Generative Modeling of Cell Images

Peter Goldsborough*
Broad Institute of MIT and Harvard
pgoldsbo@broadinstitute.org

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

Juan C Caicedo & Shantanu Singh & Anne E Carpenter
Broad Institute of MIT and Harvard
{jccaicedo, shsingh, anne}@broadinstitute.org

Abstract

We explore the application of Generative Adversarial Networks to the domain of morphological profiling of human cultured cells imaged by fluorescence microscopy. When evaluated for their ability to group cell images responding to treatment by chemicals of known classes, we find that adversarially learned representations are superior to autoencoder-based approaches. While currently inferior to classical computer vision and transfer learning, the adversarial framework enables useful and previously impossible interpretability of latent codes.

1 Introduction

Advances in high-throughput microscopy systems have enabled acquisition of large volumes of high-resolution cell images [**?**]. This paves the way for novel computational methods that can leverage these large quantities of data to study biological systems. Our work focuses on the task of morphological profiling, which aims to map microscopy images of cells to representation vectors that capture salient features and axes of variation in an unsupervised manner [**?**]. These representations ideally divide the morphological space into clusters of cells with similar properties or, in the case of induced perturbations, similar function.

Current techniques for morphological profiling broadly fall in two categories: a) Classical image processing, using specialized software like CellProfiler [**?**], to capture representations via manually-tuned segmentation and traditional computer vision pipelines, and b) transfer learning to extract features learned by deep convolutional neural networks originally trained to classify miscellaneous objects [**? ?**]. Classical computer vision approaches offer better interpretability of features but require more human tuning of the segmentation algorithms, and are limited by the feature set implemented in the image analysis software. Current transfer learning approaches have been shown to outperform classical methods in at least one dataset [**?**]. However, given that these networks were trained on natural images (in RGB), they do not discover the relations of the biologically meaningful image channels. Instead, their superior performance is likely due to the their ability to capture high-level vision features, which represent the overall structure of cells, but not necessarily intricate details of their morphological variations. Therefore, we hypothesized that learning representations specifically tuned to cell images would be valuable.

We employ Generative Adversarial Networks (GANs) [**?**] to build a generative model of cell images from the BBBC021 dataset [**? ?**]. We show that this generative model learns rich feature representations and synthesizes realistic images with interpretable properties. Our approach offers the following advantages:

*This work has been conducted as part of an internship at the Imaging Platform of the Broad Institute.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

- *Adaptability*: Unsupervised representation learning, which enables easy adaptation and incorporation of new data. This is important given that applications of morphological profiling do not have ground truth annotations, but instead aim to reveal the structure of the image collection.
- *Specialization*: Our method learns from training data and is able to extract inherent semantic relationships. Transfer learning approaches lack this ability and cannot capture the intrinsic relationships between the biologically meaningful channels.
- *Interpretability*: The generative abilities of this model enable previously impossible interpretbility by being able to visualize cells dependent on a given representation.

2 Related Work

Our work lies at the intersection of 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. Prior to this, Ljosa et al. [**?**] compared the performance of various dimensionality reduction methods for CellProfiler features on the BBBC021 benchmark [**? ?**], consisting of MCF7 cells exposed to different chemical treatments. Singh et al. [**?**] demonstrated the importance of illumination correction of those images. Pawlowski et al. [**?**] for the first time reported a representation-learning method based on deep learning that is competitive with hand-engineered features at the task of MOA prediction. Ando et al. [**?**] also applied transfer learning with a different architecture and introducing a novel feature normalization method. Further related work on this dataset spans supervised classification [**?**] or deep learning applications based on CellProfiler features [**? ?**]

Few works have applied unsupervised deep learning techniques to the task of feature extraction in morphological profiling. Pawlowski [**?**] first investigated autoencoder-based approaches but reported results far inferior to hand-tuned features and transfer-learning approaches. Our model is more related to the work by Osokin et al. [**?**] and Johnson et al. [**?**], wherein GANs were used to model cell images, although their applications did not include morphological profiling.

3 Using GANs for Interpretable Representation Learning

Goodfellow et al. [**?**] introduced GANs as a game of two *players*: a *generator* G and a *discriminator* D. The former receives samples $\mathbf{z}$ drawn from a noise prior $P(\mathbf{z})$) which it maps to values $G(z)$ that should resemble elements of some data distribution P_{data}. The discriminator must learn to distinguish such synthetic samples from real values $\mathbf{x} \sim P_{\text{data}}(\mathbf{x})$. The critic's confidence in the realism of the generator's productions is used as feedback to G, guiding it to synthesize ever more realistic replicates of samples from the data prior. This procedure is formalized in a zero-sum game, solved in a nash equilibrium found by playing the minimax game

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

Radford et al. [**?**] first specialized GANs to image synthesis by introducing Deep Convolutional GANs (DCGANs). DCGANs implement the generator and discriminator based on convolutional operations. Derivations of DCGANs such as Least Squares GAN (LSGAN) [**?**] or Wasserstein GAN (WGAN) [**?**] tackle infamous instabilities in the training procedure of early generative models. In our experiments, we found LSGAN to be most stable, in part leading to higher quality generated images than both DCGAN and WGAN.

The original GAN framework does not include an explicit means of performing inference. As such, we require extensions that allow mapping of a sample $\mathbf{x}$ drawn from the data prior to a latent representation via some encoder transformation $E(\mathbf{x})$. A common approach is to interpret the penultimate layer of the discriminator as latent space. The activations of this layer serve as a source of representation vectors. These latent codes are thought to be meaningful because the discriminator must develop a strong internal representation of its input in order to succeed at its discrimination task. Furthermore, this method imposes no computational overhead compared to the vanilla GAN models.

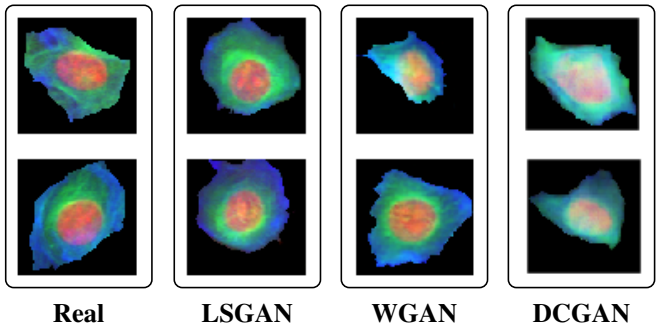

Figure 1: Examples of real MCF7 cells from the BBBC021 dataset juxtaposed with synthetic cell images generated with LSGAN, WGAN and DCGAN. The generated images of LSGAN are most realistic and consistent with biological properties of cells.

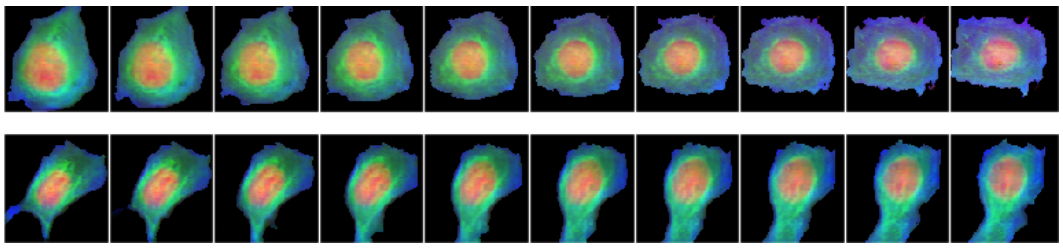

Figure 2: Interpolating between two points $\mathbf{z_1}, \mathbf{z_2} \sim P(\mathbf{z})$) results in visually smooth transitions in the corresponding synthesized images. Each row shows the transition from $\mathbf{z_1}$ to $\mathbf{z_2}$ from left to right.

4 Interpretable Synthesis of Cells

We begin by examining if GAN's reported ability to synthesize realistic natural images transfers to the domain of cell microscopy. Our data prior consists of images drawn from the BBBC021 dataset of human breast cancer cells. Each image in BBBC021 consists of three *channels* corresponding to DNA, F-Actin and B-Tubulin structures. We stack the channels and treat them as RGB images via a simple DNA $\mapsto$ R, B-Tubulin $\mapsto$ G, F-Actin $\mapsto$ B mapping. Images in BBBC021 contain many cells and are not segmented. We segment them using CellProfiler, such that P_{data} consists of segmented, single-cell images. Finally, we normalize the luminance of each channel to the range [0 – 1].

Figure 1 shows examples of images generated with LSGAN, WGAN and DCGAN architectures alongside real images. The synthetic images, particularly those produced by LSGAN, are not only highly detailed and realistic, but also consistent with their biological nature. For example, it is characteristic that B-Tubulin forms a circular halo cradling the nucleus of a cell. This property is maintained clearly in all generated images.

Current approaches to morphological profiling capture representations that are often hard to interpret. While certain CellProfiler features such as nucleic area are easy to understand, others like Zernike moments are not intuitive. For transfer learning it is nearly impossible to visualize the captured concept. In contrast, the GAN noise space $P(z)$ has been shown to be highly interpretable and reveal rich semantic structure [**?**]. We are able to demonstrate this for images of cells. Figure **??** exhibits how interpolating between two noise samples $z_1, z_2 \sim P(z)$ leads to visually smooth transitions in synthesized cells. Figure 3 shows that $P(z)$ encodes semantic relationships, enabling algebra on interpretable properties of the images. While there is no way to encode images into $P(\mathbf{z})$ with the methods presented so far, we believe more advanced models that enable this will be highly valuable if they can maintain these rich semantic properties.

5 Representation Learning for Morphological Profiling

We test the quality of representations extracted by the discriminator used to generate single cell images learned by testing their ability to cluster treatments of similar function or mechanism-of-action. We

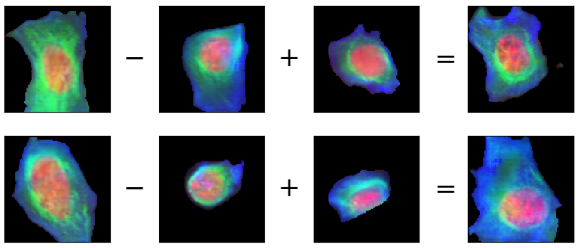

Figure 3: Vector algebra in the noise space translates to biologically valuable relationships in generated images. In the first row, subtracting the vector for a cell with small quantities of B-Tubulin (green stain) from one with high amounts yields a vector representation of "high B-Tubulin content". In the second row, the difference between a large and small nucleus encodes the semantic meaning of "larger nucleus", which can be added to vectors of other cell images to grow the size of the nucleus.

DCGAN	WGAN	LSGAN	VAE [**?**]	CP [**?**]	Deep Transfer [**?**]	Deep Transfer [**?**]
43 %	45 %	63 %	49 %	90 %	91 %	96 %

Table 1: Accuracy of various methods at the task of predicting the mechanism-of-action of treatments via nearest-neighbor classification in the BBBC021 dataset. CP refers to the results by Singh et al. [**?**] using CellProfiler Features. The Deep Transfer methods correspond to Deep Feature Transfer as proposed by Pawlowski et al. [**?**] and refined by Ando et al. [**?**].

obtain treatment profiles by averaging the extracted single cell representations found as intermediate activations of the discriminator. Further, we follow the experimental protocol of [**?**] and report an average mechanism-of-action classification accuracy of a leave-one-compound-out cross-validation using a one nearest neighbor classifier.

Table 1 compares the accuracy of our approach to classical CellProfiler features and transfer learning. We found that the quality of representations found by unsupervised learning is not yet competitive with other established methods. Nevertheless, we believe that the adversarial approach has significant benefits regarding its interpretability as outlined above. Additionally, this framework is able to adapt to the dataset at hand and extract inherent relations that are not captured by previous methods. We believe that further gains can be made, as the image quality corresponds to classification accuracy. As such the best performing method (LSGAN) also yields the highest image quality, as judged qualitatively by expert biologists and shown in Figure 1.

6 Conclusion

This work investigates the use of GANs for the domain of cell microscopy imaging, particularly morphological profiling. First, we demonstrate that GANs are able to synthesize biologically realistic images of cells. In addition, we explain that GANs induce a rich semantic structure in their noise space, providing a powerful basis for interpretability of observed properties. Second, we add an encoding component to standard GANs and evaluate the quality of learned representations via their mechanism-of-action classification performance following [**?**]. Even though adversarially learned representations are currently inferior at this task, we argue that further enhancements to GANs, especially biologically meaningful ones, will lead to higher quality latent representations.

We emphasize that this work covers only a very small fraction of possible applications of GANs to the domain of microscopical images. For example, we hope that future work investigates BiGANs [**?**] or other novel solutions to infer latent features with GANs. We believe that improved inference, combined with the interpretable nature of the GAN framework, may enable simulated experiments by performing algebra with vectors corresponding to cell lines, diseases or other perturbations. Finally, we note that deep learning generally flourishes with larger amounts of data. As such, we conjecture that the performance of the outlined approaches will improve with the use of larger datasets.

Acknowledgements

Nick Pawlowski is supported by Microsoft Research through its PhD Scholarship Program and the EPSRC Centre for Doctoral Training in High Performance Embedded and Distributed Systems (HiPEDS, Grant Reference EP/L016796/1).