Unpaired Image-to-Image Translation with Limited Data to Reveal Subtle Phenotypes

Anis Bourou, Kevin Daupin, Veronique Dubreuil, Aurélie de Thonel, Valerie Mezger-Lallemand, Auguste Genovesio

Unpaired Image-to-Image Translation with Limited Data to Reveal Subtle Phenotypes

Anis Bourou
IBENS, ENS
Université de Paris Cité
anis.bourou@ens.fr

Kévin Daupin
Université de Paris Cité
kevindaupin@gmail.com

Véronique Dubreuil
Université de Paris Cité
veronique.dubreuil@u-paris.fr

Aurélie De Thonel
Université de Paris Cité
aurelie.dethonel@u-paris.fr

Valérie Lallemand-Mezger
Université de Paris
valerie.mezger@u-paris.fr

Auguste Genovesio
IBENS, ENS
auguste.genovesio@ens.psl.eu

Abstract

Unpaired image-to-image translation methods aim at learning a mapping of images from a source domain to a target domain. Recently, these methods showed to be very useful in biological applications to display subtle phenotypic cell variations otherwise invisible to the human eye. However, while most microscopy experiments remain limited in the number of images they can produce, current models require a large number of images to be trained. In this work, we present an improved CycleGAN architecture that employs self-supervised discriminators to alleviate the need for numerous images. We demonstrate quantitatively and qualitatively that the proposed approach outperforms the CycleGAN baseline, including when it is combined with differentiable augmentations. We also provide results obtained with small biological datasets on obvious and non-obvious cell phenotype variations demonstrating a straightforward application of this method.

1 Introduction

Image-to-image translation can be defined as transferring images from a source domain to a target domain while preserving the content representations. Image-to-image translation became popular in different fields such as autonomous driving cars[19], artistic style transfer[24] and more recently in biological imaging[11]. Spotting differences in visual cell pheneotypes from medical and biological images has many applications in fundamental research, drug discovery and medicine. In microscopy experiments, hand-crafted image analysis can be used to measure the variation of phenotype produced by a perturbation when the last is visible. However, because cell-to- cell variability within an image often largely overlaps the cell-to-cell variability between phenotypes, the last is often invisible[11]. This issue prevents researchers from observing and measuring subtle phenotypic differences between two images of cells.

The advent of GANs extended the frontiers of image generation, GANs were also used in many approaches to build efficient image translation methods. In[8], the authors propose a supervised method based on conditional-GAN[15] where the condition is the paired image target. Usually, obtaining paired images may be hard or even impossible. Therefore, unpaired image translation

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

methods such as CycleGAN[23], DiscoGAN[9] and DualGAN[21] were proposed by exploiting the cycle-consistency constraint[23].

A common limitation with the methods cited above is that they all require a large amount of images in both the source and the target domains to be trained effectively. Several strategies were proposed to tackle this issue [20; 16; 13]. In general, these methods assume the existence of a large image dataset from a close domain that can be exploited using different strategies. However this assumption is not often met. In bio-imagery, for instance, it is hard to get images that resembles the images of organoids of interest while the last cannot be produced by thousands.

Self-supervised learning (SSL) [10; 4] is a paradigm that tries to overcome the problem of the lack of labelled data by obtaining a supervision signal from the data itself in order to learn a richer representation. Recently, several works proposed to leverage SSL to solve tasks in setting where little training data is available[5; 2; 12]

In this work, we present **Self-Supervised CycleGAN**(SCGAN), a model based on cycle consistency[23] and self-supervised learning for an efficient image translation with limited data.

2 Proposed Method

Given a source domain X and a target domain Y, the goal of an image translation model is to learn a mapping $G : X \rightarrow Y$ such that the output $\hat{y} = G(x)$, $x \in X$, is indistinguishable from images $y \in Y$. Our method (SCGAN) consists of two generators and two discriminators. The efficiency of the proposed approach relies mainly on the fact that we use self-supervised discriminators [12]. In the following sections, we describe the architectures and the losses we use to achieve satisfactory unpaired image translation with a low amount of images.

Generator For the generators, we adopt the architecture used in [23]. The model contains three convolutions followed by 9 residual blocks[6], then we use two fractionally-strided convolutions and finally one convolution that maps the features to RGB. We also use instance normalization[18].

Self-supervised discriminator For the discriminator we rely on the architecture proposed in [12], where a strong regularization for the discriminator is provided through a self-supervised learning strategy. Unlike regular self supervised approaches, where transformations of the same image are pushed to produce the same representation, this approach generates a self-supervised signal by learning to reconstruct a degraded version of the image. In this configuration the discriminator acts as an encoder trained with small decoders. These are simple networks made of four conv-layers which are jointly optimized with the discriminator on a simple reconstruction loss using the real samples only. For D_X, the adversarial discriminator for images from domain X, the ssl loss reads:

$$L_{ssl}(D_X) = \mathop{\mathbb{E}}_{x \sim \imath_{data(x)}} \|Dec(T'(Enc(x))) - T(x)\| \tag{1}$$

where Enc produces an intermediate feature map in D , the function T' is a simple degradation process (such as crop) on x and T is a similar degradation process on the internal representation provided by Enc. Finally, Dec is a decoder from the degraded internal representation to the degraded image $T(x)$. In this setting, the discriminator is encouraged to extract a more comprehensive and useful representation yet using a lower amount of data.

Training losses At training time, we use the least square loss introduced in LSGAN [14] as it provides a more stable training by mitigating the vanishing gradient and mode collapse issues. The two generators and the two discriminators minimize the following losses, here for G:

$$L_G = \mathop{\mathbb{E}}_{x \sim \mathbb{P}_X} [(D_Y(G(x)) - 1)^2] \tag{2}$$

$$L_{D_Y} = \mathop{\mathbb{E}}_{y \sim \mathbb{P}_Y} [(D_Y(y) - 1)^2] + \mathop{\mathbb{E}}_{x \sim \mathbb{P}_X} [(D_Y(G(x)))^2] + L_{ssl}(D_Y) \tag{3}$$

where D_Y denotes the adversarial discriminator for images from domain Y and $L_{ssl}(D_Y)$ is the self supervised loss described in the previous section. L_F for the generator F and L_{D_X} for the adversarial discriminator D_X are trained similarly and altogether the adversarial loss is:

$$L_{adv} = L_G + L_{D_Y} + L_F + L_{D_X} \tag{4}$$

In addition, the cycle-consistency constraint ensures that an image x from domain X consecutively transformed by the two generators reconstructs the original image x, i.e., $F(G(x)) \approx x$. Similarly, for each image y from domain Y, $G(F(y)) \approx y$. This constraint is enforced by the following loss:

$$L_{cyc} = \mathop{\mathbb{E}}_{x\sim_{I_X}}[\|F(G(x)) - x\|_1] + \mathop{\mathbb{E}}_{y\sim_{I_Y}}[\|G(F(y)) - y\|_1] \tag{5}$$

In addition to the adversarial and the cycle-consistency losses, it was shown in [23] that adding an identity loss [17] can enhance the performance of the CycleGAN. This loss ensures that transforming an image from a domain to the same domain will not modify it. It is defined as follows:

$$L_{id} = \mathop{\mathbb{E}}_{x\sim_{I_X}}[\|F(x) - x\|_1] + \mathop{\mathbb{E}}_{y\sim_{I_Y}}[\|G(y) - y\|_1] \tag{6}$$

The full objective loss to minimize is the summation of the three losses: the adversarial, the cyclic and the identity, it is given as follows:

$$L_{total} = L_{adv} + \lambda_1 L_{cyc} + \lambda_2 L_{id} \tag{7}$$

3 Experiment

3.1 Datasets and training

Horse2Zebra. The images of horses and zebras were downloaded from ImageNet[3]. The images were scaled to 256 × 256 pixels. The training set size of each class is: 939 (horse) and 1177 (zebra).

BBBC021. Microscopy images of MCF-7 cancer cells untreated (DMSO) and treated for 24h with Cytochalasin B at high dosages (30μM). Training was performed on 200 images from each condition.

Organoids. Microscopy images of neural organoids (mini-brains) induced from the stem cells of a rare neuro-developmental disorder. We used three marker: ZO1 for the cell-to-cell junctions, phosphohistone H3 for the dividing cells and DAPI for the nuclei. Training was performed on 56 images of healthy organoids and 83 images of diseased organoids.

Figure 1: Translating horse images to zebra images using only 40 percent of the dataset. We can see that our method outperforms CycleGAN.

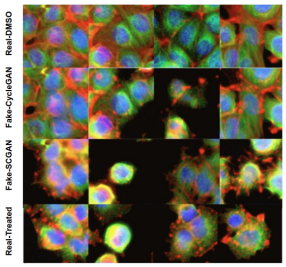
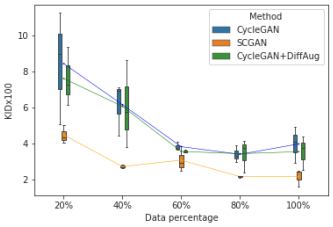
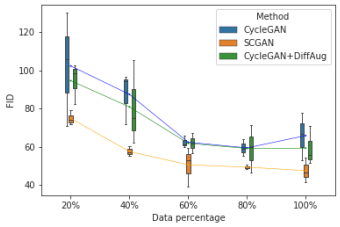


Figure 2: Translated images of untreated (DMSO) cell (first row) into drug treated cells using CycleGAN and SCGAN (2nd and 3rd rows respectively, we can see that the translation made by SCGAN is close to the real images of treated cells (4th row).

Training. For all the experiments, we set $\lambda_1 = 10$ and $\lambda_2 = 0.5$. We used the Adam optimizer with a batch size of 8 and a learning rate of 0.0002. This rate was decayed linearly once every 100 epochs. We used FID[7] and KID[1] to compare the image distributions, the lower the better. We compared SCGAN to CycleGAN with and without adding a differentiable augmentation module. We trained the models using different percentages of the horse2zebra dataset. We also trained it on biological images to illustrate the portability of our approach.



Comparing KID



Comparing FID

Figure 3: Comparing the CycleGAN, CycleGAN+DiffAug and SCGAN in terms of (a) FID and (b) KID using different percentages of the horse2zebra dataset. With each data perecentage, we train three times our model, the boxplots are plotted using KID and FID scores computed for each run.

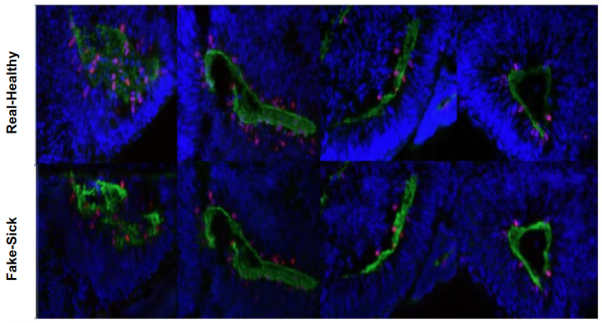


Figure 4: In the first row we have the real images of the healthy organoids and in the second row we have their translation to the other class.

3.2 Results

The FID curves in fig. 3 show the effectiveness of our method compared to the standard CycleGAN. The results based on FID and KID scores show that SCGAN improves the translation with any quantity of data for the horse2zebra dataset. While differentiable augmentations[22] can slightly improve over CycleGAN, our method remains significantly better. Qualitatively, our method achieves a better translation when trained with only 40 percent of the horse2zebra dataset (fig. 1).

In fig.2, we show that we also achieved better translation on a biological dataset where untreated cells (DMSO) were treated with a high concentration of compound (Cytochalasin B) which produces an obvious change in phenotypes with as little as 200 images from each class. Quantitatively, we obtained **FID = 77.84 and KIDx100 = 6.53** with CycleGAN and **FID = 55.36 and KIDx100 = 2.56** with our approach. This indicates that our method outperforms CycleGAN by a large margin.

In fig. 4, we then applied our approach on invisible phenotypic changes between two conditions on organoids. The translated images showcase some changes: 1) the intensity of the blue marker has diminished indicating a decrease in the number of neural cells attained with the syndrome. There are also fewer red cells in the translated images compared to the real ones indicating a decrease in cell divisions in the the diseased cells. In order to validate these subtle differences further biological experiments should of course be conducted.

4 Conclusion

In this work, we presented SCGAN, a self-supervised, cycle-consistency based method for unpaired image translation with limited size datasets. We showed through experiments that our method outperforms the standard CycleGAN with and without augmentation strategies. Furthermore, we showcased the importance of self-supervised learning as an efficient strategy to mitigate the need for large size datasets. Finally, we showed how image translation can be applied in biology to identify

subtle phenotypes when the number of the available images is limited. It can be used to guide the intuition of the experts to understand subtle biological processes or identify new therapeutic biomarkers.

5 Acknowledgments

This work was supported by ANR–10–LABX–54 MEMOLIFE and ANR–10 IDEX 0001 –02 PSL* Université Paris, ANR Visualpseudotime and was granted access to the HPC resources of IDRIS under the allocation 2020- AD011011495 made by GENCI.

References

[1] Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans, 2018.
[2] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss, 2018.
[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
[4] Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales. Self-supervised representation learning: Introduction, advances and challenges. *CoRR*, abs/2110.09327, 2021.
[5] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Boosting few-shot visual learning with self-supervision, 2019.
[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
[7] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 2017.
[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
[9] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks, 2017.
[10] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer. Revisiting self-supervised visual representation learning. *CoRR*, abs/1901.09005, 2019.
[11] Alexis Lamiable, Tiphaine Champetier, Francesco Leonardi, Ethan Cohen, Peter Sommer, David Hardy, Nicolas Argy, Achille Massougbodji, Elaine Del Nery, Gilles Cottrell, Yong-Jun Kwon, and Auguste Genovesio. Revealing invisible cell phenotypes with conditional generative modeling. *bioRxiv*, 2022.
[12] Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis, 2021.
[13] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. 2019.
[14] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2016.
[15] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.
[16] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence, 2021.
[17] Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation, 2016.
[18] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization, 2016.
[19] Lin Wang, Wonjune Cho, and Kuk-Jin Yoon. Deceiving image-to-image translation networks for autonomous driving with adversarial perturbations. *CoRR*, abs/2001.01506, 2020.
[20] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data, 2018.
[21] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation, 2017.
[22] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient GAN training. *CoRR*, abs/2006.10738, 2020.
[23] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017.
[24] Zhengxia Zou, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. Stylized neural painting. *CoRR*, abs/2011.08114, 2020.