Morpho-GAN: Generative Adversarial Networks를 사용하여 높은 형태론 데이터에 대한 비지도학습

아자맛 압두아지모프[O], 조근식[*]
[O]인하대학교 컴퓨터공학과,
[*]인하대학교 컴퓨터공학과
e-mail: azamat.abduazimov01@gmail.com[O], gsjo@inha.ac.kr[*]

Morpho-GAN: Unsupervised Learning of Data with High Morphology using Generative Adversarial Networks

Azamat Abduazimov[O], GeunSik Jo[*]
[O]Dept. of Computer Engineering, Inha University,
[*]Dept. of Computer Engineering, Inha University

● 요 약 ●

The importance of data in the development of deep learning is very high. Data with high morphological features are usually utilized in the domains where careful lens calibrations are needed by a human to capture those data. Synthesis of high morphological data for that domain can be a great asset to improve the classification accuracy of systems in the field. Unsupervised learning can be employed for this task. Generating photo-realistic objects of interest has been massively studied after Generative Adversarial Network (GAN) was introduced. In this paper, we propose Morpho-GAN, a method that unifies several GAN techniques to generate quality data of high morphology. Our method introduces a new suitable training objective in the discriminator of GAN to synthesize images that follow the distribution of the original dataset. The results demonstrate that the proposed method can generate plausible data as good as other modern baseline models while taking a less complex during training.

키워드: 생성적 적대 신경망(Generative Adversarial Network), 형태(Morphology), 딥러닝(Deep Learning)

I. Introduction

Datasets are a crucial part of Deep Learning. Major progress in this field can result from advancements in the availability of quality training datasets. Although there might not be a need for labeling, quality datasets for unsupervised learning can be hard to produce and costly. One type of these datasets is one with high morphological features (human eyes, galaxies, etc.) which usually requires meticulous lens adjustments and careful shapes measured by a human. Therefore, what if we could automate the generation of additional realistic data, that similarly reflect the true distribution of the original dataset so that we could use them as an asset in the learning process of classification systems that deal with data with morphological features.

Generative modeling of images is appropriate in our task, which is to automate the generation of data with high morphological features. Recently, exploring the use of deep learning has led to a huge number of heavily invested research in how to leverage Generative Adversarial Networks (GAN) [1] to create photo-realistic outputs learned from unlabeled data. DCGAN [2] came up with deep generative neural networks that dictate a careful hyperparameter setting to obtain natural images. HDCGAN [11] presented an architecture that exploits the use of novel activation functions in DCGAN to generate high-resolution images of human faces. Also, many variations have been proposed to train GANs stably. Recently, Wasserstein-GAN (WGAN) [7] and WGAN-GP [6] gained the learning stability using a gradient penalty. Moreover, auto-encoder based generative models have also shown their dominance by using latent variables to stabilize the convergence of GANs.

Dist-Gan [10] proposed distance constraints to address the mode collapse issue. DRAGAN [5] improved the training of GANs using a regret minimization penalty. InfoGAN [8] made modifications that maximize the mutual information between random variables and semantic features of generated data.

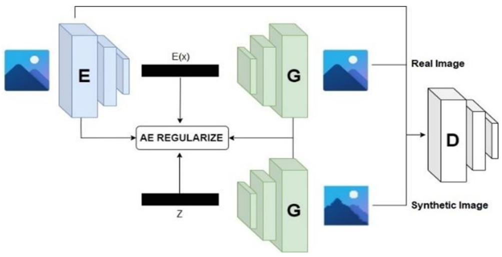

Fig. 1. Autoencoder based GAN in our model.

The contributions of this work include the following (1) Inspired by Dist-GAN and DRAGAN, we propose Morpho-GAN, a method that generates quality morphological data (2) A suitable GAN architecture, based on HDCGAN, with a new training objective.

We utilize the constraints of Dist-GAN and DRAGAN to come up with a suited training objective for Morpho-GAN. The proposed model uses HDCGAN based architecture to synthesize images with a high resolution because Dist-GAN and DRAGAN proved to be useful in dealing with usually low-resolution data.

II. Preliminaries

1. Background

Goodfellow et al. [1] proposed GAN, which comprises of a generator G and a discriminator D, training each other in an adversarial manner. GAN transforms noise latent varibles into photo-realistic images. Simultaneously, G keeps on feeding D with fake images until D can no longer tell apart the difference between a generated image and the ground truth image, after which an adversarial training reaches the convergence.

However, a fundamental challenge with GANs is to train both G and D networks in a way that none of the networks gets too powerful over one another. Unfortunately, this is the idealistic case. That is why, probability of GAN training ending up in the gradient explosion and mode collapse [10] is usually high, where G generates the same images, although given diverse input noises.

2. Related Works

2.1 DCGAN

Alec Radford et al. [2] employed a novel architecture of deep convolutional neural networks both in the generator and the discriminator for the first time. DCGAN replaced any pooling layers in both networks. The design dictates any fully connected hidden layers are discarded from the networks, too. ReLU activation function is utilized for all layers, except for the output layers of the generator (Tanh activation) and the discriminator (Sigmoid activation). Despite its achievements, DCGAN is vulnerable to mode collapse because of its high dependency on the fixed configuration of architecture and hyperparameters.

Curto et al. [11] made changes in the architecture of DCGAN by adding more deep layers into DCGAN in order to generate high-resolution face images.

2.2 Dist-GAN

Dist-GAN is a powerful method to stabilize GAN training and belongs to Autoencoder (AE) based methods. So, there are three training objectives separately: Generator objective (see Eqn. 1), Discriminator objective (see Eqn. 2), Autoencoder objective (see Eqn. 3). The encoder of AE is the generator. Distance constraints ensure the alignment of true data distribution P_x with the generated data distribution P_z.

$$\min_{\theta} L_G(\theta) = \left|\mathrm{E}_x \sigma(D_\gamma(x)) - \mathrm{E}\sigma(D_\gamma(G_\theta(z)))\right| \quad (1)$$

$$\min_{\gamma} L_D(\psi,\theta,\gamma) = -(\mathrm{E}_x \log\sigma(D_\gamma(x) + \mathrm{E}_z + \log(1 - \sigma(D_\gamma(G_\theta(z))) + \mathrm{E}_x \log\sigma D_\gamma(G_\theta(E_\psi(x))) - \lambda_1 GrP) \quad (2)$$

$$\min_{\psi,\theta} L_{AE} = \left\|x - G_\theta(E_\psi(x))\right\|_2^2 + \lambda_2\left\|f(x, G_\theta(z)) - \lambda_3 g(E_\psi(x), z)\right\|_2^2 \quad (3)$$

Here, E is the expectation, σ is the sigmoid function, x is real data, z is random noise. $\lambda_1, \lambda_2, \lambda_3$ are hyperparameters to control the scale of constraints. Eqn. 1 describes the 1st constraint that helps the generator minimize the distance between the discriminator scores assigned to real and generated images. AE is regularized through the 'latent-data distance' constraint in the 2nd term of Eqn. 3. This enforces the proportional distance differences between latent variables and the corresponding images generated from those variables. This prevents mode collapse and ensures diverse images generated from diverse noise. But the weak point of Dist-GAN is a gradient penalty inherited from WGAN-GP [6] to optimize the discriminator objective. This penalty mechanism requires multiple inner iterations to regularize D, so too heavy. Also, Dist-GAN method showed high stability with architecture based on a fixed architecture of DCGAN, which synthesizes images only size of 64x64. We need to have a much lighter gradient penalty mechanism, that can work with relaxed constraints of DCGAN as well.

III. The Proposed Model: Morpho-GAN

1. Training Objective

We propose a new training objective for the discriminator D of our autoencoder-based Morpho-GAN, inspired by [5]. The network structure is in Fig. 1. We optimize our D using the gradient penalty that applies Alternating Gradient Decent, which avoids multiple iterations to regularize D, unlike WGAN-GP [6].

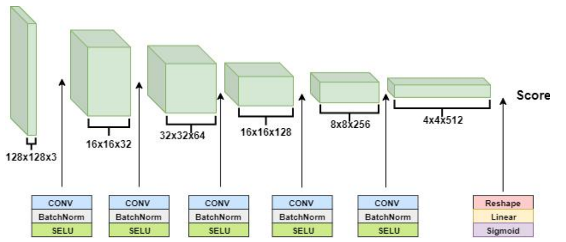

Fig. 2. Morpho-GAN discriminator network

So, a new discriminator training objective as follows:

$$\min_{\gamma} L_D(\psi,\theta,\gamma) = -(\mathrm{E}_x \log\sigma(D_\gamma(x)) + \mathrm{E}_z \log(1 - \sigma(D_\gamma(G_\theta(z))) + \mathrm{E}_x \log\sigma D_\gamma(G_\theta(E_\psi(x))) - \lambda_4 \mathrm{E}_{x\sim P_x,\alpha\sim N(0,1)}\left(\|\nabla_x D_\gamma(x+\alpha)\| - k\right)^2) \quad (4)$$

where, x is a sample real data, $x+\alpha$ is a perturbation with a small pixel-level noise α. As [8] demonstrated, DRAGAN gradient penalty uses the least computation, compared to other penalty regularized GANs, namely WGAN-GP as well. So, we can get a much simpler and faster training of our model. To implement this penalty simply, only points in local regions close to real data points are sampled, and this makes D outputs have a gradient norm-1, avoiding gradient vanishing, thus avoiding mode collapse.

2. Model Architecture of Morpho-GAN

Traditional DCGAN is only able to generate a maximum of 64x64 resolution images. When it is forced to have more layers, DCGAN based architecture usually hits the mode collapse too soon because of overpowering of D on G. Dist-GAN showed its potential using only traditional DCGAN. We empirically discover that simple Dist-GAN constraints plus our proposed discriminator objective can not fully solve the high-quality generation issue. However, Kodali et al. [5] also showed that, with DRAGAN, fixed constraints of DCGAN can be relaxed and stability still is reached. But they do not mention exact modification needed. In our case, the data we are dealing with is of high morphology, and the stability of DCGAN with deeper architecture is desirable at the same time. So the modification is to have more convolutional layers like in HDCGAN.

Morpho-GAN proposes replacing ReLUs in DCGAN with Scaled Exponential Linear Units + BatchNorm combination, inspired by HDCGAN. Curto et al. [11] used this combination to get high-quality images only for human faces. So, we make an additional modification to customize the discriminator D. We add a Gaussian noise right at the beginning of D to couple its convergence speed with that of G. Final architecture of Morpho-GAN discriminator for 128x128 resolution with more layers using the proposed modifications can be seen in Figure 2.

IV. Experiments

1. Hyperparameters

Our model empirically achieves the high performance with $\lambda_1 \sim 1, \lambda_2 \sim 1, \lambda_3 \sim \sqrt{\dfrac{100}{128*128}}, \lambda_4 \sim 10$. We use batch size 0f 64 in the training process and Adam optimizer $(\beta_1 = 0.5, \beta_2 = 0.9)$ with a learning rate of 0.0002. Weights of networks are initialized using Xavier initialization. Total number of epochs are 150.

2. Dataset

The Galaxy Zoo 2 (GZ2) dataset has a total of 141553 RGB 424x424 images. The set is divided into 61578 images for trainin and 79975 testing images. Images of the dataset are classified into 5 major but total 11 different categories. Each of the categories has attributes, and there are 37 attributes in total. These numbers represent the overall morphology of one galaxy in 37 attributes.

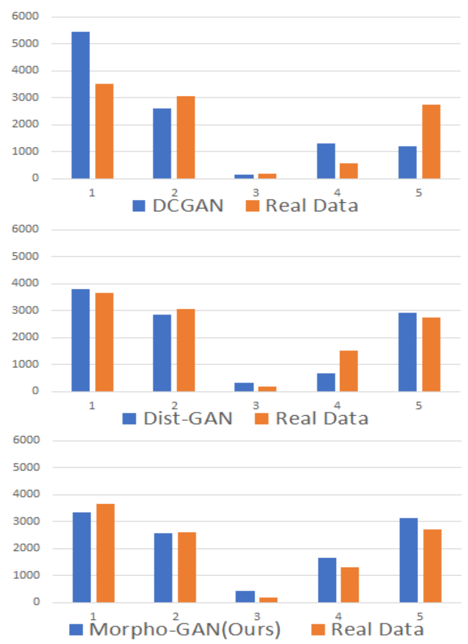

Fig. 3. Evaluation of how well models can learn the data distribution of the dataset. Images are of size 128x128.

3. Evaluation

3.1 Mode balance analysis

Fig. 4. The first column - a random generated sample. The next five columns are the nearest neighbors.

Evaluation of generated data is a challenging task. One possible quantitative evaluation method for GAN is Frechet Inception Distance (FID). To calculate the FID, they use a pre-trained Inception model to compare the distributions of generated data and real data. Inception model trained on ImageNet is not suitable for our task because 1000 categories of ImageNet do not contain a morphological data category, such as galaxy in our case.

Instead, we employ the state-of-the-art classification model trained specifically trained on Galaxy Zoo dataset with 95.2% overall accuracy [4]. We use this model to classify generated images into 5 major categories for our experiment, namely 1. complete round 2. in-between smooth 3. cigar-shaped 4. edge-on 5. spiral. We sample random 10k images from the real dataset, then generate fake samples in the same number from a generative model. We compare our method with DCGAN, Dist-GAN as a baseline on images with resolution 128x128. And we repaeat the same process for 5 times and take average counts to visualize on a diagram. We use the codes shared by the authors to implement them and make them generate 128x128 images as well. As can be seen, our Morpho-GAN (Fig. 3) closely follows the data distribution while DCGAN ends up generating too many images in the same category, which is the mode collapse. Dist-GAN shows performance similar to our model. However with simpler gradient penalty mechanism, we can say our method is computationally less heavy, with our training objective, having taken advantage of [5].

3.2 The quality and memorization analysis

It is important for a model to capture the representational features of galaxies but we do not want the model to memorize training data just to fool the discriminator. To show the data quality and learning ability of our model, we show nearest neighbors to the generated images in Fig. 4. Our method does not memorize training data, which is desirable and could generate quality 128x128 images.

V. Conclusion

We have presented Morpho-GAN, a synthetic image generation method to create data with high morphological features in this work. Inspired by several GAN training strategies, our method proposed a combined lighter training objective. A suitable architecture was build to generate quality morphological data that are not present in the dataset but closely follow the true data distribution, avoiding mode collapse.

Our future work can be extended to exploring photo-realism in images of much higher resolution and other types of data by tweaking the training model parameters in a gradual manner.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-0-01642) and supervised by the IITP (Institute for Information & communications Technology Promotion).

REFERENCES

[1] J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative Adversarial Networks," in *NIPS*, 2014.

[2] Alec Radford, Luke Metz, Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," in *ICLR*, 2017.

[3] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, Yi Zhang, "Generalization and equilibrium in generative adversarial nets (GANS)," in *ICML*, 2017.

[4] Jia-Ming Dai, Jizhou Tong, "Galaxy morphology classification with deep convolutional neural network," in *astro-ph.IM*, 2019.

[5] Naveen Kodali, Jacob Abernethy, James Hays, Zsolt Kira "On Convergence and Stability of GANs" arXiv preprint arXiv:1705.07215, 2017.

[6] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville "Improved Training of Wasserstein GANs" arXiv preprint arXiv:1704.00028, 2017

[7] Martin Arjovsky, Soumith Chintala, Léon Bottou, "Wasserstein GAN", arXiv preprint arXiv:1701.07875, 2017.

[8] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, Pieter Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", arXiv preprint arXiv:1606.03657, 2016.

[9] Mattya.chainer-gan-lib. https://github.com/pfnet-research/chainer-gan-lib

[10] Ngoc-Trung Tran, T. Bui, Ngai-Man Cheung, "Dist-GAN: An Improved GAN using Distance Constraints", in ECCV, 2018.

[11] J. D. Curtó, H. C. Zarza, T. Kim, "High-Resolution Deep Convolutional Generative Adversarial Networks", arXiv preprint arXiv: 1711.06491, 2017