Semi-Supervised Generative Adversarial Network (SGAN) for Nuclei Detection on Breast Cancer Histopathology Images

Victor M. Vargas

Seidenberg School of CSIS, Pace University.

New York City, NY.

victor@vmvargas.com

Jonathan Keller

Seidenberg School of CSIS, Pace University.

New York City, NY.

jmk4411@gmail.com

*Abstract*—We extended the work of an SGAN, combined it with the implementation of a DCGAN, made some adjustments in the model to work with a H&E breast cancer histopathology images dataset from the Case Western Reserve University. Our goal was to evaluate if the Discriminator (or semi-supervised model) could achieve a state-of-the-art at finding the probability of a given input image patch corresponding to a nucleus or not. Our results were promising but more research is necessary to conclude if this approach can outperform state-of-the-art methods such as the SSAE.

Keywords— Feature representation learning; automated nuclei detection; semi-supervised approach; Generative Adversarial Network; breast cancer histopathology;

# Introduction

Nuclei Detection allows researchers to identify each individual cell in a sample and by measuring how cells react to various treatments, the researcher can understand the underlying biological processes at work. The analysis of histopathology images is currently the standard in diagnosing Breast Cancer (BC). This fact is a convincing motivation to discover, enhanced, and automated efficient approaches to distinguish individual cancer nuclei on breast pathology images.

Getting large amounts of unlabeled medical data is generally much easier than labeled data. Unsupervised generative models with stochastic components (like GANs and VAEs) can be trained end-to-end to learn representative features in a completely unsupervised way. For that reason, both approaches could optimally leverage this amount of information.

The rest of the paper is organized as follows: A review of similar architectures and previous related works is presented in Section II. A detailed description of Semi-Supervised Generative Adversarial Network (SGAN) is presented in Section III. The experimental setup and comparative strategies are discussed in Section IV. The experiment results and discussions are reported in Section V. Conclusions and future work are presented in Section VI.

# Previous Related Work

## Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images

An interesting approach, especially in cases where object annotation to generate training data is expensive, is the integration of multiple instance learning (MIL) and deep learning. Xu et al. [1] investigated the use of a MIL-framework with both supervised and unsupervised feature learning approaches as well as handcrafted features. The results demonstrated that the performance of the MIL-framework was superior to handcrafted features, which in turn closely approaches the performance of a fully supervised method.

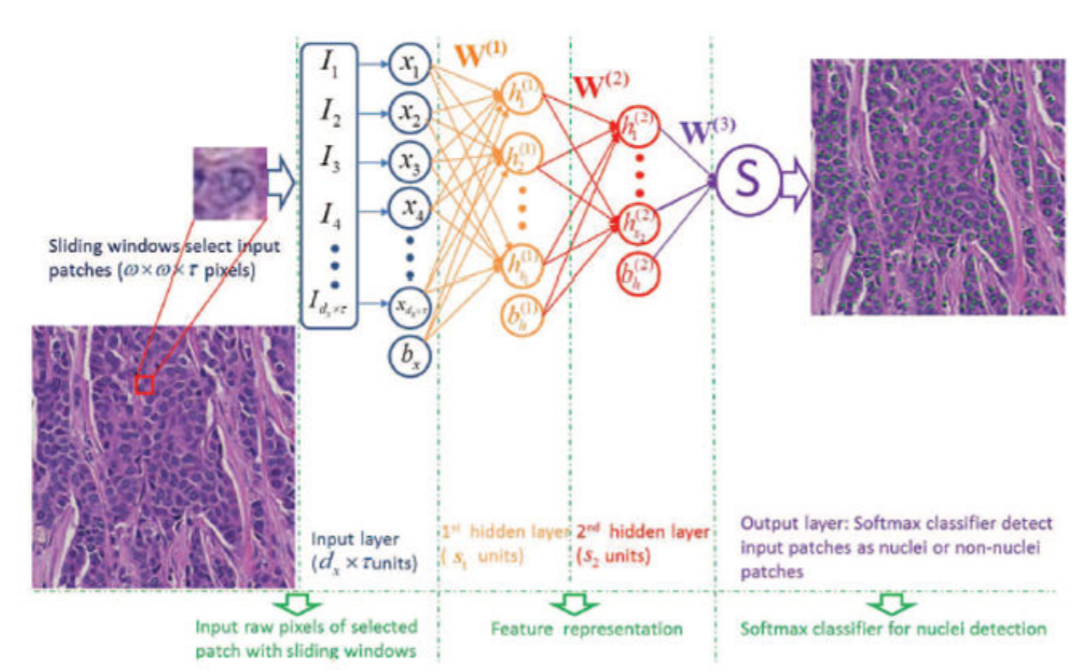


Fig. . Illustration of SSAE+SMC for nuclei detection on breast histopathology

Training an SSAE involves finding the optimal parameters θ= (W, bh, bx) simultaneously by minimizing the discrepancy between input and its reconstruction. After the optimal parameters θ are obtained, the SSAE yields a function that transforms input pixel intensities of an image patch to a new feature representation of nuclear structures.

As Fig. 1 shows, with SSAE, each training patch of pixel intensities is represented by a high-level structured representation of nuclei or non-nuclei patches (2) in the second hidden layer of the model. Note that in the SSAE learning procedure, the label information Y is not used. Hence, SSAE learning is an unsupervised learning scheme.

During detection process, each image patch detected by a sliding window is first represented by high-level feature (2). This is then fed to the SMC and produces a value between 0 and 1 that can be interpreted as the probability of the input image patch corresponding to a nucleus or not.

## Generative Adversarial Networks (GAN)

GANs [2] are based on a game theoretic scenario in which the generator network must compete against an adversary. The generator network (G) directly produces samples Its adversary, the discriminator network (D), attempts to distinguish between samples drawn from the training data and samples drawn from the generator. The discriminator emits a probability value given by , indicating the probability that is a real training example rather than a fake sample drawn from the model.

The representations that can be learned by a GAN may be used in a variety of applications, including image synthesis, semantic image editing, style transfer, image super-resolution, and classification [3]. Expanding these ideas, one can produce good output samples using a set of convolutional neural networks (Denton et al., 2015). Some years ago, (Radford et al., 2015), created surprisingly good samples from a single generator network.

## DCGAN

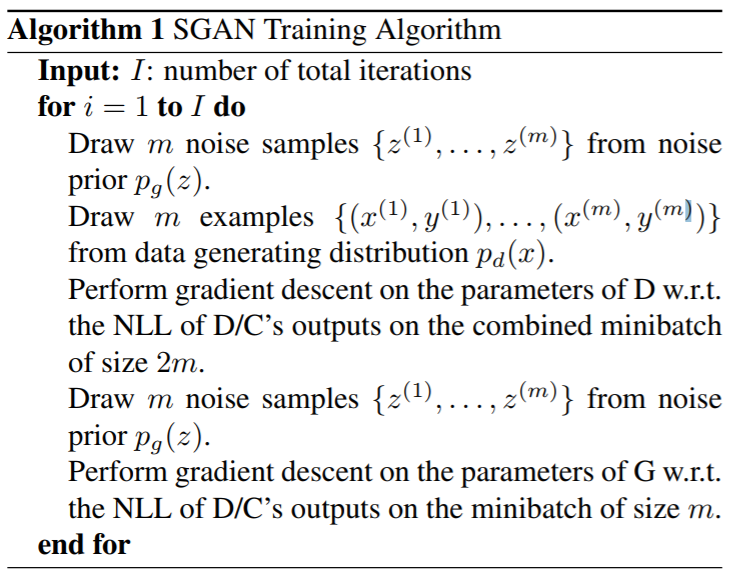
Several recent papers focus on improving the stability of training and the resulting perceptual quality of GAN samples [1, 2, 3, 4, 5]. Among these, [Radford et al. (2016](https://arxiv.org/pdf/1511.06434.pdf)) main contribution came from a set of practices that prove to stabilize the training of GAN by: (1) replacing deterministic spatial pooling functions (such as maxpooling) with strided convolutions, (2) eliminating fully connected layers on top of convolutional features, and (3) not applying Batch Normalization (Ioffe & Szegedy, 2015) to the generator output layer and the discriminator input layer.

# SGAN

[Odena, (2016)](https://arxiv.org/pdf/1606.01583.pdf) extended Generative Adversarial Networks (GANs) to the semi-supervised context by forcing the discriminator network to output N+1 different output classes, N different “real” classes, and an additional fake class (anything that came from the generator).

Using generative models on semi-supervised learning tasks is not a new idea - Kingma et al. (2014)[6] expand work on variational generative techniques (Kingma & Welling, 2013[7]; Rezende et al., 2014[8]) to do just that. However, Odena, (2016) described a new extension called SGAN that improves classification performance on restricted data sets over a baseline classifier with no generative component.

According to Odena, 2016 Training an SGAN is similar to training a GAN. We simply use higher granularity labels for the half of the minibatch that has been drawn from the data generating distribution. D is trained to minimize the negative log likelihood with respect to the given labels and G is trained to maximize it, as shown in Algorithm 1.



# Experimental Setup

We conducted experiments on MNIST to see whether the classifier component of our customized SGAN would perform correctly and similar to the original implementation. Afterwards, we conducted experiments on the TMI Dataset (the Breast Cancer histopathology images).

The experiments in this paper were conducted with <https://github.com/vmvargas/GAN-for-Nuclei-Detection/>, which borrows heavily from <https://github.com/eriklindernoren/Keras-GAN/tree/master/sgan> and which contains more details about the experimental setup.

## Parameter setting

To train the model with the TMI dataset, the patch size was initially defined as 34×34 pixels but due to G input constraints, every image from the training and testing dataset was downscale to 32×32 = 1024 pixels which is big enough to contain a nucleus within the patch under 40X optical magnification resolution images. Each patch size has three color channels. Therefore, τ = 3. Therefore, there are dx = s0 = 32×32×3 = 1024×3 input units in the input layer.

Regarding the optimizer, we used the Adam optimizer (Kingma & Ba, 2014) with tuned hyper-parameters. We use 0.0002 as the learning rate. And 0.5 as momentum term β1, which helped stabilize training. Both, G and D used the same optimizer. Figure X and X depicts the architecture of both models. Both are very similar to the DCGAN. That means that we used all architecture guidelines suggested by Radford:

* Replace any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator).
* Use batchnorm in both the generator and the discriminator.
* Remove fully connected hidden layers for deeper architectures.
* Use ReLU activation in generator for all layers except for the output, which uses Tanh.
* Use LeakyReLU activation in the discriminator for all layers.

A notably difference relies on G first layer which is an 8×8×128 this size assures that the generator outputs an image with equal size as the training images.

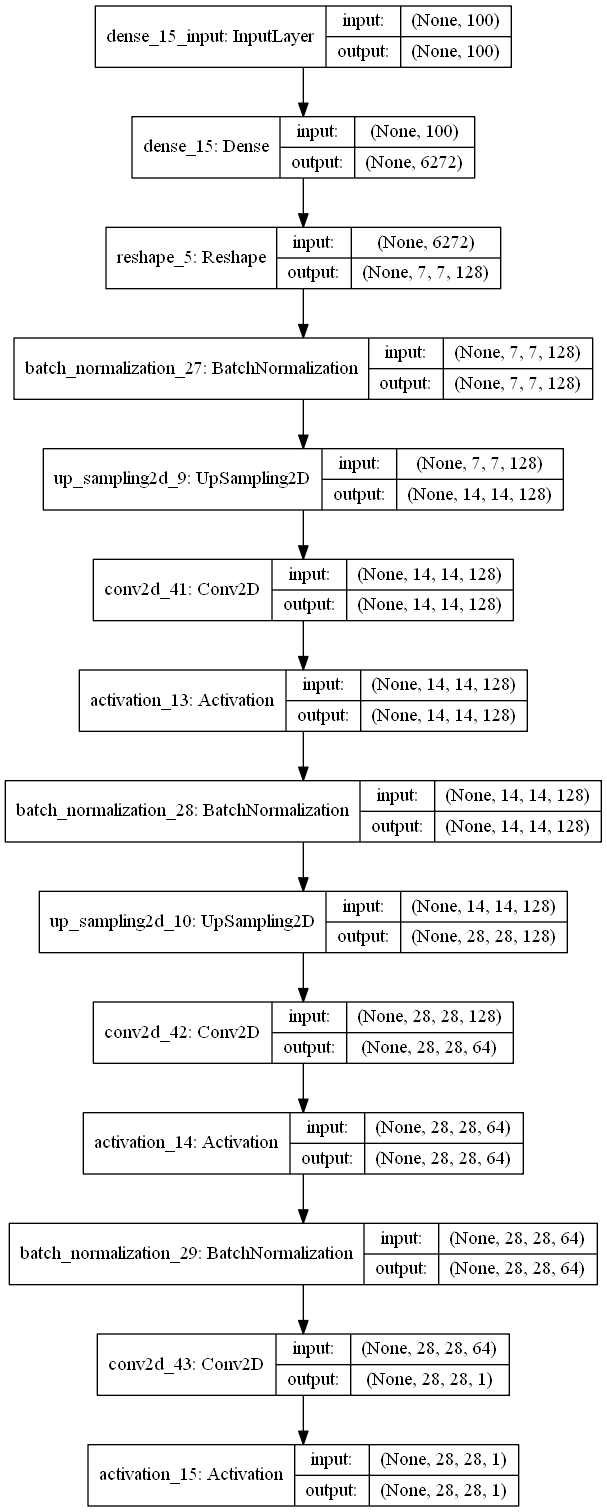


Figure G

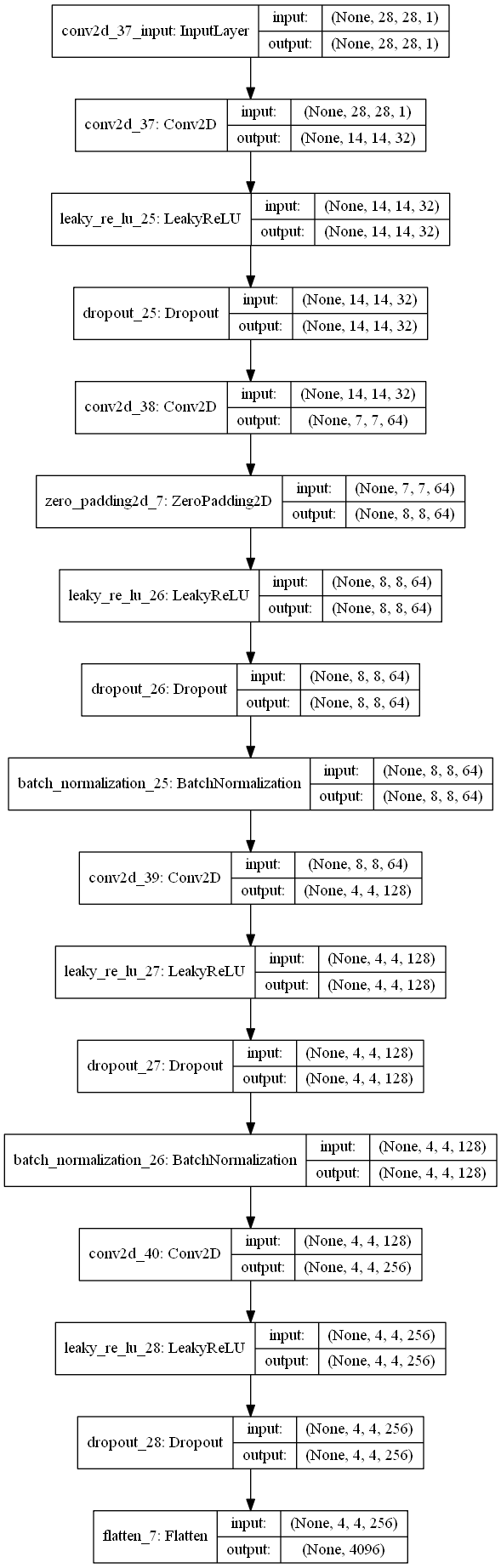


Figure D

Evaluating D [loss: 0.8791, bi-loss: 0.6904, cat-loss: 1.0678, bi-acc: 90.20%, cat-acc: 94.10%]

Overall accuracy: 94.100000%

Classification report:

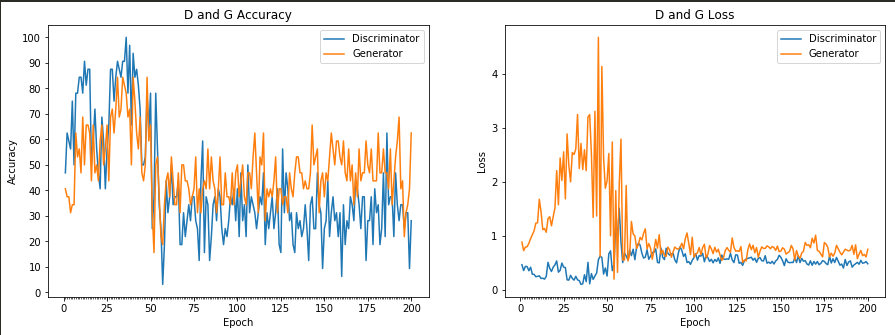
precision recall f1-score support

Non-nunclei 0.95 0.93 0.94 500

Nuclei 0.93 0.95 0.94 500

avg / total 0.94 0.94 0.94 1000





Training time: 6.2 minutes

192: Training D [loss: 0.4888, acc: 28.12% ] - G [loss: 0.8374, acc: 68.75%]

Evaluating D [loss: 0.8791, bi-loss: 0.6908, cat-loss: 1.0674, bi-acc: 85.60%, cat-acc: 93.70%]

193: Training D [loss: 0.5131, acc: 34.38% ] - G [loss: 0.5816, acc: 40.62%]

Evaluating D [loss: 0.8785, bi-loss: 0.6900, cat-loss: 1.0670, bi-acc: 91.50%, cat-acc: 93.60%]

194: Training D [loss: 0.4780, acc: 34.38% ] - G [loss: 0.6678, acc: 43.75%]

Evaluating D [loss: 0.8787, bi-loss: 0.6907, cat-loss: 1.0668, bi-acc: 86.30%, cat-acc: 93.90%]

195: Training D [loss: 0.5511, acc: 25.00% ] - G [loss: 0.7340, acc: 21.88%]

Evaluating D [loss: 0.8788, bi-loss: 0.6905, cat-loss: 1.0670, bi-acc: 86.30%, cat-acc: 93.60%]

196: Training D [loss: 0.4971, acc: 31.25% ] - G [loss: 0.6384, acc: 31.25%]

Evaluating D [loss: 0.8792, bi-loss: 0.6911, cat-loss: 1.0673, bi-acc: 80.70%, cat-acc: 93.70%]

197: Training D [loss: 0.5044, acc: 31.25% ] - G [loss: 0.6614, acc: 34.38%]

Evaluating D [loss: 0.8792, bi-loss: 0.6910, cat-loss: 1.0674, bi-acc: 82.30%, cat-acc: 94.00%]

198: Training D [loss: 0.5279, acc: 9.38% ] - G [loss: 0.6140, acc: 40.62%]

Evaluating D [loss: 0.8790, bi-loss: 0.6904, cat-loss: 1.0676, bi-acc: 89.20%, cat-acc: 94.10%]

199: Training D [loss: 0.4869, acc: 28.12% ] - G [loss: 0.7555, acc: 62.50%]

Evaluating D [loss: 0.8791, bi-loss: 0.6904, cat-loss: 1.0678, bi-acc: 90.20%, cat-acc: 94.10%]

## TMI Dataset

In order to be able to compare the experimental results with the SSAE model, the experimental setup for the dataset was almost identical to [Xu et al](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4729702/pdf/nihms750311.pdf). (2014). We use the same dataset of 537 H&E stained histopathological images were obtained from digitized glass slides corresponding to 49 lymph node-negative and estrogen receptor-positive breast cancer (LN-, ER+ BC) patients at Case Western Reserve University. The training data includes 14421 nuclear and 28032 non-nuclear patches. There are 516 testing data in this dataset. The size of each testing image is 2200\*2200 pixels. In each image, a Region of Interest (ROI) of 400\*400 pixels is chosen for validation. For each ROI, an expert proceeded to meticulously place a dot in the center of each nucleus. Consequently, quantitative evaluation of the different models was limited to these 500 ROIs across the 500 images.

The complete access to the full dataset is provided with this link: <http://engineering.case.edu/centers/ccipd/data>.

## Training the SGAN for Nuclei Detection

The training procedure was for 1250 epochs over the MNIST dataset, with a batch size of 32 images. For each epoch, we selected a random half batch of images from the training set (16 images) and another random half samples from a Gaussian distribution, the latest half batch is pass to the G and rendered as images. D classified real samples as 1s and fakes (samples coming from G) as 0s

to train the discriminator

To compute accuracy, we took the maximum of the outputs not corresponding to the FAKE label. For each model, we did a random search on the learning rate and reported the best result

MENTION TMI SETUP

AND Model Architecture.

# Experimental Results

SGAN versus SSAE.

Qualitative results.

Quantitative results.

Sensitivity Analysis.

Computational Consideration.

# Conclusion and future work

Conclusion.

We are excited to explore the following related ideas:

* …

##### Acknowledgment

...

##### References

1. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., & Madabhushi, A. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Transactions on Medical Imaging, 35(1), 119-130. [7163353]. DOI: 10.1109/TMI.2015.2458702
2. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Networks. ArXiv e-prints, June 2014.
3. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," in IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, Jan. 2018.
4. Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. arXiv preprint arXiv:1506.05751, 2015.
5. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.
6. Ian J Goodfellow. On distinguishability criteria for estimating generative models. arXiv preprint arXiv:1412.6515, 2014.
7. Daniel Jiwoong Im, Chris Dongjoo Kim, Hui Jiang, and Roland Memisevic. Generating images with recurrent adversarial networks. arXiv preprint arXiv:1602.05110, 2016.
8. Donggeun Yoo, Namil Kim, Sunggyun Park, Anthony S Paek, and In So Kweon. Pixel-level domain transfer. arXiv preprint arXiv:1603.07442, 2016.
9. Kingma, D. P and Welling, M. Auto-Encoding Variational Bayes. ArXiv e-prints, December 2013.
10. Kingma, Diederik P., Rezende, Danilo Jimenez, Mohamed, Shakir, and Welling, Max. Semi-supervised learning with deep generative models. CoRR, abs/1406.5298, 2014. URL <http://arxiv.org/abs/1406.5298>.
11. Litjens, G. et al. A survey on deep learning in medical image analysis. arXiv preprint arXiv:1702.05747v2 (2017).
12. Rezende, D., Mohamed, S., and Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. ArXiv e-prints, January 2014.