ICLR Reproducibility Challenge, Ambient GAN

Christopher Culley, Deekshana Veluchamy & Aneeshaa S Chowdhry
Department Electrical Engineering and Computer Science
University of Southampton,
Southampton,
UK

Abstract

As part of the International Conference on Learning Representations reproducibility challenge we attempted to reproduce some of the experiments in *AmbientGAN: Generative models from lossy measurements*; a twist on the traditional baseline generative adversarial network where it is assumed that the underlying state vector is not directly observed– instead the model must learn from images with various noise transformations applied. Our reproduced results confirm that the ambient approach training on noisy-images indeed gives improvement over GAN methods that do not account for the noise as well as baseline GANs trained after de-noising approaches.

1 Introduction

The ability to reproduce findings is key to good scientific practice. This core tenant has not always been forthcoming in the field of artificial intelligence – where often either full details of the methods used are not provided or the code used to generate experiments not shared (Hutson (2018); Peng (2011)).

To overcome this, the International Conference on Learning Representations (ICLR) have invited researchers to engage in a reproducibility challenge. Here experiments and findings of accepted papers to the conference are attempted to be reproduced to determine if the conclusions of the paper are supported by the reproduced results.

To this end, in this work we aim to reproduce and explore the empirical results of *AmbientGAN: Generative models from lossy measurements* (Bora et al. (2018)). This paper is structured as follows, first we give an overview of the problem the paper aims to overcome, and its approach. Next we give details of the experiments we have run, followed by their results. We then conclude with a discussion.

2 Ambient Gan Overview

In the paper the authors try to tackle the problem of noisy image inputs – which they pose to be how most data is found in real-world settings – in training generative models. Their solution, which they show empirically to generate more realistic images than the traditional Generative Adversarial Network (GAN) approach (Goodfellow et al. (2014)), enables the recovery of underlying distribution from images that have been corrupted by a measurement noise process. A critical assumption behind the success of the framework and theory is that the measurement process is known, and satisfies some technical conditions.

To do, as with traditional GANs, they use noise passed into a generator to produce a full image. The AmbientGAN next passes this image into a measurement transformation block, which adds noise to the image. This noisy image is in turn then passed, along with an image from the training dataset (which is also noisy) to the discriminator that is tasked with distinguishing between these simulated measurements of generated samples, and true measurements of real samples. We show this architecture schematically in Figure 1 (bottom). With this mechanism, AmbientGAN is able to generate high-quality samples from substandard images.

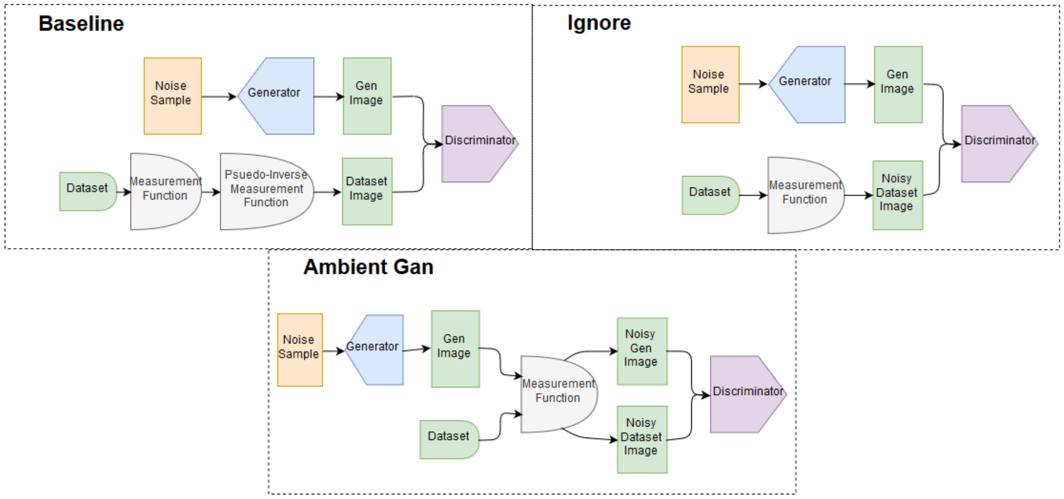

Figure 1: Here we show the three models used in the experiments with the AmbientGAN (shown bottom) trained on noisy images with the generated images passed through a measurement transformation to make them noisy also. The baseline model (top-left) takes the traditional GAN approach with the noisy training images first de-noised using pseudo-inverse measurement transformations. The ignore experiments (top-right) make no attempt to de-noise the training images or account for the noise in the generator side.

3 Experiment Details

For the experiments in the paper a number of noisy measurement functions and dataset combinations were explored. Unfortunately no indication is given in the paper of computational processing time for the experiments or the machine used. Empirically we found some of the approaches much more computationally intensive than others and for this reason concentrate on one dataset-measurement function combination which, since the gains in using AmbientGAN were shown to be across all dataset and measurement function combinations, we hope suffices to capture the reproducability of the successes.

In order to quantify any utility of the AmbientGAN approach the authors compare the inception score with the baseline and ignore models, trained on the same data distribution. The ignore model uses a traditional GAN structure where the discriminator is tasked to identify the fake from the transformed image and the one produced by the generator (see Figure 1 top-right). Baseline is a slightly more sophisticated way of measuring the success of Ambient. Here the measurement function is applied and then a pseudo-inverse function thereafter, to be passed into the discriminator (see Figure 1 top-left). This technique does require that a psuedo-inverse function exists, for the case of block-pixel, we follow the author in using in-painting.

We use MNIST the dataset from which to learn the underlying distribution and block-pixel as a measurement function. In block pixel each pixel in a given image the RGB values are set to 0 with some probability p. The authors use two methods of in-painting, average blurring and variation in-painting, for the Baseline model which give almost identical results. We opt for a variation of Biharmonic in-painting (Damelin & Hoang (2018)), which we find empirically gives stronger results than both the variation in-painting and average blurring, in an attempt to fully test the ambient results.

We note that although no definitive set of experiments are defined in the paper, we infer from the results graph that the p values tested were $p \in [0.0, 0.1, 0.5, 0.8, 0.9, 0.95, 0.99]$ and we run experiments for each of these p values for 25 epochs for ambient, baseline and ignore models following the altered Conditional DCGAN architecture (Radford et al. (2015)) capturing the inception score metric comparison. We note also, in the case of the baseline model we produced pre-processed datasets rather than apply the measurement function and its inverse on-the-fly. This was for computational reasons - allowing PC downtime.

4 Results and Discussion

Figure 2 gives the full set of results using the three models. We can see that the results given by the authors are consistent with our findings. Figure 3 gives a set of example images generated by ambient, baseline and ignore.

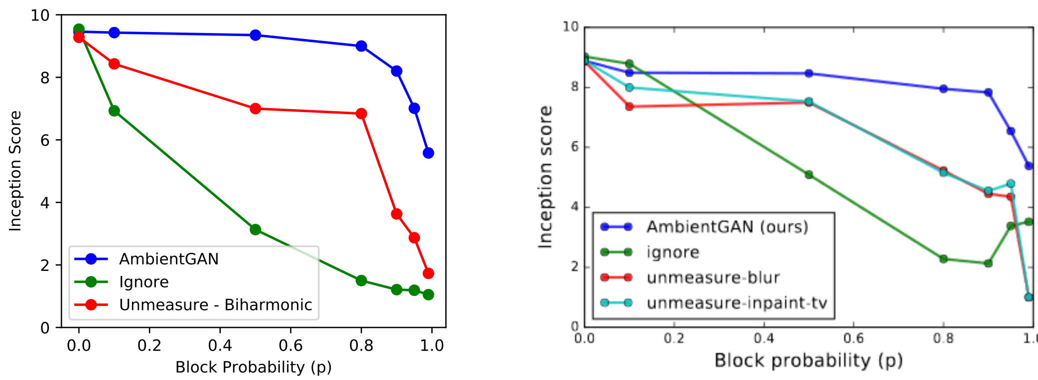

Figure 2: The set of inception scores captured by each of the models trained on MNIST in our reproduced setting (left) with the results given in the original AmbientGAN paper (right). We can see consistency in the Ambient superiority over both the ignore and baseline measures. Our attempts at using a strong unmeasure Baseline – whilst stronger than the inpaint-tv – are still outperformed by the AmbientGAN

Whilst these result demonstrate strong reproducibility from the original paper we note at this point an area of attention not fully clear in the authors work, namely the psuedo-inverse used. For example, convolve-plus-noise – where a Gaussian blurring is applied with additional noise, the stated psuedo-inverse used is a Wiener deconvolution but no details of parameters used are given. We can only infer therefore that an unsupervised version is used with this method being computationally quite expensive.

Although outside of the initial purposes of this paper we notice a drawback in the Ambient approach which could be addressed in future works. Namely, how, in a real setting, the network would know both which measurement transformation was used – since each of the networks in the original paper are trained on disjoint transformation functions. Since the premise of the work is to be able to work with noisy data in producing models able to generate clear representations, this point is particularly pertinent and it's not apparent if such methods exist. For the case of block-pixel with general colour images, the mask could perhaps be inferred by the pixels that are black and isolated. For the other transformations - such as Gaussian noise – its not apparent how the ambient approach could be used without knowing how the images are transformed. In general, in a real-setting its unlikely that this information would be forthcoming and image distributions might contain a mixed set of noisy images. Any future approaches would need to account for this uncertainty which is universal to the problem Ambient attempts to solve.

5 Conclusion

In this work we have reproduced AmbientGAN models and tested them in the same setting as described in the original paper. Our results have shown strong consistency with the authors reported. An area of future attention would be to detail with more completeness the parameters and methods

Figure 3: Example set of generated MNIST digits with a block-pixel transformation p = 0.9 trained with Ambient (left), Baseline with Bi-harmonic in-painting (center) and Ignore (right)

used in inverse-transformations as well as the computational running time and machine used to generate the results. More generally, work is needed to understand how this successful technique could be utilized in a real-setting under noise transformation uncertainty.

6 Code Availability and Contribution

In this contribution we have re-written ambient, ignore and baseline models using the Conditional DCGAN (Radford et al. (2015)) model as starting point with alterations as required. We also recreated the transformation functions and their inverses as per the instructions in the paper and ran experiments with block-pixel varying p to reproduce the results in the paper. All code implementations are written in Python with PyTorch.

We note that due to computation restrictions – for the baseline approaches we created a inverse-transformed dataset from the MNIST originals to train on rather than train on-the-fly.

Full code is available at - `www.github.com/COMP6248-Reproducability-Challenge/AmbientGanReproduction`

References

Ashish Bora, Eric Price, and Alexandros G Dimakis. Ambientgan: Generative models from lossy measurements. In *International Conference on Learning Representations (ICLR)*, 2018.

SB Damelin and NS Hoang. On surface completion and image inpainting by biharmonic functions: Numerical aspects. *International Journal of Mathematics and Mathematical Sciences*, 2018, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Matthew Hutson. Artificial intelligence faces reproducibility crisis. *American Association for the Advancement of Science*, 2018.

Roger D Peng. Reproducible research in computational science. *Science*, 334(6060):1226–1227, 2011.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.