×256. Bicubic interpolation is adopted to scale the image to its original size.
- **Mirroring**: Half of the NIR images will be mirrored based on the original images after size scaling and contrast adjustment.

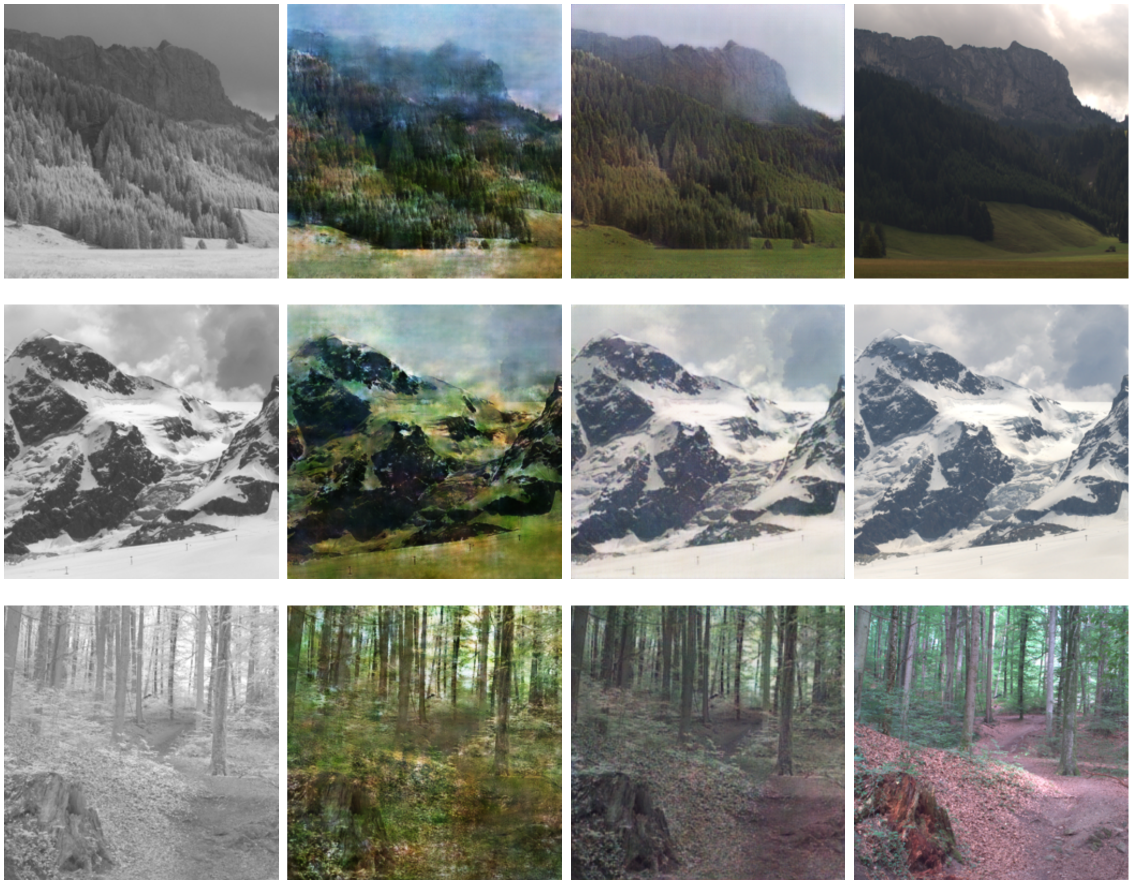
TABLE I: Comparison between the proposed method and CycleGAN with UNet256.

No.	CycleGAN: UNet256			Proposed Method		
	PSNR	SSIM	AE	PSNR	SSIM	AE
1	16.60	0.45	3.72	17.81	0.53	3.17
2	16.58	0.51	7.78	21.30	0.64	3.38
3	12.08	0.54	15.33	12.73	0.56	9.40
4	15.80	0.22	8.05	20.17	0.40	5.18
5	12.45	0.53	8.02	17.47	0.63	6.15
6	11.78	0.28	9.92	13.58	0.36	6.05
7	13.68	0.41	7.10	14.91	0.50	5.92
8	12.46	0.29	10.18	15.50	0.40	7.28
9	10.53	0.44	10.22	12.94	0.58	5.13
10	13.49	0.39	7.85	13.63	0.49	8.87
11	15.29	0.35	9.31	19.74	0.57	6.14
12	17.10	0.55	8.23	14.52	0.61	6.76
13	8.73	0.50	9.17	26.09	0.91	1.31
14	13.60	0.44	12.65	15.65	0.54	4.68
15	13.01	0.40	12.36	17.66	0.49	8.23
16	15.07	0.54	4.79	18.89	0.65	2.78
17	14.30	0.41	11.56	20.93	0.62	12.99
18	11.02	0.41	12.60	15.05	0.58	3.30
Avg	13.53	0.43	9.38	17.14	0.52	5.93

III. Experimental Results

We have trained our proposed CycleGAN using a PC with dual Intel Xeon Gold 6240 (36 cores) CPUs and two NVIDIA Geforce RTX 2080 Ti GPUs under Ubuntu operating system. All the images have the same size of 256×256 and are normalized to the range of (0,1). $\lambda_1 = 100$ and $\lambda_2 = 60$ are used in Eq. 5. Training for one epoch using paired data is 32s in average, while it takes 93s using unpaired data in average for each epoch. As mentioned in Section II-D, three-phase training scheme is executed. In the first phase from 1 to 250 epoch, only paired data is used for the training. In the second phase (250 to 550 epoch), paired data and unpaired data will be trained alternately. In the third phase (550 to 800 epoch), the learning rate will be decayed from one to zero gradually.

Based on the 18 provided validation paired data, we evaluated our method by comparing with the state-of-the-art method (CycleGAN with UNet256 [8] with the same total loss using in our proposed method) in terms of PSNR, SSIM [12] and AE. The evaluation results are given in Table I. It can be found that the proposed approach significantly outperforms the one using CycleGAN with Unet256 given that the loss functions are set to the same. The proposed method achieves 17.14, 0.56 and 5.93 in terms of the averaged PSNR, SSIM and AE, respectively. In addition, some colorized NIR images from the validation data, together with the NIR images and the ground truths from the validation data, are shown in Fig. 3. It is obvious that the colorized NIR images using our proposed method have effectively captured the textures and edges information as well as learning the color information correctly.



(a) original NIR image (b) CycleGAN with UNet256 (c) Proposed method (d) Groud truth

Fig. 3: Comparison of the colorized NIR images using CycleGAN with Unet256 and the proposed approach.

IV. Conclusion

In this paper, we proposed a novel NIR image colorization approach for the Grand Challenge held by IEEE VCIP 2020. It is built upon CycleGAN with cross-scale dense connections for both paired and unpaired data. Data augmentation techniques are adopted to relieve the issue of limited data for training. We also proposed a novel alternating training strategy for performance improvement. Evaluation has been conducted on the validation data. The results indicate that the proposed method performs quite well for NIR image colorization. Evaluated using the validation data, the proposed method achieves 17.14, 0.56 and 5.93 in terms of the evaluation metrics of averaged PSNR, SSIM and AE, respectively. The proposed method consistently outperforms the current state-of-the-art under the three evaluation metrics.

References

[1] P. Govardhan and U. C. Pati, "NIR image based pedestrian detection in night vision with cascade classifcation and validation," in Proc. IEEE Int. Conf. Adv. Commun., Control Comput. Technol., May 2014, pp. 1435-1438.

[2] E. R. Hunt, W. D.Hively, S. J. Fujikawa, D. S. Linden, C. S. Daughtry, and G. W. McCarty, "Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring," Remote Sensing, vol.2, no.1, pp. 290-305.2010.

[3] M. Limmer and H. P. A. Lensch, "Infrared colorization using deep convolutional neural networks," in Proc. 15th IEEE Int. Conf. Mach. Learn.Appl. (ICMLA), Dec. 2016, pp. 61-68.

[4] Z. Dong, S.-I. Kamata, and T. P. Breckon, "Infrared image colorization using a S-shape network," in Proc. IEEE Int. Conf. Image Process., Oct. 2018, pp. 2242-2246.

[5] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla. "Infrared image colorization based on a triplet DCGAN architecture," In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 212–217. IEEE, 2017.

[6] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, "Learning to colorize infrared images,", In International Conference on Practical Applications of Agents and Multi-Agent Systems, pages 164–172. Springer, 2017.

[7] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2223-2232. 2017.

[8] Mehri, Armin, and Angel D. Sappa. "Colorizing near infrared images through a cyclic adversarial approach of unpaired samples." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[9] Sun T, Jung C, Fu Q, et al. NIR to RGB Domain Translation Using Asymmetric Cycle Generative Adversarial Networks[J]. IEEE Access, 2019, 7: 112459-112469.

[10] Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1501-1510. 2017.

[11] Zhang, Jiawei, Yuzhen Jin, Jilan Xu, Xiaowei Xu, and Yanchun Zhang. "Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation." arXiv preprint arXiv:1812.00352, 2018.

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