PHENDIFF: REVEALING INVISIBLE PHENOTYPES WITH CONDITIONAL DIFFUSION MODELS

Anis Bourou[1,2], Thomas Boyer[1,2*], Kévin Daupin[2], Véronique Dubreuil[2], Aurélie De Thonel[2] Valérie Mezger[2] and Auguste Genovesio[1⋆]*

[1]IBENS, ENS, Université PSL
[2]Université Paris Cité

ABSTRACT

Over the last five years, deep generative models have gradually been adopted for various tasks in biological research. Notably, image-to-image translation methods showed to be effective in revealing subtle phenotypic cell variations otherwise invisible to the human eye. Current methods to achieve this goal mainly rely on Generative Adversarial Networks (GANs). However, these models are known to suffer from some shortcomings such as training instability and mode collapse. Furthermore, the lack of robustness to invert a real image into the latent of a trained GAN prevents flexible editing of real images. In this work, we propose PhenDiff, an image-to-image translation method based on conditional diffusion models to identify subtle phenotypes in microscopy images. We evaluate this approach on biological datasets against previous work such as CycleGAN. We show that PhenDiff outperforms this baseline in terms of quality and diversity of the generated images. We then apply this method to display invisible phenotypic changes triggered by a rare neurodevelopmental disorder on microscopy images of organoids. Altogether, we demonstrate that PhenDiff is able to perform high quality biological image-to-image translation allowing to spot subtle phenotype variations on a real image.

***Index Terms*—** Diffusion models, image-to-image translation, microscopy images, phenotypes

1. INTRODUCTION

Spotting subtle visual cell variations in biological images has many applications, from understanding diseases processes to identifiying novel biomarkers for drug discovery and diagnostic. When considering microscopic experiments, this task can be difficult. In this context, the cell-to-cell variability within a condition often hides the cell-to-cell variability between conditions. This issue represents a burden for the researchers because it prevents them from observing and quantifying properly subtle visual phenotypic differences between cells in two different conditions. Recently, generative models have been explored to overcome this issue by leveraging image-to-image translation methods.

Image-to-image translation consists in transferring an image from a source domain to a target domain. This technique was adapted to many applications. Recently, image-to-image translation was applied to spot subtle phenotypic changes in microscopy images [1, 2]. The key idea is that image-to-image translation methods can discard the biological variability and keep only the differences due to the phenotypic variations. Using these models is a great opportunity for biology to discover sounder and more subtle phenotypes and biomarkers.

So far, image-to-image translation models used in the context of spotting cell phenotype variations mainly rely on Generative Adversarial Networks (GANs) [1, 2]. However, these models are known to suffer from important issues such as training instability and mode collapse. Specifically, methods such as CycleGAN [3] employ two GANs which makes them even more unstable. Besides, existing methods based on conditional GANs such as PhenExplain [1] do not make it possible to spot phenotypic differences on real image directly. Recently, diffusion models have drawn a great attention thanks to the quality of images they produce and the stability of their training.

In this work we propose **PhenDiff**, a method based on conditional diffusion models to translate real images of cells from one condition to the other in order to spot invisible cell phenotype differences. The code is available on GitHub [4].

In the following sections, we briefly describe the state-of-the-art in image translation, present our method, and report quantitative and qualitative evaluations on biological images.

2. RELATED WORK

2.1. Image-to-Image Translation

Image-to-image translation has many applications in various fields. This has led to the development of a plethora of methods mostly based on GANs [5]. In [6] the authors proposed pix2pix, a conditional GAN supervised by pairs of images in two modalities. However, it is difficult or even impossible to collect paired images datasets. To alleviate this limitation,

* Equal contribution
⋆ Correspondence: auguste.genovesio@ens.psl.eu

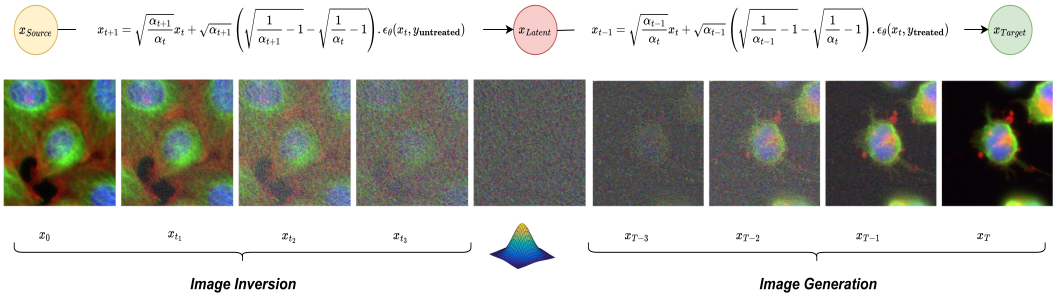

Fig. 1. PhenDiff consists of 2 phases: an *inversion* stage where a *true* image from the source condition is inverted into a Gaussian sample, and a *generation* phase where this sample is turned back into a *(synthetic)* image in the target condition. In this example a real image of untreated cells from the BBBC021 dataset is transformed into an image of cells treated by Latrunculin B.

unpaired image-to-image translation methods were proposed such as CycleGAN [3], DiscoGAN [7] and DualGAN [8] by leveraging the cycle consistency constraint [3]. In addition to GANs, other generative models were used, and more recently diffusion models were explored for this task [9, 10].

2.2. Diffusion Models

Diffusion models [11, 12] have emerged recently and rapidly became the new state-of the art family of generative models. This is due to the fact that they have shown outstanding capabilities in various generation tasks. Diffusion models can be approached using three predominant formulations, namely: denoising diffusion models [11, 12], score based generative models [13] and stochastic differential equations [14]. These approaches share a common principle: data are progressively perturbed with gradually increasing random noise; the goal of the models is then to successively learn to remove noise in order to generate new data samples.

2.3. Identifying Phenotypes in Biological Images

Identifying visual cell changes in different conditions in microscopy images is important in biology. Hand-crafted features were proposed in the past to obtain cellular profiles after a tidy cell detection step [15]. With the rise of deep learning and its success in image analysis tasks, convolutional encoders [16] were used to extract rich features directly from whole images[17, 18]. Importantly, GANs [5] were recently adopted to reveal subtle phenotypes in biological images. In PhenExplain [1] a conditional StyleGAN was trained to learn the subtle cell morphological differences between various conditions such as anti-cancer drugs at low concentration, a neuronal mutation or infected red blood cells, but the last displayed these differences on synthetic images because precise GAN inversion remains an open problem. In SCGAN [2], the authors used an improved version of CycleGAN to perform image-to-image translation from a real image with a limited amount of data. The authors replaced the discriminators of the CycleGAN with self-supervised ones to alleviate the need of large training datasets. While it offered to display subtle differences on real images, it still suffered from GAN limitations such as mode collapse and training instability. In PhenDiff, we use conditional diffusion models [11, 10] which alleviate these issues, to perform image-to-image translation.

3. METHOD

PhenDiff is built on *Denoising Diffusion Implicit Models* (DDIMs) [12]. It comprises two operations: image inversion and conditional image generation as shown in fig. 1. These two operations are performed using the same neural network, namely a UNET [19] which is previously trained as a noise predictor. A similar approach was proposed in DDIBs [10] where the authors proposed an image-to-image translation method based on diffusion models that relies on two diffusion models trained independently on each domain. DDIBs first obtain latent encodings for source images with the source diffusion model, and then decode such encodings using the target model to construct target images. Our method can be seen as a special case of DDIBs where we train a single conditional diffusion models on both domains simultaneously.

In this section we first provide an overview of diffusion models and then dive in the details of this approach.

3.1. Background

Denoising Diffusion Probabilistic Models (DDPMs) are one of the earliest and most successful diffusion models. They are latent variable models that make use of two Markovian processes: a forward process that turns data to noise and a backward process that use the later to learn to turn noise to data.

Formally, given a data distribution $x_0 \sim q(x_0)$, the forward process increasingly perturbs data by adding a Gaussian noise to it at successive timestamps. We can factorize the joint

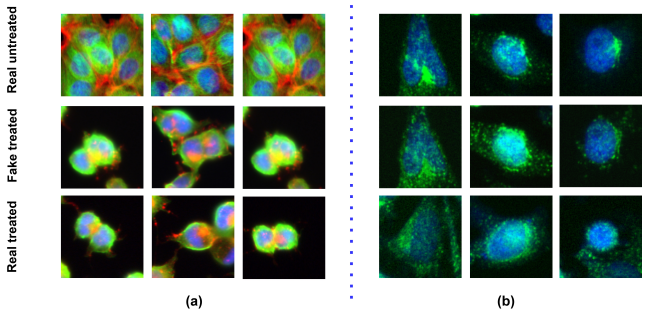

Fig. 2. Untreated cells images (1st row) of the BBBC021 and Golgi datasets (a and b respectively) are translated to images of treated cells (2nd row, Latrunculin B and Biotin were used to treat BBBC021 and Golgi cells respectively). The last row displays real images of treated cells for comparison. We can see that PhenDiff is able to learn the phenotypes relative to drug treatments and apply them on specific cell samples.

distribution of $x_1, x_2, ..., x_T$, the noised images at timestamps $1, 2, ..., T$ conditioned on x_0, as follows:

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{1}$$

where the transition kernel $q(x_t|x_{t-1})$ incrementally transform the data distribution $q(x_0)$ into a Gaussian one, the transition kernel is given as follows:

$$q(x_t|x_0) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{2}$$

where β_t are hyperparmaters previously selected.

In the backward process the noise is gradually removed by using a learnable transition kernel given by:

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{3}$$

Similar to latent variable models, diffusion models can be trained using the Variational Lower Bound (VLB). In *DDPMs* [11], the authors derived the following simplified objective function to minimize:

$$\mathbb{E}_{t,x_t,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2] \tag{4}$$

where ϵ_θ is a learnt function that predicts $\epsilon \sim N(\mathbf{0}, \mathbf{I})$ given x_t and t. Usually, it is parmeterized by a UNet[19] network. DDPMs were able to achieve high quality images. Compared to GANs, diffusion models ensures a more stable training and more diverse samples. However, they require many iterations at inference time to generate good images. To speed up the process, Denoising Diffusion Implicit Models (DDIMs) were introduced. Indeed, they can be seen as a generalization of DDPMs where a non-Markovian forward process is considered. The authors showed that using these non-Markovian processes lead to the same surrogate objective function as DDPMs. In addition to the fast sampling, DDIMs enjoy other compelling proprieties such as an exact inversion which we take advantage of in our approach.

3.2. Conditional Image Generation

Sampling from a diffusion model corresponds to gradually removing noise from noised images. As described in [20], when using the DDIMs formulation, given x_t, a noised image at the timestamp t, the denoised version of it at the timestamp $t-1$ is given by the following formula:

$$x_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}}x_t + \gamma\epsilon_\theta(x_t, t, y) \tag{5}$$

where $\gamma = \left(\sqrt{\frac{1}{\alpha_{t-1}} - 1} - \sqrt{\frac{1}{\alpha_t} - 1}\right)$, $\epsilon_\theta(x_t, y)$ is the noise predicted by the UNet and $\alpha_t = \prod_{i=1}^{t}(1-\beta_i)$, we repeat this operation starting from x_T which corresponds to a pure Gaussian noise to x_0 which is the generated image. From this equation we can see that generation is deterministic, meaning that repeating the process from the same Gaussian noise produces the same image. The class label y is added to the UNet to make the generation conditional.

3.3. Image Inversion

Image inversion is the task of finding the latent code that generates a given real image, it plays a major role in image editing models. GANs inversion methods are based either on optimization or on learning an image-to-latent encoder [21]. Despite recent progresses, GANs inversion remains challenging due to the reduced dimensionality of the latent space in comparison to the image pixel space.

Diffusion models are less sensitive to this issue because the Gaussian noise and the generated image are of the same dimension. Some diffusion model based image translation methods relied on perturbing the initial image with Gaussian noise. Thanks to DDIMs, the inversion is more accurate. Indeed, the inverted latent code can be obtained in the limit of small steps using the following formula:

$$x_{t+1} = \sqrt{\frac{\alpha_{t+1}}{\alpha_t}}x_t + \bar{\gamma}\epsilon_\theta(x_t, t, y) \tag{6}$$

where $\bar{\gamma} = \left(\sqrt{\frac{1}{\alpha_{t+1}} - 1} - \sqrt{\frac{1}{\alpha_t} - 1}\right)$.

4. EXPERIMENTS

4.1. Datasets

BBBC021: Microscopy images of MCF-7 cancer cells untreated (DMSO) and treated for 24h with Cytochalasin B at high dosages. The training set size of each condition is 1685 images [22].

Dataset	Method	FID↓	KID↓	Precision↑	Recall↑
BBBC021	CycleGAN	50.47	0.046	0.011	0.062
	PhenDiff	**36.92**	**0.036**	**0.1513**	**0.272**
Golgi	CycleGAN	12.13	0.01	0.68	0.55
	PhenDiff	**8.89**	**0.008**	**0.71**	**0.56**

Table 1. Comparison of the performance of PhenDiff to CycleGAN trained on the BBBC2021 and Golgi datasets based on FID, KID, Precision and Recall. The metrics are computed using the fake images of treated cells (obtained from the translation of images of untreated cells) and the real images of treated cells. Best results are in **bold**.

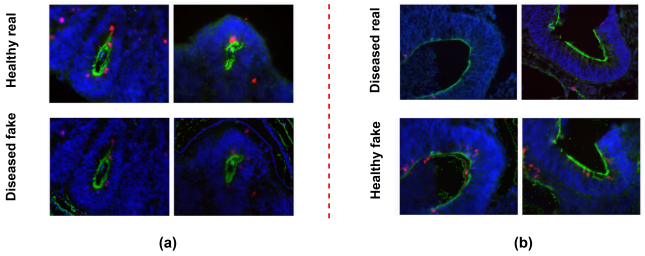

Fig. 3. In (a), we translated real images of healthy organoids to fake images of sick ones. We observe a decrease in the number of divisions (marked in red) and also a decrease in the density of nuclei (marked in blue). On the contrary, in (b) we translate images of sick organoids to images of healthy ones, which displays the reverse phenotypic change, namely an increase in the number of divisions and an increase in nuclei density.

Organoids: Microscopy images of neural organoids induced from the stem cells of a rare neuro-developmental disorder. Training was performed on 56 images of healthy organoids and 83 images of diseased organoids [23].
Golgi: Microscopy images of HeLa cells untreated (DMSO) and treated with Biotin. We used 5000 images for each condition [24].

4.2. Training

We used CycleGAN as a baseline method. We did not include SCGAN in the comparison because this method was conceived to achieve better results that CycleGAN only in low data regimes (around 100 images per class). Indeed, it was showed in [2] that CycleGAN and SCGAN are equivalent when enough data is available. Wa also did not compare our method to PhenExplain because it does not perform image-to-image on real images.

To evaluate our method, we used the FID , KID, Precision and Recall scores to assess the fidelity and the diversity of the generated images. First, we trained our method on Golgi and BBBC021 datasets that display obvious phenotypic variations. To showcase the efficacy of PhenDiff, we translated real untreated images to fake treated ones and compared them to the real treated images. Finally, we trained our method on the organoid datasets to identify subtle phenotypes that are invisible to the human eye.

4.3. Results

Qualitative results provided in fig. 2 show that the method is able to learn the effect of the drug in both BBBCO21 and Golgi datasets. Indeed, we can observe in the generated images that an obvious change in phenotypes induced by these drugs can be artificially applied to an image of real cells.

Quantitatively, we can observe in table 1 that our method performs significantly better than CycleGAN. In fact, we can see that Phendiff achieves better FID, KID, Precision and Recall scores, this indicates that the translated images produces better quality and more diverse images compared to CycleGAN.

In Fig. 3, we applied our approach on invisible phenotypic variations that may occur between two conditions on organoids. The translated images showcase the following changes: 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 cells labeled with a red marker 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.

5. CONCLUSION

In this work, we presented PhenDiff, an image-to-image translation method based on conditional diffusion model to decipher invisible phenotype variations in biological images. We showed through experiments that our method outperforms CycleGAN in terms of the quality and the diversity of the translated images. Finally, we showed how PhenDiff can be applied in biology to identify subtle phenotypes that are invisible to the human eye. Altogether, generative models can be very effective in guiding the intuition of experts to understand subtle biological processes or to identify new therapeutic biomarkers.

6. ACKNOWLEDGMENTS

This work was supported by the RUBINeuroStress ANR-19-CE16-0030, the Fondation de la Recherche Médicale (FRM Équipe labellisée Equ201903007924), ANR–10–LABX–54 MEMOLIFE and ANR–10 IDEX 0001 –02 PSL* Université Paris and was granted access to the HPC resources of IDRIS under the allocation 2020- AD011011495 made by GENCI. AB was funded for a PhD Fellowship by the Fondation de la Recherche Médicale (FRM Equ201903007924). TB was funded by the Data Intelligence Institute of Paris. KD by the Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI).

7. REFERENCES

[1] Alexis Lamiable et al., "Revealing invisible cell phenotypes with conditional generative modeling," *Nature Communications*, vol. 14, 2022.

[2] Anis Bourou and Auguste Genovesio, "Unpaired image-to-image translation with limited data to reveal subtle phenotypes," 2023.

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2020.

[4] Thomas Boyer, "Phendiff," github.com/ Warmongering-Beaver/PhenDiff, 2023.

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," 2014.

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks," 2018.

[7] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim, "Learning to discover cross-domain relations with generative adversarial networks," 2017.

[8] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," 2018.

[9] Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai, "Bbdm: Image-to-image translation with brownian bridge diffusion models," 2023.

[10] Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon, "Dual diffusion implicit bridges for image-to-image translation," in *The Eleventh International Conference on Learning Representations*, 2023.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," 2020.

[12] Jiaming Song, Chenlin Meng, and Stefano Ermon, "Denoising diffusion implicit models," 2022.

[13] Yang Song and Stefano Ermon, "Generative modeling by estimating gradients of the data distribution," 2020.

[14] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, "Score-based generative modeling through stochastic differential equations," 2021.

[15] Anne Carpenter et al, "Cellprofiler: Image analysis software for identifying and quantifying cell phenotypes," *Genome biology*, vol. 7, pp. R100, 02 2006.

[16] Zewen Li, Wenjie Yang, Shouheng Peng, and Fan Liu, "A survey of convolutional neural networks: Analysis, applications, and prospects," 2020.

[17] Ihab Bendidi, Adrien Bardes, Ethan Cohen, Alexis Lamiable, Guillaume Bollot, and Auguste Genovesio, "No free lunch in self supervised representation learning," 2023.

[18] Umar Masud, Ethan Cohen, Ihab Bendidi, Guillaume Bollot, and Auguste Genovesio, "Comparison of semi-supervised learning methods for high content screening quality control," 2022.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[20] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or, "Null-text inversion for editing real images using guided diffusion models," 2022.

[21] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang, "Gan inversion: A survey," 2022.

[22] Peter D. Caie et al., "High-Content Phenotypic Profiling of Drug Response Signatures across Distinct Cancer Cells," *Molecular Cancer Therapeutics*, vol. 9, no. 6, pp. 1913–1926, 06 2010.

[23] Daupin K. et al. de Thonel A., Ahlskog J.K., "CBP-HSF2 structural and functional interplay in Rubinstein-Taybi neurodevelopmental disorder," *Nat Commun 13, 7002 (2022)*, 2022.

[24] Gaelle Boncompain et al., "Targeting ccr5 trafficking to inhibit hiv-1 infection," *Science Advances*, vol. 5, no. 10, pp. eaax0821, 2019.