IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 2, 2019, accepted July 27, 2019, date of publication August 7, 2019, date of current version August 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933671

NIR to RGB Domain Translation Using Asymmetric Cycle Generative Adversarial Networks

TIAN SUN, CHEOLKON JUNG, (Member, IEEE), QINGTAO FU, AND QIHUI HAN
School of Electronic Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Cheolkon Jung (zhengzk@xidian.edu.cn)

This work was supported in part by the Huawei Innovation Research Program under Grant HF2018095025, in part by the National Natural Science Foundation of China under Grant 61872280, and in part by the International S&T Cooperation Program of China under Grant 2014DFG12780.

ABSTRACT Near infrared (NIR) images have clear textures but do not contain color. In this paper, we propose NIR to RGB domain translation using asymmetric cycle generative adversarial networks (ACGANs). The RGB image (3 channels) has richer information than the NIR image (1 channel), which makes NIR-RGB domain translation asymmetric in information. We adopt asymmetric cycle GANs that have different network capacities according to the translation direction. We combine UNet and ResNet in generator and use the feature pyramid networks (FPNs) in discriminator. With the help of a 128 × 128 large receptive field, we capture rich spatial context information with a multiscale architecture. Experimental results show that the proposed method achieves natural looking NIR colorization results with high generalization ability, i.e. feasible in category unaware cases, and outperforms state-of-the-art ones in realistic colorization and resistance to unregistration.

INDEX TERMS NIR colorization, asymmetric cycle GAN, domain translation, feature pyramid networks, ResNet, UNet.

I. INTRODUCTION

In low light condition, color (RGB) cameras capture noisy images with loss of color and texture. Near infrared (NIR) cameras are commonly equipped in public to consider the low light condition. NIR images contain details and textures even in low light condition. Thanks to this property, NIR images are usually used in nighttime for object detection systems [1]–[3] or human assistant systems [4]. NIR images are also used as an important cue for RGB color correction, detail enhancement and haze removal by fusion [5]–[13]. However, NIR images are of one channel (no color information), and are much different from human visual perception. Meanwhile, color information is not always available especially in low light condition. Therefore, NIR to RGB domain translation is required. NIR to RGB domain translation, also called NIR colorization, is an ill-posed problem that projects a single channel NIR image into a three channel RGB image. This projection is difficult because the RGB domain distribution is too complex and hard to be constrained by a certain rule. In this work, we address the NIR to RGB domain translation problem and generate plausible RGB images from only NIR images. The generated RGB images are not necessary to be exactly the same as common truth. However, the textures in the NIR domain should be transferred to the RGB domain, and colors of the translated RGB image should be natural looking.

A. RELATED WORK

The NIR colorization is similar to gray image colorization, but it has some significant differences from it. Although the NIR colorization is not a spotlight in computer vision yet, gray image colorization has already received much attention in recent years [15]–[22]. In early image colorization work, the color domain ambiguity problem is not effectively handled. Thus, human interactions or reference color images are necessary [20]–[22]. Example based method is also presented in [23], which is similar to the input reference color image. With the introduction of deep learning, convolutional neural networks (CNNs) are successfully applied to the single gray image based colorization [16], [19], [24], [25]. In general,

The associate editor coordinating the review of this manuscript and approving it for publication was Malik Jahan Khan.

spatial context is the main cue for the learning based image translation approaches. Image pairs in two domains are supposed to have the same distribution in edges and textures. According to the local spatial context information, color estimation is conducted by a deep neural network. In recent years, generative adversarial networks (GANs) play an important role in this area because the cross domain translation is an ill-posed problem [26]. The mapping target is highly ambiguous, that one input can be registered to multi-probable targets. Generation based methods are proven to be effective in this situation because GANs approximate an unknown distribution with generator instead of learning a sample-wise relationship. GANs have a discriminator D which discriminates the real and fake samples, and have a generator G to generate a fake sample with a random input trying to fool the descriminator. G and D have an adversarial training objective defined by (1) in [26], and are trained alternately. Furthermore, GANs are extended to a conditional style to control the generator output [27].

$$\min_G \max_D V(D, G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

One of the most remarkable generation based image colorization work is Pix2Pix [17]. Pix2Pix used UNet [28] and L1-norm to enhance the local performance. Although Pix2Pix performs well in the grayscale colorization, it is not qualified in the NIR colorization. The NIR colorization has some differences from the grayscale colorization. First, the grayscale colorization is mostly seen as chrominance estimation, and the luminance between grayscale and RGB images is supposed to be identical. However, for NIR imaging, the light source is totally different from RGB imaging. The wavelength for NIR imaging ranges [780nm, 1000nm], while the wavelength for RGB imaging ranges [380nm, 780nm]. Moreover, the difference is not only from the wavelength but also from additional active light sources. These differences break the pixel-wise consistency in textures between two domains. Up to the present, there are several researches to deal with the NIR colorization problem [29], [30]. Dong *et al.* [30] proposed an end-to-end network for NIR colorization based on UNet. Suarez *et al.* [31] proposed a triplet deep convolutional GAN (DCGAN) model [32] and improved its colorization performance by the validation on training loss [33]. In addition to the light source difference, capturing RGB-NIR pairs in the same view point simultaneously is a difficult task. The existing open datasets often fail in pixel-wise registration. Some approaches solve this problem by removing the unregistered outliers [30], [31], [33]. However, data cleaning is labour intensive, and data reduction is undesirable for deep learning. Fortunately, the cycle GAN [14], which is designed for training GANs with unpaired samples, has a good resistance against unregistration. Fig. 1 illustrates the cycle GAN framework. The cycle GAN calculates the cyclic loss $L(A, F(G(A)))$ instead of single directional loss $L(G(A), B)$ to gain a resistance against unregisterd sample pairs. However, the cyclic loss can not guarantee a high resolution result. To gain a plausible result with high resolution, Liu *et al.* presented unsupervised image to image translation networks (UNIT) [34] based on the shared latent space assumption [35]. The shared latent space assumption supposes that a pair of corresponding images in different domains can be mapped to the same latent representation in a shared-latent space. UNIT uses variational autoencoder (VAE) to learn the projection from the sample space to the latent space, and utilizes generator to learn the inverse projection. By sharing the parameters in the ending layers of VAE and beginning layers of generator, UNIT forces the samples in different domains projected in the same shared latent space. Inspired by UNIT, bicycle GAN [36] and DRIT [37] achieve better results by using additional projection modules. With respect to the subjective assessment in DRIT, the present shared latent space based methods yield more diverse and plausible results compared with cycle GAN. However, cycle GAN still wins reality in the domain translation.

B. CONTRIBUTIONS

To achieve realistic colorization and resistance to unregistration, we use a cycle GAN based framework. Different from traditional cycle GAN, we build an asymmetric model which has different architectures according to the translation direction. Most of image to image translation methods such as cycle GANs UNIT and DRIT are designed for general domain translations. In their methodology, two translation directions are modeled by the same architecture and the projections are preferred to be invertible. However, in NIR colorization, the RGB domain has more information than the NIR domain. The RGB to NIR projection is a much easier problem than the inverse projection. Hence, we propose an asymmetric model architecture which has different complexities in two projection directions.

Compared with existing methods, the main contributions of this paper are as follows:

- We adopt cycle GANs for the NIR to RGB domain translation to deal with the unregistration problem between two domains.
- We utilize an asymmetric structure to consider different network capacities according to translation direction.
- We combine UNet and ResNet in generator to increase the network depth, while we use FPN in discriminator to capture spatial context information.

The rest of this paper is organized as follows: In Section II, we introduce the proposed networks for generator and discriminator including their architectures and training procedures. In Section III, we perform several experiments to prove the proposed method. Section III.A compares the proposed method with state-of-the-art ones in both NIR colorization and image to image translation under scene type aware and unaware conditions, and shows both the quantitative and visual comparison results. In Section III.B, we make ablation experiments to discuss the contributions of each component. At last, we conclude this work in Section IV.

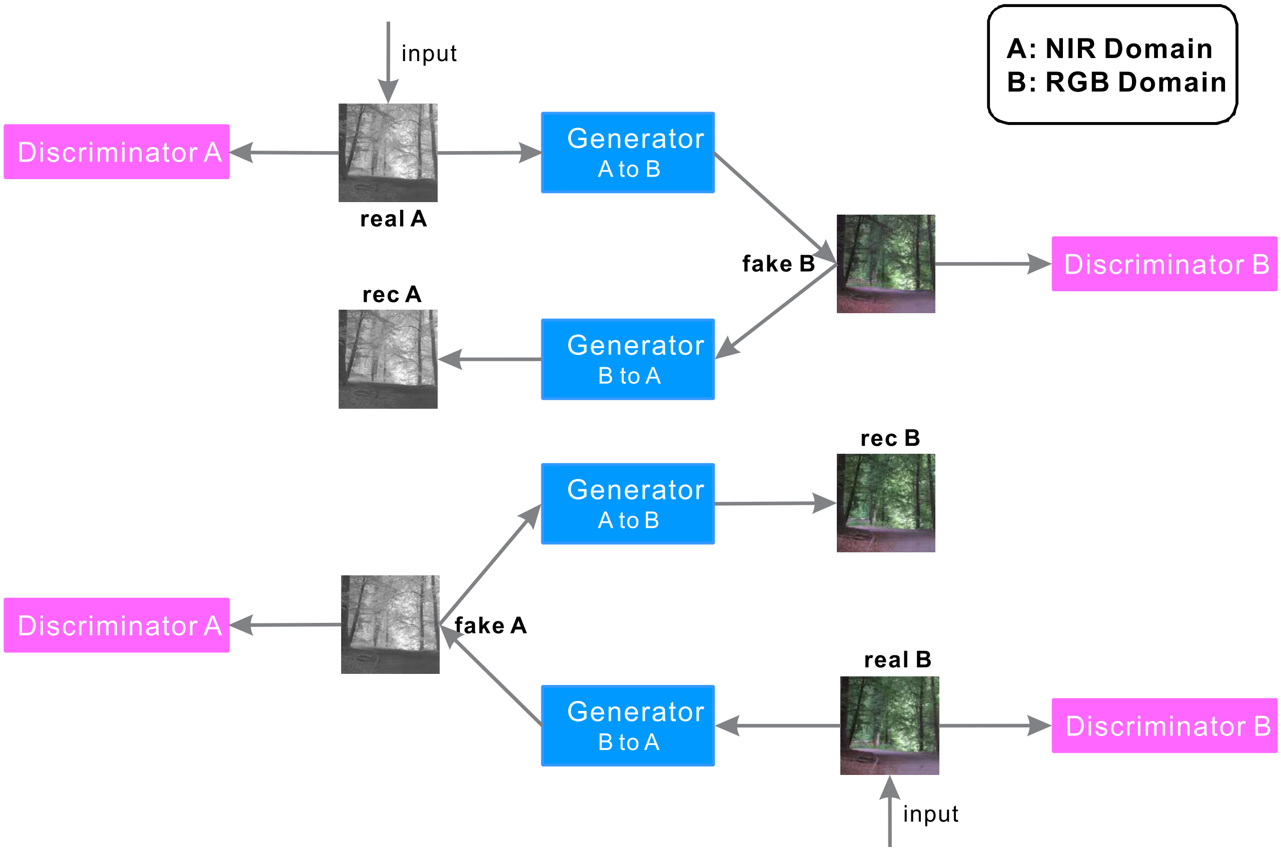

FIGURE 1. NIR to RGB domain translation using asymmetric cycle GAN, redrawn from [14]. Single directional model often calculates L_1 norm $L(\textit{fake B}, \textit{real B})$ as training loss. When *real A* is not registered with *real B*, the training loss is confused by unregistration but the cyclic loss $L(\textit{real A}, F(G(\textit{real A})))$ is unaffected by unregistration.

II. PROPOSED METHOD

Different from the original cycle GAN, the proposed cycle GAN follows an asymmetric style. Since the NIR colorization is a mapping problem from single-channel to multi-channel, NIR→RGB has larger complexity and ambiguity than RGB→NIR. However, the original cycle GAN was designed for general image domain translation, and it is symmetric. In NIR colorization, the RGB domain has more information than the NIR domain. RGB to NIR projection is a much easier problem than the inverse projection. It is inefficient to use the same complex architecture in the inverse direction. Hence, we propose an asymmetric model architecture which has different complexities in two projectional directions. Thus, we specialize traditional cycle GAN into an asymmetric style as shown in Fig. 2. We assign different network architectures to generator and discriminator according to mapping directions.

A. ASYMMETRIC GENERATORS

Generators in both directions use UNet modules because UNet effectively preserves low level and high resolution features. In NIR colorization, we regard the textures as low semantic level features. UNet can effectively preserve the edges and textures. Thus, it is powerful enough for the RGB→NIR translation. However, the colorization according to context is a high level semantic work. It is implicit to understand the scene at the object level and assign color to it. The standard UNet is not enough to model this problem, and thus we add a ResNet block in the UNet architecture for G_{AB} to enlarge its capacity. The ResNet blocks are embedded in the innermost layer of UNet because this style spends more capacities in capturing high level features. The detailed architecture of G_{AB} is shown in Fig. 3. Besides, pooling layers

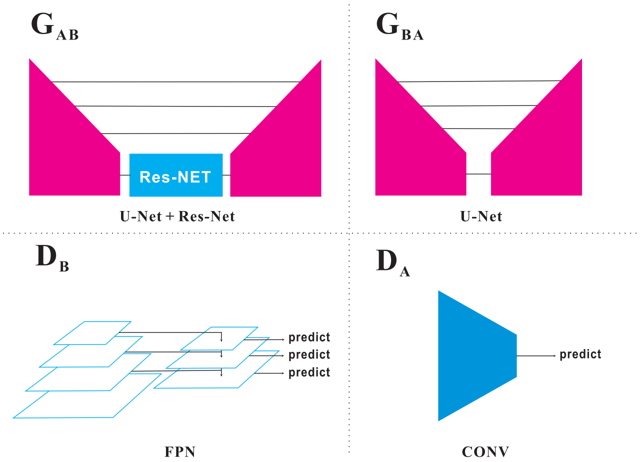

FIGURE 2. Asymmetric network architecture for the proposed cycle GAN.

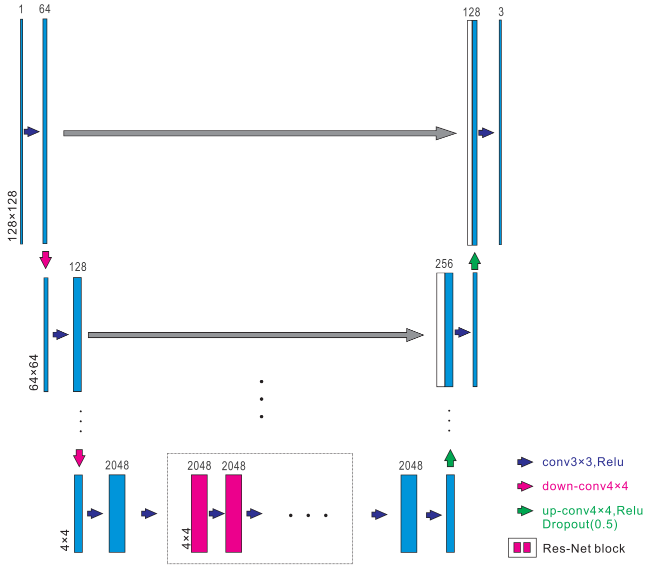

FIGURE 3. Network architecture of the generator. We combine UNet and ResNet into the generator.

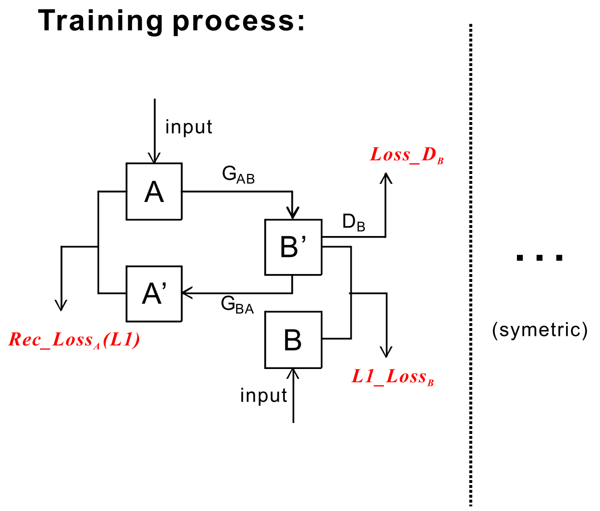

FIGURE 4. Training process and loss calculations. A: Real NIR. A′: Fake NIR. B: Real RGB. B′: Fake RGB.

are substituted by dilated convolutions [38], and dropout layers [39] follow upsampling convolution instead of the random input. It makes the generator a fully convolutional style as illustrated in Fig. 3.

B. ASYMMETRIC DISCRIMINATORS

The discriminator also follows the asymmetric style. The RGB to grayscale domain translation is a relatively easy task. Grayscale image colorization can even be modeled by a linear mapping. Similar to RGB→grayscale, RGB→NIR is not a very complex task. Thus, CNN is enough for D_{BA}. In D_{AB}, we use the feature pyramid network (FPN) strategy for a better discrimination. FPN successfully captures both semantics and scales from convolutions. During the convolutional process, the semantic level increases as the receptive field increases. Objects are difficult to find the proper scale and semantic level, especially for small objects. FPN has top-down and bottom-up paths. The bottom-up path is a general CNN forward process. In the top-down path, high level features are upsampled and added to the low level ones. This procedure brings high level semantic information to a small scale feature map. FPN is proved to be effective in local context based applications such as object detection, object recognition, and segmentation [40]. The NIR colorization is a local context based application, and thus it benefits from FPN.

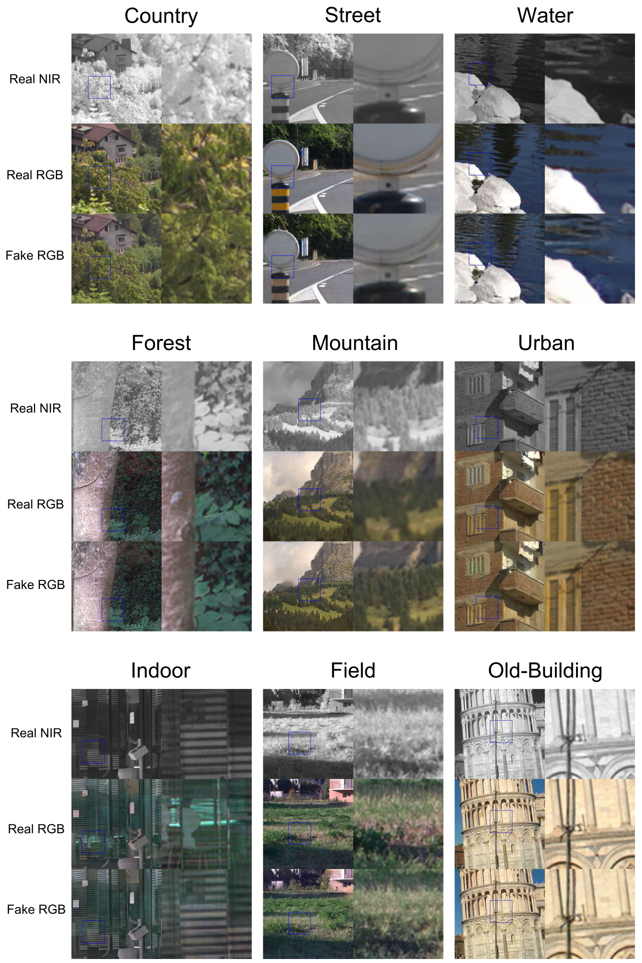

FIGURE 5. Domain translation results in the category aware dataset. In each block, the left column is the original patch and the right column is the zoomed patch (4 times).

C. MODEL TRAINING

The training process of the proposed cycle GAN is illustrated in Fig. 4. Total loss consists of two directional GAN loss and reconstruction loss as follows:

$$\begin{aligned} Total\_loss = \min_{G}\max_{D}\{&L_{GAN}(G_{AB}, D_B, A, B)\\ &+\lambda_1 L_{GAN}(G_{BA}, D_A, A, B)\\ &+\lambda_2 L_{rec}(G_{AB}, G_{BA}, A, B)\}, \end{aligned} \tag{2}$$

The main difference from the original cycle GAN loss is L_1 norm between the original and fake images. L_{GAN} is defined

TABLE 1. Performance comparison in terms of angular error (AE) and structural similarity (SSIM).

	AE		SSIM	
	Urban	*Old-Building*	*Urban*	*Old-Building*
Conditional GAN [31]	5.77	5.96	0.84	0.86
Stacked Conditional GAN [33]	5.04	4.78	0.90	0.91
UNIT [34]	16.02	14.65	0.52	0.52
Cycle GAN [14]	8.41	8.15	0.81	0.83
Proposed Method	5.05	5.30	0.90	0.89

AE is measured by degree (the smaller the better). Although the evaluation data are pixelwise registered, UNIT [34], Cycle GAN [14], and the proposed method are trained on the coarsely registered data (globally calibrated by [42]) instead of the pixelwise registered data. However, [31], [33] are trained on the pixelwise registered data.

TABLE 2. Performance comparison in terms of the training time and stability to unregistration.

	Training time per epoch (hours)	*Stability to unregistration*
Triplet DCGAN [31]	—	×
Stacked Conditional GAN [33]	—	×
UNIT [34]	315.9	✓
Cycle GAN [14]	15.9	✓
Proposed Method	27.9	✓

Time per epoch means time cost for training on the whole training data. UNIT [34] and DRIT [37] use multi-models architecture, and thus they have higher computational cost than the others. Triplet DCGAN [31] and Stacked Conditional GAN [33] do not concern data unregistration. Once coarse registered samples are fed into them, these models are cracked, and the similar case is shown in Fig. 7.

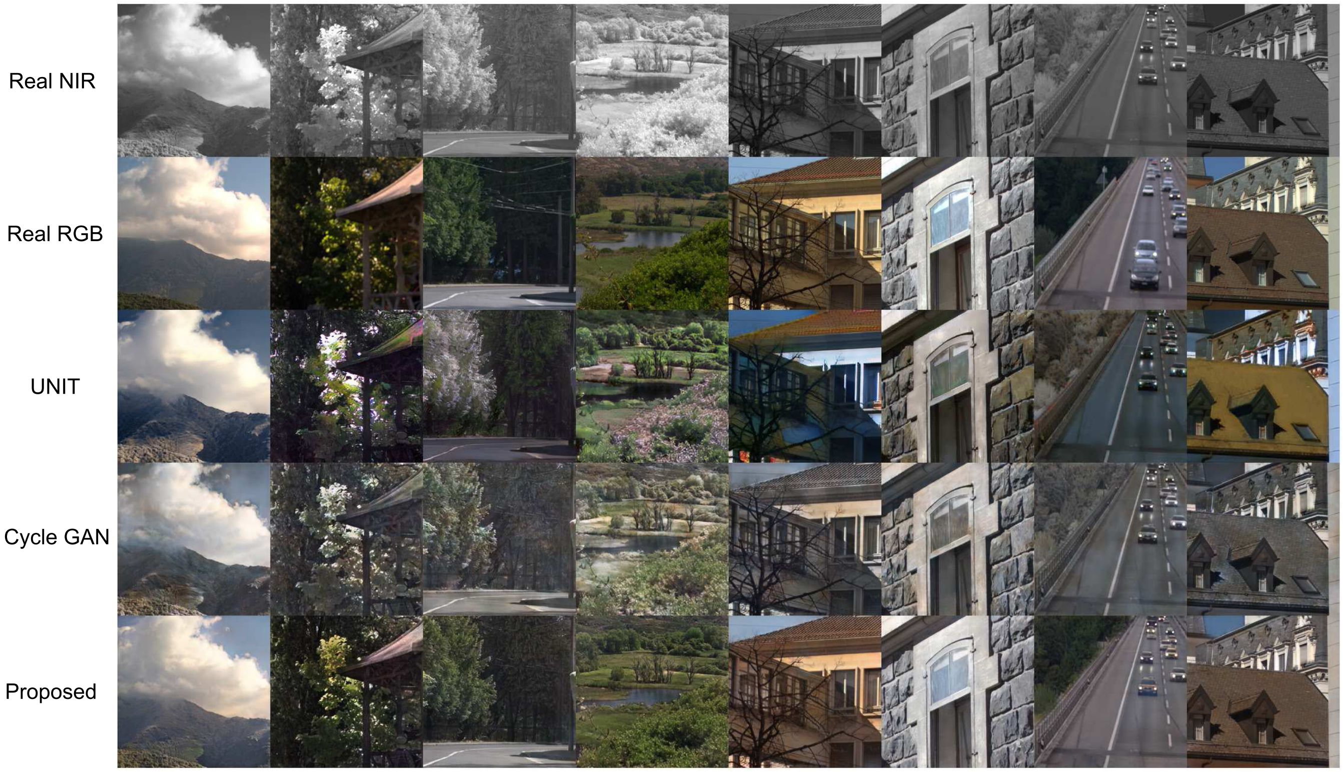

FIGURE 6. Domain translation comparison in the category unaware dataset.

as follows:

$$L_{GAN}(G_{AB}, D_B, A, B) = E[\log D_B(B)] + \lambda_3 E[\log(1 - D_B(G_{AB}(A)))] + \lambda_4 ||A - G_{AB}(A)||_1 \quad (3)$$

$$L_{GAN}(G_{BA}, D_A, A, B) = E[\log D_A(A)] + \lambda_3 E[\log(1 - D_A(G_{BA}(B)))] + \lambda_4 ||B - G_{BA}(B)||_1 \quad (4)$$

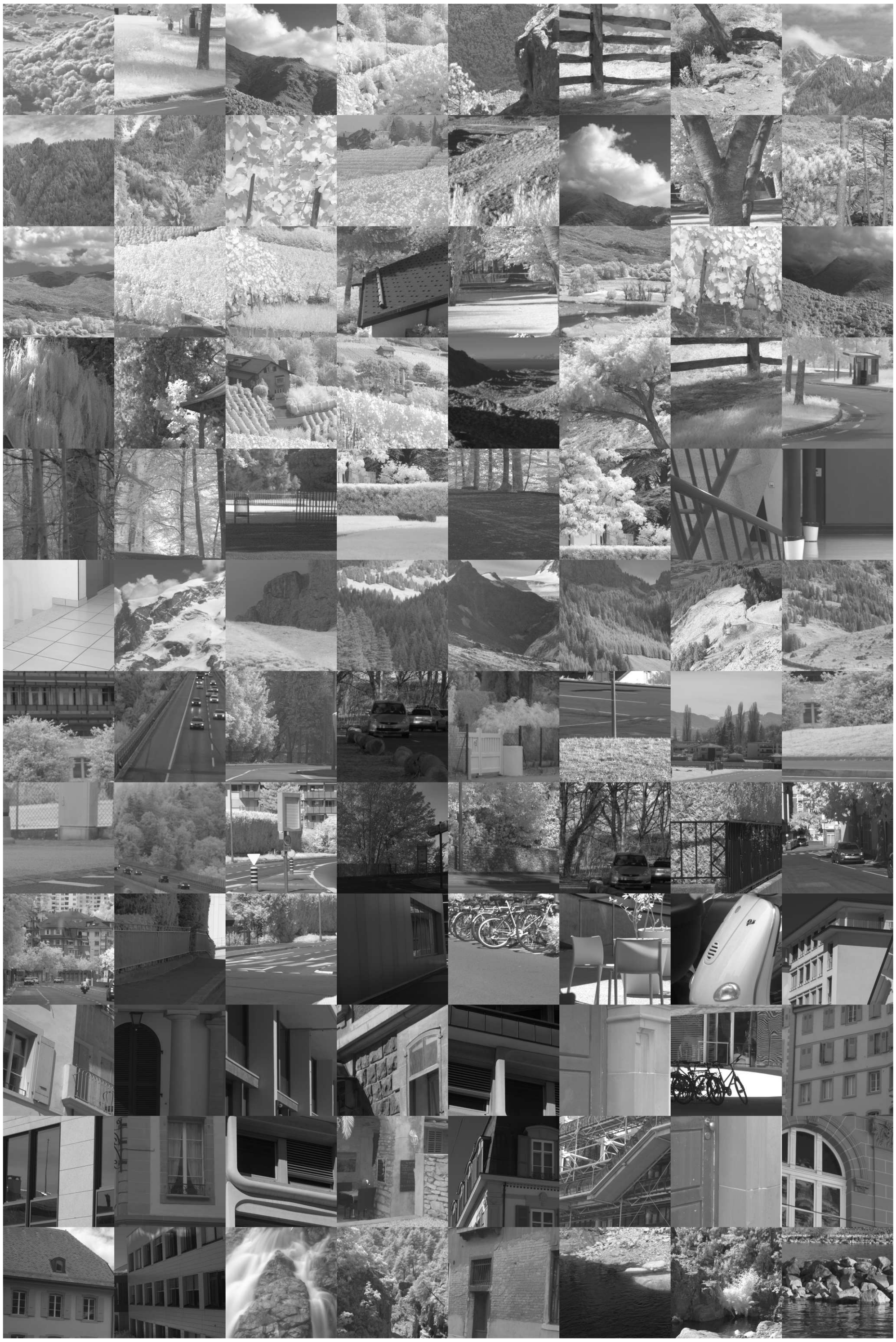

FIGURE 7. **Input NIR images.**

The reconstruction loss follows the original cycle GAN as follows:

$$L_{rec}(G_{AB}, G_{BA}, A, B) = E[||G_{BA}(G_{AB}(A)) - A||_1] + E[||G_{AB}(G_{BA}(B)) - B||_1] \quad (5)$$

L_{rec} and L_{GAN} are jointly optimized in *Total_loss*. Due to the unregistration between RGB and NIR, the two losses sometimes contradict each other. Parameters λ_1 and λ_2 control the trade-off between generation accuracy and robustness to the unregistration. GAN based models are difficult

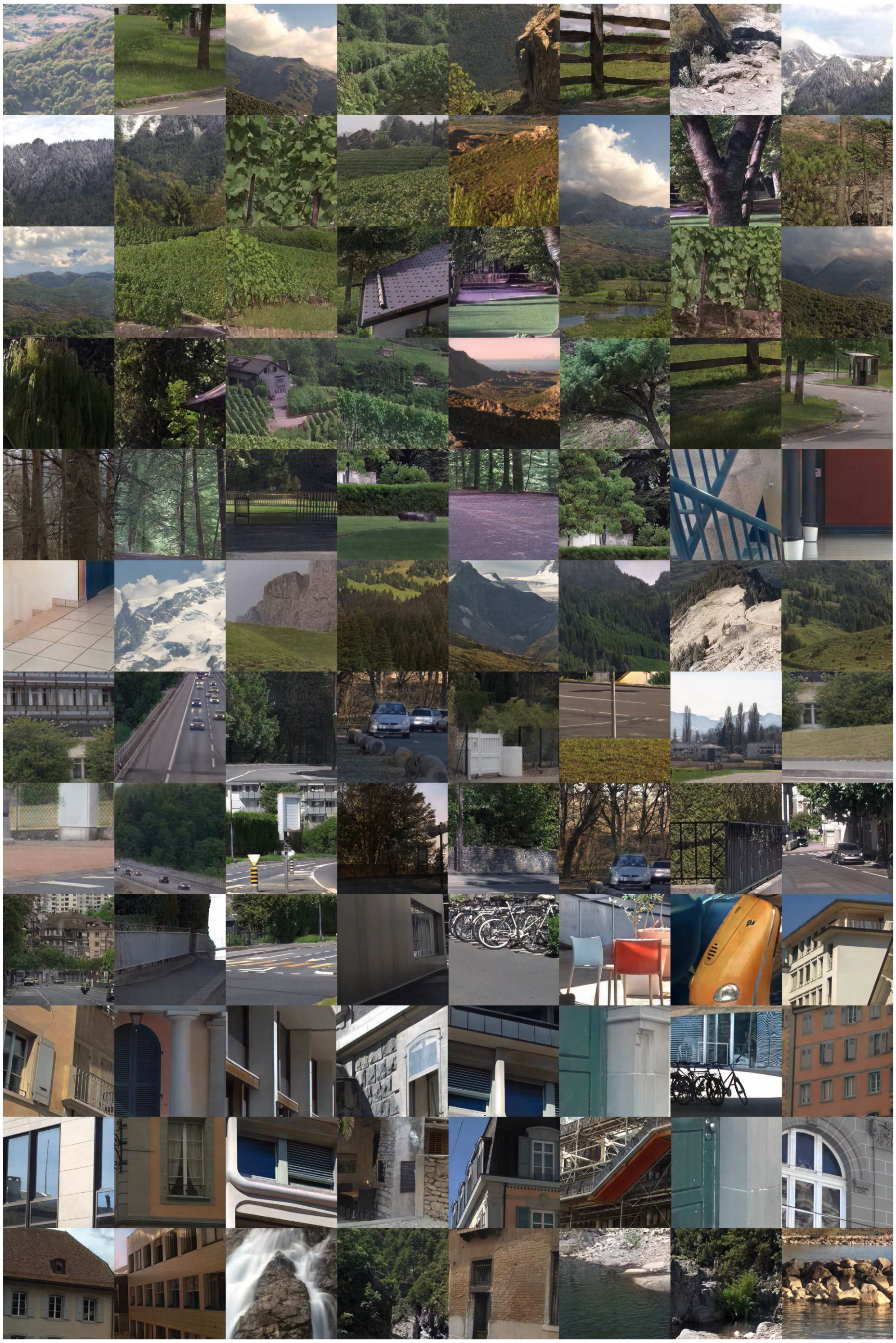

FIGURE 8. NIR colorization results of Fig. 7 by the proposed method.

for training. Thus, batch normalization is used to help the training convergence [41].

III. EXPERIMENTAL RESULTS

We train the proposed asymmetric cycle GAN using a PC with Intel i7 3.6GHz CPU and one NVIDIA GTX 1080Ti GPU. Training for one epoch in each scene category takes 3.1 hours in average. We use IVRL RGB-NIR scene dataset [42], which contains 477 image pairs with resolution of 1024 × 680 captured from 9 categories of scenes. The dataset is available at https://ivrl.epfl.ch/research-2/research-downloads/supplementary_material-cvpr11-index-html/.

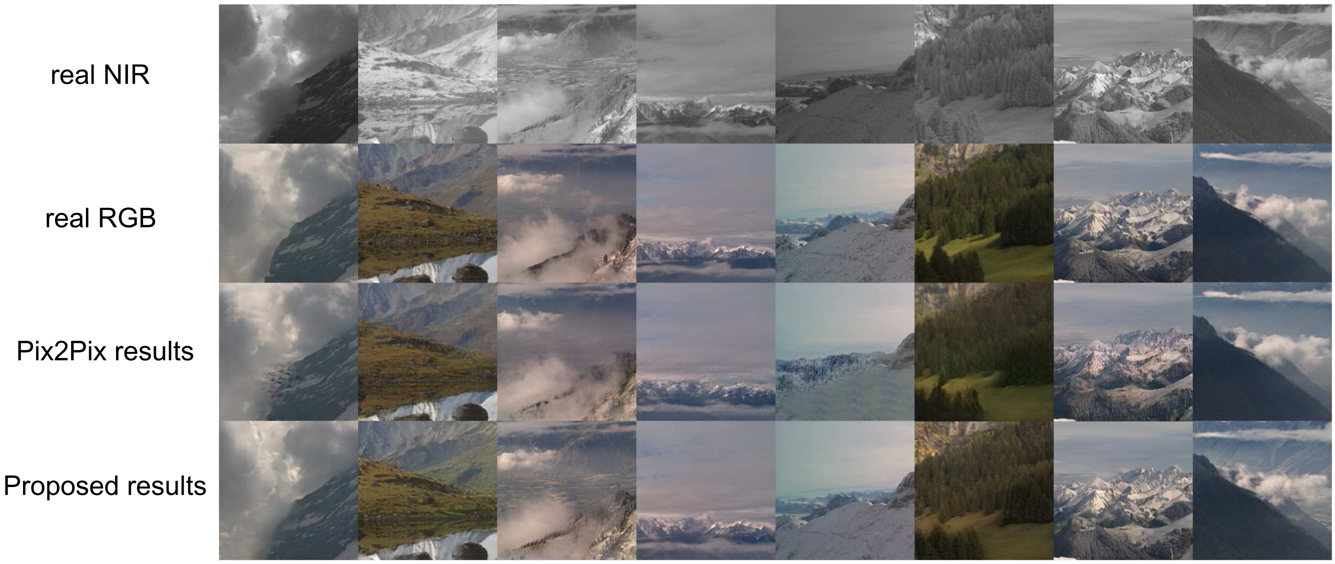

FIGURE 9. Performance comparison between the proposed asymmetric cycle GAN and Pix2Pix. We use coarsely registered data to show the robustness to the unregistration.

TABLE 3. Performance comparison in terms of angular error (AE) and structural similarity (SSIM).

Category Aware Experiments	AE		SSIM	
	Urban	*Old-Building*	*Urban*	*Old-Building*
Cycle GAN + FPN + UNet128×128	5.71	6.14	0.88	0.87
Cycle GAN + CONV + UNet128×128	8.21	8.92	0.79	0.79
Cycle GAN + FPN + UNet64×64	10.81	11.27	0.70	0.67
Cycle GAN + FPN + UNet32×32	12.78	12.33	0.64	0.62

The first row is the proposed method, while the second row is the use of fully convolution discriminator instead of FPN. In the third and fourth rows, we shrink the receptive field size.

Image pairs in this dataset are coarsely registered by [42] using a global calibration method, and thus the pixelwise registration are not ensured. We compare the evaluation results of the proposed method with those of the state-of-the-art methods: Triplet DCGAN [31], Stacked Conditional GAN [33], Cycle GAN [14], and UNIT [34]. Reference [33] randomly crops the dataset into 64 × 64 patches. We crop the images into 146,8000 patches whose size is 256 × 256 to fit 128 × 128 receptive field. For a fair comparison, this sampling ratio is adjusted to 0.0079, i.e. no more than 0.0083 [33].

A. MAIN EXPERIMENTS

We perform two experiments to prove the effectiveness of the proposed method. Experiment I is performed on a scene category aware dataset. The models are separately trained and evaluated for each scene category. In Experiment II, we perform evaluations in a category unaware dataset. Since the category is unknown, this experiment verifies generalization ability of the proposed method. The different scenes in [42] are merged into the training dataset of Experiment II. Data distribution gets randomized without the scene category constraint. Note that more samples with large diversity lead to better generalization performance. Experiments I and II have 16 training epochs. The training steps keep consistency in the first 8 epochs while fading to zero linearly in the second 8 epochs. Since GAN produces plausible results instead of pixel-wise precise results, quantitative evaluations for the NIR colorization are difficult. When NIR is unregistered with RGB, the ground truth is not really true. In this case, pixel-wise assessment does not make sense. Therefore, we remove the unregistered pairs in quantitative evaluations. For quantitative evaluations, we use angular error (AE) and structural similarity (SSIM) [43] as evaluation metrics. It has been reported that AE is the most similar to the human vision [44]. The parameters defined in Eqs. (2), (3), and (4) are empirically set, aiming to keep intermediate results (fake A, fake B, rec A, rec B) by the same step, where $\lambda_1 = 1, \lambda_2 = 0.6$, $\lambda_3 = 0.6, \lambda_4 = 0.6$.

1) SCENE CATEGORY AWARE

We crop 5000 patches for visual evaluations, while we crop another 5000 correctly matched patches for quantitative evaluations. Fig. 5 shows some NIR colorization results in Experiment I. Some contradictions between ground truth RGB and NIR images appear. We provide performance comparison

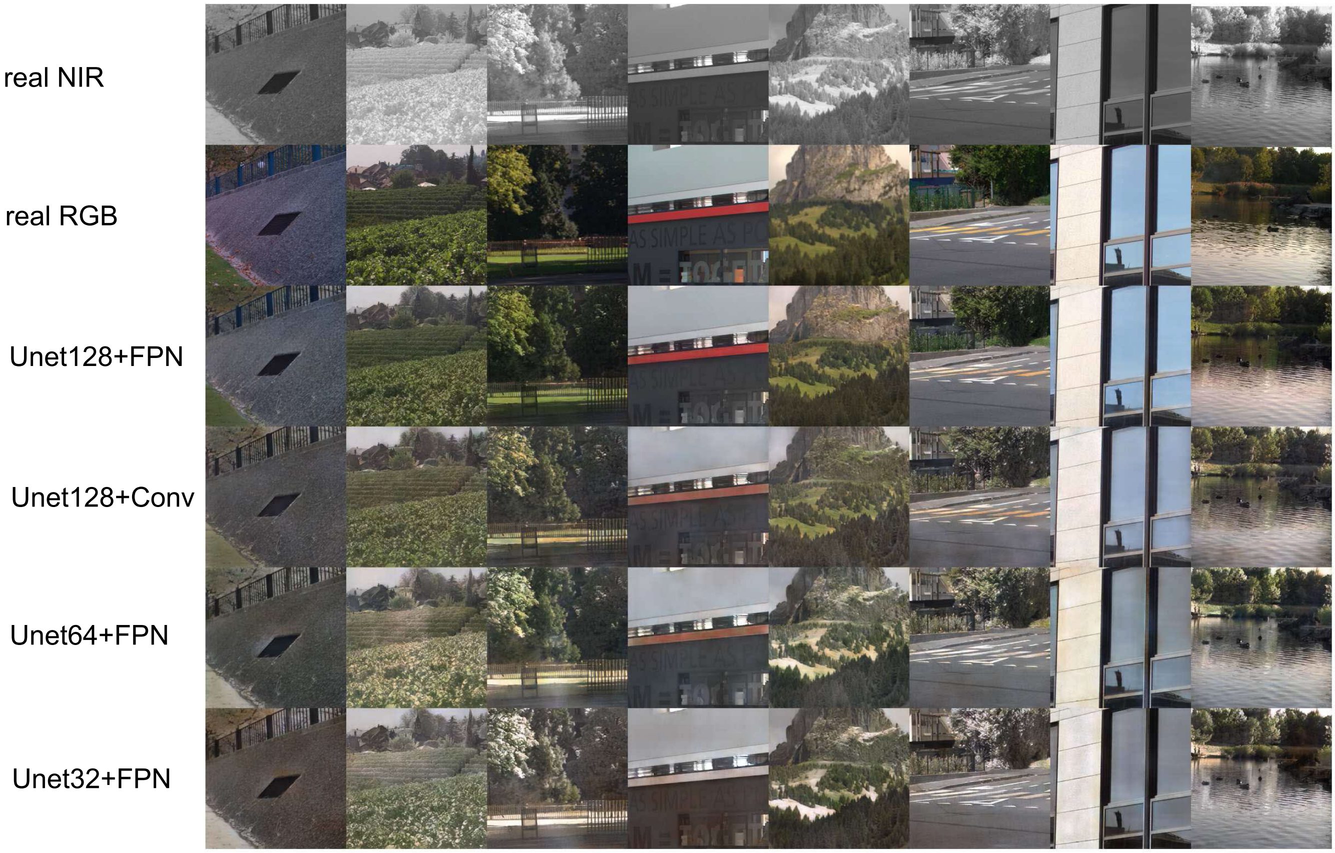

FIGURE 10. Performance comparison under different receptive field sizes.

in Table 1. The proposed model is trained in an unregistered dataset. Small translative distortions often appear in fake RGB images. The human eyes are not very sensitive to the translative distortion in NIR colorization. Since the MSE is sensitive to the translative distortion, we do not report the MSE evaluations. Table 2 shows performance comparisons in terms of the training time per epoch and stability to unregisteration.

2) SCENE CATEGORY UNAWARE

RGB images of the category unaware dataset have larger ambiguity than those of the category aware dataset. Artificial object scenes are more difficult in the NIR colorization than natural scenes due to their ambiguity. Fig. 6 shows a visual comparison among the proposed method, Cycle GAN [14], and UNIT [34]. The main difference between Cycle GAN and the proposed method is the asymmetric strategy. Since conventional quantitative measurements are not very effective in the coarsely registered and unregistered data, we provide more NIR images and their colorization results in Figs. 7 and 8.

B. ABLATION STUDIES

We conduct three ablation studies to analyze the contribution of each component. The ablations are carried on the same dataset as the main experiments. The training dataset is coarsely registered by [42], i.e. the pixelwise registration is not guaranteed. All quantitative measurement are performed on the manually registered subset whose pixels are precisely registered.

1) COMPARISON WITH Pix2Pix

We compare the visual performance with Pix2Pix [17] on the mountain subset. Since the mountain subset is not well registered compared with other categories, the comparison verifies the effectiveness of a cyclic loss in solving the pixelwise unregistration problem. The proposed asymmetric cycle GAN jointly estimate the image translation by G_{AB} and G_{BA}, but Pix2Pix has only unilateral projection. For a fair comparison, we make $G_{pix2pix}$ the same network depth as $G_{CycleAB} + G_{CycleBA}$ by additional ResNet blocks. As mentioned before, the quantitative measurements are not available in the unregistered case. The results are shown in Fig. 9. The results show that Pix2Pix is confused by unregistration especially in sharp regions. However, our asymmetric cycle GAN produces natural looking results in this case.

2) FPN vs CONVOLUTION

We use FPN in the RGB domain discriminator. To prove the contribution of FPN, we perform both visual and quantitative comparison between FPN discriminator and fully convolutional discriminator without a multiscale solution. The quantitative comparison is found in Table 3 (see the first and second rows). As a result, FPN discriminator achieves 2.6 degree gain in average AE and 0.9 gain in average SSIM over fully convolutional discriminator. The visual comparison is demonstrated in Fig. 10. Note that all networks are trained on the scene category unaware dataset [42].

3) RECEPTIVE FIELD SIZE

It is intuitively supposed that a larger receptive field captures richer context information. We expect that the NIR colorization can benefit from a large receptive field. To this end, we narrow the receptive field size by 128→64→32 (a larger receptive field than 128 × 128 is not supported by the current amount of data), and compare their performances on the scene category unaware dataset. The visual comparison is shown in Fig. 10. The experimental results shows that the colorization performance declines as the receptive field size decreases. Moreover, the scene category unaware NIR colorization is difficult by a 32 × 32 receptive field.

IV. CONCLUSION

In this paper, we have proposed a novel asymmetric cycle GAN for the NIR to RGB domain translation. Since the RGB image with three channels has richer information than the NIR image with one channel, we have built cycle GAN in an asymmetric manner. Thus, the proposed asymmetric cycle GAN is very robust to the data unregistration caused by luminance difference between RGB and NIR. Experimental results demonstrate that the proposed method achieves competitive performance to state-of-the-art methods in both category aware and unaware datasets. Moreover, the proposed method has a good generalization by adapting to the scene category unaware case. Thus, it can be enhanced by increasing a larger receptive field and the amount of data. FPN also contributes to the improvement of the NIR colorization performance by effectively considering spatial context information.

ACKNOWLEDGMENT

The authors would like to thank Dr. Chen Su and Dr. Hubert Liu in Huawei Technologies for their helpful advice and discussion.

REFERENCES

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

[2] N. M. Uzunov, M. Bello, G. Moschini, P. Rossi, A. Rosato, M. B. Rondina, I. M. Montagner, D. Boldrin, and P. C. Muzzio, "Spatial resolution performance and object detection improvement with a multiple-wavelength NIR light transmission scanner," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Oct. 2010, pp. 2587–2590.

[3] S. M. Mavadati, M. T. Sadeghi, and J. Kittler, "Fusion of visible and synthesised near infrared information for face authentication," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 3801–3804.

[4] H. Honda, R. Timofte, and L. Van Gool, "Make my day—High-fidelity color denoising with near-infrared," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 82–90.

[5] M. Oliveira, A. D. Sappa, and V. Santos, "Unsupervised local color correction for coarsely registered images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 201–208.

[6] M. Oliveira, A. D. Sappa, and V. Santos, "A probabilistic approach for color correction in image mosaicking applications," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 508–523, Feb. 2015.

[7] H. Su and C. Jung, "Multi-spectral fusion and denoising of RGB and NIR images using multi-scale wavelet analysis," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1779–1784.

[8] M. Awad, A. Elliethy, and H. A. Aly, "A real-time FPGA implementation of visible/near infrared fusion based image enhancement," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3968–3972.

[9] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[10] R. Wu, D. Yu, J. Liu, H. Wu, W. Chen, and Q. Gu, "An improved fusion method for infrared and low-light level visible image," in *Proc. 14th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2017, pp. 147–151.

[11] A. Elliethy and H. A. Aly, "Fast near infrared fusion-based adaptive enhancement of visible images," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 156–160.

[12] B. Ahn, T. Bae, and I. Kweon, "Haze removal using visible and infrared image fusion," in *Proc. 8th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Nov. 2011, pp. 813–814.

[13] C.-H. Son and X.-P. Zhang, "Near-infrared fusion via color regularization for haze and color distortion removals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3111–3126, Nov. 2018.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.

[15] A. Deshpande, J. Lu, M. Yeh, M. J. Chong, and D. Forsyth, "Learning diverse image colorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2877–2885.

[16] S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, and K. Murphy, "PixColor: Pixel recursive colorization," *CoRR*, vol. abs/1705.07208, pp. 1–17, May 2017.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*. [Online]. Available: https://arxiv.org/abs/1611.07004

[18] A. Deshpande, J. Lu, M.-C. Yeh, M. J. Chong, and D. Forsyth, "Learning diverse image colorization," *CoRR*, vol. abs/1612.01958, pp. 1–9, Dec. 2016.

[19] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, p. 110, 2016.

[20] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, 2004.

[21] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1120–1129, May 2006.

[22] B. Sheng, H. Sun, S. Chen, X. Liu, and E. Wu, "Colorization using the rotation-invariant feature space," *IEEE Comput. Graph. Appl.*, vol. 31, no. 2, pp. 24–35, Mar./Apr. 2011.

[23] A. Deshpande, J. Rock, and D. Forsyth, "Learning large-scale automatic image colorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 567–575.

[24] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning representations for automatic colorization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 577 –593.

[25] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 649–666.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Jun. 2014, *arXiv:1406.2661*. [Online]. Available: https://arxiv.org/abs/1406.2661

[27] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, pp. 1–7, Nov. 2014.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, pp. 1–8, May 2015.

[29] 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.

[30] 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.

[31] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Infrared image colorization based on a triplet DCGAN architecture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 212–217.

[32] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR*, vol. abs/1511.06434, pp. 1–16, Nov. 2015.

[33] P. L. Suárez, A. D. Sappa, B. X. Vintimilla, and R. I. Hammoud, "Near infrared imagery colorization," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 2237–2241.

[34] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *CoRR*, vol. abs/1703.00848, pp. 1–9, Mar. 2017.

[35] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," *CoRR*, vol. abs/1606.07536, pp. 1–9, Jun. 2016.

[36] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 465–476.

[37] H.-Y. Lee, H.-Y. Tseng, Q. Mao, J.-B. Huang, Y.-D. Lu, M. Singh, and M.-H. Yang, "DRIT++: Diverse image-to-image translation via disentangled representations," *CoRR*, vol. abs/1905.01270, pp. 1–14, May 2019.

[38] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *CoRR*, vol. abs/1511.07122, pp. 1–13, Nov. 2015.

[39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[40] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 936–944.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, pp. 1–11, Feb. 2015.

[42] M. Brown and S. Susstrunk, "Multi-spectral SIFT for scene category recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 177–184.

[43] 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.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] A. Gijsenij, T. Gevers, and P. M. Lucassen, "A perceptual comparison of distance measures for color constancy algorithms," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer-Verlag, 2008, pp. 208–221.

TIAN SUN received the B.S. and M.S. degrees in electronic engineering from Xidian University, China, in 2009 and 2012, respectively, where he is currently pursuing the Ph.D. degree. His research interests include image processing, multimedia security, 3D imaging, and machine learning.

CHEOLKON JUNG (M'08) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Sungkyunkwan University, South Korea, in 1995, 1997, and 2002, respectively. He was with the Samsung Advanced Institute of Technology (Samsung Electronics), South Korea, as a Research Staff Member, from 2002 to 2007. He was a Research Professor with the School of Information and Communication Engineering with Sungkyunkwan University, from 2007 to 2009. Since 2009, he has been with the School of Electronic Engineering, Xidian University, China, where he is currently a Full Professor and the Director of the Xidian Media Lab. His main research interests include image and video processing, computer vision, pattern recognition, machine learning, computational photography, video coding, virtual reality, information fusion, multimedia content analysis and management, and 3DTV.

QINGTAO FU received the B.S. degree in telecommunication engineering and the M.S. degree in information and communication engineering from Xidian University, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree. His research interests include image processing and video coding.

QIHUI HAN received the B.S. degree in automation engineering from Henan Polytechnic University, in 2013, and the M.S. degree in electronic engineering from Xidian University, China, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include image processing, computational photography, virtual reality, and deep learning.