
CytoGAN: Synthesis and Representation Learning of Microscopy Cell Images

Peter Goldsborough
Facebook, Inc.
Menlo Park, CA, USA
psag@fb.com

Nick Pawlowski
Imperial College London
London, UK
n.pawlowski16@imperial.ac.uk

Juan C. Caicedo
Imaging Platform
Broad Institute of MIT and Harvard
Cambridge, MA, USA
jccaicedo@broadinstitute.org

Shantanu Singh
Imaging Platform
Broad Institute of MIT and Harvard
Cambridge, MA, USA
shsingh@broadinstitute.org

Anne E Carpenter
Imaging Platform
Broad Institute of MIT and Harvard
Cambridge, MA, USA
anne@broadinstitute.org

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Advances in high-throughput microscopy systems have enabled acquisition of large volumes of high-resolution cell images, paving the way for novel computational methods that can leverage these large quantities of data, both in order to assist biologists with tasks such as morphological profiling, which aims to capture quantitative differences in the phenotypes of cellular perturbations, as well as automate certain image analysis pipelines, like mechanism-of-action (MOA) prediction for drug compounds, in their entirety, end-to-end.

Morphology-based profiling can be stated as the task of mapping microscopic cell images to representations that capture the most salient features and axes of variation of each cell's phenotype in an unsupervised manner, ideally partitioning the representation space cleanly into clusters of cells with similar visible properties and in the case of drug compounds, similar function. Two current methods that tackle this challenge successfully are specialized software like CellProfiler, which extracts representations via manually tuned features and traditional computer vision pipelines, as well as, more recently, transfer learning approaches using features learned by deep convolutional neural networks originally trained to classify miscellaneous objects.

Being an unsupervised learning problem, morphological profiling naturally invites investigation into the effectiveness of suitable approaches employing deep neural network architectures. Indeed, there have been attempts at capturing representations from cell images using autoencoders and deep variational methods, albeit with limited success. A recent addition to the unsupervised learner's

toolbox is the adversarial framework introduced by Goodfellow et al. Generative adversarial networks (GANs) in particular have shown to excel at the task of synthesizing realistic natural images. In this work, we present findings on the effectiveness of GANs for two tasks in the domain of cell microscopy imaging: First, we show that GANs can generate highly realistic and biologically consistent cell images while providing greater intuition and interpretability over the synthesis mechanism than other generative methods. Second, we investigate ways to capture the internal representations of adversarial networks trained on fluorescence microscopy images of cultured human cells as a solution to morphological profiling and evaluate the quality of learned features for the task of predicting the mechanism-of-action of drug compounds.

2 Related Work

There exists a respectable body of work in the two lines of research that our work intersects: automated morphological profiling, and representation learning with deep neural networks, in particular generative architectures.

Caicedo et al. recently outlined the state and challenges of the morphological profiling problem in [], where the authors describe best practices for profiling pipelines using both traditional as well as deep learning methods. Prior to this, Ljosa et al. contrast the quality of representations composed of manually selected features with various augmentations and evaluate them at the task of MOA prediction on the BBBC021 dataset, consisting of microscopic images of human breast cancer cells (MCF7) treated with different compounds at differing concentrations. Both [Singh] and [Mike Ando] build on this work by showing how correcting illumination biases in images of the same dataset and applying statistical whitening transformations to learned representations, respectively, can ameliorate results reported by [Ljosa et al]. In [Nick], Pawlowski et al. for the first time report of a representation-learning method based on deep learning that is competitive with hand-picked features at the task of MOA prediction. In this work, feature layers from Inception-v3 networks trained on assorted consumer images are treated as cell representations, leveling prediction accuracy reported by [Ljosa].

Outside the domain of biological images, solutions to unsupervised representation learning have been proposed by [Hinton, Salakuthdinov] in the form of autoencoders and its many variations such as denoising autoencoders [], variational autoencoders [] and stacked autoencoders []. Transfer learning of convolutional network features as employed by [Nick] had first been investigated by [someone]. Preceding all of the above, Hinton et al. studied the effectiveness of restricted boltzman machines and deep belief networks at the representation learning task. More recently, the ladder network architecture was proposed by Rasmus to learn representations in a semi-supervised setting with promising results.

While both morphological profiling and unsupervised representation learning have been studied extensively in isolation, few works have synthesized the two. Nick et al. first investigated autoencoders and variational autoencoders to capture semantic representations of microscopic cell images in the BBBC021 dataset but reported results far inferior to hand-tuned features or transfer-learning approaches. The work most similar to ours is [], where Osokin et al. first study the task of cell image synthesis using generative adversarial networks and show that generated images are visually realistic and the latent space semantically smooth and biologically interpretable.

3 Cell Image Synthesis

In this section, we present our methodologies and findings for the task of generating realistic cell images with adversarial architectures. We begin by introducing the general GAN framework, outline the nature of the data we train on, and finally discuss our results as well as the qualitative and quantitative differences between a number of GAN models we tried.

Generative adversarial networks were first described in a seminal work by Goodfellow et al. [Foo]. The architecture they present consists of two *players*, a *generator* network G and a *discriminator* or *critic* network D . The former receives samples $\mathbf{z} \sim P(\mathbf{z})$ drawn from a noise prior $P(\mathbf{z})$, typically a multi-variate uniform or normal distribution, which it maps to values $G(\mathbf{z})$ from the generator distribution P_{model} . The discriminator network must learn to distinguish such synthetic samples from real values \mathbf{x} drawn from a data prior $P_{data}(\mathbf{x})$, with the goal of maximizing the distance between

the generated distribution P_{model} and data prior P_{data} . Feedback from the discriminator is in turn provided to the generator, which follows the opposite goal of fooling the discriminator into believing its synthesized samples stem from P_{data} . This formulation leads to a zero-sum game where the loss L_G of the generator network is the negative of the discriminator’s loss L_D , with the game’s solution lying in a nash equilibrium found by playing the minimax game

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{data}} [\log(D(\mathbf{x}))] + \mathbb{E}_{z \sim P_{noise}} [\log(1 - D(G(\mathbf{z})))]$$

We evaluate the GAN framework at the task of synthesizing realistic and biologically consistent cell images. Our data prior P_{data} consists of microscopy images from the BBBC021 dataset of MCF7 human breast cancer cell lines [] that were treated with 113 different drug molecules at 8 concentrations. Three *channels* are available for each image in BBBC021, corresponding to DNA, F-Actin and B-Tubulin structures obtained from fluorescence microscopy. We stack these channels and treat them as RGB images via a simple $DNA \mapsto Red, Actin \mapsto Green, B-Tubulin \mapsto Blue$ mapping. Images in BBBC021 contain multiple cells and are not segmented. We segment them using the CellProfiler software, such that P_{data} consists of segmented, single-cell images. Finally, we normalize the luminance of each channel to the range $[0 - 255]$ to effectively brighten the image, which improved the quality of generated images in our experience.

We evaluated two *families* of GAN architectures. The first includes the Deep Convolutional GAN (DCGAN) model proposed by Radford et al., which models the generator network as a series of convolution and deconvolution (up-sampling?) operations that successively transform and up-sample the noise sample z into an image $G(z)$ that ideally resembles those in P_{data} . The discriminator D is a conventional deep convolutional neural network that maps images x to a probability $D(x)$ representing the discriminator’s confidence that the image stems from P_{data} . In general, the GAN game is notoriously unstable and difficult to optimize, as either player can overpower the other, leading to the discriminator giving the generator no chance to improve, or the generator exploiting only a single mode of the data distribution – a phenomenon known as *mode collapse* (this sentence needs to be better!!). We experienced this instability first-hand with the DCGAN model, with mode-collapse occurring frequently and very early on in the training process, leading to dissatisfying results. Derivations of the DCGAN such as the Least Squares (LSGAN) [cite] or Wasserstein (WGAN) [cite] architecture address the inherent instability of the GAN game and we were able to train both models to convergence. We report best qualitative results for the LSGAN model, which produced highly realistic and biologically consistent synthetic images, as shown in Figure 1.

The second family of models we successfully trained to convergence that we deem noteworthy is the class of Boundary Equilibrium GAN [cite] models proposed by [Foobar et al]. BEGAN replaces the discriminator in DCGAN with an autoencoder and uses the reconstruction loss of this autoencoder as a measure of whether the image stems from P_{data} or P_{model} .

Besides generating realistic cell images, we additionally found that interpolations between noise samples drawn from P_{noise} were visually smooth. We draw two variables z_1, z_2 from the noise prior P_{noise} and sample images for noise samples at various distances between z_1 and z_2 . In all of our experiments, we used a multi-variate Gaussian distribution for the noise distribution, thus we use the spherical linear interpolation strategy from [cite] to transition between z_1 and z_2 . Figure 2 shows that images change in a smooth fashion between the two points of the noise space.

Radford et al. originally reported on intriguing relationships found between regions of the noise distribution, where algebra on vectors in the noise space translated directly to corresponding semantic changes in generated images. This is similar to results studied by Mikolov et al. for word embeddings, where distances between words in the embedding space proved surprisingly rich and linear in nature. We present similar findings in Figure 3 (have to make).

4 Representation Learning

In this section, we investigate the ability of GANs to infer vector representations of images for the purpose of profiling. We evaluate the quality of learned representations in both quantitative and qualitative experiments. First, we discuss the performance of GAN representations in the well-studied task of MOA prediction. Then, we show that representation learned by GANs have interesting semantic relationships.

The original GAN framework proposed by Goodfellow et al. does not incorporate an explicit means of inference. As such, we require extensions to traditional GAN architectures that allow us to map images drawn either from P_{model} or P_{data} to latent representations via some *encoding* function $E(x)$. The simplest such extension is to add a hidden layer to the end of the discriminator network D that induces a latent space. The intuition being that the discriminator must learn very strong internal representations in order to perform its task of separating synthetic from real images. Another benefit of this strategy is that it adds almost no computational overhead, since all but the final layer of the discriminator is shared with the encoder. Since the BEGAN architecture models the discriminator as an autoencoder, it has a bottleneck layer that naturally provides latent representations. We compute MOA classification accuracy for LSGANs with an added hidden layer, as well as BEGANs, and report our findings in Table 1.

We found that the latent representations learned by either of the two architectures did not outperform other current methods such as transfer learning or mean CellProfiler profiles. Even though MOA classification accuracy is not yet competitive, we believe that representation learning of cell images provides benefits over established methods:

- Less manual tuning and faster than CellProfiler,
- More tuned to actual data since it is trained on cell images, providing more room for biological extensions,
- Results are better than for other generative models such as VAEs

5 Conclusion

We investigate use cases of generative adversarial networks (GANs) for the domain of cell microscopy imaging. First, we show that GANs have the ability to synthesize highly realistic images of cells that respect natural phenomena and explain how this is biologically useful (NOT YET). Second, we break ground on the ability of GANs to learn representations of cell images and establish their performance at the task of drug compound mechanism-of-action prediction. We discuss that even though the representations learnt by adversarial models are currently inferior at this task, GANs provide valuable benefits over each of the competing approaches.

We emphasize that this work has shed light only on the tip of the iceberg of how the adversarial framework is useful for the field of computational biology. Due to the generality of the adversarial framework, there remain many more avenues to explore. For example, we hope that future work investigates recently proposed bidirectional GAN models or other, novel approaches to inferring representations under the GAN umbrella. Finally, we note that deep neural network architectures such as those described in this work flourish with large amounts of data. As such, we conjecture that the performance of approaches outlined in this work will improve if and when larger biological image datasets become available.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to small (9 point) when listing the references. **Remember that you can go over 8 pages as long as the subsequent ones contain *only* cited references.**

- [1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.
- [2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural Simulation System*. New York: TELOS/Springer-Verlag.
- [3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.