

# Insta(nt) Pet Therapy: GAN-generated Images for Therapeutic Social Media Content

Tanish Jain

Stanford University

tanishj@stanford.edu

## Abstract

*The positive therapeutic effect of viewing pet images online has been well-studied. However, it is difficult to obtain large-scale production of such content since it relies on pet-owners to capture photographs and upload them. I use a Generative Adversarial Network-based framework for the creation of fake pet images at scale. These images are uploaded on an Instagram account where they drive user engagement at levels comparable to those seen with images from accounts with traditional pet photographs, underlining the applicability of the framework to be used for pet-therapy social media content.*

## 1. Introduction

Several studies have highlighted the positive impact of viewing images of pets on social media apps such as Instagram [7] [12]. Zhou et al., for instance, demonstrate “the positive effect of an online pet on subjective well-being” [12], underlining its therapeutic effect. However, content may not be necessarily tailored for this purpose, and users can only see these images as frequently as the pet owner chooses to post them.

In this paper, I build on prior work to explore two different GAN architectures and their ability to generate pet images. Importantly, the images must be of a quality that can drive user engagement on social media, which can indicate the presence of the desired therapeutic effect.

I add to the existing literature by using a new dataset which results in better model performance. Additionally, I use the generated images for a novel task by using them directly on an Instagram account and evaluate their ability to create engagement. The results are promising: GAN-generated images are able to reach comparable or better levels of engagement when compared to real images. This showcases the applicability of the approach for the generation of therapeutic social media content in the context of online pet watching.

## 2. Related Work

Prior work has shown that therapy dog pages on Instagram perform well in terms of generating user engagement, which is backed by the positive effects of “online pet watching” [12]. Blaine and Kremer [1], for example, discuss this phenomenon; however, the difficulty of generating this in-demand content is highlighted, especially since it relies on pet owners consistently uploading photographs of their pets on an online platform. In this project, I attempt to solve this problem with the use of GAN-generated images, which may be produced on a large scale effectively.

Although the application itself is novel, several prior works have looked at the problem of using GANs for generating images of dogs. Shangguan et al. [9] utilize a DCGAN to generate images of dogs. The authors utilize the Stanford Dogs dataset, which I replace in my approach for reasons discussed in Section 3. Kumari et al. [5] also use a DCGAN to generate dog images. Their model is pretrained on ImageNet, which is shown to work well and produce a more diverse set of images (namely, by producing dogs of various breeds). Both these approaches, however, generate low-resolution images, which would be unfit for an Instagram-based solution.

Compared to the problem of generating dog images, the problem of generating cat images has been relatively less explored. While the high-level problem is not fundamentally different, the performance of models in prior work across these two tasks is not consistent, which may underline a problem with the choice of the training dataset. Gupta and Gupta [3] use two models - DCGAN and LS-GAN - to generate cat images. They note that the DCGAN produces realistic images, which is critical for the proposed application.

Prior work on the generation of animal images suggests a strong preference for the use of DCGAN; however, concerns around its ability to produce high-resolution images remain. Jain et al. [4] have shown that an alternative model, BigGAN, can produce better quality results than DCGAN. However, they note that the model training can be much

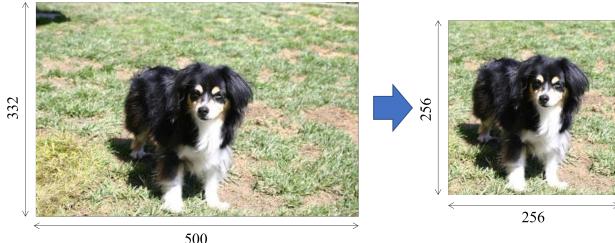


Figure 1. Acceptable images are cropped/resized during preprocessing.

more time-consuming and resource-intensive, which may explain the preference for DCGAN.

In my project, I attempt to build on this prior work and combine individual improvements to comprehensively address some of the underlined challenges. I use a new dataset to address issues arising from the choice of dataset. This is further discussed in Section 3. I use both a DCGAN and a BigGAN given their individual advantages and compare their results. I augment data by using images from two datasets, based on its benefits noted in [9], and fine-tune a pretrained model in the case of BigGAN, instead of training from scratch, to address concerns around computational constraints.

### 3. Dataset

The project uses the “Dogs vs Cats” Kaggle dataset<sup>1</sup> as well as the “Cats and Dogs Breeds Classification Oxford Dataset”<sup>2</sup> for this task. The former contains 25,000 images in total, of which exactly half are images of dogs, and half are images of cats. The latter contains 7,393 images in total, where the ratio of cat images to dog images is approximately 2:1.

These datasets have not been used for pet image generation in existing literature, which therefore presents an opportunity to evaluate the impact of the choice of the dataset on the quality of generated images. Other popular datasets, and especially the Stanford Dogs dataset, which has been a popular choice for using with GANs before, poses some challenges. This includes the fact that the images in the dataset often contain wider shots which may contain less interesting features, and the dogs themselves may form a small part of the image. The new datasets also contains images of both dogs and cats, the two most popular pet types on the internet [7].

Nevertheless, there are a few limitations of the chosen datasets. In the “Cats vs. Dogs” dataset, although the number of cat and dog images is the same, and diversity of the cat images is lower than that of the dog images in terms



Figure 2. Sample images from curated dataset after preprocessing.

of the number of unique breeds and frequency of images for each breed, even when accounting for the fact that dog breeds are more numerous in general. Additionally, images may be of different sizes and may contain multiple animals and humans. However, these images are annotated, making it possible to crop and filter images using data preprocessing. The “Cats and Dogs Breeds Classification Oxford Dataset” contains much higher quality images overall in terms of resolution and focus on the animal; however, it contains fewer images. The data from both these datasets is combined to ensure data diversity while ensuring that enough images of acceptable quality are retained after preprocessing.

For the preprocessing step, images with low-resolution (less than 256 px in either dimension) and humans were dropped entirely. Then, using image annotations, each image was cropped into a square shape focused on the pet and resized to either  $64 \times 64$  or  $256 \times 256$  pixels, depending on the model. The data was augmented by adding laterally-inverted copies of each image. Finally, the images were normalized. In total, 36,232 images were available for use after preprocessing, with dog and cat images represented equally.

### 4. Methods

I utilize two existing GAN architectures - DCGAN and BigGAN - to develop a model for the specific task of generating high-quality pet images for Instagram. These models are trained using the curated dataset discussed in Section 3. Additionally, the user engagement generated by these images is evaluated.

#### 4.1. DCGAN

For the baseline model, I use a Deep Convolutional Generative Adversarial Network (DCGAN) architecture. DCGAN contains a Generator and a Discriminator like a traditional GAN, but both Generator and Discriminator models are based on convolutional neural networks. [6]. This was chosen as the baseline as it has been widely used in prior work [9][5] and may be trained using moderate computa-

<sup>1</sup><https://www.kaggle.com/c/dogs-vs-cats/data>

<sup>2</sup><https://www.kaggle.com/zippyz/cats-and-dogs-breeds-classification-oxford-dataset>

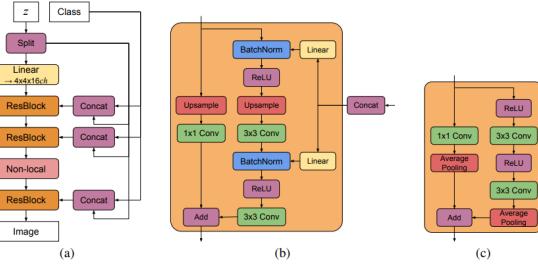


Figure 3. (a) A typical architectural layout for BigGAN’s G. (b) A Residual Block (ResBlock up) in BigGAN’s G. (c) A Residual Block (ResBlock down) in BigGAN’s D. Source: Brock et al. [2]

tional resources [4].

The implementation is largely based on the model and techniques described by Radford et. al. in their paper introducing DCGAN [6]. The model was implemented using [6] and the DCGAN implementation in Coursera’s “Build Basic Generative Adversarial Networks” course [11].

To train this model, I use a version of the binary cross-entropy loss, defined by:

$$V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))]$$

Where D and G represent Discriminator and Generator.  $G(z)$  is a sample generated by a random vector and x is real sample data. The optimal weight value is obtained by minimizing the loss function.

## 4.2. BigGAN

BigGAN is a GAN architecture based on Self-Attention GAN that is capable of synthesizing high-fidelity images, shown in Figure 3 [2]. An important component in the model is the use of the “truncation trick” which when used during image generation results in an improvement in image quality [4]. This model also uses orthogonal regularization:

$$R_\beta(W) = \beta \|W^\top W \odot (1 - I)\|_F^2$$

Where  $\beta$  is a hyperparameter and  $W$  is weight matrix, which improves the possibility of the truncation trick in generating high quality images [4]. However, compared to DCGAN, training the model is computationally expensive and time-consuming.

I rely on the official implementation provided by the authors, which has been pretrained on ImageNet.<sup>3</sup> The use of a pretrained model can result in significant time and resource savings, making it ideal given the resource constraints. The pretrained model is then fine-tuned with my curated dataset.

<sup>3</sup><https://github.com/ajbrock/BigGAN-PyTorch>

## 4.3. Evaluation

### 4.3.1 Model Evaluation

I use the Inception Score (IS) for evaluating model performance. The IS incorporates measures of both fidelity and diversity [8], which is relevant for my task.

Additionally, since prior work on related tasks uses IS, it is possible to compare results for my models.

### 4.3.2 User Engagement Evaluation

To evaluate the applicability of the approach for the specific task of generating images for social media use, it is critical to measure user engagement generated by the images using a metric. Importantly, this metric must be simple, application-specific (in this case, suitable for Instagram) and use publicly-available data to allow for comparison with content from other pages on the platform.

Yew et al. [10] propose the “Instagram Engagement Score” or IES using marketing case studies, which is a simple yet effective metric for measuring engagement. I use a scaled version of this metric, given by the formula:

$$\frac{\text{Instagram Engagement Score}}{\text{Number of Likes} + \text{Number of Comments}} = \frac{\text{Number of Followers}}{\text{Number of Followers}}$$

Critically, this metric can capture engagement across pages of different follower counts and uses publicly available data (i.e., the number of likes, comments, and followers). The metric may be applied to measure both image-level and page-level engagement. In this study, the image-level IES (or i-IES) is measured exactly 24 hours after an image is posted, using the follower count of the page and likes and comments count of the image of interest. The page-level IES (or p-IES) is measured at any given instant, using the follower count of the page and the sum of likes and comments count of the 10 most recent relevant images on the page.

## 5. Experiments

### 5.1. Image Generation

Images were generated using both the DCGAN and BigGAN models discussed in Section 4. For both models, data was preprocessed and augmented using the approach outlined in Section 3. Although the models were originally trained with dog images only, they were later trained with both dog and cat images in order to expand the scope of the project.

As the baseline, a DCGAN model was trained with the dataset from scratch. Following the recommendations in [6], I used the Adam optimizer for both the generator and

the discriminator, with learning rate,  $lr = 0.0002$  and  $\beta_1 = 0.5$ . The DCGAN was trained with smaller-sized  $64 \times 64$  images using a single GPU setup on the Farmshare 2 environment of the Stanford Research Computing Center. The model took a long time to train, and the fidelity of images, unsurprisingly, improved with the number of epochs.

The generated images had an IS score of 10.399 (dog images only) and 7.484 (dog and cat images) after 500 epochs. Notably, this is higher than the IS score achieved by models in prior work [4], potentially underlining the impact of the dataset choice.

Nevertheless, the images generated were low-resolution and easily distinguishable as artificial by humans, making them unfit for use as social media content. A BigGAN model was used next given its ability to produce high quality images [2]. While BigGAN can produce better-quality images, it also requires significantly more time and resources to train [4]. As a result, I used an implementation pretrained on ImageNet provided by Brock et al. [2]. This is then fine-tuned with my application-specific dataset using the larger  $256 \times 256$  images.

Following the recommendations in [2] and given the resource constraints, I use the Adam optimizer for both the generator and the discriminator, with learning rates,  $lr = 0.0001$  and  $0.0004$ , respectively with a batch size of 256. The model was trained (fine-tuned) for 10,000 iterations using a 4-GPU setup on Farmshare 2.

The model generated images that were not only higher-resolution but also more realistic than the baseline approach. Additionally, the model had an IS of 94.331 (dog images only) and 67.500 (dog and cat images), which presents a significant improvement over the baseline as well as the computationally-efficient approaches in prior works discussed in Section 2.

## 5.2. Managing an Instagram Page

An Instagram page was set up to upload generated images, using the handle @logans\_pawsome\_friends. The page follows the life of the titular golden retriever, Logan, who meets his various animal friends and posts their images. Since the goal is to evaluate engagement produced by GAN-generated content, no indication is provided about the images being synthetic, outside of subtle hints, such as the name of characters (such as LoGAN) and cryptic captions (“A fun day out with Snowy - it was *unreal!*”). A screenshot of the page is shown in Figure 4.

While images generated by the model are highly realistic, they may still be distinguishable by humans. As a result, images for posting are manually picked. As of the writing of this report, 15.2% of all generated dog samples have been considered fit for posting on the page. An example of the selection process is shown in Figure 5. Surprisingly, none of the cat images generated were fit for posting; while many

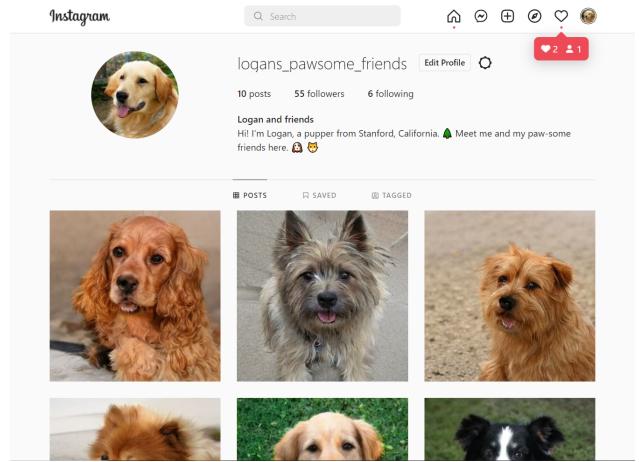


Figure 4. Screenshot of @logans\_pawsome\_friends Instagram page.



Figure 5. The top images were selected manually from the set of GAN-generated images for posting on the Instagram page; above, only the left-most image was considered fit for uploading.



Figure 6. Images which have the highest engagement score (i-IES) so far.

images are high-quality, they are still distinguishable by humans.

For evaluating the engagement of the images of @logans\_pawsome\_friends with other Instagram pages, three categories of pages are delineated: “low popularity” pages with under 100 followers, “medium popularity” pages with between 100 and 10,000 followers, and “high popularity” pages with over 10,000 followers. @logans\_pawsome\_friends is currently a low popularity page, but growing rapidly. In just over two weeks of the creation of the page, it has garnered 55 followers completely organically, without any promotion, either on Instagram or otherwise.

The user engagement of the page, as measured by the

Model	IS
<i>DCGAN - 350 epochs [4] (Dog Images)</i>	5.281
<i>BigGAN - 50 epochs [4] (Dog Images)</i>	8.173
<i>DCGAN (350 epochs) (Dog Images)</i>	7.879
<i>DCGAN (Baseline) (Dog Images)</i>	10.399
<i>DCGAN (Baseline) (Dog and Cat Images)</i>	7.484
<i>BigGAN (Pretrained) (Dog Images)</i>	94.331
<i>BigGAN (Pretrained) (Dog and Cat Images)</i>	67.500

Table 1. Summary of model performance in terms of inception score. My implementations which use the curated dataset described in Section 3 are italicized.

Instagram Engagement Score (p-IES) is 1.254 at the time of writing this paper.

### 5.3. Posting Low-Fidelity Images

While only the most realistic images were selected for posting on the Instagram page initially, I experimented with uploading a random image generated by the model directly to the page. This was potentially risky because lower-quality images may hurt engagement for the specific image as well as the page in general. It would, nonetheless, be useful to measure the impact of image fidelity on user engagement.

The first such image to be uploaded, however, was taken down by Instagram shortly after uploading for violating its terms of use based on user reports. Interestingly, the image was flagged as “Bullying or Harassment” rather than “Spam,” perhaps indicating that the user viewed the disfigured image of the dog as animal abuse rather than as a fake image. Regardless, due to the likelihood of images being reported and the risk of the page being banned altogether, the experiment was subsequently abandoned.

## 6. Analysis

My models yielded results that were either comparable or better than related tasks in prior work. The results are summarized in Table 1.

A key difference between prior work and my implementation is the choice of the dataset and related preprocessing. This underlines the fact that the model performance is sensitive to the quality of the training data. An important limitation is that the training time varies between the prior work and present study, and not all results are strictly comparable (except the first item and the third item in the table). Another limitation of this comparison is that it does not account for differences in hardware used for training the models.

While no prior work used images of both dogs and cats, using the comparison between the performance of my models on different data, it is clear that the models performed



Figure 7. The model performs poorly in terms of generating eyes. Since the feature is more distinctive for cats than dogs, the top two images were considered fit for posting on Instagram, while the bottom two were not.

better when only generating dog images. This may be because of two reasons: (a) the addition of a new class of data makes the task inherently more complex, and (b) cat images, in particular, may be more difficult to generate. While (a) is a well-understood problem, I speculate that (b) may have a role to play because of the quality of results generated by the model.

Specifically, in terms of interpretable features, the model tends to perform poorly in terms of generating realistic eyes for animals. This is true for both dog images and cat images. However, for dogs, the eyes are often uniformly colored and contain indistinct irises; as a result, these images may nevertheless seem indistinguishable from real images to humans. On the other hand, cats have distinctive irises. As a result, imperfect eyes can result in images not being suitable for human consumption altogether. Examples of this are shown in Figure 7. This underlined a key failure of the model, as it was unable to generate high-quality cat images.

The page-level engagement for @logans\_pawsome\_friends was compared against average (mean) page-level engagements for high-popularity, medium-popularity and low-popularity pages. For each category,  $n = 20$  pages were sampled. The distribution of the number of followers roughly corresponded to the standard normal distribution centered around the midpoint of the range of followers for each category. The upper limit for the range in the case of “high-popularity” pages was

Page / Page Type	p-IES
High-popularity pages (average)	0.025
Medium-popularity pages (average)	0.754
Low-popularity pages (average)	0.612
@logans_pawsome_friends	1.254

Table 2. Comparison of page-level engagement.

set to 9.9M followers, which is the number of followers on the @jiffpom page, the most popular page with comparable content. The results are summarized in Table 2.

@logans\_pawsome\_friends drives more engagement than the average in each category. Additionally, it also has significantly higher engagement than the most popular page, @jiffpom, with p-IES = 0.026. This may be at least partly due to the fact that as the number of followers grows on a page, it is likely that a large number of followers are passive, with engagement driven only by a small subset of active users [1]. This may also explain why engagement drops for high-popularity pages in general. However, @logans\_pawsome\_friends, a low-popularity page, also drives more engagement compared to the average low-popularity page, which may be attributed to the fact that it is able to deliver content more consistently, which can drive engagement.

The comparisons may suffer from sampling bias, especially in the low-popularity category, since these pages were sampled through Instagram-delivered results based on popular hashtags. Additionally, given the relatively small amount of images on @logans\_pawsome\_friends presently, it is not possible to isolate the effects of other variables that may drive engagement including, but not limited to, hashtags used with the images and time of the day of uploading the images. Regardless, the results show that GAN-generated images are capable of driving user engagement at levels that are at least comparable to those achieved by real images.

## 7. Conclusion

In this study, I evaluated the quality of pet images generated by two different GAN architectures, DCGAN and BigGAN, and evaluated their applicability for use as therapeutic social media content. The model performance results were consistent with prior work presented in [4] in that the BigGAN model generated better quality images. However, my experiments showed that the results may be further improved by the choice of the dataset. Additionally, using a pretrained model can have significant benefits, both in terms of training time and quality of images generated. This contributed to the significant improvement in the BigGAN-generated images when compared to Jain et al.’s work [4].

The dog images generated by the models were consistently better than the cat images; as a result, only dog images were used for evaluating user engagement with these images. These images were uploaded to @logans\_pawsome\_friends, a newly-created Instagram account, and the engagement generated by the images was compared to that of similar pages’ content. @logans\_pawsome\_friends achieved higher engagement than pages with real pet images on average, which underlines the applicability of the approach. This study, therefore, provides a basis for the creation of GAN-powered pet therapy pages, which can engage users at levels comparable to traditional pages, while allowing for large-scale content creation.

In the future, this study may be extended by exploring new techniques or augmenting the discussed techniques (by using new data or training for longer) to generalize the task to other pets (such as cats). Additionally, further data can be collected from the Instagram page over an extended period of time to understand the individual impact of different variables on driving user engagement and to reinforce confidence in the applicability of this approach as the page grows in popularity.

## References

- [1] Matthew Blaine and Jacalyn Kremer. Are we failing at instagram? *Social Media: The Academic Library Perspective*, page 101, 2018. [1](#), [6](#)
- [2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018. [3](#), [4](#)
- [3] Richa Gupta and Vanshika Gupta. Performance analysis of different gan models: Dc-gan and ls-gan. In *2021 7th International Conference on Signal Processing and Communication (ICSC)*, pages 222–227. IEEE, 2021. [1](#)
- [4] Ayush Jain, Ayushi Bansal, and Yogesh Kakde. Performance analysis of various generative adversarial network using dog image dataset. In *2020 International Conference for Emerging Technology (INCET)*, pages 1–6. IEEE, 2020. [1](#), [3](#), [4](#), [5](#), [6](#)
- [5] Nandini Kumari, Shamama Anwar, and Vandana Bhattacharjee. Dcgan based pre-trained model for image reconstruction using imagenet. In *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pages 1–6. IEEE, 2021. [1](#), [2](#)
- [6] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. [2](#), [3](#)
- [7] Elizabeth Riddle and Jill RD MacKay. Social media contexts moderate perceptions of animals. *Animals*, 10(5):845, 2020. [1](#), [2](#)
- [8] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques

- for training gans. *Advances in neural information processing systems*, 29, 2016. 3
- [9] Zhongkai Shangguan, Yue Zhao, Wei Fan, and Zhehan Cao. Dog image generation using deep convolutional generative adversarial networks. In *2020 5th International Conference on Universal Village (UV)*, pages 1–6. IEEE, 2020. 1, 2
  - [10] Roy Ling Hang Yew, Syamimi Binti Suhaidi, Prishtee See-woochurn, and Venantius Kumar Sevamalai. Social network influencers’ engagement rate algorithm using instagram data. In *2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA)*, pages 1–8, 2018. 3
  - [11] Sharon Zhou, Eda Zhou, and Erik Zelikman. Build basic generative adversarial networks: Deep convolutional gan (dcgan). 3
  - [12] Zekun Zhou, Duo Yin, and Quan Gao. Sense of presence and subjective well-being in online pet watching: The moderation role of loneliness and perceived stress. *International journal of environmental research and public health*, 17(23):9093, 2020. 1