SinGAN-Based Representative Data Augmentation

Dennis J. McWherter, Jr.
Columbia University
djm2242@columbia.edu

Jiageng Li
Columbia University
jl5421@columbia.edu

Uzay Macar
Columbia University
oum2000@columbia.edu

Abstract

Data augmentation improves generalizability and reduces overfitting in deep neural networks. However, manually designed augmentations often produce unrealistic and unrepresentative images of their underlying dataset. Our approach applies SinGAN [1] to a set of **representative** images from a training set. This produces realistic samples to augment the training data. Our results suggest these augmentations can produce improved accuracy compared to other state of the art approaches.

1 Introduction

Data augmentation improves generalizability and reduces overfitting in deep neural networks. However, manually designed augmentations such as affine transformations, contrast changes, Gaussian noise, dropout of regions, random cropping, and padding [2] make data augmentation an implicit form of feature engineering, and often produce unrealistic and unrepresentative images of their underlying dataset. Additionally, our results demonstrate that these manual techniques do not always have a large or sufficient affect on the final accuracy of a model.

Generative Adversarial Networks (GANs) demonstrate promise for generating more realistic augmented data samples. A GAN can generate unique out-of-sample images trained from a representative corpus. These generated results bolster the model training process with more information than simple transformations. Several researchers have already used GANs for data augmentation to improve model accuracy across a variety of fields including biomedics [3] and across different modalities such as vision [4, 5, 6] and sound [7, 8].

Motivated by GAN-based data augmentation techniques, our work applies SinGAN to augment our dataset and improve overall model accuracy. Our approach first takes a number of **representative** image samples from a training dataset. [1] It then applies SinGAN [1] to these representatives which produces a variety of substantially similar images to each representative. Finally, these generated samples are used to augment the original training datasets. Our preliminary results demonstrate that these augmentations produce better results than "standard" data augmentations.

1.1 Related Works

Hauber et al. [6] proposed one of the first, known end-to-end data augmentation strategies for computer vision tasks, aligning image pairs within each class with the assumption that spatial transforms between images belong to a large class of diffeomorphisms. Their end model learns class-specific transformations in a Riemanian submanifold. BAGAN [4]– or Balancing GAN– is yet another GAN-based data augmentation mechanism. It uses an autoencoder and GAN to generate additional samples to correct dataset imbalances. It initializes the GAN Using the trained encoder and decoder and then runs the adversarial process. This approach generates realistic sample images to augment the original dataset. Antoniou et al. proposed a Data Augmentation GAN (DAGAN) that

[1] Code available at: https://github.com/DennisMcWherter/SinGANBasedDataAugmentation

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

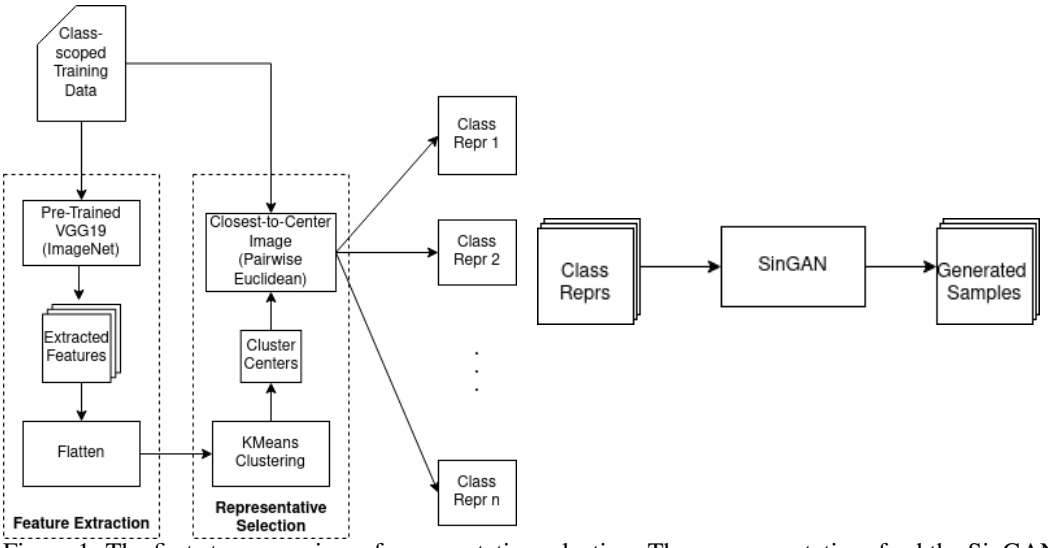


Figure 1: The first step comprises of representative selection. These representatives feed the SinGAN generation step. After training SinGAN on each representative (one model per input), augmented data samples are directly generated from the resulting model.

learns how to generate a synthetic image using a lower-dimensional representation of a real image [5]. DAGAN takes data from a source domain and learns to take any data item and generalise it to generate other within-class data items.

Park et al. [8] proposed a simple data augmentation method for automatic speech recognition where they warped and masked blocks of frequency channels. There exists significant previous work on manual feature engineering for speech-based tasks, yet recent research on this modality also demonstrate the promise of GANs. Chatziagapi et al. [7], for example, addressed the problem of data imbalance for the task of automatic speech recognition via GANs. They adapted and improved a conditional GAN architecture to generate synthethic spectrograms for the minority class and demonstrated improved validation scores on the IEMOCAP emotion recognition dataset.

Our work focuses on assessing the success of GANs, and SinGAN in particular, for data augmentation in computer vision tasks. SinGAN learns an unconditional generative model trained on a single source image. By learning an image's patch distributions at varying resolutions, the GAN can generate a variety of modified images. Additionally, the output images retain significant semantic context.

2 Methods

At a high level, our approach consists of choosing class-based representative samples, training SinGAN based generators on these representatives, generating random samples using the generators, and training a separate purpose-driven model using the augmented training set. Thus, our primary contribution demonstrates that SinGAN can produce high-quality augmented data samples. Additionally, we describe a method for choosing samples to generate. This selection makes the process computationally feasible for large datasets. The full approach is showcased in Figure 1.

2.1 Dataset

We use the Tiny ImageNet dataset [2]– a smaller version of the conventional image classification ImageNet dataset [9]– to run our experiments. This reduced dataset enables faster prototyping and iteration times. As opposed to the 1000 categories of ImageNet, Tiny Imagenet consists of 200 categories (i.e. classes) where each category has 500 training images, 50 validation images, and 50 test images. The number of classes is still relatively high, so our process reduces the dataset to 22 different classes.

[2]More information on dataset: `https://tiny-imagenet.herokuapp.com`

Previous works on this dataset exhibit training procedures that rely on a wide range of data augmentation strategies. Abai et al. modified the resolution of input images accordingly to half (0.5) and quarter (0.25) scales, and applied coarse dropout, rotation, additive Gaussian noise, cropping and padding, horizontal and vertical flip, translation, shearing, and contrast normalization [10]. They reported an increase of 9.5% in validation accuracy with augmentations, and a maximum validation accuracy of 62.73%. Le et al. used far less augmentations, randomized crops and random horizontal flips, yet still reached a maximum validation accuracy of 61.23% thanks to choices of dropout, model ensembling, and parametric rectifiers (PReLU) [11]. Similarly, Barnes et al. applied sequential data augmentation operations of horizontal flip and cropping to reduce overfitting [12]. They have utilized a pretrained ResNet18 architecture and achieved a maximum validation accuracy of 60.20%.

2.2 Experimental Data Preparation

We first filter the dataset to 22 random categories. That is, we have 20 categories with the same number of samples and 2 imbalanced categories. The labeled dataset is randomly split into a training set and a holdout set. We applied a 70/30 data split, respectively. The holdout set applies only in final model evaluation and intends to simulate "real-world" data for model evaluation.

Data is randomly discarded from the training set to simulate class-imbalance. The discard only occurs on the 2 imbalanced categories. The two categories we have selected have dataset-specific identifiers as `n01882714` and `n04562935`, which respectively correspond to *koala* and *water tower* images.

2.3 Choosing Representatives

Training a SinGAN model for all training samples is both time and cost prohibitive. As a result, a smaller number of samples are chosen to represent a particular class. These representatives source SinGAN models to produce augmented data samples.

Representative selection occurs on a per-class basis. Thus, our method for selecting representatives has three phases as displayed in Figure 1. First, we extract image features using the VGG-19 model [13] pre-trained on ImageNet. We then take the extracted features and apply K-Means clustering on the flattened result. Finally, we find the "closest" (defined by Euclidean distance) real images in the dataset to the cluster centers. These closest images are selected as our representatives.

2.4 Augmentation

We train SinGAN for each selected representative. These representatives are highly similar to their source image but clearly differ. After producing these new samples, we add them to our training set.

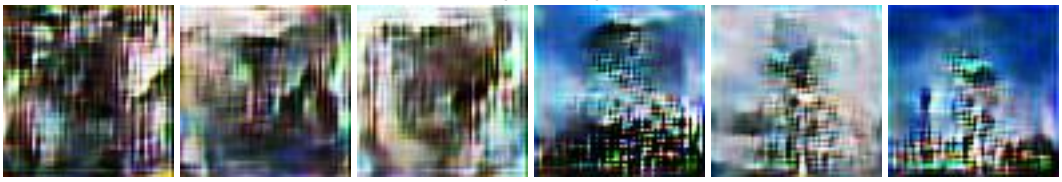
In our current implementation, we produce 50 images for each representative. Since we chose 5 representatives for each class, this amounts to a total of 300 additional training images across the two classes. Hand-picked SinGAN generated images, as well as original images from the aforementioned two categories, and BAGAN and DAGAN generated images via similar approaches are show in Figure 2.

2.5 Training

After augmenting our training set, we perform a standard training run over the augmented data as depicted in Figure 3.

We decided to use MobileNet [14] architecture to train our classification network. MobileNet is an efficient general purpose architecture designed with a good balance of accuracy, size and latency. To compare the effectiveness of SinGAN based augmentation, comparing to no augmentation or other augmentation methods, the classification network does not need to be state of the art. We only need to ensure the same architecture is applied when training for different comparison cases.

(a) Original Images



(b) BAGAN Generated Images

(c) DAGAN Generated Images

(d) SinGAN Generated Images

Figure 2: Hand-picked samples from (i) original TinyImageNet Dataset, (ii) BAGAN generated images, (iii) DAGAN generated images, and finally (iv) SinGAN generated images, where left half in each row shows *koala* images and right half shows *water tower* images.

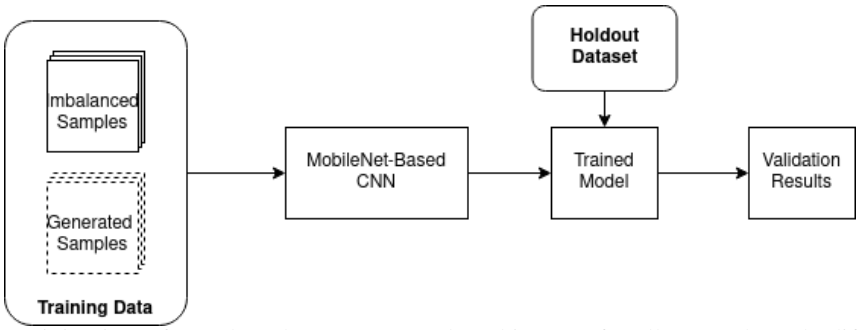


Figure 3: Training is performed on the same network architecture for all runs. The only difference is the training data between each run.

3 Results

3.1 Evaluation Methodology

We would like to measure whether using SinGAN generated images to augment the training data improves the accuracy of the classification network. We leverage well-known network architectures such as MobileNet to train against the datasets. We compare the results against a baseline of no augmentations as well as the results from BAGAN.

We use IBM's implementation of BAGAN [3] with slight modifications to load our dataset. Unlike SinGAN, BAGAN was trained on the entire training corpus rather than the selected representatives. For sample parity, BAGAN also generates 150 samples per test class for augmented data.

We use the official implementation of DAGAN [4] with slight modifications to load our dataset. We chose to train DAGAN on the entire training corpus, much like BAGAN. For sample parity, DAGAN also generates 150 samples per test class for agumented data.

We evaluate our results against training benchmarks for our experiments. Particularly, we use the Tiny-ImageNet-200 dataset.

3.2 Comparisons

We have trained three different model configurations: (i) model with no additions, (ii) model with standard data augmentations of random zoom, brightness modification, and horizontal flip respectively with probabilities 0.05, 0.5, and 1.0, and (iii) model with SinGAN-based data augmentations. On the other hand, we have evaluated these configurations once on the entire validation set, and once only on the 2 handpicked categories from the validation set which were augmented by SinGAN.

Table 1: Performance of Models on Entire Validation Set

Configuration	Validation Accuracy
MobileNet (Vanilla)	0.6100
MobileNet + Standard Aug.	0.6007
MobileNet + BAGAN Aug.	0.5985
MobileNet + DAGAN Aug.	0.6384
MobileNet + SinGAN Aug.	**0.6378**

Table 2: Performance of Models on 2 Handpicked Categories from Validation Set

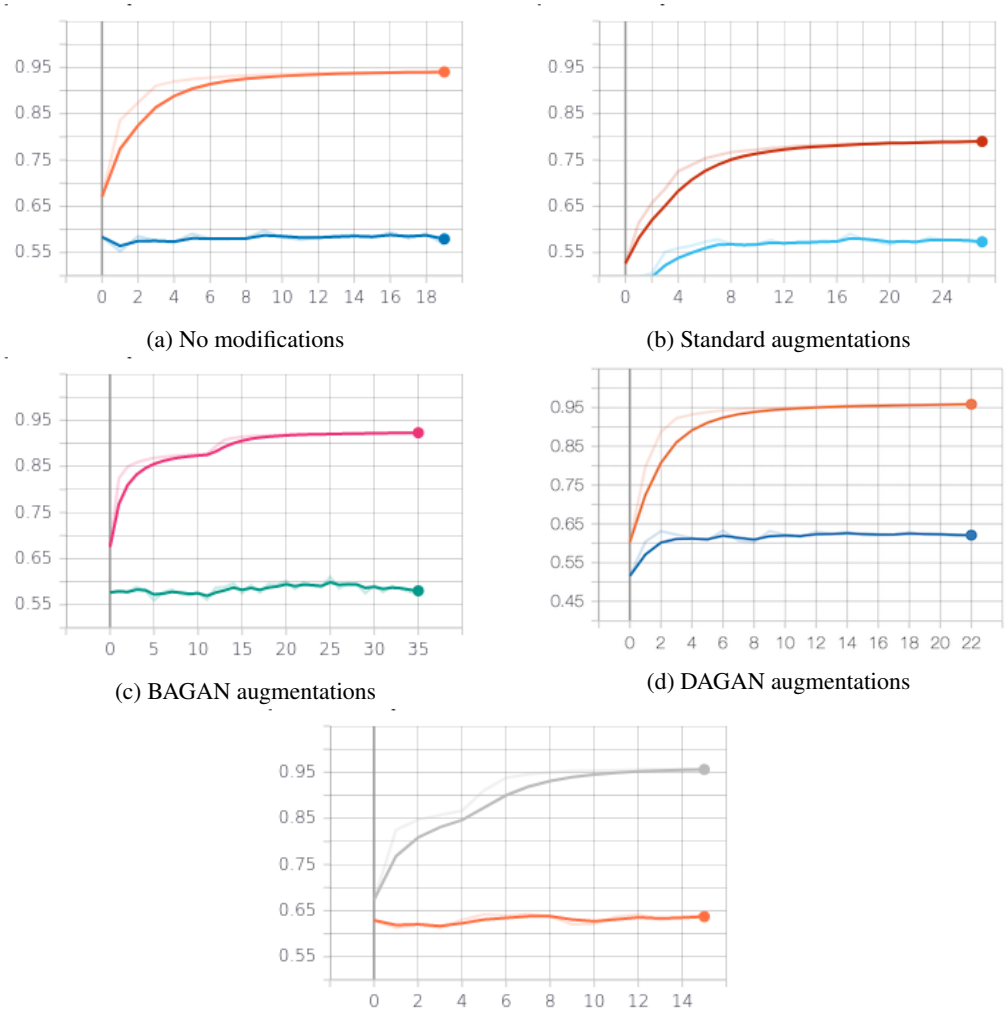
Configuration	Validation Accuracy
MobileNet (Vanilla)	0.3428
MobileNet + Standard Aug.	0.3816
MobileNet + BAGAN Aug.	0.7668
MobileNet + DAGAN Aug.	0.7032
MobileNet + SinGAN Aug.	**0.7562**

These results suggest that datasets augmented with SinGAN-generated samples outperform standard augmentations and BAGAN-based augmentations. We observe that the SinGAN-based method outperforms BAGAN in general accuracy on the hold-out set by 3.93%. Additionally, the SinGAN-augmented model trained in about half as many epochs as shown in Figure 4.

While BAGAN performed the best on class-specific validation, we note that both augmented approaches perform significantly better on these classes than the imbalanced approaches. This demonstrates that the augmentations are actually providing a positive affect on the imbalanced classes. Moreover, since the accuracy of the SinGAN-augmented approach is better than the other models on the holdout set, it does not appear that the approach forces the model to overfit for specific classes.

[3] Code available at: https://github.com/IBM/BAGAN

[4] Code available at: https://github.com/AntreasAntoniou/DAGAN



(a) No modifications

(b) Standard augmentations

(c) BAGAN augmentations

(d) DAGAN augmentations

(e) SinGAN augmentations

Figure 4: Training and validation accuracy per epoch.

4 Conclusion

Our work confirms that GAN-based data augmentation strategies in general have the potential to yield better performance when combined with vanilla classifiers, as all training configurations involving GAN architectures showed significant improvements over the MobileNet vanilla classifier. On top of this, augmenting an imbalanced dataset using SinGAN-generated samples on a set of class representatives shows promising results and is able to at least partially alleviate the data imbalance problem. We were able to observe a dramatic improvement in category-specific classification and an appreciable increase in generalization accuracy when compared to other approaches. We also show that even a simple clustering algorithm like K-means, when used for class representative selection, has the promise to yield improvements over the baseline while significantly reducing the training time.

5 Future Work

In the future, we would like to perform a more robust evaluation using a mechanism such as k-fold validation. Additionally, we would like to explore the affect of representative selection against other mechanisms (i.e. random). The methods of representative selection could also be extended to more complex algorithms such as expectation maximization (EM). Finally, we would like to draw

representatives from all 200 classes in the TinyImageNet dataset and train on a fully augmented TinyImageNet, and train on the full ImageNet dataset as well to prove that our approach scales to larger datasets.

References

[1] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image, 2019.

[2] Alexander B. Jung, Kentaro Wada, Jon Crall, Satoshi Tanaka, Jake Graving, Christoph Reinders, Sarthak Yadav, Joy Banerjee, Gábor Vecsei, Adam Kraft, Zheng Rui, Jirka Borovec, Christian Vallentin, Semen Zhydenko, Kilian Pfeiffer, Ben Cook, Ismael Fernández, François-Michel De Rainville, Chi-Hung Weng, Abner Ayala-Acevedo, Raphael Meudec, Matias Laporte, et al. imgaug. `https://github.com/aleju/imgaug`, 2020. Online; accessed 01-Feb-2020.

[3] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58:101552, 2019.

[4] Giovanni Mariani, Florian Scheidegger, Roxana Istrate, Costas Bekas, and Cristiano Malossi. Bagan: Data augmentation with balancing gan, 2018.

[5] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks, 2017.

[6] Søren Hauberg, Oren Freifeld, Anders Boesen Lindbo Larsen, John W. Fisher III, and Lars Kai Hansen. Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation, 2015.

[7] Aggelina Chatziagapi, Georgios Paraskevopoulos, Dimitris Sgouropoulos, Georgios Pantazopoulos, Malvina Nikandrou, Theodoros Giannakopoulos, Athanasios Katsamanis, Alexandros Potamianos, and Shrikanth S. Narayanan. Data augmentation using gans for speech emotion recognition. In *INTERSPEECH 2019*, 2019.

[8] Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, and Quoc V. Le. Specaugment: A simple data augmentation method for automatic speech recognition. *Interspeech 2019*, Sep 2019.

[9] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[10] Zoheb Abai and Nishad Rajmalwar. Densenet models for tiny imagenet classification, 2019.

[11] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. 2015.

[12] Zach J Barnes. Techniques for image classification on tiny-imagenet. 2017.

[13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[14] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.