Learning to Colorize Infrared Images

Patricia L. Suárez[1], Angel D. Sappa[1,2](✉), and Boris X. Vintimilla[1]

[1] Facultad de Ingeniería en Electricidad y Computación, CIDIS,
Escuela Superior Politécnica del Litoral, ESPOL,
Campus Gustavo Galindo, 09-01-5863 Guayaquil, Ecuador
{plsuarez,asappa,boris.vintimilla}@espol.edu.ec
[2] Computer Vision Center, Edifici O, Campus UAB,
Bellaterra, 08193 Barcelona, Spain

Abstract. This paper focuses on near infrared (NIR) image colorization by using a Generative Adversarial Network (GAN) architecture model. The proposed architecture consists of two stages. Firstly, it learns to colorize the given input, resulting in a RGB image. Then, in the second stage, a discriminative model is used to estimate the probability that the generated image came from the training dataset, rather than the image automatically generated. The proposed model starts the learning process from scratch, because our set of images is very different from the dataset used in existing pre-trained models, so transfer learning strategies cannot be used. Infrared image colorization is an important problem when human perception need to be considered, e.g., in remote sensing applications. Experimental results with a large set of real images are provided showing the validity of the proposed approach.

Keywords: CNN in multispectral imaging · Image colorization

1 Introduction

Image acquisition devices have largely expanded in recent years, mainly due to the decrease in price of electronics together with the increase in computational power. This increase in sensor technology has resulted in a large family of images, able to capture different information (from different spectral bands) or complementary information (2D, 3D, 4D); hence, we can have: HD 2D images; video sequences at a high frame rate; panoramic 3D images; multispectral images; just to mention a few. In spite of the large amount of possibilities, when the information needs to be provided to a final user, the classical RGB representation is preferred. This preference is supported by the fact that human visual perception system is sensitive to (400–700 nm); hence, representing the information in that range help user understanding. In this context, the current paper tackles the near infrared (NIR) image colorization, trying to generate realistic RGB representations. Different applications could take advantage of this contribution—infrared sensors can be incorporated for instance in driving assistance applications by

F. De la Prieta et al. (eds.), *Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection - 15th International Conference, PAAMS 2017*,
Advances in Intelligent Systems and Computing 619, DOI 10.1007/978-3-319-61578-3_16

providing realistic colored representation to the driver, while the image processing can be automatically performed by the system in the infrared domain (e.g., semantic segmentation at the material level avoiding classical problems related with the color of the object surface).

The NIR spectral band is the closest in wavelength to the radiation detectable by the human eye; hence, NIR images share several properties with visible images. The interest of using NIR images is related with their capability to segment images according to the object's material. Surface reflection in the NIR spectral band is material dependent, for instance, most pigments used for material colorization are somewhat transparent to NIR. This means that the difference in the NIR intensities is not only due to the particular color of the material, but also to the absorption and reflectance of dyes.

The absorption/reflectance properties mentioned above are used for instance in remote sensing applications for crop stress and weed/pest infestations. NIR images are also widely used in video surveillance applications since it is easier to detect different objects from a given scene. In these two contexts (i.e., remote sensing and video surveillance), it is quite difficult for users to orientate when NIR images are provided, since the lack of color discrimination or wrong color deploy. In this work a neural network based approach for NIR image colorization is proposed. Although the problem shares some particularities with image colorization (e.g., [1–3]) and color correction/transfer (e.g., [4,5]) there are some notable differences. First, in the image colorization domain—gray scale image to RGB—there are some clues, such as the fact that luminance is given by grayscale input, so only the chrominance need to be estimated. Secondly, in the case of color correction/transfer techniques, in general three channels are given as input to obtain the new representation in the new three dimensional space. In the particular problem tackled in this work (NIR to visible spectrum representation) a single channel is mapped into a three dimensional space, making it a difficult and challenging problem. The manuscript is organized as follows. Related works are presented in Sect. 2. Then, the proposed approach is detailed in Sect. 3. Experimental results with a large set of images are presented in Sect. 4. Finally, conclusions are given in Sect. 5.

2 Related Work

The problem addressed in this paper is related with infrared image colorization, as mentioned before somehow it shares some common problems with monocromatic image colorization that has been largely studied during last decades. Colorization techniques algorithms mostly differ in the ways they obtain and treat the data for modeling the correspondences between grayscale and color. Non-parametric methods, given an input grayscale image, firstly they define one or more color reference images (provided by a user or automatically retrieved) to be used as source data. Then, following the image analogy framework, color is transferred onto the input image from analogous regions of the reference image(s). Parametric methods, on the other hand, learn prediction functions from large datasets of color images at training time, posing the problem as either regression onto continuous color space or classification of quantized color values.

Welsh et al. [6] describe a semi-automatic technique for colorizing a grayscale image by transferring color from a reference color image. They examine the luminance values in the neighborhood of each pixel in the target image and transfer the color from pixels with matching neighborhoods in the reference image. This technique works well on images where differently colored regions give rise to distinct luminance clusters, or possess distinct textures. In other cases, the user must direct the search for matching pixels by specifying swatches indicating corresponding regions in the two images. It is also difficult to fine-tune the outcome selectively in problematic areas.

The approaches presented above have been implemented using classical image processing techniques. However, recently Convolutional Neural Network (CNN) based approaches are becoming the dominant paradigm in almost every computer vision task. CNNs have shown outstanding results in various and diverse computer vision tasks such as stereo vision [7], image classification [8] or even difficult problems related with cross-spectral domains [9] outperforming conventional hand-made approaches. Hence, we can find some recent image colorization approaches based on deep learning, exploiting to the maximum the capacities of this type of convolutional neural networks. As an example, we can mention the approach presented on [3]. It proposes a fully automatic approach that produces brilliant and sharpen image color. They model the unknown uncertainty of the desaturated colorization levels designing it as a classification task and use class-rebalancing at training time to augment the diversity of colors in the result. On the contrary, [10] presents a technique that combines both global priors and local image features. Based on a CNN a fusion layer merges local information, dependent on small image patches, with global priors, computed using the entire image. The model is trained in an end-to-end fashion, so this architecture can process images of any resolution. They leverage an existing large-scale scene classification database to train the model, exploiting the class labels of the dataset to more efficiently and discriminatively learn the global priors. In [11], a recent research on colorization, addressing images from the infrared spectrum, has been presented. It uses convolutional neural networks to perform an automatic integrated colorization from a single channel NIR image to RGB images. The approach is based on a deep multi-scale convolutional neural network to perform a direct estimation of the low RGB frequency values. Additionally, it requires a final step that filters the raw output of the CNN and transfers the details of the input image to the final output image.

Generative Adversarial Networks (GANs) are a class of neural networks which have gained popularity in recent years. They allow a network to learn to generate data with the same internal structure as other data. GANs are powerful and flexible tools, one of its more common applications is image generation. It is a framework presented on [12] for estimating generative models via an adversarial process, in which simultaneously two models are trained: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to 1/2 everywhere. In addition on [13] is explained Some techniques to improve the efficiency of the generative adversarial networks, one of them called, the virtual batch normalization, which allows to significantly improve the network optimization using the statistics of each set of training batches. The disadvantage is that this process is computationally expensive. Our proposal is based on designing a generative adversarial deep learning architecture that allows the colorization of images of the near infrared spectrum, so that they can be represented in the visible spectrum. The following section will explain in detail the network model used.

3 Proposed Approach

This section presents the approach proposed for NIR image colorization. A GAN network based architecture is selected due to its fast convergence capability. The network is intended to learn to generate new samples from an unknown probability distribution. In our case, in order to obtain a true color, the GAN framework is reformulated for a conditional generative image modeling tuple. In other words, the generative model $G(z;\theta_g)$ is trained from a near infrared image in order to produce a colored RGB image; additionally, a discriminative model $D(z;\theta_d)$ is trained to assign the correct label to the generated colored image, according to the provided real color image, which is used as a ground truth. Variables (θ_g) and (θ_d) represents the weighting values for the generative and discriminative networks.

The GAN network has been trained using Stochastic AdamOptimazer since it prevents overfitting and leads to convergence faster. Furthermore, it is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters. Besides, GANs provide a powerful technique for generating plausible-looking natural images with high perceptual quality. The model was trained with the following hyper-parameters: learning rate 0.0001 and 0.0002 for the generator and the discriminator networks respectively; epsilon = 1e-08; exponential decay rate for the 1st moment momentum 0.5 for discriminator and 0.4 for the generator; weight initializer with a standard deviation 0.0081; weight decay 1e-5; leak relu 0.2 and patch's size of 64×64. The convolutional architecture of the baseline model is conformed by convolutional, de-convolutional, relu, leak-relu, fully connected and activation function tanh and sigmoid for generator and discriminator networks respectively. Additionally, every layer of the model uses batch normalizaton for training any type of mapping that consists of multiple composition of affine transformation with element-wise nonlinearity and do not stuck on saturation mode. Figure 1 presents an illustration of the proposed GAN architecture.

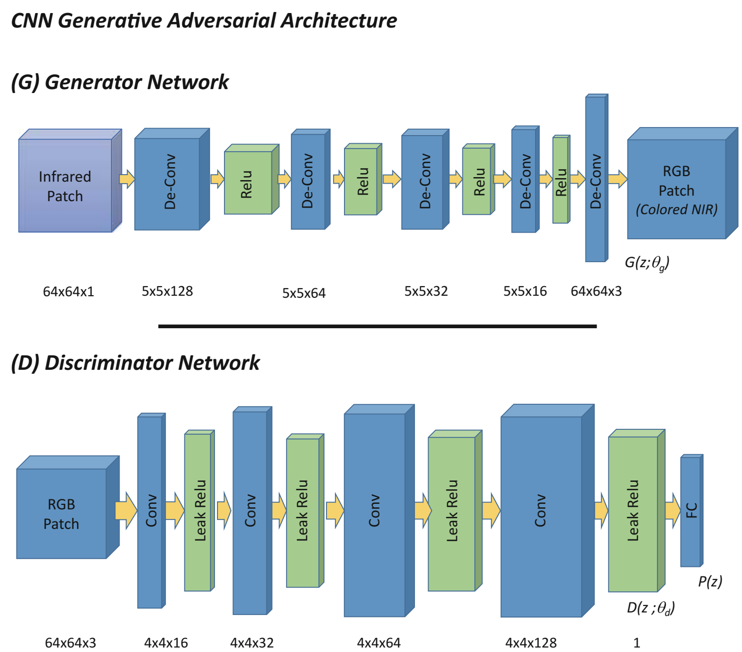

Fig. 1. Illustration of the network architecture used for NIR image colorization.

The generator (G) and discriminator (D) are both feedforward neural networks that play a min-max game between one another. The generator takes as an input a near infrared image patch of 64×64 pixels, and transforms it into the form of the data we are interested in imitating, in our case a RGB image. The discriminator takes as an input a set of data, either real image (z) or generated image (G(z)), and produces a probability of that data being real (P(z)). The discriminator is optimized in order to increase the likelihood of giving a high probability to the real data (the ground truth given image) and a low probability to the fake generated data (wrongly colored NIR image), as introduced in [12]; thus, it is formulated as follow:

$$\nabla_{\theta_g} \frac{1}{m}[\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))], \tag{1}$$

where m is the number of patches in each batch, x is the ground truth image and z is the colored NIR image generated by the network. The weights of the discriminator network (D) are updated by ascending its stochastic gradient. On the other hand, the generator is then optimized in order to increase the probability of the generated data being highly rated:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)}))). \tag{2}$$

where m is the number of samples in each batch and z is the colored NIR image generated by the network. Like in the previous case, the weights of the generator network (G) are updated by descending its stochastic gradient.

4 Experimental Results

The proposed approach has been evaluated using NIR images and their corresponding RGB obtained from [14]. The *urban* category has been considered; it contains 116 images of (1024×680 pixels). From these images 64,200 pairs of patches of (64×64 pixels) have been cropped both, in the NIR images as well as in the corresponding RGB images. Additionally, 12,800 pairs of patches of (64×64 pixels) have been also generated from the given *urban* dataset for validation. It should be noted that images are correctly registered, so that a pixel-to-pixel correspondence is guaranteed (Fig. 2).

Fig. 2. Pair of images (1024×680 pixels) from [14], *urban* category: (*top*) NIR images to colorize; (*bottom*) RGB images used as ground truth.

The GAN network proposed in the current work for NIR image colorization has been trained using a 3.2 eight core processor with 16 Gb of memory with a NVIDIA GeForce GTX970 GPU. On average every training process took about 8 h. The obtained results (RGB_{NIR}) where qualitatively and quantitatively evaluated with respect to the corresponding RGB images provided in the given data set, which are used as ground truth (RGB_{GT}). The quantitative evaluation consists of measuring at every pixel the angular error between the obtained result (colorized NIR image) and the corresponding RGB image provided in the given data set as ground truth values:

$$AngularError = \cos^{-1}\left(\frac{\mathrm{dot}(RGB_{NIR}, RGB_{GT})}{\mathrm{norm}(RGB_{NIR}) * \mathrm{norm}(RGB_{GT})}\right) \tag{3}$$

Fig. 3. Set of **good results** obtained from the proposed approach (**average color angular error: 5.7**°): (*top*) Original NIR patches to be colorized (64×64 pixels); (*middle*) Results from the proposed approach; (*bottom*) Ground truth images.

This angular error is computed over every single pixel of the whole set of images used for validation, obtaining the following results: mean angular error = $9.86°$; standard deviation = $5.31°$. As aforementioned, these are mean values, so in order to visually appreciate these results, two sets of image patches colored with the proposed approach have been generated—due to space limitation just three images per set are presented[1]. In the first family, patches with small angular error are presented (in this case average angular error is below to $6°$), see Fig. 3. In the second family, patches with larger angular errors are depicted (in this case average angular error is higher than $15°$), see Fig. 4. In this second case, although the angular error is larger than before, the global color is some how obtained; the main problem with these patches lies on the texture present in the scene.

[1] The whole set of image patches used for training and validation, as well as the obtained results, are available by contacting the authors.

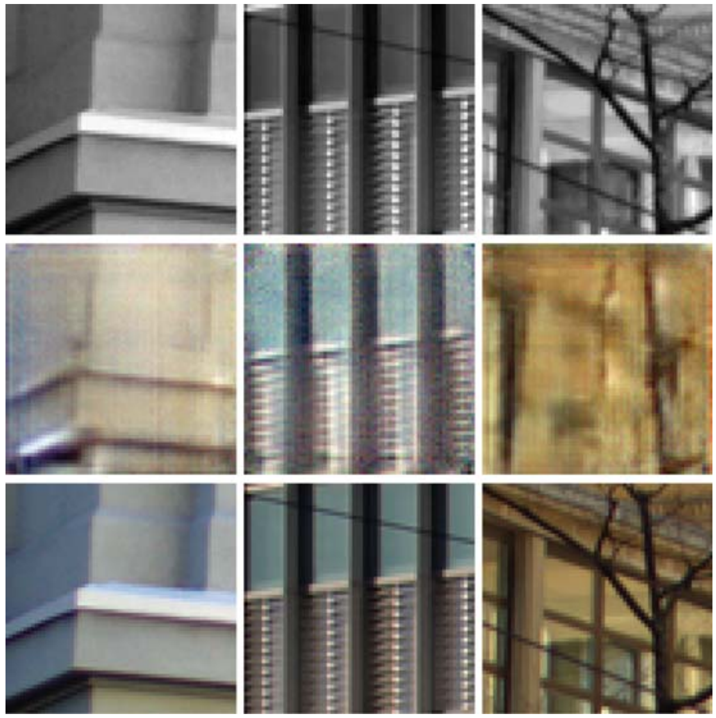

Fig. 4. Set of **bad results** obtained from the proposed approach (**average color angular error: 15.28°**): (*top*) Original NIR patches to be colorized (64×64 pixels); (*middle*) Results from the proposed approach; (*bottom*) Ground truth images.

5 Conclusions

This paper tackles the challenging problem of NIR image colorization by using a novel Generative Adversarial Network architecture model. Results have shown that in most of the cases the network is able to obtain a reliable RGB representation of the given NIR image. Future work will be focused on evaluating others network architectures, like autoencoders, which have shown appealing results in recent works. Additionally, increasing the number of images to train, in particular the color variability, will be considered. Finally, the proposed approach will be tested in other image categories.

Acknowledgments. This work has been partially supported by the ESPOL projects: "Pattern recognition: case study on agriculture and aquaculture" (M1-DI-2015) and "Integrated system for emergency management using sensor networks and reactive signaling" (G4-DI-2014); and by the Spanish Government under Project TIN2014-56919-C3-2-R.

References

1. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: European Conference on Computer Vision, pp. 577–593. Springer (2016)
2. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 415–423 (2015)
3. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European Conference on Computer Vision, pp. 649–666. Springer (2016)
4. Oliveira, M., Sappa, A.D., Santos, V.: Unsupervised local color correction for coarsely registered images. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 201–208. IEEE (2011)
5. Oliveira, M., Sappa, A.D., Santos, V.: A probabilistic approach for color correction in image mosaicking applications. IEEE Trans. Image Process. **24**, 508–523 (2015)
6. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: ACM Transactions on Graphics (TOG), vol. 21, pp. 277–280. ACM (2002)
7. Zbontar, J., LeCun, Y.: Stereo matching by training a convolutional neural network to compare image patches, arXiv preprint arXiv:1510.05970 (2015)
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. CoRR abs/1409.4842 (2014)
9. Aguilera, C.A., Aguilera, F.J., Sappa, A.D., Aguilera, C., Toledo, R.: Learning cross-spectral similarity measures with deep convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–9 (2016)
10. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. In: ACM Transactions on Graphics (Proceedings of SIGGRAPH 2016), p. 35 (2016)
11. Limmer, M., Lensch, H.: Infrared colorization using deep convolutional neural networks, arXiv preprint arXiv:1604.02245 (2016)
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680 (2014)
13. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems, pp. 2226–2234 (2016)
14. Brown, M., Süsstrunk, S.: Multi-spectral SIFT for scene category recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 177–184. IEEE (2011)