SynCGAN: Using Learnable Class Specific Priors to Generate Synthetic Data for Improving Classifier Performance on Cytological Images

Soumyajyoti Dey[1], Soham Das[1], Swarnendu Ghosh[1], Shyamali Mitra[1],
Sukanta Chakrabarty[2], and Nibaran Das[1(✉)]

[1] Jadavpur University, Kolkata 700032, WB, India
{soumyajyoti.cse.rs,shyamalimitra.iee,nibaran.das}@jadavpuruniversity.in,
soham.das0000@gmail.com, swarbir@gmail.com
[2] Theism Medical Diagnostics Centre, Kolkata 700030, India
drsukantachakraborty@gmail.com

Abstract. One of the most challenging aspects of medical image analysis is the lack of a high quantity of annotated data. This makes it difficult for deep learning algorithms to perform well due to a lack of variations in the input space. While generative adversarial networks have shown promise in the field of synthetic data generation, but without a carefully designed prior the generation procedure can not be performed well. In the proposed approach we have demonstrated the use of automatically generated segmentation masks as learnable class-specific priors to guide a conditional GAN for the generation of patho-realistic samples for cytology image. We have observed that augmentation of data using the proposed pipeline called "SynCGAN" improves the performance of state of the art classifiers such as ResNet-152, DenseNet-161, Inception-V3 significantly.

Keywords: Conditional generative adversarial networks (CGAN) · Synthetic data generation · Cytology image classification · Deep learning

1 Introduction

The modern machine learning algorithms such as deep learning, have been greatly dependent on the availability of a large amount of high-quality data. But for various niche domains such as medical imaging large quantities of data are generally unavailable due to various constraints, such as lack of patients, infrastructural inadequacy, noisy environments, lack of experts for annotations and so on. However, with the advent of *g*enerative *a*dversarial *n*etworks (GANs) [6], an avenue for high quality data generation has opened. In its base form, GANs

are capable of generating samples from a randomly sampled prior which demonstrates likeliness to a predefined data distribution. However, without proper guidance, the generation process can result in eccentric outputs. However, conditional GANs (CGANs) [12], on the other hand, use a semantically sensible prior for guiding the data generation process to generate more accurate and meaningful samples. In the proposed work, we explore the ability of CGANs to work with learnable priors for efficient data generation to improve classifier performance on cytology images. In most practical cases the number of available data samples is too limited for deep learning approaches to thrive. Thus data augmentation serves as a primary tool for improving learning ability. Though annotating pixel specific masks for cytology images is a difficult and expensive job, however, with adequate expertise and a decent amount of labor it is possible to annotate at least a small batch of samples for a better semantic representation. The proposed approach makes use of such semantic masks to serve as a prior for CGANs. While generating fully detailed cytology images without priors is much difficult, the generation of segmentation masks from scratch is a much simpler task given that the output distribution is binomial. Our proposed approach makes use of this factor to create learnable segmentation masks that can guide CGANs for synthetic data generation. Some relevant studies are discussed in the next section. The proposed methodology is provided in the subsequent section followed by experimental setups, results, and discussions in Sect. 4 and future scopes are discussed in the conclusion thereafter.

2   Related Works

Most common methods for data augmentation involved affine transformations [5] such as translation, rotation, scaling, shear, flipping and so on. Also, it has been noticed that training with added noise results in a much more robust classifier. The introduction of generative adversarial networks (GANs) has brought a shift in the paradigm of generative processes in computer vision. Several approaches have made use of GANs for data augmentation. Adar et al. [3] proposed a GAN based liver lesion data augmentation technique where after the extraction of ROI for classification was done using CNN. Dataset was augmented in two ways: i) the ROIs were augmented by affine transformations ii) the synthetic data was generated from ROIs using DCGAN (Deep Convolution Generative adversarial Network) and ACGAN(Auxiliary Classifier GAN). DCGAN showed greater performance compared to ACGAN. Shin, et al. [13] proposed a GAN based model to segment tumor of brain MRI images of two traditional datasets: ADNI and BRATs. Normal brain MRI images were segmented using an image to image translation model using CGAN [9]. Synthetic abnormal brain MRI scans were obtained from labels (tumors) by incorporating some changes in the label (e.g. increasing the size, changing the position of the label, or placing the tumor in a healthy brain MRI segmentation map). The synthetic images were used in data augmentation for training the model. Improved performance of tumor segmentation was observed by adding the synthetic data to the real data but

without using normal data augmentation methods. Tom et al. [15] simulated
patho-realistic ultrasound images of the IVUS dataset using deep generative
models. Tissue echogenicity maps were generated from the ground truth of the
dataset. From these maps simulation of ultrasound images was produced using a
physics-based simulator. Two-staged GAN was used to generate patho-realistic
ultrasound images and stability of training state. In the first stage, images from
the simulator were taken as input to GAN from which low-resolution images
were generated. In the second stage, these low-resolution images from the first
stage of GAN were transformed into high-resolution images. Bissoto et al. [2]
suggested a GAN based model to generate high-resolution images of skin lesion
of the ISIC challenge dataset. Classifiers were trained on real data as well as on
synthetic data.

3   Proposed Methodology

The goal of the current work is to generate realistic cytology images similar
to images collected during FNAC (Fine Needle Aspiration Cytology) test. The
cytology images were collected from Theism Medical Diagnostics Centre, Dum-
dum, Kolkata. These cytological data were mainly collected by FNAC test, and
were captured using an Olympus microscope at 40X magnification in the pres-
ence of the professional practitioners. Around 156 cytology images were collected
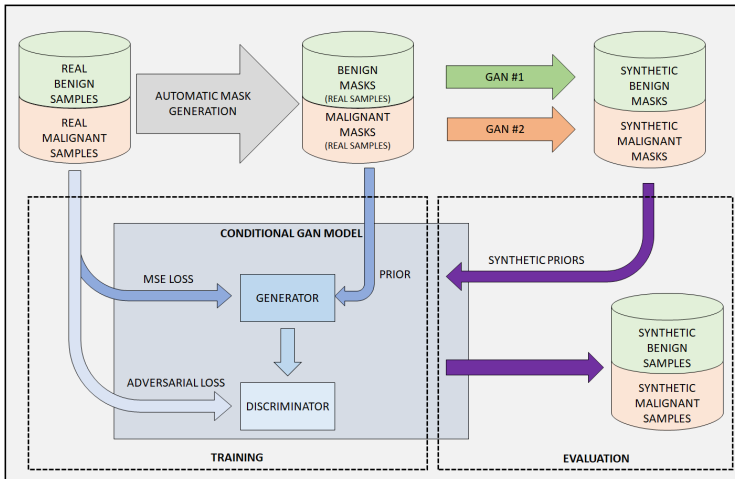among which 77 were benign samples and 79 were malignant samples (Fig. 1).

Fig. 1. SynCGAN: the proposed data augmentation pipeline

3.1 The Data Augmentation Pipeline

The proposed method of synthetic data generation consists of 3 phases. Firstly, segmentation masks are collected from real images using an unsupervised method. Secondly, a CGAN is trained on these pairs of images and auto-generated segmentation masks. Two sets of synthetic segmentation masks are generated using a GAN for each of the two classes. Finally, these synthetic segmentation masks are used to guide the previously trained CGANs to generate patho-realistic samples for data augmentation.

3.2 Mask Generation

The proposed methodology requires a set of pixel-level annotations to guide a CGAN for data generation. Due to the lack of hand-annotated samples, an unsupervised approach was used for nuclei segmentation. There have been many developments in the field of image segmentation lately [4]. For our work, first, the contrast of the RGB cytology image is increased by the histogram stretching method. The image is then converted to a grayscale image. To eliminate the irrelevant portions, adaptive thresholding [10] based segmentation algorithm is adopted and the RGB image is converted to a binary segmented mask. But the red blood cell, cytoplasm which had similar high local contrast is distinguished using the Gaussian Mixture clustering algorithm. Finally, the refined binary segmented mask is extracted. The presence of an unsupervised mask generation technique alleviates the necessity of large amount of training data (Fig. 2).
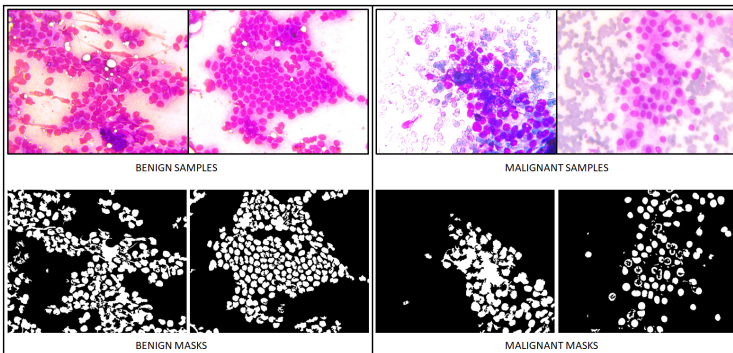

Fig. 2. Auto generated segmentation masks using [10] from real samples

3.3 Training the CGAN to Generate Patho-Realistic RGB Images from Segmentation Masks

Generating RGB images from scratch using a traditional GAN is difficult as the generation procedure can be represented as a prediction of 256-dimensional

multinomial distribution across three channels for each pixel. However, the segmentation mask is simply a pixel specific binomial distribution which is much easier to predict when starting without a predefined prior. Thus a CGAN [12] must be trained which takes the segmentation masks generated in the previous step as a prior and a generator loss is reduced against the corresponding RGB image. We derive inspiration from the pix2pix network [9]. For the generator, we use a modified UNet like architecture. Normally the UNet architecture used transposed convolution for upscaling the feature maps. However, that results in checkerboard artifacts due to overlap of kernels during the fractional stride. Instead, bilinear interpolation opted for upscaling the feature map followed by a 3×3 convolution layer for refinement. The discriminator network has been directly implemented using the PatchGAN discriminator as demonstrated in [9]. The discriminator attempts to detect real and fake samples from the dataset and the generator respectively. The objective function V can be written as:

$$\min_G \max_D V(D, G) = \mathbb{E}_{(x,y) \sim p_{data}(x,y)}[\log(D(x|y))]$$
$$+ \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D(G(y)))]. \tag{1}$$

Here G and D refers to the generator and discriminator. x represents the RGB sample, y represents the corresponding auxiliary representation which serves as a prior for the generator. The x and y samples are drawn from the input data distribution $p_{data}(x, y)$ that consists of RGB images and their auxiliary representations or segmentation masks in the current scenario. In our case, the auxiliary representations are the automatically generated segmentation masks as described in Sect. 3.2.

At every iteration, the discriminator and the generator are trained alternatively as was performed previously. During the training of the discriminator, the segmentation mask and its corresponding RGB image are concatenated on its channel dimension. It is then passed through the discriminator (Patch GAN) [9] and discriminator loss of the real image is calculated as below

$$D_{loss} = -\log(D(x)) + \log(1 - D(G(y))), \tag{2}$$

where x represent samples from the input database and y refers to segmentation masks of those samples. A binary cross-entropy loss function is used to calculate the adversarial loss. While training the generator, the segmentation masks are passed through the generator network and the loss is calculated. The loss has two components denoting the adversarial loss exhibited by the discriminator and the mean square error between the generated sample and the actual RGB image from the dataset that corresponds to the mask y.

$$G_{loss} = -\log(D(G(y))) + \lambda \operatorname{MSE}(G(y), x|y) \tag{3}$$

where λ is the weight of Mean Squared Error (MSE) loss. The weight of the mean squared error loss is set to 100 based on empirical analysis on a small validation set. x and y represent samples from the RGB image and the segmentation mask dataset.

3.4   Training the GAN to Generate Segmentation Masks

A conditional GAN(CGAN) usually generates synthetic samples conditioned by some predefined priors. In the current scenario, the CGAN has been trained to generate samples from segmentation masks highlighting the spatial distribution of nuclei across the cytoplasm. To generate patho-realistic synthetic samples during the evaluation phase a class-specific prior distribution is necessary. For that purpose, we train a GAN model [6] to generate binary segmentation masks based on a randomly drawn seed from a gaussian distribution. While models like CycleGAN [16] can be used for image translation, it is not suitable for synthetic data generation. The most straight forward method to generate synthetic samples would have been to train an end-to-end GAN. However, it has been noticed in ablation studies that without a prior the quality of outputs is very poor. The primary reason being the complexity of predicting the intensity value of a pixel. Given the output image has three channels, each pixel exists within a search space with 256^3. However, a binary segmentation mask is a much easier output to predict given that each pixel belongs to a binomial distribution. On the other hand, the shape information encoded within these segmentation masks is quite informative about the class of the samples, namely, benign or malignant. We train two separate GANs trained on segmentation masks belonging to each of the predefined classes.

The objective of the GAN network is simply defined as

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x))]$$
$$+ \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z)))] \tag{4}$$

Here x refers to samples drawn from the input data distribution $p_{data}(x)$. z refers to randomly sampled priors from a Gaussian distribution $p(z)$. G and D refers to the generator and discriminator network. The architecture of the GAN used in the current work is very similar to the one described in the previous section. It consists of a generator inspired from UNet whos transposed convolutions have been replaced with bilinear interpolation for upscaling and a convolution layer for refinement. The discriminator is derived from the PatchGAN discriminator as demonstrated in [9]. During the training phase, the discriminator loss is given by,

$$D_{Loss} = -(\log(D(x)) + \log(1 - D(G(z)))) \tag{5}$$

and the generator loss is given by:

$$G_{Loss} = -\log(D(G(z))) \tag{6}$$

However, due to very low number of samples, the discriminator was too overpowering and saturates at a very early stage. To deal with this issue some additional measures were taken as described below [11].

– **Label smoothing:** The labels for real and fake samples are set as 1 and 0 by default. To enforce some fuzziness in the system, a random number between

0.9 and 1 was taken for real samples and a random number between 0.1 and 0 was taken for fake samples while training. However, this is unnecessary while training the generator as we want to bottleneck the learning curve of the discriminator and not the generator.

– **Randomly flipping labels:** To even further confuse the discriminator at some random iterations real samples are labeled as fake and vice versa. This confusion results provide some breathing space for the generator so that it can learn the requisite features.

Other models such as Wasserstein GAN [1] can further improve results.

3.5   Evaluating on the Trained Model to Generate Synthetic Images

The final phase generates class specific patho-realistic synthetic samples. According to the pipeline discussed in Sect. 3.1, at first the class-specific GANs are used to generated segmentation masks (refer Sect. 3.4). Then the generated segmentation masks are fed into the trained CGAN model (refer Sect. 3.3) to obtain the RGB synthetic images. The synthetic data distribution is modeled as,

$$p_{data(z)} = \mathbb{E}_{z \sim p(z)}[G_{CGAN}(G_{GAN}(z))] \tag{7}$$

where, G_{CGAN} and G_{GAN} refers to the generators of the CGAN and GAN model described in Sect. 3.3 and 3.4 respectively.

While a simple CGAN trained on a prior denoting the class of samples could also be viable, having a richer representation like a mask provides more flexibility in terms of variations in the synthetic dataset.

4   Experiments and Results

The main objective of the work is to generate class-specific synthetic data similar to microscopic cytology images that can boost the performance of standard image classification algorithms. For our experiments we have used three classifiers namely, ResNet-152 [7], Inception-V3 [14] and DenseNet-161 [8]. Each of these networks has a proven track record in tough image classifications tasks such as the ILSVRC. In the original dataset, there were 156 images in total. Out of which 77 were benign samples and 79 were malignant samples. For the benefit of a cleaner calculation, a total of 150 images were selected with 75 images from each class. The dataset was randomly divided in the ratio of 3:1:1 into training, validation, and testing set with an equal number of samples in each class. The ratio of synthetic data to original training data was maintained at 2:1. Thus 90 synthetic training images were generated per class (Table 1).

4.1   Experimental Setup

The performance metric defining the goodness of the synthetic data generation pipeline, referred to as SynCGAN in the current work, is given by its impact

Table 1. Number of samples in the original and synthetic dataset

Dataset	Training	Validation	Testing
Original images	90	30	30
Synthetic images	180	–	–
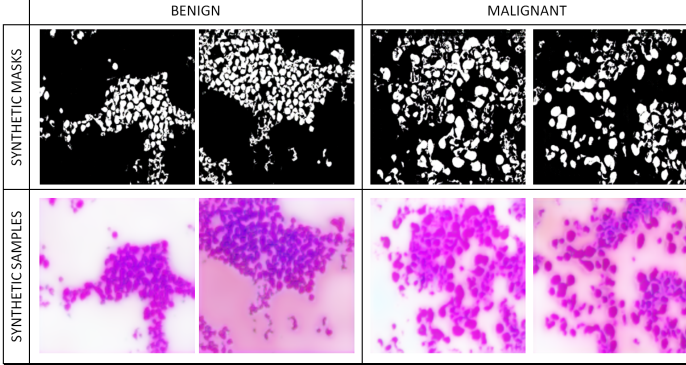
Fig. 3. Synthetic masks generated by class specific GANs (top) are fed as priors to the conditional GAN during evaluation phase to generate patho-realistic samples (bottom) for benign (left) and malignant (right) classes separately

on the test accuracy obtained using the three previously mentioned network, namely, ResNet-152, Inception-V3, and DenseNet-161. The experiment was conducted to analyze the impact of synthetic data augmentation on several grounds (Fig. 3).

1. Impact of augmentation using data generated by SynCGAN,
2. Performance of SynCGAN generated data augmentation against traditional data augmentation,
3. Performance of SynCGAN generated data augmentation against GAN generated data augmentation.

The CGAN model trained for a maximum of 200 epochs and the best model was saved based on minimum generator loss on the validation dataset. While the GAN model was trained for almost until the generator loss saturated(1600 epochs). For both the cases adam optimizer was used. All the experiments were conducted on Nvidia GTX 1060 GPU.

4.2   Observations and Analysis

The first observation as shown in Table 2, shows that augmentation of data generated with the proposed SynCGAN improves the performance of classifiers. When compared with traditional augmentation techniques like random horizontal and vertical flipping, random rotation and addition of Gaussian noise, the

proposed method of augmentation has a higher impact. When traditional data augmentation was combined with SynCGAN based augmentation, the performance was either at par or lower than exclusive SynCGAN based augmentation.

Table 2. Performance of classifiers while using the dataset with and without augmentation. Orig: Original Data, Prop: Data generated using the proposed SynCGAN pipeline, Trad: Data generated using traditional augmentation techniques

Classifier	Orig	Orig+Prop	Orig+Trad	Orig+Trad+Prop
ResNet-152	73.33	**76.67**	70.00	**76.67**
DenseNet-161	80.00	**86.67**	83.33	84.67
Inception-V3	73.33	**80.00**	63.33	76.67

Secondly, as a control to our proposed model, we implemented a purely GAN based pipeline for data augmentation, which performed far below the proposed model as shown in Table 3. This GAN based architecture also had a generator and discriminator network similar to our proposed model for a fair comparison.

Table 3. Performance of classifiers while using the dataset with and without augmentation. Orig: Original Data, Prop: Data generated using proposed SynCGAN pipeline, GAN: Data generated using vanilla GAN.

Classifier	Orig	Prop	Orig+Prop	GAN	Orig+GAN
ResNet-152	73.33	73.33	**76.67**	50.00	63.33
DenseNet-161	80.00	63.33	**86.67**	50.00	60.00
Inception-V3	73.33	66.67	**80.00**	56.67	66.67

5   Conclusion

In the present work, a CGAN based data augmentation technique has been proposed using class specific priors that improves the performance of various state of the art CNNs such as ResNet-152, DenseNet-161, and Inception-V3 on cytology images corresponding to FNAC tests. Unlike a normal GAN, we have used learnable segmentation masks as class-specific priors to guide a conditional GAN for more robust synthetic data generation. It is to be noted that the method is quite dependent on the mask generation algorithm and hence extensive studies may be performed using other available nuclei segmentation techniques for further analysis. Furthermore, the method may be further generalized to adapt to other types of cytological data.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint arXiv:1701.07875 (2017)
2. Bissoto, A., Perez, F., Valle, E., Avila, S.: Skin lesion synthesis with generative adversarial networks. In: Stoyanov, D., et al. (eds.) CARE/CLIP/OR 2.0/ISIC - 2018. LNCS, vol. 11041, pp. 294–302. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01201-4_32
3. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)
4. Ghosh, S., Das, N., Das, I.S., Maulik, U.: Understanding deep learning techniques for image segmentation. ArXiv abs/1907.06119 (2019)
5. Ghosh, S., Das, N., Nasipuri, M.: Reshaping inputs for convolutional neural network: some common and uncommon methods. Pattern Recogn. **93**, 79–94 (2019)
6. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
10. Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J.: Computer-aided diagnosis of breast cancer using gaussian mixture cytological image segmentation. J. Med. Inf. Technol. **17** (2011)
11. Kumar, S., Gupta, M.D.: C+GAN: complementary fashion item recommendation. arXiv preprint arXiv:1906.05596 (2019)
12. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
13. Shin, H.-C., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (eds.) SASHIMI 2018. LNCS, vol. 11037, pp. 1–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00536-8_1
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

15. Tom, F., Sheet, D.: Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1174–1177. IEEE (2018)
16. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)