Manga Colorization using Generative Adversarial Nets

Souhaib Attaiki
Ecole Normale Supérieure de Paris

Manga Colorization using Generative Adversarial Nets

Souhaib Attaiki
souhaib.attaiki@ens-paris-saclay.fr

Abstract

Image colorization is a famous image to image translation problem, where the goal is to, given a gray scale image, generate a plausible colorization. To do this, many of the previous approaches were either assisted approaches based on significant human contribution, or automatic approaches that use convolutional neural networks and formulate the problem as a classification or regression problem, which has led to desaturated colorization. We investigate the use of Conditional Deep Convolutional Generative Adversarial Network (DCGAN) for colorization, in particular, we are interested in the paired-approach presented by Isola et al. in [13], and the unpaired approach by Zhu et al. in [9] to translate images from the gray scale domain to the RGB domain, our final goal is to colorize black and white manga pages, based on the corresponding anime adaptation. Our results are evaluated using a perceptual study.

1. Introduction

The problem studied is the following: from a black and white image, we want to hallucinate a plausible colorization without it being necessarily similar to the reality of the ground truth, and which can potentially fool a human observer.

Due to cost and fast production purposes, mangas[1] are still published in black and white, even if, as we show in section 5, the fan base are interested in colorized manga as they are easy to read and are better to differentiate smaller details.

Colorizing a black and white image is a simple task for a human. However, interpretation depends on each person, objects can take many colors, and most importantly, it takes a long time to produce good results.

Since the anime adaptation of the manga (which is in color) is based on the same characters, scenario and scenes, we think that the latter presents sufficient elements to colorize the manga, and thus, automate the colorization process, and solve the interpretation issue.

In this paper, we will study the use of DCGANs to colorize mangas, using two different approaches. The first is based on the work of Isola *et al.* in [13], which assumes that we have paired sets of black and white images, and colored images. As it is not usually the case, we will harness the use of cycle-gans[9], to translate images from the manga domain, to the anime domain, which is colored.

2. Related work

Due to the page limit of the report, the reader interested in the related work can consult Appendix A.

3. Generative Adversarial Networks

In 2014, Goodfellow *et al.* proposed a new type of generative models: Generative Adversarial Networks (GANs)[7], that can generate images from random noise, and which are considered by Lecun as "*the most interesting idea in machine learning in last ten years*". GANs are composed of two components: the generator (**G**) and the discriminator (**D**). The role of **G** is to capture the data distribution by mapping a noise distribution to the data distribution and successfully fooling **D**, which amounts to minimizing $\log(1 - D(G(z)))$ where z is a random noise. The role of **D** is to distinguish whether a sample is real or generated by **G**. The hole architecture is trained following two-player minimax game with value function $V(G, D)$:

$$\begin{aligned}\min_G \max_D V(G, D) = \mathbb{E}_{x\sim p_{data}(x)} \big[\log(D(x))\big] \\ + \mathbb{E}_{z\sim p_z(z)} \big[\log(1 - D(G(z)))\big]\end{aligned} \tag{1}$$

Since their apparition, many variant of GANs have been published. In this project, we are interested in Conditional Deep Convolutional Generative Adversarial Network (DCGAN)[6], which instead of generating images from random noise, conditional GAN is given a condition to generate an output image. For example, gray-scale image is the condition for colorization. In this study, we are interested in *pix2pix*[13] and *cycle-gan*[9].

pix2pix harnesses the power of GANs to learn a mapping from an input image (domain A) to an output image (domain B). The Generator (**G**) is trained the apply

[1]Manga are comics created in Japan or by creators in the Japanese language, conforming to a style developed in Japan in the late 19th century [10]

some transformations to the input image to get an image in the domain of the output image. The discriminator (**D**) is trained to compare a given input image and an unknown image, and to guess if the unknown image is real or generated by the generator. To force **G** to generate images near ground truth, an identity loss was introduced: $\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}\left[||y - G(x,z)||_1\right]$. The total loss is the sum of the identity loss and the adversarial conditional loss $\mathcal{L}_{cGAN}$:

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}\left[\log(D(x,y))\right] \\ + \mathbb{E}_{x,z}\left[\log(1 - D(x, G(x,z)))\right]\end{aligned} \tag{2}$$

cycle-gan in the other hand is an unpaired-approach unlike *pix2pix*. The network is trained to translate images from domain **X** to domain **Y**, and *vice versa*, without having *pairs* of images in the both domains. To do that, two generators $G : X \rightarrow Y$ and $F : Y \rightarrow X$ are trained to learn the mapping from one domain to the another, and two discriminators D_X and D_Y are trained to guess if an image is real or generated. In order to force the network to map an individual input x to a desired output y (because if not, network can map the same set of input images to any random permutation of images in the target domain), a *cycle consistency loss* was introduced:

$$\begin{aligned}\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}\left[||F(G(x)) - x||_1\right] \\ + \mathbb{E}_{y\sim p_{data}(y)}\left[||G(F(y)) - y||_1\right]\end{aligned} \tag{3}$$

The total loss is then

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F)$$

where λ is a hyperparameter controlling the relative importance of the two objectives.

4. Datasets

Details of the used datasets are provided in table 1. The colored One Piece dataset is obtained from *powermanga*[4], which is a fan community website. Both *Death note* manga scans and animes are obtained from legal sources. Examples of the used datasets are shown in figure 1.

Preprocessing For the colorization task using *pix2pix*, the images are converted from RGB to the L*A*B* color space. The network is trained to predict, given the L* channel (which represent the gray scale part of the image), to predict the A* and B* channels that encode the color. The three channels are then concatenated and converted to RGB to produce the final image.

For the colorization using *cycle-gan*, the same gray scale image is stacked 3 times in order to be able to use the identity loss (because the color images have three channels).

Concerning death note videos, images are sampled every two seconds to avoid having identical frames.

All the images are then resized to 256×256.

5. Experiments and results

For our experiments, we used the architecture from [15] as our **baseline**, which is implemented and trained at [1], and is considered as the state of the art for the colorization task using only CNN's. We used also the *pix2pix* implementation available at [5] for our paired colorization experience. And we finally implemented **from scratch** the *cyclegan* architecture from [9] using *python3* and *PyTorch*. The entire preprocessing step was done using *OpenCV* and *Pilow*. All the code is available on our github repository [3].

CycleGAN architecture For our generator, we implemented a coder /decoder network with 9 ResNet blocks in the middle. Our descriminator is 70×70 patch GAN. Details of the architecture are provided in figure 11. During training, we used the same hyperparameters as the original paper.

5.1. Experiments

5.1.1 Reproduction result

In order to verify if our implementation is working, we have tried to reproduce some results of the original paper. For this purpose, we used the *horse2zebra* dataset. We did not train our network for 200 epochs due to our limited computing power, but we were able to obtain satisfactory results similar to those of the original paper (figure 2).

5.2. Application to flower colorization

Before moving to colorizing mangas, we tested our network on the simple image colorization task using the *ImageNet flower synset*. To do this, two sets of disjointed images were used, one of which was converted to grayscale. The network has therefore been trained to make the image translation from domain A (grayscale) to domain B (RGB) and vice versa. Examples of our results are shown in figure 3. We can see that the images produced look very real and full of colors. Unlike the classic method based on regression, which produces grayish images due to euclidean loss, which has the effect of taking the average color, we find that the colors of our images are bright and indistinguishable from real color photos. An in-depth evaluation will be done in the section 5.5.

5.3. Manga Colorization, paired approach

Before moving to the unpaired manga colorization, we tried to compare our results against *pix2pix*, to see how close we are to this "*upper bound*" by using only unpaired images. To do this, we used the *pix2pix* implementation available online, as well as images of the manga one piece colorized by the fans. The network is trained to translate the image in grayscale (channel L*) to the two channels that represent the colors (A* and B*). The three channels are then assembled and converted to RGB. Examples of our results are shown in figure 4. We can see that the colorization obtained is flawless, and can be considered perfect for manga readers. This sets the bar very high for unpaired colorization.

5.4. Manga colorization: Unpaired approach

As we discussed earlier, manual manga colorization takes a long time, and in most cases, colorized images of the manga do not exist to do the paired method. To remedy this, images from the corresponding anime adaptation, which is based on the same character and scene, were used to translate images from the manga domain to the anime domain and vice versa.To test our approach, we used images from the popular *Death Note* manga and the corresponding anime.

First experience In the first implementation, we cut the scan image into three parts according to the height of the page (figure 5) (the same method was used for manga colorization with pix2pix), and we feed them to the network with the anime images. Examples of our results are shown in figure 6.

This first experience suffers from many problems (see figure 9). First, since the page is divided into three parts, and each part contains several panels, the network believes that it is a single image and tries to merge the different panels, which gives a colorization where the different images are not distinguished. Secondly, the writing on the manga page disturbs the colorization. Since the writing does not exists in the anime, the network tends to replace it with black. We also noticed that the anime to manga network has started to add a kind of writing in the generated images, but this prevents the convergence of the cycle loss because the manga to anime network does not know how to remove them. Finally, we noticed that the network does not succeed in coloring images that do not contain much detail, and tends to colorize them in black.

Second experience The purpose of the second experiment is to solve the problems of the first. To solve the first problem mentioned above, we implemented an algorithm to detect panels using *OpenCV*. This algorithm is based on the fact that the panels are separated by white and surrounded by a black contours. Our algorithm first detects all the contours, and is based on rules of thumb that have been designed to leave only blocks corresponding to panels. The results of this algorithm are shown in the figure 7. These panels are therefore used as an input for our second experiment. Results are shown in figure 8.

We can clearly see that the results of the second experiment are much better than the first, in terms of colorization and the resolution of the pannel problem. However, we always get a bad colorization if the image contains text.

Third experience A third experience was conducted to measure the effect of removing text. For that, a dataset without text was created, by manually removing it from 100 images. Unfortunately, this has not been successful. On the one hand, because we used a very small dataset, and on the other hand, because we didn't have enough computing power, so we couldn't train the network for long time.

5.5. Evaluation

To evaluate our results, two methods are used. First, we used the perceptual study from [9], and compared our results, to the baseline used [15], which considered as the state of the art using CNN's only, and was trained on 1.2M images from ImageNet. And to evaluate the colorization of manga, we conducted a survey on the *r/manga* and *r/deathnote* subreddits [2].

Perceptual study We conducted a perceptual study to, first, evaluate the colorization of our algorithm on the flower synset dataset, and second, compare it to the colorization of our baseline. Each test was conducted on 19 persons and using 15 images. For the first test, participants were asked to give a score between 1 (poor colorization) and 5 (good colorization) for each of the images. For the second test, the participants were led to choose, which of the colorizations seemed better to them, ours or that of the baseline (see figure 10 for comparison). Results are summarized in tables 2 and 3. We can see that in the two tests, participants preferred our colorization, and gave it a higher score. However, it should be noted that our network is trained to color only flowers, which is not the case for our baseline.

Reddit survey In the reddit survey, redditors were asked if they prefer to read black and white manga, or our colorized version, also, how they find it. 100 % of responses preferred the colorized version, and found that it's allows to differentiate small details and easy to read, however, they suggest more improvement to the algorithms, especially when some details are lost if they are colored by a dark color.

6. Conclusion

In this study, we were able to colorize grayscale image using GANs, in a way that was appreciated by the participant of the perceptual study. Concerning the colorization of mangas, we have seen that the paired approach gives good results if we have manually colored images, otherwise, the unpaired approach must be improved. The next steps are to use a big dataset of images without texts, finetune the hyperparameters as we used the defaults values and train for a long period.

A. Related work

For a long time, colorization required human supervision. The scribbling method, introduced by Levin *et al.* [11], requires manual specification of the desired colors for certain regions. These scribbling colors are propagated assuming that adjacent pixels with similar luminance should have a similar color. Another human assisted method is colorization transfer [12] [14], where besides the gray scale image, the model needs another colored image for reference so that it can match the information between the two.

Despite the quality of the results, previous methods requires a lot of human assistance. Thus, with the advent of deep learning and Convolutional Neural Networks, fully automated colorization models were proposed. The classical approaches have either formulated the problem as a regression problem, or as a classification problem. The state of the art of the regression methods is proposed by Iizuka *et al.*[8]. Their architecture is composed of four main components to extract local and global features, and it's trained end to end using a L2 loss. Zhang *et al.* proposed a classification architecture which considered as the state of the art in the classification setting. Their approach is based on the fact that not all colors are equally distributed, so they tailored a loss function adapted to the problem. They discretise the color space, and use a classification loss to predict the class for each pixel (possible color), and re-weight the loss at training time to emphasize rare colors ("class rebalancing").

Dataset	Description
Horse2Zebra	2500 images
ImageNet flower synset	1100 images
Colored One Piece	600 colored pages
Death note (The manga)	4 tomes, 750 pages
Death note (The anime)	7 episodes, 140 minutes

Table 1. Details about used datasets

Ours	Baseline
3.65 / 5	2.6 / 5

Table 2. Colorization average score

Ours	Baseline
79 %	21 %

Table 3. Participants preferred colorization

References

[1] Colorful image colorization. https://demos.algorithmia.com/colorize-photos/.

[2] Manga subreddit. https://www.reddit.com/r/manga/.

[3] Mangan: automatic manga colorization using gans. https://github.com/pvnieo/ManGAN.

[4] Powermanga fan comunity. http://www.powermanga.org.

[5] Pytorch cyclegan and pix2pix implementation. https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

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

[7] M. M. B. X. D. W.-F. S. O. A. C. an Goodfellow, Jean Pouget-Abadie and Y. Bengio. Generative adversarial nets. *In Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] S.-S.-E. I. H. Iizuka, S. 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)*, 2016.

[9] P. I. J.-Y. Zhu, T. Park and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *International Conference on Computer Vision*, 2017.

[10] J. A. Lent. Illustrating asia: Comics, humor magazines, and picture books. honolulu, hawaii: University of hawaii press, 2001.

[11] L. D.-W.-Y. Levin. Colorization using optimization, 2004.

[12] G. C. M. Hofmann, B. Scholkopf. Automatic image colorization via multimodal predictions, 2008.

[13] T. Z. P. Isola, J.-Y. Zhu and A. A. Efros. Imageto-image translation with conditional adversarial networks. *Conference on Computer Vision and Pattern Recognition*, 2017.

[14] A. Yong Sang Chia, S. Zhuo, R. Gupta, Y.-W. Tai, D. Cho, P. Tan, and S. Lin. Semantic colorization with internet images, 2011.

[15] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *ECCV*, 2016.

Figure 1. Examples of the used datasets. From top to bottom, left to right: horse2zebra, ImageNet flower synset, Colored One piece, Death note (manga), Death note (anime)

Figure 2. Our results on the *horse2zebra dataset*

Figure 3. Our results on colorizing the *ImageNet flower synset*. Left: grayscale image, right: colorized image

Figure 4. Our results on colorizing the paired One piece images. Left: grayscale image, right: colorized image

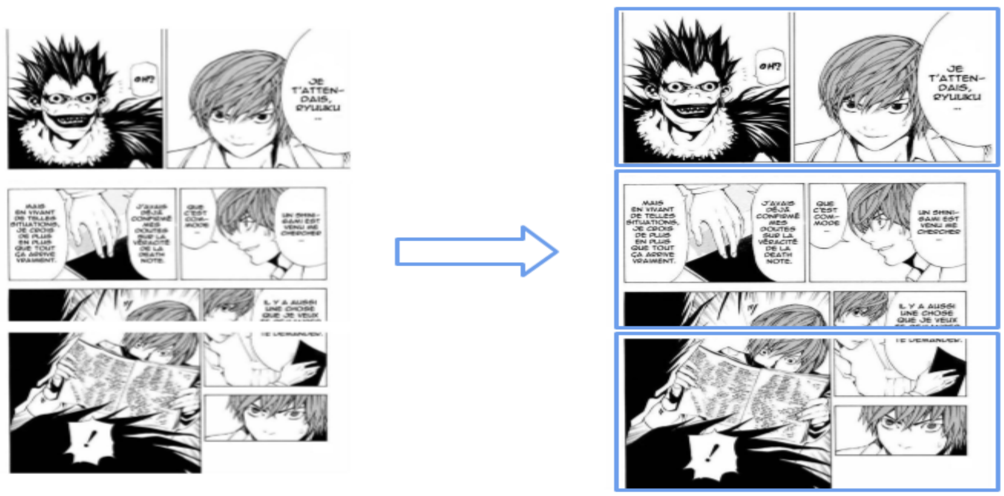

Figure 5. First experience cutting method

Figure 6. First unpaired colorization experience results. Left: manga image, right: colorized manga image

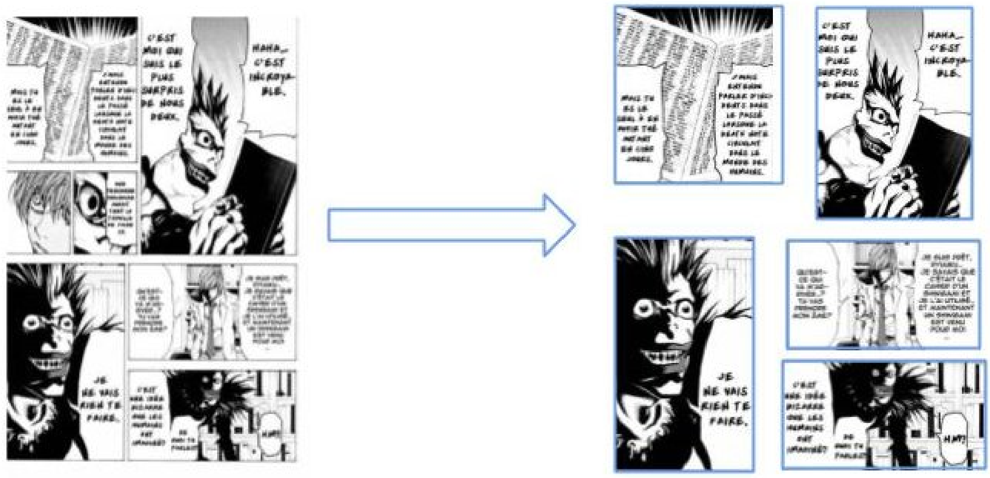

Figure 7. Second experience cutting method

Figure 8. Second unpaired colorization experience results. Left: manga image, right: colorized manga image

Figure 9. Common failures. Top: manga to anime, bottom: anime to manga

Figure 10. Comparison of our colorization and baseline's. Left: baseline colorization, middle: grayscale image, right: our colorization

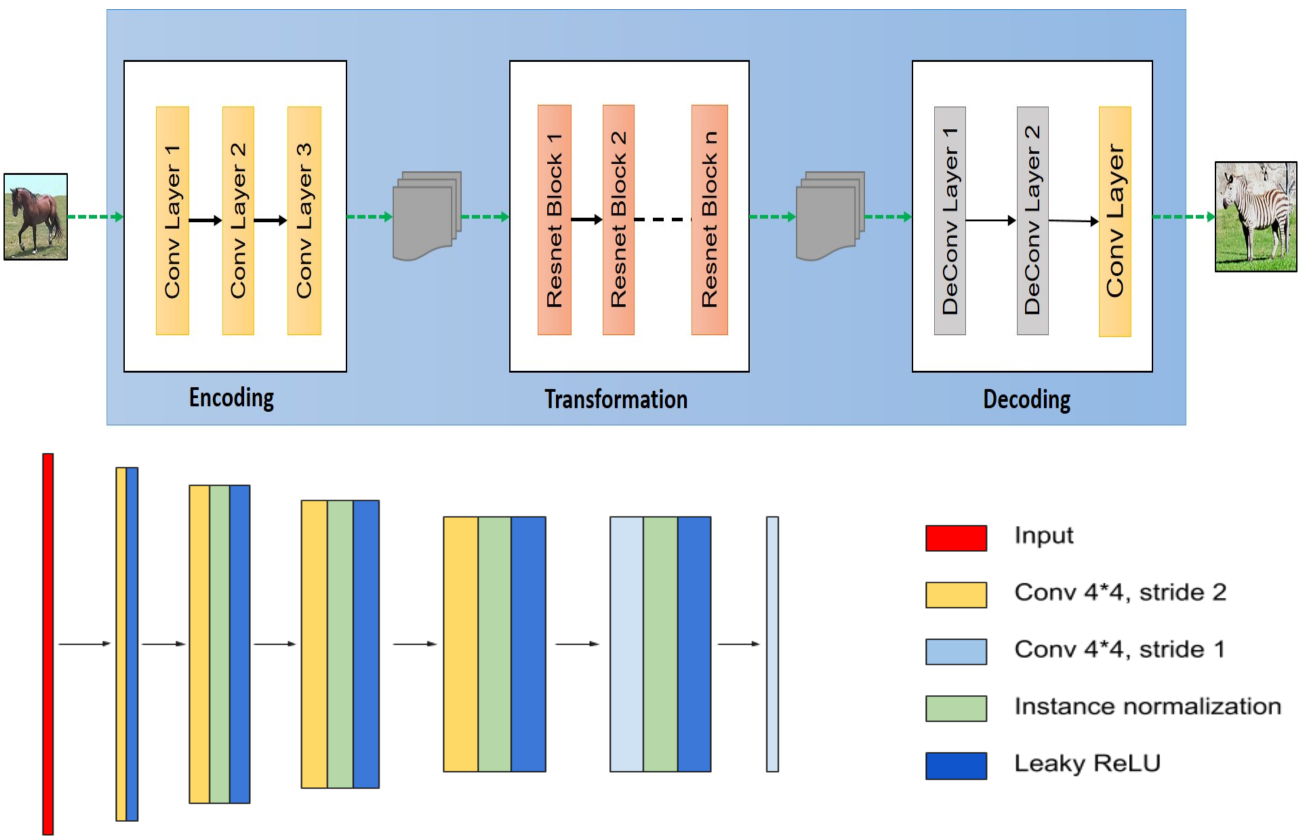

Figure 11. Architecture of the implemented generator (top, source: *https://hardikbansal.github.io/CycleGANBlog/*), and the implemented discrimination (bottom)