Colorizing Near Infrared Images through a Cyclic Adversarial Approach of Unpaired Samples

Armin Mehri[1]
[1]Computer Vision Center
Campus UAB,
Bellaterra, Barcelona, Spain
amehri@cvc.uab.es

Angel D. Sappa[1,2]
[2]Escuela Superior Politécnica del Litoral, ESPOL,
FIEC, CIDIS, Campus Gustavo Galindo,
Guayaquil, Ecuador
sappa@ieee.org

Abstract

This paper presents a novel approach for colorizing near infrared (NIR) images. The approach is based on image-to-image translation using a Cycle-Consistent adversarial network for learning the color channels on unpaired dataset. This architecture is able to handle unpaired datasets. The approach uses as generators tailored networks that require less computation times, converge faster and generate high quality samples. The obtained results have been quantitatively—using standard evaluation metrics—and qualitatively evaluated showing considerable improvements with respect to the state of the art.

1. Introduction

In recent years, image acquisition devices have expanded significantly due to the increase in computational power and the reduction in electronics prices. Improving the sensor technology has lead to a large family of cameras capable of capturing information from various spectral bands or additional information (3D, 4D); hence nowadays we can have: panoramic 3D images; multispectral images; HD 2D images; video sequences at a high frame rate; and many others. Regardless of the large number of possibilities, the fact that the human visual system is sensitive to (400-700 nm) the classical RGB representation is preferred if the information needs to be provided to the final user. Therefore, for better user understanding, representing the information in the range of 400-700nm is preferred [24]. Out of this spectral range the NIR band is one of the most widely used band.

The NIR spectral band is the nearest band to the human eye perception system, so NIR images share various attributes with visible spectrum images. The use of NIR images is concerned with their ability to segment images according to the material of the object. For example, most coloring matter utilized for colorization of materials are slightly transparent to NIR. In other word, the distinction within the NIR intensities is not solely due to specific color of the material, however conjointly to the absorption and coefficient of reflection of the materials of a given object. The aforementioned attributes (absorption and reflectance) are interesting for applications such as video surveillance, detection and remote sensing for crop stress. In these two contexts (i.e., video surveillance and remote sensing), it is quite troublesome to orient once near infrared images are provided to the final users, as color discrimination is lacking or incorrect color deployment. Hence, obtaining realistic RGB image representations from NIR images is a needed in most of these applications.

NIR image colorization shares some similarities with those approaches proposed in the literature for gray scale image colorization or color transfer functions (e.g., [20], [5], [29]). In spite of the similarity with these approaches, due to the nature of NIR images, their colorization is more challenging. In recent years several approaches for NIR image colorization have been proposed (e.g., [24], [16], [25]). Most of them are learning based approaches where couple of registered NIR and RGB images are provided during the training stage. The limitation with all these approaches is related with the need of these couple of paired images (NIR-RGB). In general, although there are Single Sensor Cameras (e.g., [23]) where RGB and NIR information is acquired at the same time, NIR images are taken by one camera while the corresponding RGB image by another camera. This means that there are shifts between the acquired images, or in some cases even worse since just the NIR images are provided.

In the last few years, Generative Adversarial Networks (GANs) have drawn attention in many field of computer vision to help researchers to build powerful models, where there were difficulty by using only simple Convolutional Neural Networks (CNNs). Nevertheless, most of the GAN techniques [13] have focused on supervised context. The

unpaired (NIR-RGB) problem mentioned above, can be tackled by a GAN architecture in the unsupervised context under a cyclic structure (CycleGAN) [31]. CycleGAN learns to map images from one domain (source domain) onto another domain (target domain) when paired images are unavailable. This functionality makes models appropriate for image to image translation/colorization in the context of unsupervised learning. In the current paper a novel CycleGAN architecture is proposed for colorizing unpaired NIR-RGB images; the main contributions of our proposed model, compared with baseline [31], are as follows: i) it utilizes tailored generators, which can work better in colorization context and have less computation time and less parameters size; ii) it converges faster than the baseline approach [31]; finally, iii) it produces higher quality images.

This paper is organized as follow. In section 2, a summary of related works is given. In section 3, the proposed approach is presented in details. Then, experimental results are depicted in section 4 and finally conclusions are provided in section 5.

2. Related Work

Colorization problem has been studied during last decades, several techniques have been proposed to unravel this difficult task. Some of the methods proposed in the literature follow a semi-automatic approach, which means they need user interactions or to employ some user-defined search table. Other approaches, mainly learning based approaches, are based on having aligned image pairs (NIR-RGB), which in most of the cases are not available. The issues mentioned within the current paper are expounded with infrared image colorization, as mentioned above, it somehow shares some common issues with monochromatic approaches to image colorization. Monochromatic image colorization algorithms vary in the ways they obtain and process data for modeling between gray-scale and RGB images.

Colorization approaches can be usually classified into two groups: parametric and non-parametric. At the training time, parametric techniques try to learn predictive functions from large color image datasets, posing the problem either as a classification of quantified color values or as a regression to continuous color space. On the other hand, in the non-parametric techniques a gray-scale image is provided as an input and then one or more color provided as source images by user or automatically; then color from reference images transferred statistics onto homogeneous regions of the input image, such as Welsh et al. [27], Gupta et al. [9], Irony et al. [12]. All the papers mentioned before are example-based approaches, which works as semi-automatic methods to transfer color statistics from reference images onto input gray-scale images. Although good results are obtained, there is a big drawback with all these techniques that is related with the requirement that input and reference images should share the same content, actually both of them should be perfectly registered, which is not the case in most of the real scenarios.

GAN networks are a kind of Convolutional Neural Network (CNN) that are able to generate samples from a given latent space, this network has been introduced by Goodfellow et al. [7]. The mentioned GAN architecture build up by a series of linear layers (fully-connected layers) and so insufficient to complex dataset. The model consists of two networks a Discriminator (D) and a Generator (G), which going to against each other. In other words, the discriminator try to distinguish the real samples from fake samples that have been generated by the generator. On the other hand, the generator job is to fool the discriminator with the generated samples (fake images) to be classified as real images. Both networks, D and G, are simultaneously optimized. As mentioned above, the main issue of the standard GAN was limited to simple datasets so shortly afterwards researcher proposed the new architecture for GAN to address this limitation, DCGAN (Deep Convolutional Generative Adversarial Network) [21] has been introduced by Radford et al. and changed the standard for most of GAN architectures. DCGAN architecture became one of the most popular and successful network design for GAN. In DCGAN instead of using series of linear layers that is only suitable for simple datasets, convolution layers without max pooling or fully connected layers are considered and furthermore convolutional stride for the down-sampling and transposed convolution for the up-sampling are used that made DCGAN architecture appropriate for complex dataset. DCGAN standard has been applied in various computer vision problems such as image colorization [24], image enhancement [15], style transfer [6], data augmentation [1] and many others.

Previous approaches are useful when paired images are provided for the training process. In the case of unpaired images, architectures such as CycleGAN [31] or Dual-GAN [28] have been proposed by learning mapping between different visual domains jointly, each as a separate generative adversarial network. Via a cycle-consistency loss ensures that applying each mapping followed by its reverse yields the identity map (i.e., *"if we translate from one domain to another and back again we must arrive where we started"*).

Regardless of the used architecture, Generative Adversarial Networks usually suffer from multiple challenges during training that needs more attention than Convolutional Neural Networks (CNNs). Such as mode collapse, convergence properties, diminished gradient and highly sensitive to the hyper-parameter selections. Arjovksy et al. [2] illustrated that the discriminator in standard GAN cannot be trained well or with a high learning rate; otherwise gradient vanish may show off and generator not able to generate samples anymore and learning will stop. They

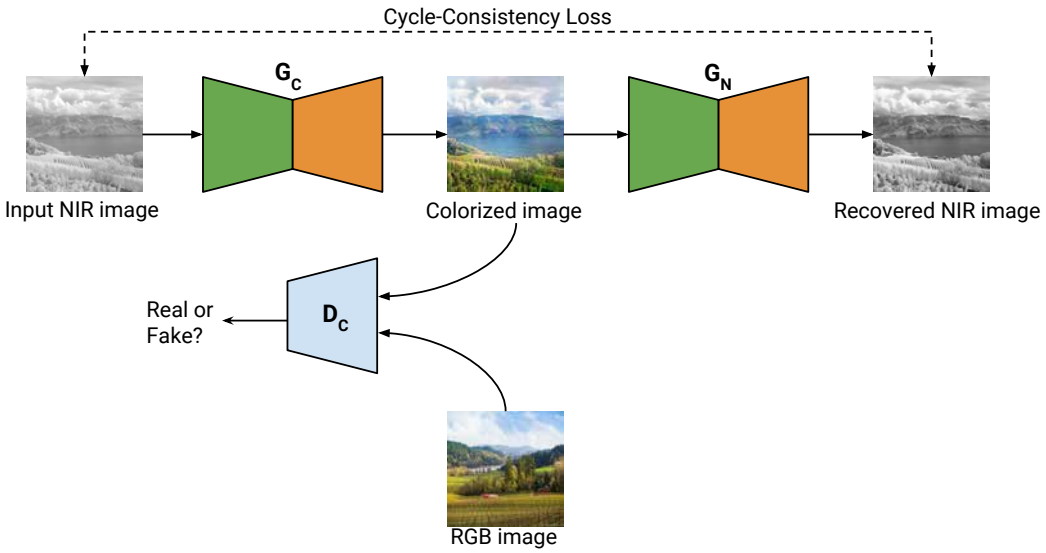

Figure 1. Illustration of the CycleGAN architecture used for NIR image colorization with unpaired NIR-RGB datasets.

also proved that the standard GAN loss function cannot accurately deal with inappropriate distributions, for example those with disjoint supports, often found during training stage of GAN. To solve the mentioned challenges many different GANs have been proposed by using vary loss functions during training or using different D during the learning process such as LSGAN [18], WGAN [3] and many others. Even though the proposed approaches have been relatively successful solving these challenges (training stability, data quality, etc.), Lucic et al. [17]'s large - scale research suggests that such approaches are not improving standard GAN consistently. In addition, some of the best proposed approaches, like WGAN-GP [8], requires far more computational comparing with standard GAN. Alexi [14] illustrated that a relativistic discriminator based on integral probability metrics (IPM), is essential to make GANs similar to divergence minimization and generate reasonable forecasts on the basis of a previous knowledge. In such discriminator, half of the images in the mini-batch consider as fake. The proposed approach prove that GANs are able to generate higher quality samples, less computational and more stable than the previous approaches.

3. Proposed Approach

This section explains in details the approach proposed for colorizing NIR images. As mentioned in the section 2, most of recent work on colorization have proposed the usage of a deep convolutional generative adversarial network on aligned paires of images, which in most of the cases do not represent a real scenario. In the current work the usage of a CycleGAN to colorize NIR images to a RGB representation is proposed (see Fig. 1), when an aligned paired dataset does not exist. In order to handle inputs and outputs in both generators, a model that is feed with three channels is proposed. This model will receive as an input three channels, which could correspond to: i) a given NIR image three times (this is in the G_C case); or ii) a RGB image (this is the G_N case). A loss function different to the one proposed in [31] is used to minimize the overall classification error in the training process, which improves the generalization capability of the model.

The proposed architecture built up by a series of convolutional and transposed convolutional layers; relu and leaky relu as non-linear activation functions; for generators and discriminators respectively. Moreover, every layer of D uses the spectral normalization and instance normalization in G. Also, it is worth to mention pooling layer have not been used in the networks, instead strided convolutions used in order to keep as much features as possible, since a pooling layer is down-sampling the feature depends on stride number, which leads to data in features map to loss. Dropout layers are used in the terms of noise to few layers of generators in order to prevent overfitting and modal collapse. Added noise in few layers of G leads network to generate the necessary variability of the training set, to be able to generalize the learning of the colorization process.

Both networks (G, D) are based on feed-forward deep neural networks, which play a min-max game against each other. The near infrared image given to G as an input data with the image size 256×256 pixels, and networks try to transforms the given sample (NIR image) onto the interested form of the data we concerned, a RGB representation.

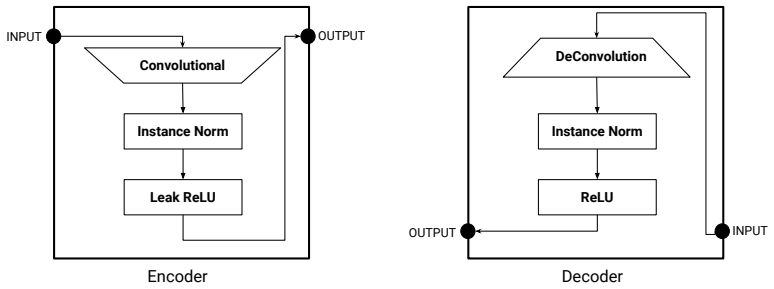

Figure 2. Illustration of the Encoder and Decoder structures of the proposed approach.

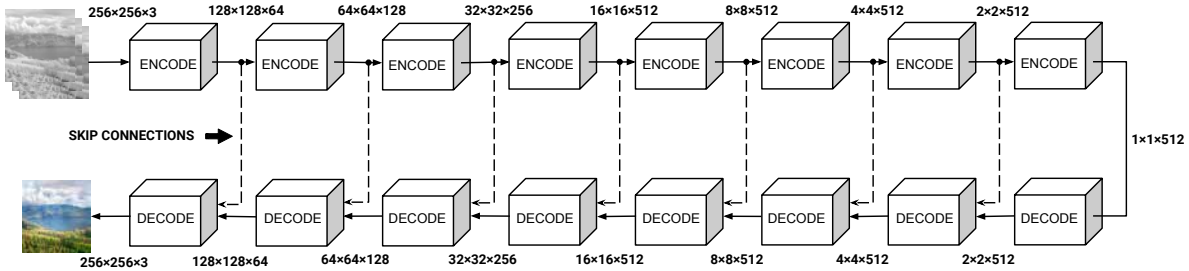

Figure 3. Illustration of the structure of generators for NIR image colorization.

On the other hand, D takes a set of data, either a real sample or a produced sample, and produces a probability of that data being real. The network D is optimized in order to increase the likelihood of giving a high probability to the real data and a low probability to the generated data (i.e., if the probability is near to 1 it means the NIR image is correctly colored, while if probability is near to 0 it means that NIR image is wrongly colored).

3.1. U-net as Generators

ResNet architecture [10] has been used as a part of the generators in the CycleGAN [31], showing that it is quite powerful in transfiguring one image to another image and has been achieved the reasonable results in style transfer, photo enhancement, season transfer and other several applications. Unfortunately, it could not achieve acceptable results in learning the color between domains and transferring the learned color without affecting the samples' shapes, since the network after some number of epochs starts to transfigure between domains; hence the net will stop to learn the color and also networks need plenty of data in RGB domain. The first contribution of current work with respect to baseline model [31] is to select the generators, which are able to generate better samples and learn accurately the colors of the different objects in images and works better when not enough data are available in both domains. The architecture proposed in this paper (U-net [22]) also leads to have less computational time in training process. U-net based architecture [22], proposed as generators of the models, has showed efficiency on a wide range of approaches, especially in the colorization problems (e.g., [13],[30]).

The Unet architecture build up based on three components: 1) encoder where the input passes through a series of down-sampling layers (i.e., convolutional layers to extract the feature samples); 2) bottleneck layer, which helps the model to share all information pass through all the layers so low-level information will be available directly among the net. To give the generator a means to circumvent the bottleneck for information like this, in the pix2pix model [13] the skip connections, following the general shape of a U-Net, which simply concatenates all channels at layer i with those at layer $n - i$; 3) decoder, the last component of U-net, which do the reverse process of the encoder (i.e., back to the normal image from the extracted feature by pass through the series of transposed convolutional layers (upsampling)). Also, U-net shows that is quite powerful in the case of understanding the color from one domain and transferring onto another domain (NIR images). U-net architecture with skip connections is illustrated in Fig. 3.

3.2. Loss Functions

In the proposed model a multi-term loss function ($\mathcal{L}_{final}$) has been used by combination of RaLSGAN loss, Cycle Consistency loss, Structural Similarity loss (SSIM) and Identity loss. The combination of these loss functions leads to achieve better image quality for human perceptual criteria as presented in the experimental result section.

The RaLSGAN loss function [14] is applied to both generators G_C, G_N and their discriminators D_C, D_N of the

model respectively:

$$\mathcal{L}^{G_i}_{RaLSGAN} = \mathbb{E}_{x_f \sim \mathbb{P}}[(C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}} C(x_r) - 1)^2] + \mathbb{E}_{x_r \sim \mathbb{P}}[(C(x_r) - \mathbb{E}_{x_f \sim \mathbb{Q}} C(x_f) + 1)^2] \quad (1)$$

$$\mathcal{L}^{D_i}_{RaLSGAN} = \mathbb{E}_{x_r \sim \mathbb{P}}[(C(x_r) - \mathbb{E}_{x_f \sim \mathbb{Q}} C(x_f) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{Q}}[(C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}} C(x_r) + 1)^2] \quad (2)$$

where $\mathbb{P}$ and $\mathbb{Q}$ are the distributions of real and generated data respectively; $C(x_r)$ and $C(x_f)$ are the probability of D for real and fake data.

The Cycle Consistency loss function is defined as follow:

$$\mathcal{L}_{cyc}(G_C, G_N) = \mathbb{E}_{n \sim p_{data}(n)}[||G_N(G_C(n)) - n||] + \mathbb{E}_{c \sim p_{data}(c)}[||G_C(G_N(c)) - c||] \quad (3)$$

where n and c correspond to domain images (n for NIR images and c for color images).

The Structural Similarity Index (SSIM) [26] has been used during training process, where the aim of using such loss function is to help the learning model to generate a visually improved image. The structural loss function defined as below:

$$\mathcal{L}_{SSIM} = \frac{1}{NM} \sum_{p=1}^{P} 1 - SSIM(p). \quad (4)$$

The Identity loss function employed to regularize the generator. The aim of using such loss function is if something already looks like from the target domain, should not map it into a different image.

$$\mathcal{L}_{identity}(G_C, G_N) = \mathbb{E}_{c \sim P_{data}(c)}[||G_C(c) - c||] + \mathbb{E}_{n \sim P_{data}(n)}[||G_N(n) - n||]. \quad (5)$$

The final objective $\mathcal{L}_{final}$ is obtained as below:

$$\mathcal{L}_{final} = \mathcal{L}_{RaLSGAN} + \lambda \mathcal{L}_{Cycle} + \mathcal{L}_{SSIM} + \gamma \mathcal{L}_{Identity} \quad (6)$$

where λ, γ are the weights to Cycle Consistency and Identity loss function, which play as regularization terms impacting on the optimization of the model. Assigning a bigger weights lead the model to have better reconstruction loss and model will make smaller changes. On the other hand, a smaller weights increase the risk of artifacts and lead the model to bring more dramatic changes with respect to input images.

3.3. Spectral Normalization

The performance control of the discriminator is an ongoing challenge in training Generative Adversarial Networks. The density ratio estimates by the discriminator in high-dimensional spaces is often imprecise and unstable during learning phase, so generators do not learn the multimodel structure of the target. To solve the mentioned issue [19] proposed the normalization method. It helps to stabilize the training of discriminators by applying spectral normalization. Hence, discriminator becomes more stable and the network converge faster in less number of epochs. In the current work spectral normalization has been used so that the network learns the structure of images much better and generates better image quality comparing to baseline model [31].

3.4. Better Cycle Consistency

Cycle Consistency loss function is one of the main features of CycleGAN, which simply motivates generators to prevent needless changes and generates images that share structural similarity with inputs. Also, the Cycle Consistency helps a lot to make training phase stable in the early stages, but becomes a problem in later stages to generate realistic images. Since Cycle Consistency is a form of regularization we propose to progressively decrease the weight of cycle loss after half way of training process. Nevertheless, λ (in eq. (6)) needs to be checked to not become 0 in order to prevent the generators become unstable and unconstrained.

3.5. Two Time-Scale Update Rule

Training GANs, unlike of CNNs, needs more attention since mode collapse may occur in the learning process, when the generator generates a restricted variety of samples, or even the same sample, regardless of the input and prevent GAN to learn the target distribution. In [11] the authors propose the two time-scale update rule (TTUR), which improves the general performance, convergence speed and helps to prevent the mode collapse of GANs. TTUR has been applied to the proposed approach with ADAM stochastic optimizer to risk reduction of mode collaps and also to make sure that the discriminators converge in the training process.

4. Experimental Results

The proposed approach has been evaluated by using NIR images and their corresponding RGB, which were used as ground truth. The data set has been obtained from [4]; it contains pairs of NIR-RGB images of 1024×680 pixels each from different categories. It should be mentioned that dataset images are correctly registered and a pixel-to-pixel correspondence is guaranteed for quantitative and qualita-

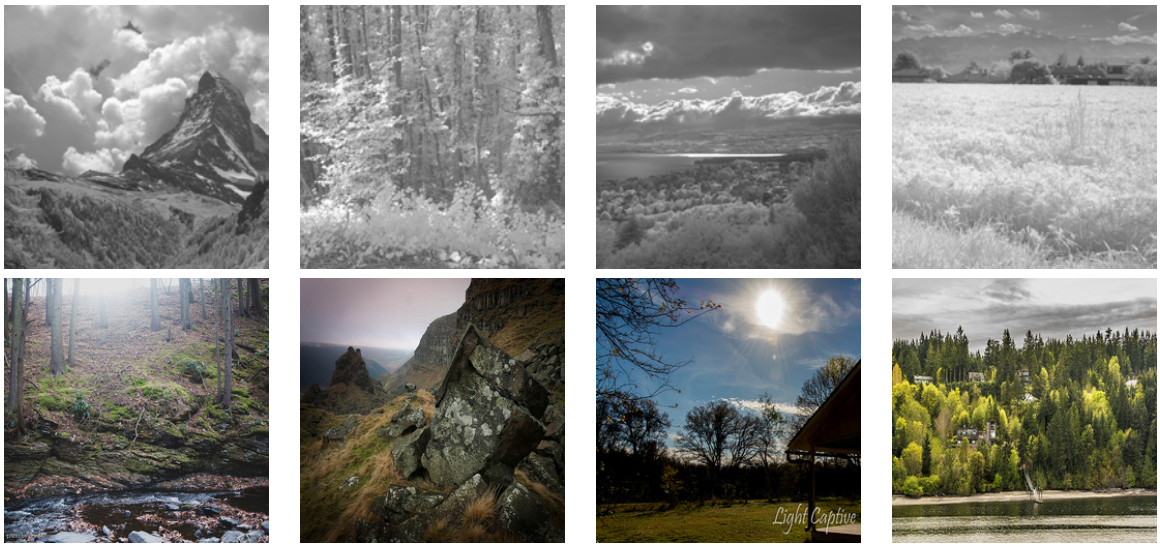

Figure 4. Unpaired set of images (256×256 pixels each) used for training the proposed approach and CycleGAN [31]; ($top - row$) NIR images from [4]; ($bottom - row$) RGB images collected from internet.

tive evaluation. Only categories with similar scenarios have been chosen for training the proposed model. The selected categories are as follow: country (50 pairs of images), field (51 pairs of images), forest (52 pairs of images) and mountain (50 pairs of images). The objective is to train the network in scenarios that contain similar objects.

The NIR images from the mentioned categories have been used during training while the corresponding RGB images (ground truth) have not been used neither during the training nor testing phases; they are only used for quantitative and qualitative evaluations. The RGB images used during the training process have been collected from internet (700 images); all the collected images correspond to scenarios similar to those from the aforementioned categories. Each pair of the original NIR and RGB images (from [4]) has been split up into two smaller images of 680×680 pixels each, resulting in a total of 406 pairs. From this set 68 pairs of images have been randomly selected and keep aside for evaluating the performance of the proposed approach. The rest of NIR images have been resized to 256×256 pixels, which was the size used to feed the network. All the RGB images collected from internet have been also resized up to 256×256 pixels each. In order to increase the number of images for training a data augmentation process has been applied (horizontal flipping and random crop). Figure 4 shows just four pairs of these unpaired (NIR-RGB) images used for training.

Results obtained with the proposed approach have been compared with results obtained using the baseline model presented in [31]. Quantitative and qualitative results from this NIR image colorization are presented in next sub sections. The proposed network has been trained using a 3.2 eight core processor with 62GB of memory with a NVIDIA GeForce GTX TITAN X GPU; on average the training process took near to 13 hours to complete 200 epochs. The model has been trained by using ADAM stochastic optimizer due to several advantages, slight memory requirements, it is computationally effective, also leads network to converge faster compared with the other stochastic optimizer and it prevents from overfitting. Dataset has been normalized from range of (0, 255) to (-1, 1); normal weights initialized with mean 0 and standard derivation 0.2 used in the proposed approach. The hyper-parameters were tuned during training stage as follows: learning rate 0.0003 and 0.0009 for generators and discriminators respectively; weight decay 1e-8 for generators, exponential decay rate 0.50, 0.999 for the first and second momentum (beta1, beta2); leak relu 0.2; cycle consistency weight 100; dropout with 0.5 probability.

4.1. Evaluation Metrics

In order to assess the performance of the proposed approach average Angular Error (AE) is considered. It is a widely used evaluation measure in color constancy research. AE is defined as the average angular distance between every obtained RGB pixel ($RGBo_{i,j}$) with the corresponding ground truth ($RGBg_{i,j}$). AE is used as an evaluation metrics since this measure is quite similar to the human spectator. AE is defined as:

$$AE = cos^{-1}\left(\frac{dot(RGB_o, RGB_g)}{norm(RGB_o) \times norm(RGB_g)}\right). \quad (7)$$

Additionally, Frchet Inception Distance (FID) [11] has

	AE	FID
CycleGAN [31]	13.87	146.77
Prop. App.	**10.04**	**105.21**

Table 1. Comparative results between proposed approach and CycleGAN using evaluation metrics from Sec. 4.1.

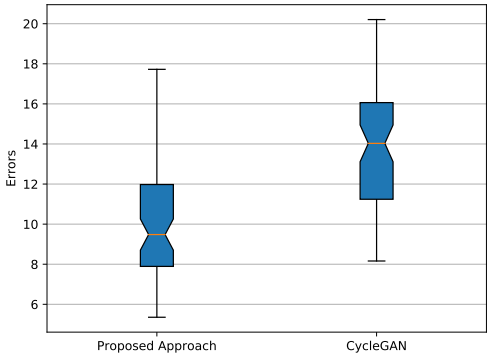

Figure 5. Avarage AE distribution for both approaches.

been used for comparing the similarity between obtained images and ground truth images in an embedded space. The FID is computed by using the Inception model up to a specific layer. Hence, in the case of two sets of multivariate Gaussians the FID between two distributions is obtained by calculating their means and covariances:

$$FID(X_g, X_o) = ||\mu_g - \mu_o||_2^2 + Tr(\Sigma_g + \Sigma_o - 2(\Sigma_g\Sigma_o)^{\frac{1}{2}}) \quad (8)$$

where the X_g is the set of real images (ground truth) and X_o is the set of obtained images. Lower FID results show the obtained images are more similar to ground truth images.

4.2. Quantitative Results

Table 1 presents the quantitative results based on average AE and FID with the scenarios mentioned above (set of 68 pairs used as ground truth) for both models (proposed approach and baseline model [31]). According to the obtained average AE, the proposed approach improves CycleGAN in about 40%. In the case of FID metrics, the proposed approach gets an improvement of almost 39% with respect to CycleGAN. It can been seen that the proposed approach has smaller errors than the baseline model [31]. These results show that Relativistic loss and SSIM loss functions help to enhance the performance of the original CycleGAN [31]. Figure 5 depicts the box plot for the average AE of both approaches.

4.3. Qualitative Results

Figure 6 depicts some illustrative results for comparisons, both with respect to CycleGAN [31] and the corresponding ground truth. These images correspond to the set of 68 pairs of images mentioned above that have not been used neither during the training nor during the validation stages. Each column shows the given NIR images, colorized with baseline model [31], colorized with the proposed approach and the ground truth respectively. It should be mentioned that all categories are trained simultaneously and also our colorized NIR images look quite better than the baseline model when compared with the ground truth.[1]

5. Conclusions

This paper proposes a novel architecture by using a Cycle-Consistent Adversarial Network in the context of colorization. The proposed approach address the challenging problem of colorizing NIR images when the ground truth is not available during the learning phase (i.e., in the unsupervised learning context) by using the appropriate generators and loss functions. Experimental results have shown that the NIR images colorized with proposed approach are visually better than those obtained with the CycleGAN baseline model as well as lower quantitative values are obtained.

Acknowledgment

This work has been partially supported by the Spanish Government under Project TIN2017-89723-P, the "CERCA Programme / Generalitat de Catalunya" and the ESPOL project PRAIM (FIEC-09-2015). The authors gratefully acknowledge the support of the CYTED Network: "Ibero-American Thematic Network on ICT Applications for Smart Cities" (REF-518RT0559) and the NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] A. Antoniou, A. Storkey, and H. Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.

[2] M. Arjovsky and L. Bottou. Towards principled methods for training generative adversarial networks. arxiv. 2017.

[3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.

[4] M. Brown and S. Süsstrunk. Multi-spectral sift for scene category recognition. In *CVPR 2011*, pages 177–184. IEEE, 2011.

[5] Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015.

[1] Additional results are provided at http://bit.ly/2VQG4B0

Figure 6. Colorized NIR images obtained with CycleGAN and with the proposed approach. RGB images (ground truth) are provided for qualitative evaluation. The numbers below the images show the average AE between the obtained colorized image and the ground truth.

[6] M. Chidambaram and Y. Qi. Style transfer generative adversarial networks: Learning to play chess differently. *arXiv preprint arXiv:1702.06762*, 2017.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017.

[9] R. K. Gupta, A. Y.-S. Chia, D. Rajan, E. S. Ng, and H. Zhiyong. Image colorization using similar images. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 369–378. ACM, 2012.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

[12] R. Ironi, D. Cohen-Or, and D. Lischinski. Colorization by example. In *Rendering Techniques*, pages 201–210. Citeseer, 2005.

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[14] A. Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018.

[15] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[16] M. Limmer and H. P. Lensch. Infrared colorization using deep convolutional neural networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 61–68. IEEE, 2016.

[17] M. Lučić, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet. Are gans created equal? a large-scale study. 2017.

[18] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[19] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

[20] M. Oliveira, A. D. Sappa, and V. Santos. A probabilistic approach for color correction in image mosaicking applications. *IEEE Transactions on image Processing*, 24(2):508–523, 2015.

[21] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

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

[23] X. Soria, A. D. Sappa, and A. Akbarinia. Multispectral single-sensor RGB-NIR imaging: New challenges and opportunities. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2017.

[24] P. L. Suárez, 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.

[25] P. L. Suárez, 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.

[26] 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.

[27] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to greyscale images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 277–280. ACM, 2002.

[28] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2849–2857, 2017.

[29] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[30] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros. Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999*, 2017.

[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.