Exploring Generative Adversarial Networks(GANs)

Rupali Tripathi
19BCE1508
Computer Science and Engineering
Vellore Institute of Technology,
Chennai

Felicia Sharon
19BCE1215
Computer Science and Engineering
Vellore Institute of Technology,
Chennai

Hima Rani Mathews
19BCE1532
Computer Science and Engineering
Vellore Institute of Technology,
Chennai

***Abstract*— GANs (Generative Adversarial Networks) are a new emerging technology that may be used for both supervised and unsupervised learning. They do it by implicitly modelling high-dimensional data distributions. In this project different types of GANs are used for different applications. Synthesis of Realistic image from text description is a challenging task in the field of computer vision. They accomplish this by implicitly modelling high-dimensional data distributions. Different forms of GANs are employed in this research for various applications. In the subject of computer vision, creating a realistic image from a text description is a difficult task. To create 256 ×256 photo-realistic images conditioned on text descriptions, a StackGAN is utilised. To improve the diversity of the synthesised images and stabilise the training of the conditional-GAN, a new Conditioning Augmentation approach that encourages smoothness in the latent conditioning manifold is applied.Dealing with multi-class imbalanced datasets is another area where GAN is widely employed. For synthesising high-dimensional picture data, a Semi-supervised Generative Adversarial Network (SGAN) is used as a data generator. In addition, to boost classification performance, a k-Nearest Neighbor (kNN) based selection method is used to add the most relevant samples to the original imbalanced database. Finally, the GAN and SMOTE's performance is compared.**

I. Introduction

GANs (generative adversarial networks) allow us to learn deep representations without needing a lot of training data. They accomplish this by using a competitive approach utilising two networks to derive backpropagation signals. GANs can learn representations that can be used for picture synthesis, semantic image editing, style transfer, image super-resolution, and classification, among other things.

Photo-realistic images from text are a difficult challenge with many applications, including photo editing, computer-aided design, and so on. Generative Adversarial Networks (GAN) have recently demonstrated promise in synthesising real-world images. Conditional GANs can produce images that are highly connected to the text meanings when given text descriptions. However, teaching GAN to generate high-resolution photo-realistic images from text descriptions is quite tough. Adding more upsampling layers to state-of-the-art GAN models for generating high-resolution (e.g.256×256) images produces training instability. To encourage smoothness in the latent conditioning manifold, we suggest an unique Conditioning Augmentation approach. Small random perturbations in the conditioning manifold are allowed, which increases the diversity of the synthesised image.

Data mining experts presented two sets of approaches to improve the classification performance of imbalanced data sets. Algorithm level and data level algorithms are two types of unbalanced learning solutions. Cost-sensitive learning and recognition-based learning, for example, modify the classifier itself to bias towards the minority class without modifying the original data. The cost of different types of misclassification is given proportional weights in cost sensitive learning. The purpose of this sort of learning is to keep the total cost as low as possible. To produce a balanced data distribution, oversampling and undersampling procedures are used to create or eliminate samples at the data level. When dealing with unbalanced data, traditional synthetic oversampling methods delivered state-of-the-art results. Those approaches, on the other hand, are built for low-dimensional feature space samples in binary classification scenarios and struggle to deal with high-dimensional data samples such as images, audio signals, and time series.

Deep generative models are a source of inspiration for alternative imbalanced learning methods for dealing with increasingly intricate imbalanced data. Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), two of the most prominent models for learning data distribution in an unsupervised method, have already succeeded in creating a variety of complicated data, such as handwritten digits, faces, home numbers, and CIFAR images. The prospects of using the Semi-supervised Generative Adversarial Networks (semi-GAN) model in imbalanced learning fields are investigated in this research. Semi-GAN takes picture data as input, and the k-Nearest Neighbor (kNN) approach is used to choose synthetic candidates for the minority class. The generation results are compared to the classic synthetic oversampling method-SMOTE based on accuracy.

II. Literature survey

In [1] a simple yet effective Stacked Generative Adversarial Networks is proposed. To generate high-resolution images with photo-realistic details. It decomposed the text-to-image generative process into two stages

Stage-I GAN: it sketches the primitive shape and basic colours of the object conditioned on the given text description, and draws the background layout from a random noise vector, yielding a low-resolution image.

Stage-II GAN: it corrects defects in the low-resolution image from Stage-I and completes details of the object by reading the text description again, producing a high-resolution photo-realistic image.

In [2] they proposed a novel framework Semantic-Spatial Aware GAN, which is trained in an end-to-end fashion so that the text encoder can exploit better text information. Their network performed 2 basic functions i.e (1) learns semantic-adaptive transformation conditioned on text to effectively fuse text features and image features, and (2) learns a mask map in a weakly-supervised way that depends on the current text-image fusion process in order to guide the transformation spatially.

In [3] they extended Generative Adversarial Networks (GANs) to the semi-supervised context by forcing the discriminator network to output class labels. They trained a generative model G and a discriminator D on a dataset with inputs belonging to one of the N classes. At training time, D is made to predict which of N+1 classes the input belongs to, where an extra class is added to correspond to the outputs of G. They showed that this method can be used to create a more data-efficient classifier and that it allows for generating higher quality samples than a regular GAN.

In [4] Based on the SMOTE method, this paper presented two new minority over-sampling methods, borderline-SMOTE1 and borderline-SMOTE2, in which only the minority examples near the borderline are over-sampled. For the minority class, experiments show that their approaches achieve better TP rate and F-value than SMOTE and random over-sampling methods.

III. Problem statement

In this project we will explore 2 types of GAN for 2 different types of applications. The first type of application is using StackGAN to generate 256×256 photo-realistic images conditioned on text descriptions. The second type is using a Semi-supervised GAN to generate image samples of the imbalanced 28*28 image dataset (Fashion-mnist). The goal of this approach is to solve class imbalance in an image dataset and to use knn as the criteria to add samples to the dataset. The balanced dataset will also be compared to that of the one made by SMOTE approach by building a decision tree classifier on them and comparing their accuracies.

IV. Proposed work

Text to image translation

For synthesising photo-realistic images from text descriptions, we offer a unique Stacked Generative Adversarial Networks. It breaks down the challenging task of creating high-resolution images into smaller, more manageable chunks, considerably improving the state of the art. For the first time, StackGAN creates 256x256 quality images with photorealistic details from text descriptions. A new Conditioning Augmentation strategy is proposed to boost the diversity of the generated samples while stabilising conditional GAN training.

Extensive qualitative and quantitative studies show the overall model's performance as well as the influence of individual components, providing useful knowledge for future conditional GAN model creation.

Solving multi-class imbalance

In the proposed work we plan to build a semi-supervised GAN, which would not only identify whether the generated image sample is fake or real but also classify it and generate accuracies and loss rates simultaneously. Further we will also be comparing SGAN and SMOTE based on the decision tree model which will be built on the balanced dataset.

V. Experimental Setup and results

A. Text to image translation

For synthesizing photo-realistic images from text descriptions, we are using the CUB_200_2011 dataset. Caltech-UCSD Birds 200 (CUB-200) is an image dataset with photos of 200 bird species (mostly North American) with a total of 11788 images as shown in Figure 1.

Figure 1. A sample from CUB_200_2011 dataset

Using the dataset, we created high resolution images with a Stacked Generative Adversarial Network. It divides the entire process into two stages. Stage – I GAN, which focuses on the primitive features like color, shape, textures etc, that are mentioned in the text embedding and creates a layout with help of random noise in order to create diverse outputs. Thus, it yields a low resolution 64X64 image. In Stage -II GAN takes the output from Stage – I and corrects its defects and focuses on the rest of the features and produces a high-resolution photo-realistic image 256X256. Hence using the Stacked Generative Adversarial Network model we created high resolution images from text

description. After 10 or 20 epochs we will get a very high resolution image. In the project we have run the program for just one epoch which gives a low resolution image for the input. The result of Stage-I after the first epoch is shown in Figure 2.

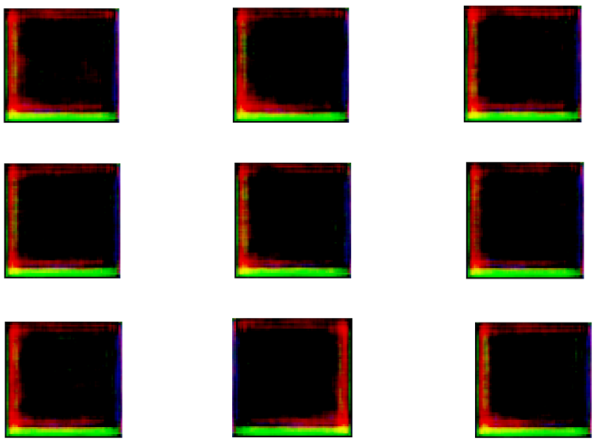

Figure 2. The result of the Stage-I GAN Generator after the first epoch. Using this input Stage-II GAN Generator generates a High Resolution Image.

StackGAN Architecture

Instead of generating a high resolution image conditioned on a text description directly we simplify the task by dividing this process into two stages as shown in Figure 3. In Stage-1 we focus on the primitive features like shape and color. To generate diverse output for different input we perform conditioning augmentation on the text embedding with a random noise of length 100. This thus adds a randomness to the output created. Taking these inputs, the Stage-I generator upsample the low resolution image generated from 4X4 to 64X64. Stage-I discriminator acts as a binary classifier then classifies the image generated by Stage-I generator with the compressed text embedding. Here the discriminator down samples the image from 64X64 to 4X4 and performs binary classification on the down sampled image. If the image generated is real with real caption then the discriminator gives a value '1' . If it's a fake image with real caption or Wrong image with real caption then the discriminator gives value as '0'.

Stage-II focuses on other features than the primitive ones which are vital in generating the photo realistic images. Stage-II GAN is developed on the results from Stage-I GAN. Taking the other features ignored in earlier Stage-I we generate a more realistic image. Taking the 64X64 image from the Stage-I and the text we again perform conditional augmentation and extract the ignored features. With the ignored feature details, the generator produces a high resolution image of size 256X256 after downsampling and upsampling the image. This is again then checked by Stage-II discriminator and produces binary outputs. After running forty to fifty epochs we get a very good quality high resolution image.

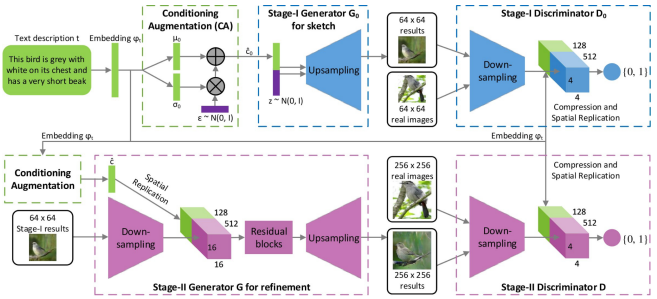

Figure 3. Architecture of StackGAN

B. Solving multi-class imbalance using SGAN

Semi-supervised GAN

The semi-supervised GAN, or SGAN, model is an extension of the GAN architecture that involves the simultaneous training of a supervised discriminator, unsupervised discriminator, and a generator model. The result is both a supervised classification model that generalizes well to unseen examples and a generator model that outputs plausible examples of images from the domain.

The discriminator in a traditional GAN is trained to predict whether a given image is real (from the dataset) or fake (generated), allowing it to learn features from unlabeled images. The discriminator can then be used via transfer learning as a starting point when developing a classifier for the same dataset, allowing the supervised prediction task to benefit from the unsupervised training of the GAN.

In the Semi-Supervised GAN, the discriminator model is updated to predict K+1 classes, where K is the number of classes in the prediction problem and the additional class label is added for a new "fake" class. It involves directly training the discriminator model for both the unsupervised GAN task and the supervised classification task simultaneously. The discriminator is trained in two modes: a supervised and unsupervised mode.

Unsupervised Training: In the unsupervised mode, the discriminator is trained in the same way as the traditional GAN, to predict whether the example is either real or fake.

Supervised Training: In the supervised mode, the discriminator is trained to predict the class label of real examples.

The Discriminator has two different outputs:

Softmax Activation for classifying labeled data into their correct classes i.e this is supervised discriminator.

Custom Activation for classifying into real or fake. We'll see about the custom activation in implementation.

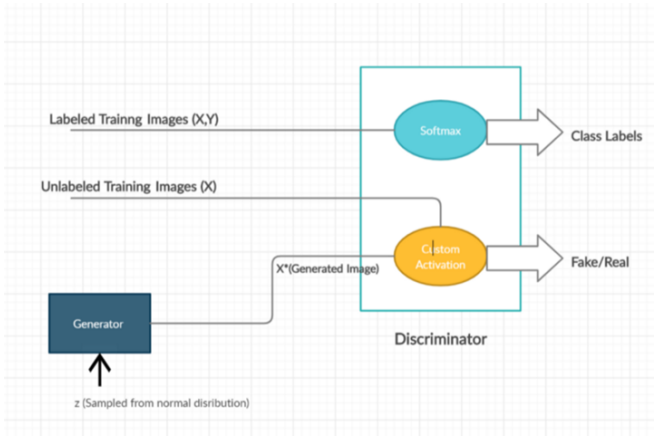

Figure 4. Architecture of Semi-supervised GAN

Synthetic Minority Oversampling Technique(SMOTE)

In order to provide more information to the training data, synthetic oversampling methods create new samples to balance the data set and achieve better learning performance. As one of the most popular oversampling methods to cope with imbalanced data, the 'Synthetic Minority Oversampling Technique' (SMOTE) aims to create "synthetic" examples based on original samples instead of adding replicated ones to the minority class. SMOTE generates samples according to the similarities among existing minority instances. For specific feature samples xi in a feature set S, SMOTE and the K-nearest neighbors of xi in the feature space. To generate a synthetic sample, one of these K-nearest neighbors is randomly selected, then calculate the euclidean distance between these two samples, and at last add the multiplication result of the distance with a random number between [0,1] to the original feature instance:

$$x_{new} = x_i + (\hat{x}_i - x_i) * \delta$$

Compared with random oversampling, SMOTE reduces the overfitting problem to a certain degree and enlarges the minority data in a way that benefits the learning process.

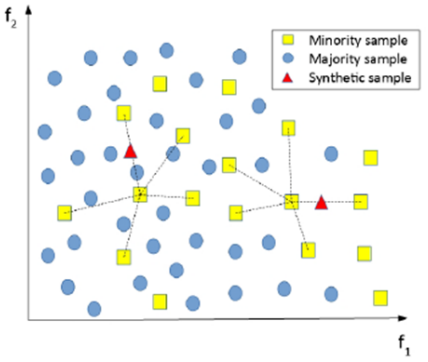

Figure 5. Synthetic Data Generation Based on SMOTE

K-Nearest Neighbors(KNN)

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. The K-NN algorithm assumes the similarity between the new case/data and available cases and puts the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using K- NN algorithm. The K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. The KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

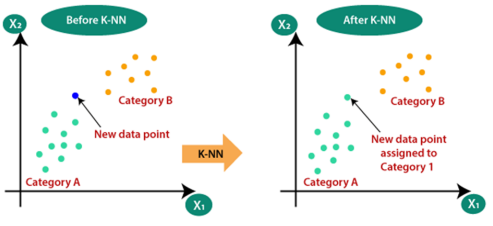

Figure 6. Working of K-NN

The dataset used is a perfectly balanced Fashion-mnist dataset, a sample of which is shown in Figure 3.

Figure 7. A sample from Fashion-mnist dataset

First, the Fashion-mnist dataset is loaded from keras library and the testing and training datasets are combined, and they are reshaped and converted into float32 datatype. Since Fashion-mnist is a perfectly balanced dataset with 7000 samples per class, it'll be imbalanced by taking only 869 samples for classes 0,4 and 8 each. After this KNN model is built on the processed dataset. Then the Semi-supervised GAN - SGAN is built. In the summarise_performance() function the already built KNN model will be used as a criteria to add generated samples in the imbalanced dataset. Then a Decision tree model will be built on the balanced dataset and the accuracy and the no.of misclassified samples are noted. The results for SGAN can be seen in Figure

```
1  Counter(d['labels'])

Counter({0: 869,
         1: 7000,
         2: 7000,
         3: 7000,
         4: 869,
         5: 7000,
         6: 7000,
         7: 7000,
         8: 869,
         9: 7000})
```

Figure 8 . Imbalanced Fashion-mnist dataset

```
Counter({0: 7000,
         1: 7000,
         2: 7000,
         3: 7000,
         4: 7000,
         5: 7000,
         6: 7000,
         7: 7000,
         8: 7000,
         9: 7000})
```

Figure 9. Balanced Fashion-mnist dataset

Figure 10. Images generated by SGAN at the end of Epoch 100

Similarly, the same pre-processing(combining testing and training datasets, reshaping the dataset to 2D dataset and converting the pixels into float32 data type and bringing them into the range [-1,1])is done for SMOTE. The SMOTE model is built and the imbalanced dataset is made balanced. Then a Decision tree model will be built on the balanced dataset and the accuracy and the no.of misclassified samples are noted. The results of using SMOTE are shown in Figure

Figure 11. Images generated by SMOTE

A Decision tree classifier is built on the balanced dataset by the respective models and the accuracies and the no.of misclassified samples are compared.

```
Misclassified examples: 985
Accuracy: 0.930
```

Figure 12. Performance of Decision tree model built on the balanced dataset by using SGAN.

```
Misclassified examples: 1357
Accuracy: 0.903
```

Figure 13. Performance of Decision tree model built on the balanced dataset by using SMOTE

VI. Conclusion

The exploration of StackGAN and SGAN, showed us that GANs are very powerful tools that are better with respect to similar performing neural networks. Where StackGAN performs way better than other GANs like attentionGAN because of its 2 stage structure. SGAN performs better than SMOTE, a synthetic image generating algorithm because of the discriminatory power of the GAN which produces great results due to less overfitting issues. GANs can be very useful in generating synthetic images which are photo-realistic and can help solve problems like class imbalance and generating images from just text descriptions.

Acknowledgment

We wish to express our sincere thanks and deep sense of gratitude to our project guide, Dr.S.Geetha, School of Computer Science and Engineering for her consistent encouragement and valuable guidance offered to us throughout the course of the project work.

We are extremely grateful to Dr. R. Ganesan, Dean, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, for extending the facilities of the School towards our project and for his unstinting support. We express our thanks to our Head of the Department for his support throughout the course of this project.

We also take this opportunity to thank all the faculty of the School for their support and their wisdom imparted to us throughout the courses. We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

References

[1] Zhang, Han, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N. Metaxas. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 5907-5915. 2017.

[2] Hu, Kai, Wentong Liao, Michael Ying Yang, and Bodo Rosenhahn. "Text to Image Generation with Semantic-Spatial Aware GAN." arXiv preprint arXiv:2104.00567 (2021).

[3] Odena, Augustus. "Semi-supervised learning with generative adversarial networks." arXiv preprint arXiv:1606.01583 (2016).

[4] Han H., Wang WY., Mao BH. (2005) Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: Huang DS., Zhang XP., Huang GB. (eds) Advances in Intelligent Computing. ICIC 2005. Lecture Notes in Computer Science, vol 3644. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11538059_91