Infrared Image Colorization based on a Triplet DCGAN Architecture

Patricia L. Suárez[1] plsuarez@espol.edu.ec

Angel D. Sappa[1,2] sappa@ieee.org

Boris X. Vintimilla[1] boris.vintimilla@espol.edu.ec

[1]Escuela Superior Politécnica del Litoral, ESPOL,
Facultad de Ingeniería en Electricidad y Computación, CIDIS,
Campus Gustavo Galindo, 09-01-5863, Guayaquil, Ecuador

[2]Computer Vision Center, Edifici O, Campus UAB,
08193, Bellaterra, Barcelona, Spain

Abstract

This paper proposes a novel approach for colorizing near infrared (NIR) images using a Deep Convolutional Generative Adversarial Network (GAN) architecture. The proposed approach is based on the usage of a triplet model for learning each color channel independently, in a more homogeneous way. It allows a fast convergence during the training, obtaining a greater similarity between the colored NIR image and the corresponding ground truth. The proposed approach has been evaluated with a large data set of NIR images and compared with a recent approach, which is also based on a GAN architecture where all the color channels are obtained at the same time.

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 (400-700nm); 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.

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. 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., [6], [3], [15]) and color correction/transfer (e.g., [8], [9]) 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 Section 2. Then, the proposed approach is detailed in Section 3. Experimental results with a large set of images are presented in Section 4. Finally, conclusions are given in Section 5.

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.32

2. Related work

Colorization techniques have been largely studied in recent years. Several methods have been proposed to solve this challenging task. However, most of them are not fully automatic, some techniques require some user interactions or utilize user-defined search table. The problem addressed in the current paper is related with infrared image colorization, as mentioned above, somehow it shares some common problems with monocromatic image colorization approaches proposed 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.

Coarsely speaking colorization techniques can be classified into parametric and non-parametric approaches. Parametric methods 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. Non-parametric methods, on the other hand, given an input grayscale image, firstly they define one or more color reference images (provided by an 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).

Welsh et al. [13] 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 [14], image classification [12] or even difficult problems related with cross-spectral domains [1] 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 [15]. It proposes a fully automatic approach that produces brilliant and sharpen colored images. They model the unknown uncertainty of the desaturated colorization levels, designing it as a classification task and using class-rebalancing at training time to augment the diversity of colors in the result.

On the contrary, [5] 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. A recent research on a colorization technique, focused on images of the infrared spectrum, has proposed to use convolutional neural networks to perform an automatic integrated colorization from a single channel NIR image to RGB images [7]. In this paper the author proposes 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.

Deep Convolutional Generative Adversarial Networks (DCGANs) are a class of neural networks that 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 their most common applications is image generation. In the GAN framework [4], generative models are estimated 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 min-max 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 [10] some techniques to improve the efficiency of the generative adversarial networks have been proposed; one of them, referred to as the virtual batch normalization, allows to significantly improve the network optimization using the statistics of each set of training batches.

Recently, [11] proposes a NIR image colorization using a DCGAN architecture. In that work, a colorization model is obtained based on a GAN architecture. On the contrary to that work, in the current paper a triplet based colorization model is proposed to generate the colorized images, in the same scheme of architectures of DC Generative Adversarial Networks. The proposed model generates three instances, each corresponding to one of channels of the (R,G,B) image. This model shares learning parameters and its output is then measured by its probability of being as similar as possible to the image given as ground truth. The details of the implementation are presented in the following section.

3. Proposed approach

This section presents the approach proposed for NIR image colorization. As mentioned above, a recent work on colorization [11] has proposed the usage of a deep convolutional adversarial generative learning network. It is based on a traditional scheme of layers in a deep network. In the current work we also propose the usage of a DCGAN but in a triplet learning layer architecture scheme. These models have been used to solve other types of problems such as learning local characteristics, feature extraction, similarity learning, face recognition, etc. Based on the results that have been obtained on this type of solutions, where improvements in accuracy and performance have been obtained, we propose the usage of a learning model that allows the multiple representation of each of the channels of an image of the visible spectrum (R, G, B). Therefore, the model will receive as input the same image of the near infrared spectrum (NIR), with a Gaussian noise added in each channel of the image to generate the necessary variability of the training set, to be able to generalize the learning of the colorization process. A global loss function is used to minimize the overall classification error in the training set, which can improve the generalization capability of the model.

A DCGAN network based architecture is selected due to several reasons: $i)$ its fast convergence capability; $ii)$ the capacity of the generator model to easily serve as a density model of the training data; and $iii)$ sampling is simple and efficient. The network is intended to learn to generate new samples from an unknown probability distribution. In our case, the generator network has been modified to use a triplet to represent the learning of each image channel independently; at the output of the generator network, the three resulting image channels are recombined to generate the RGB image. This will be validated by the discriminative network, which will evaluate the probability that the colorized image (RGB) is similar to the real one, which is used as ground truth. Additionally, the generator model, in order to obtain a true color, the DCGAN 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 plus some Gaussian noise, 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 DCGAN 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. Our image dataset was normalized in a (-1,1) range and an additive Gaussian Distribution noise with a standard deviation of 0.00011, 0.00012, 0.00013 added to each image channel of the proposed triplet model. The following hyperparameters were used during the learning process: learning rate 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 of 0.00282; weight decay 1e-5; leak relu 0.2 and patch's size of 64×64.

The Triplet 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 normalization 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. It is very important to maintain the spatial information in the generator model, there is not pooling and drop-out layers and only the stride of 1 is used to avoid downsize the image shape. To prevent overfiting we have add a l1 regularization term (λ) in the generator model, this regularization has the particularity that the weights matrix end up using only a small subset of their most important inputs and become quite resistant to noise in the inputs, this characteristics is very useful when the network try to learn which features are contributing to the learning proccess. Figure 1 presents an illustration of the proposed Triplet GAN architecture.

The generator (G) and discriminator (D) are both feedforward deep neural networks that play a min-max game between one another. The generator takes as an input a near infrared image blurred with a Gaussian noise 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 [4]; thus, it is formulated as follow:

$$max_D \mathrm{V}(D,G) = \bigtriangledown_{\theta_g} \frac{1}{m}\sum_{i=1}^{m}[\mathrm{log}D(x^{(i)}) \\ +\mathrm{log}(1-D(G(z^{(i)})))], \tag{1}$$

where m is the number of patches in each batch, x is the

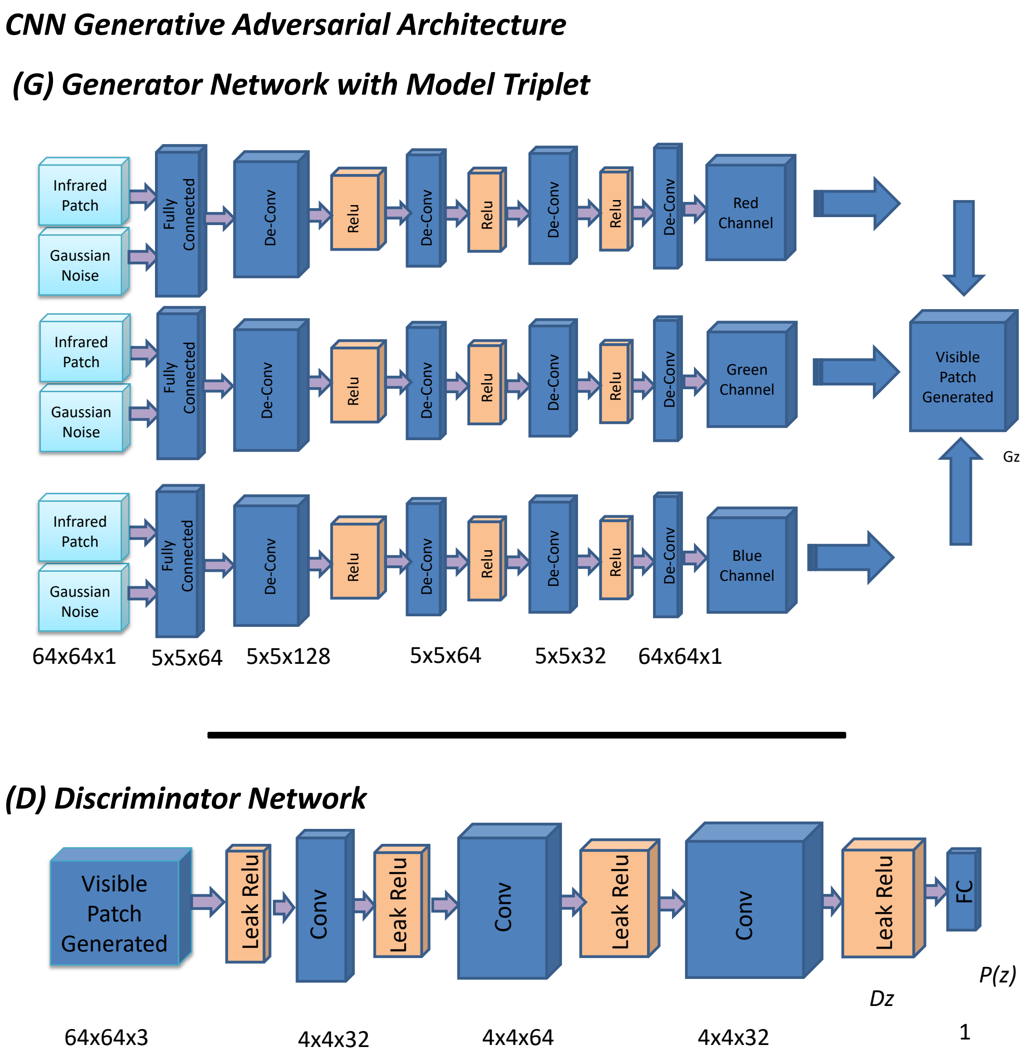

Figure 1. Illustration of the network architecture used for NIR image colorization.

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:

$$min_G \mathrm{V}(D,G) = \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)}))), \quad (2)$$

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 [2]. The *urban* and *old-building* categories have been considered for evaluating the performance of the proposed approach. These categories have been selected since they look quite similar; the intention is to evaluate the capability of the network to be used in scenarios containing similar objects, which have not been used during the training stage. Figure 2 and Fig. 3 presents three pairs of images from each of these categories. The *urban* category contains 58 pairs of images of (1024×680 pixels), while the *old-building* contains 51 pairs of images of (1024×680 pixels). From each of these categories 250.000 pairs of patches of (64×64 pixels) have been cropped both, in the NIR images as well as in

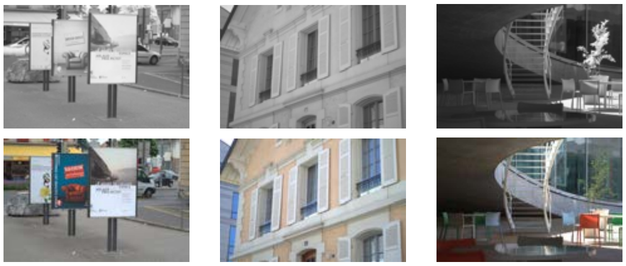

Figure 2. Pair of images (1024×680 pixels) from [2], *urban* category: (*top*) NIR images to colorize; (*bottom*) RGB images used as ground truth.

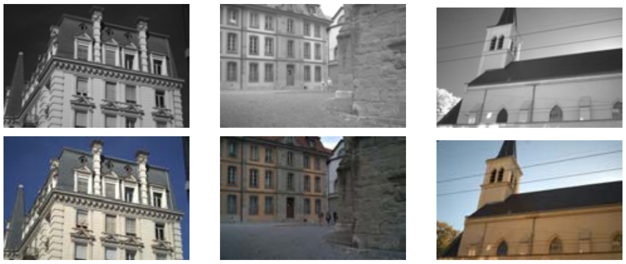

Figure 3. Pair of images (1024×680 pixels) from [2], *old-building* category: (*top*) NIR images to colorize; (*bottom*) RGB images used as ground truth.

the corresponding RGB images. Additionally, 2500 pairs of patches of (64×64 pixels) have been also generated for validation. It should be noted that images are correctly registered, so that a pixel-to-pixel correspondence is guaranteed.

The DCGAN network proposed in the current work for NIR image colorization has been trained using a 3.2 eight core processor with 16Gb of memory with a NVIDIA GeForce GTX970 GPU. On average every training process took about 28 hours. The proposed architecture has been evaluated using three different training schemes. Firstly, the DCGAN network has been trained with the *urban* category and evaluated with both *urban* and *old-building* categories. The same process was applied but by training with the *old-building* category and testing with both *urban* and *old-building* categories. Finally, the DCGAN network has been trained with both data sets and evaluated independently in each of them, *urban* and *old-building* categories. The same scheme has been applied to the GAN model presented in [11] and compared with the results obtained with the proposed approach.

Colored images are referred to as (RGB_{NIR}) while the corresponding RGB images, provided in the given data set, are referred to as (RGB_{GT}) and used as ground truth. The quantitative evaluation consists of measuring at every pixel the angular error (AE) between the obtained result (colorized NIR image) and the corresponding RGB image provided in the given data set as ground truth value :

Table 1: Average angular errors obtained with the proposed Triplet based DCGAN architecture.

Training	Evaluation	
	urban	*old-building*
urban	4.8	8.6
old-building	9.8	7.1
both categories	7.4	8.2

Table 2: Average angular errors obtained with the approach presented in [11].

Training	Evaluation	
	urban	*old-building*
urban	8.6	12.5
old-building	11.7	10.6
both categories	9.9	11.4

$$AE = \cos^{-1}\left(\frac{\text{dot}(RGB_{NIR}, RGB_{GT})}{\text{norm}(RGB_{NIR}) * \text{norm}(RGB_{GT})}\right) \quad (3)$$

This angular error is computed over every single pixel of the whole set of images used for validation. Table 1 presents the average angular errors (AE) obtained with the three schemes mentioned above. The same evaluation scheme has been used with the approach presented in [11]; the results obtained with that approach are presented in Table 2. It can be appreciated that in all the cases the results with the proposed DCGAN are better that those obtained with [11].

Qualitative results are presented in Fig. 4 and Fig. 5. Figure 4 shows NIR images from the *urban* category colorized with the DCGAN network trained with images from that category. On the contrary, Fig. 5 depicts NIR images from the *old-building* category colorized with the DCGAN network trained with images from the *urban* category. It should be noticed that although the weights of the network have been obtained from a different category, colorized images look quite similar to the ground truth ones. Colorization results from other training schemes are similar.

5. Conclusions

This paper tackles the challenging problem of NIR image colorization by using a novel Deep Convolutional 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. Additionally, comparison with a recent approach shows the advantages of the proposed DCGAN architecture. Future work will be focused on evaluating other network architecture, like autoencoders, conditional GAN, which have shown appealing results in recent works. Finally, the proposed approach will be tested in other image categories trying to exploit the transfer learning approaches.

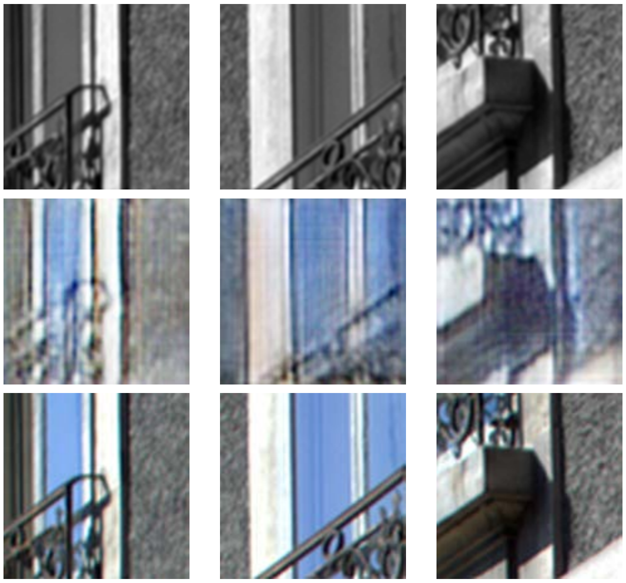

Figure 4. (*top*) NIR images from the ***urban* category**. (*middle*) Images colorized with the DCGAN network **trained with *urban* images**. (*bottom*) Ground truth images.

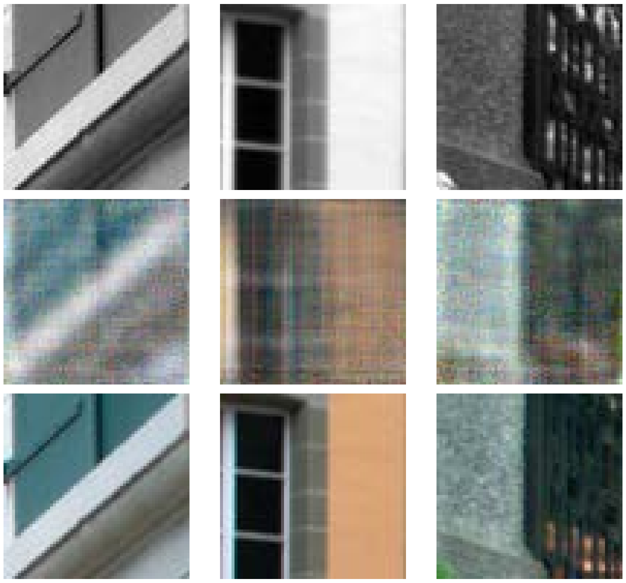

Figure 5. (*top*) NIR images from the ***old-building*** category. (*middle*) Images colorized with the DCGAN network **trained with *urban* images.** (*bottom*) Ground truth images.

6. Acknowledgment

This work has been partially supported by the ESPOL projects PRAIM and KISHWAR; and by the Spanish Government under Project TIN2014-56919-C3-2-R.

References

[1] C. A. Aguilera, F. J. Aguilera, A. D. Sappa, C. Aguilera, and R. Toledo. Learning cross-spectral similarity measures with deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2016. 2

[2] M. Brown and S. Süsstrunk. Multi-spectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 177–184. IEEE, 2011. 4, 5

[3] Z. Cheng, Q. Yang, and B. Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015. 1

[4] 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. 2, 3

[5] 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 Transactions on Graphics (Proc. of SIGGRAPH 2016)*, 35(4), 2016. 2

[6] G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization. In *European Conference on Computer Vision*, 2016. 1

[7] M. Limmer and H. Lensch. Infrared colorization using deep convolutional neural networks. *arXiv preprint arXiv:1604.02245*, 2016. 2

[8] M. Oliveira, A. D. Sappa, and V. Santos. Unsupervised local color correction for coarsely registered images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 201–208. IEEE, 2011. 1

[9] 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. 1

[10] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2226–2234, 2016. 2

[11] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla. Learning to colorize infrared images. In *Proceedings of the 15th Int. Conf. on Practical Applications of Agents and Multi-Agent Systems*, 2017. 2, 3, 5

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. 2

[13] 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. 2

[14] J. Zbontar and Y. LeCun. Stereo matching by training a convolutional neural network to compare image patches. *arXiv preprint arXiv:1510.05970*, 2015. 2

[15] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666. Springer, 2016. 1, 2