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 Cité
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 is defined as transferring images from a source domain to a target domain while preserving the content. Image-to-image translation became popular in various fields[16; 20] and recently in biological imaging[10]. Because cell-to-cell variability within an image often largely overlaps the cell-to-cell variability between phenotypes, the last is often invisible[10]. Therefore, spotting differences in visual cell pheneotypes from medical or biological images has many applications in fundamental research, drug discovery and medicine.

The existing image-to-image translation methods[19; 8; 1; 12; 2] require a large number of images in both the source and the target domains to be trained effectively. Several strategies were proposed to tackle this issue [17; 14; 12]. In general, these alternative methods still assume the availability of a large image dataset from a close domain that can be exploited using different strategies. However this assumption is often not met.

Self-supervised learning (SSL) [9; 5] is a paradigm that attempts to overcome the problem of the lack of labelled data by obtaining a supervision signal from the data itself in order to learn a useful representation. Recently, several works proposed to leverage SSL to solve tasks in settings where little training data is available[6; 3; 11]. In this work, we present **Self-Supervised CycleGAN**(SCGAN), a

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

model based on cycle-consistency[19] and self-supervised learning for an efficient image-to-image translation with limited data.

2 Proposed Method

Given a source domain X and a target domain Y, the goal of an image-to-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 with the following architectures:

Generator For the generators, we adopt the architecture used in [19]. The models contain three convolutions followed by 9 residual blocks[7], then we use two fractionally-strided convolutions and finally one convolution that maps the features to RGB. We also use instance normalization[15].

Self-supervised discriminator For the discriminator we rely on the architecture proposed in [11], where a strong regularization for the discriminator is provided through a self-supervised learning strategy. 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:

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

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 $T(x)$. Furthermore, in order to stabilize the training of the generators and the discriminators, we use the least square loss [13].

3 Experiments

3.1 Datasets and training

Horse2Zebra. The images of horses and zebras were downloaded from ImageNet[4]. 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.

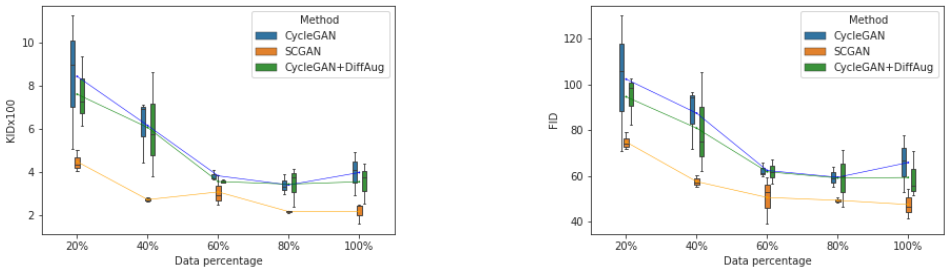
Organoids. Microscopy images of neural organoids (mini-brains) induced from the stem cells of a rare neuro-developmental disorder. We used three marker: ZO1 for the cell-to-cell junctions, phosphohistone H3 for the dividing cells and DAPI for the nuclei. Training was performed on 56 images of healthy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

3.2 Results

The FID and KID curves in fig. 1 show that SCGAN improves the translation with any quantity of data for the horse2zebra dataset. While differentiable augmentations[18] 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. 3).

In fig.4, 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. 2, we then applied our approach to decipher invisible phenotypic changes between two conditions on organoids. The translated images showcase some changes: 1) the intensity of the blue



Comparing KID Comparing FID

Figure 1: Comparing the CycleGAN, CycleGAN+DiffAug and SCGAN in terms of (a) FID and (b) KID using different percentages of the horse2zebra dataset. We trained three times our model, for each data size.

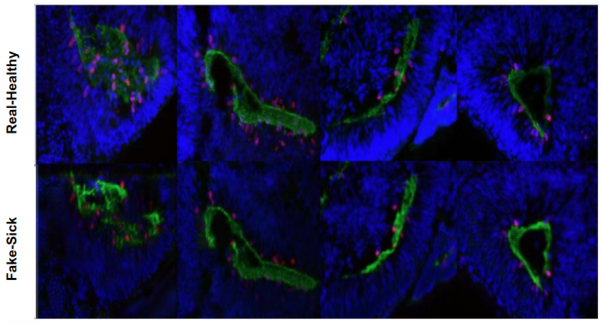


Figure 2: In the first row we have the real images of the healthy organoids and in the second row we have their translation to the other class.

marker has diminished indicating a decrease in the number of neural cells attained with the syndrome. There are also less red cells in the translated images compared to the real ones indicating a decrease in cell divisions in the sick cells. In order to validate these subtle differences further biological experiments should of course be conducted.

Figure 3: Translating horse images to zebra images using only 40 percent of the dataset. We can see that our method outperforms CycleGAN.

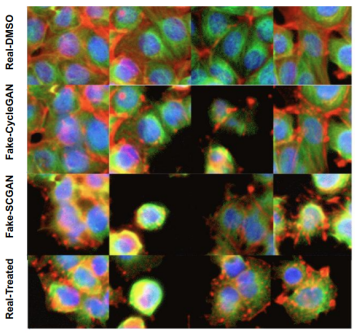


Figure 4: Translated images of untreated (DMSO) cell (first row) into drug treated cells using CycleGAN and SCGAN (2nd and 3rd rows respectively, we can see that the translations made by SCGAN closely approach the real images of treated cells (4th row).

4 Conclusion

In this work, we presented SCGAN, a self-supervised, cycle-consistency based method for unpaired image-to-image translation with limited size datasets. We showed through experiments that our method outperforms the standard CycleGAN with and without augmentation strategies used to overcome data scarcity. We then showed how image translation can be applied in biology to identify subtle phenotypes when the number of available images is limited. This approach can be used to guide the intuition of the experts to understand subtle biological processes or to 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] Matthew Amodio and Smita Krishnaswamy. Travelgan: Image-to-image translation by transformation vector learning, 2019.
[2] Sagie Benaim and Lior Wolf. One-sided unsupervised domain mapping, 2017.
[3] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss, 2018.
[4] 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.
[5] Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales. Self-supervised representation learning: Introduction, advances and challenges. *CoRR*, abs/2110.09327, 2021.
[6] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Boosting few-shot visual learning with self-supervision, 2019.
[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
[8] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks, 2017.
[9] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer. Revisiting self-supervised visual representation learning. *CoRR*, abs/1901.09005, 2019.
[10] 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.
[11] Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis, 2021.
[12] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. 2019.
[13] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2016.
[14] 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.
[15] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization, 2016.
[16] 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.
[17] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data, 2018.
[18] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient GAN training. *CoRR*, abs/2006.10738, 2020.
[19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017.
[20] Zhengxia Zou, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. Stylized neural painting. *CoRR*, abs/2011.08114, 2020.