DA SinGAN: Data Augmentation using SinGAN

Yusheng Hu (yh3097), Shaohua Tang (st3207), Andy Guo (ag4063)

Abstract

In this report we propose to use Single Image Generative Adversarial Net(SinGAN) to generate artificial training data as part of the data augmentation. GANs have proved to be quite successful at generating artificial data, and SinGAN further facilitates the process by allowing an input size of 1 image. In theory, the high quality realistic images produced from SinGAN can be treated as real training data and used as a better approach of data augmentation. The process can be extremely useful in situations such as imbalanced datasets, or just for when original data is extremely scarce. We test our hypothesis on simple binary image classification. Our benchmark is a simple CNN, and we perform data augmentations with both SinGANs and traditional methods, and compare each result.

1 Introduction

In machine learning, if the model were the engine, data would be the fuel. The performance of the engine not only depends on the amount of fuel available, but also the quality of fuel. Likewise, acquiring enough but also high quality data for machine learning projects are not always so easy. One may say that noise is not necessarily bad. Indeed, a good dataset would have as many types of good cases as there are noises, such that it should cover different cases and be a good representation of reality. Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. In past years, researchers use cropping, padding, and horizontal flipping as techniques for data augmentation. The goals of data augmentation include but not limited to oversampling minority cases, reducing data collection cost, keeping the privacy of the original sensitive data, etc.

We aim to adopt new techniques for data augmentation. SinGAN is an unconditional generative model that is improved based on Generative Adversarial Nets(GANs). Unlike conventional GANs, SinGans takes in one training data, and learns the internal distribution to generate high quality and realistic image samples. Suppose we run SinGAN for the existing dataset, for every single original image comes various artificial images. SinGAN could drastically improve the efficiency of data augmentation, creating more diverse data while maintaining the quality of the dataset. Therefore we would like to explore the possibility to use SinGAN in data augmentation.

One may question the necessity of this seemingly complex approach for data augmentation. Well, first, the lightweight nature of SinGAN makes the data augmentation process fairly straightforward and easy to do, so long as there is enough computing power. Second, due to the unique nature of SinGAN and the ability to create realistic images, we want to know if artificial data from SinGAN could perhaps replace the original, in many cases, for privacy reasons. Medical data such as MRI/CT scans could specifically benefit from SinGAN generated data for confidentiality of the record.

Due to the special circumstances we have currently during the pandemic, collaboration has not been the easiest. To test our hypothesis, we took on simple image classifications, and focused on the classic cat vs dog classification. We believe there are enough models on cat vs dog such that it would be easy to benchmark our results. Our experiments consist of simple classification with no data augmentation, classification with traditional data augmentation, and classification using SinGAN.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

2 Related Work and reference

2.1 SinGan

While Generative Adversarial Nets(GANs) have come a long way in modeling high dimensional distributions of visual data, SinGANs took it to a new height.(1) The idea behind SinGAN is that the model would take in one natural image, and use unconditional generation to produce realistic image samples. By learning the internal statistics of patches within a single image, and pushing the information through a pyramid of fully convolutional light-weight GANs, each accountable for patches at different scales, the SinGAN model could produce diverse high quality images containing new features and structures while preserving the semantics of the training image.

2.2 Data Augumentation Using GANs

Using GANs to do data augmentation is not an entirely novel idea. Tanaka et al. (2) proposed to use GANs to generate numeric dataset, aiming to train a classifier without the original dataset, or to oversample the minority class in an imbalanced dataset. In a balanced dataset, Tanaka et al.concluded that using only GAN generated synthetic data led to better accuracy for the classifier than the original dataset. For an imbalanced dataset, they concluded that while the GAN generated synthetic data did improve performance, the improvement was not as significant as that from Synthetic Minority Oversampling TEchnique(SMOTE) or adaptive synthetic (ADASYN) sampling approach. In their research, Tanaka et al. performed all their experiments with simple general GAN architecture with minimal modifications. By using SinGAN, we could potentially further improve the classification in terms of both accuracy and efficiency.

3 Problem

SinGANs are extremely convenient in generating realistic and high quality images. Because it only takes one natural image for SinGAN to generate diverse image samples, theoretically SinGAN would then drastically facilitate the process of data augmentation, making image classification more light-weight and efficient. Our hypothesis is that we can use SinGAN to do data augmentation for any classification problems, taking in only a fraction of the original dataset, yet still being able to achieve the same accuracy.

3.1 SinGAN Random Image Samples

SinGAN consists of a pyramid of GANs, in which both training and inference are done from coarse level to fine level scale by scale. At each scale, the patch generator G_n learns to generate image samples where all the overlapping patches cannot be distinguished from the patches in the down-sampled training image X_n by the discriminator D_n. The effective patch size decreases as we go up the pyramid. The generation of an artificial image sample starts at the coarsest level, and sequentially goes through all generators to the finest level. In essence, at each level of generator G_n, it adds some more details to that were not included in the previous layer, which is relatively more coarse. After training SinGAN on a single image, the model can generate random image samples that depict new structures and configuration while preserving the original patch distribution. (1)

4 Dataset and Model Parameters

4.1 Dataset

We prepare 3 datasets.

1. Original Dataset
   This dataset is the original dataset without data augmentation. It contains 2 classes, cat and dog, each has 1000 images selected from cifar10 dataset.
2. SinGAN Augmented Dataset
   This dataset is the dataset with SinGAN data augmentation. We select 100 images from the original dataset, trained using SinGAN, and random sample 20 images for each image

trained. Each class contains 1000 original images and 2000 generated images, which is 3000 images in total.

3. Traditional Augmented Dataset

   This dataset is the dataset with traditional data augmentation. For each image in the original dataset, we generate a flip image (vertical mirroring) and a flop image (horizontal mirroring). Each class contains 1000 original images and 2000 generated images, which is 3000 images in total.

4.2 Model Parameters

In this section, we will provide a more detailed explanation about the parameters of the binary classification model. Since the trained image size is 32*32 pixels, we chose regular CNN for the binary classification model and in this CNN, we included two convolutional layers: one 3*16 and one 16*32. We selected cross-entropy-loss as the loss function and Adam gradient descent as the optimizer. The learning rate is set to 0.001.

Since in the selected data, we generated 2000 images by SinGAN out of 100 training images. To reduce the over fitting effect of SinGAN random samples images. We chose a new loss function $(1 - \frac{1}{N})f(X) + \frac{1}{N}f(Y)$ that gives higher weight on the original training image and less weight on the SinGAN-generated images.In the equation, we set N as 5 and function f represents the cross-entropy-loss function, X represents the original CIFAR-10 training data and Y represents the SinGAN generated images.The we preform the same experiment on the new loss function.

5 Result

We ran a total of 20 experiments. For each experiment, we train the 3 different datasets for 30 epoches. Then we find the average test accuracy and standard deviation over the 20 experiments. The results are shown in the table below.

	Without Data Augmentation	SinGAN Data Augmentation	Traditional Data Augmentation
Average	67.71	65.97	68.17
Standard Deviation	0.66	0.77	0.92

From the figure below, the average test accuracy from training with traditional flip-and-flop data augmentation is about 2% higher than that without data augmentation. On the other hand, the average test accuracy from training with SinGAN data augmentation is consistently the worst among those with flip-and-flop and without data augmentation. In conclusion, we can see that SinGAN data augmentation method does not increase the accuracy of CIFAR-10 binary classification model.

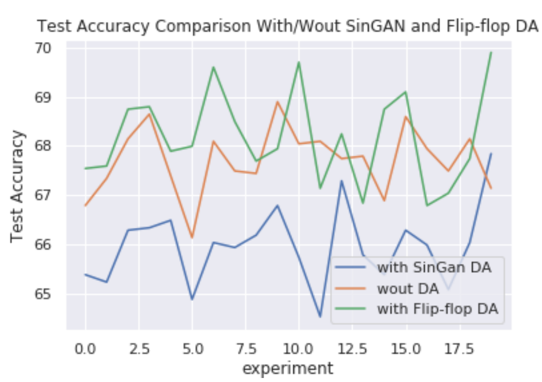

Figure 1: Average Test Accuracy Comparison between Original Dataset, SinGAN Augmented and Traditionally Augmented

Then, we implemented the new loss function mentioned in the Model Parameters section and try to eliminate the over fitting effect of SinGAN random samples images. The result is shown in the table below. From the result, we can see that when we give a lower wight on the SinGAN-generated image, the accuracy increases 0.41% but not significant enough to show the difference. It is hard to conclude that the number of SinGAN generated images caused the model to have lower accuracy.

	With regular cross-entropy-loss	With weighted cross-entropy-loss
Average	65.42	65.83
Standard Deviation	0.58	0.23

6 Result Analysis

The results did not completely match our expectations, and we were indeed a bit perplexed. As we look deeper into how SinGAN works, we believe there are various factors that led to our result. Now suppose we present you a small low quality image of a dog that you can barely make sense of silhouette. SinGAN would only be capable of producing an equivalent level of image, such that it would not be able to generate a large high resolution picture of a dog. We believe this is the bottleneck of our experiment. We chose 32*32 images as our training data, and tried to use SinGAN to generate more images from them. The coarsest scale is not very big to begin with, so that there can be only so many layers of GANs to pass through. The limited number of layers means each finer level would not have a lot more details than the previous level, and as a result, our generated images were identical to the original image to our human eyes–we could barely tell a difference.

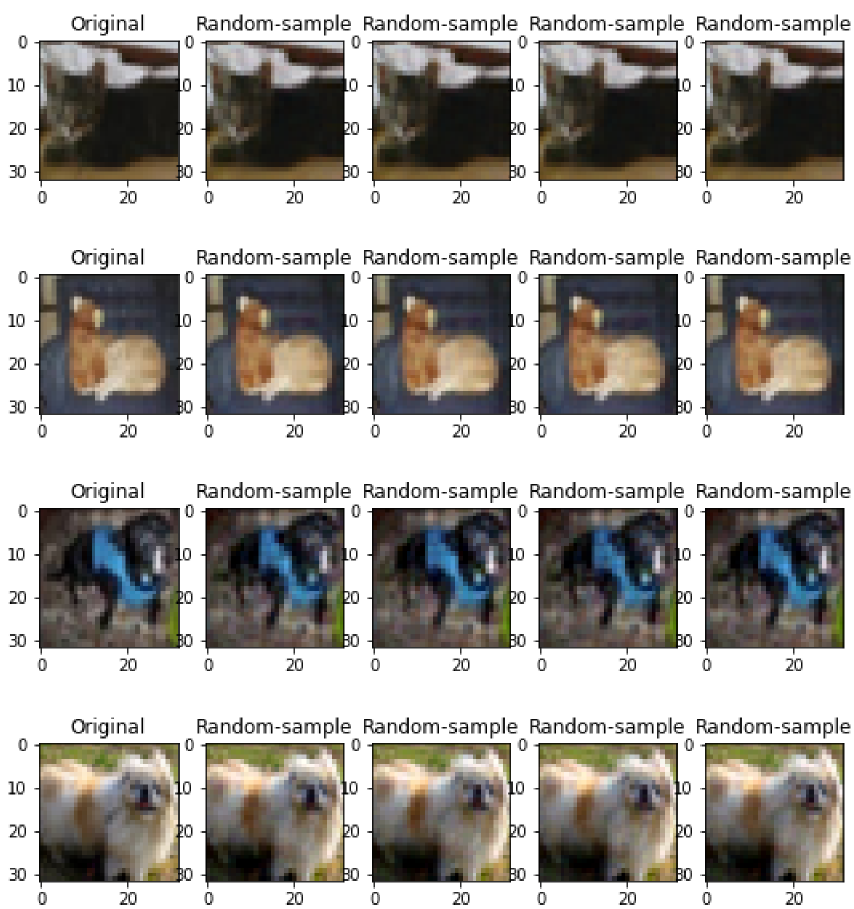

Figure 2: Comparison between Original image and SinGAN random-sample image for CIFAR-10

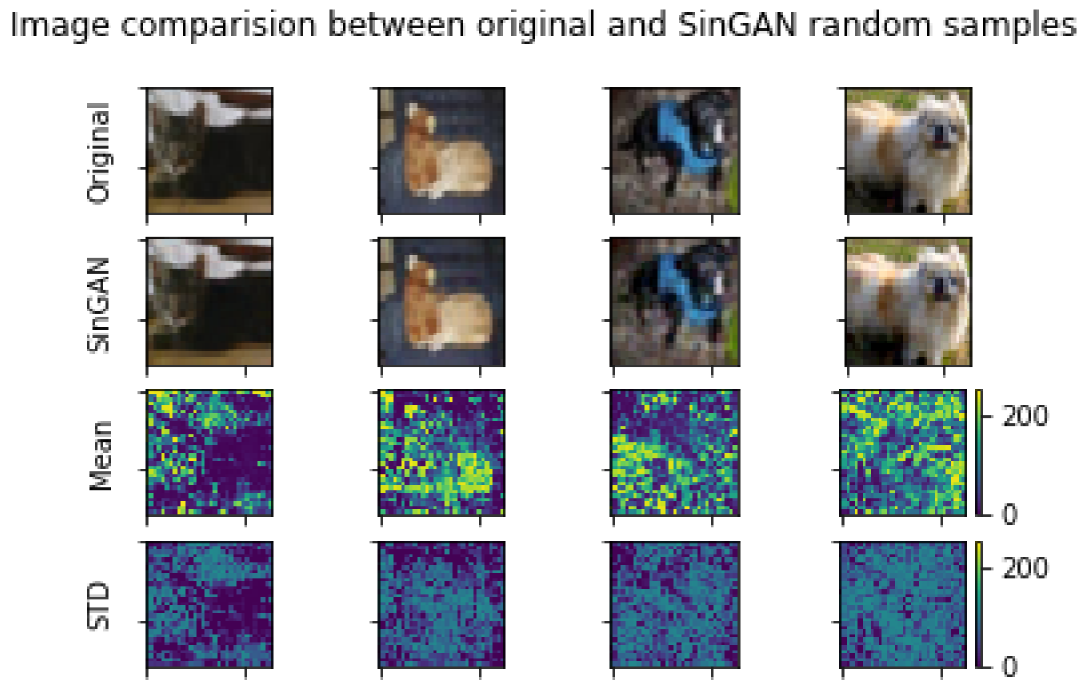

Figure 3: Statistics of Absolute difference between Original image and SinGAN random-sample image for CIFAR-10

To test how much of the original image was altered in the corresponding SinGAN generated-ones, we wrote a simple program to look at the statistical differences of the generated images compared to the original. We select one sample training image from each of the CIFAR-10 cat and dog classes. And then we selected their 20 coresponding SinGAN-generated images. The comparsion result is shown above in the Figure 3. The first and second row shows the original and its corresponding SinGAN-genearated image. The third rows shows the graph of absolute difference of 20 SinGAN generated images and the original image. We can see that the difference is variable on different part of the image. The average mean value is 112.23, which means that the SinGAN random-sample image is noticeably different from the original image. But the difference seems not distinguishable to human eyes. However, such difference still might damage the binary classification model to lower the prediction accuracy.

In order to test if the SinGAN images damage the trainning model. We trained the CNN model on only 1000 of each dog and cat CIFAR-10 images and test the model accuracy on SinGAN generated images. we found out the total test accuracy on test set is 99.8%. Only 8 images out of 4000 are predicted to the wrong class.

Running time on SinGAN was another major issue for our experiments. Generating random image samples on one 32*32 image takes 5 minutes. Therefore, to augment a dataset of 100 images on SinGAN takes 500 minutes, approximately 8.3 hours. While we could tweak on the learning rate, as we did, the running time and the results remain about the same. For our experiment, we used SinGAN to train 100 images per class, with a total of 2 classes. Unfortunately, our computing power prevented us from multi-processing on both classes at the same time.

7 Conclusion and Future Work

From our preliminary result, we find out that SinGan-generated images do not have a significant influence on the improvement of testing accuracy. We increased our training size, and trained 100 images for each class through SinGAN. Comparing our results from the previous milestone, we conclude that the size of training set from SinGAN has little impact on the performance.

For any future work, we should shift our focus on the size of training images, as we believe larger images with more patch information could lead to better performance from SinGAN, producing more diverse and useful artificial images. CIFAR-10 includes only 32*32 pixels and might be too limited to generate any useful data. We could find more classification images dataset in ImageNet and investigate the data augmentation of SinGan on larger images if time and computing power were not a limiting factor. Finally, we could potentially implement object detection on an image and move the object in the image and implement SinGan to generate better background.

8 Code Repository

Link to GitLAB respository: https://gitlab.com/vcaptainv/singan-data-augment/

References

[1] T. R. Shaham, T. Dekel, and T. Michaeli, "Singan: Learning a generative model from a single natural image," *CoRR*, vol. abs/1905.01164, 2019. [Online]. Available: http://arxiv.org/abs/1905.01164

[2] F. H. K. dos Santos Tanaka and C. Aranha, "Data augmentation using gans," *CoRR*, vol. abs/1904.09135, 2019. [Online]. Available: http://arxiv.org/abs/1904.09135