This Fish Does Not Exist

Callan Hoskins
B.S., Computer Science
Stanford University
Palo Alto, CA
chosk@stanford.edu

***Abstract*—Advancements in GANs have made it easier than ever to make a creative AI agent. NVIDIA's StyleGAN2 model architecture has proven its ability to generate realistic-looking, high-quality fake images. In this mini-project, I successfully apply StyleGAN2 to a novel dataset of fish images and compare the results from StyleGAN2 variations. I also made a website, This Fish Does Not Exist, showcasing the fish generated.**

***Index Terms*—machine learning, GANs, generative, emotions, emotion recognition, FER, deep learning**

I. Introduction

Generative Adversarial Networks (GANs) are the *de facto* standard in generating articial images. Since the invention of GANs in 2014, they have become more intricate, effective, and just plain better. Recent improvements have enabled GANs to generate incredibly lifelike images.

The application that first caught my (and many others' interest) was in making realistic images of people's faces. Others have applied GAN to the night sky, cats, AirBnb photos, pictures of Homer Simpson, and more. In this mini-project, I use a novel dataset to create a GAN capable of generating realistic images of salt-water fish.

II. Related Work

NVIDIA's StyleGAN [1] was a breakthrough improvement in generator architecture when it was released in 2018. Building on the earlier ProGAN, it built on the foundation of progressive growing and incorporated other design features to become the SOTA for certain types of generative applications. In particular, the authors of StyleGAN added a) bilinear sampling, b) a mapping network and adaptive instance normalization, c) translation of latent point input, d) Gaussian noise input in each generator block, and e) mixing regularization to progressive growing. Their results were fantastic.

But StyleGAN wasn't perfect. The following year, the authors released StyleGAN2 [2], which improved upon StyleGAN by adding weight demodulation, lazy regularization, path length regularization, a larger network, and removing progressive growing. It replaced StyleGAN as the SOTA.

In 2020, Liu et al. proposed a "lightweight" variant of StyleGAN2 that performs better than StyleGAN2 on a limited computational and data budget [3]. The model uses skip-layer channel-wise excitation and self-supervised loss of autoencoding to achieve success in the few-shot setting.

A. Applications

The original StyleGAN and StyleGAN2 papers showed the model architectures applied to high-resolution images of faces. The website This Person Does Not Exist showcases some examples of faces generated in this method; it went viral. Since then, these models have been applied by others to cats, vessels, feet, artwork, and more. But no one has applied it to fish... that is, until now.

III. Dataset

In this paper, I also introduce a novel dataset: collected images from FishBase. FishBase is an online forum where fish enthusiasts can post images they capture of fish. The photos vary in their quality, but most are pretty good pictures of fish.

Using BeautifulSoup and Selenium, I scraped most of the images from FishBase – almost 60,000 in total. The images were grouped by the family of fish in the photograph.

However, fish differ vastly between families, and some of the fish are *really* weird-looking. To make it a little easier for the GAN to converge, I chose a subset of the data that had a more coherent look.

Fig. 1. *This* is a fish??? An example image from the original set of 60k FishBase images.

The subset I chose spans only a few families (Pomecetridae, Cichlidae, and Labridae are some of the largest) but includes 17,880 images (see below for examples from this subset).

I zero-padded each image on the top and bottom to make it a square, then resized it to a 256x256 image; since most images were high-resolution, this involved downsampling.

I would like to publicize this dataset, but it's currently private because I don't know the copyright laws involved in releasing a dataset like this.

IV. Methods

I used ready-to-go StyleGAN implementations written in Pytorch (shoutout `@lucidrains`).

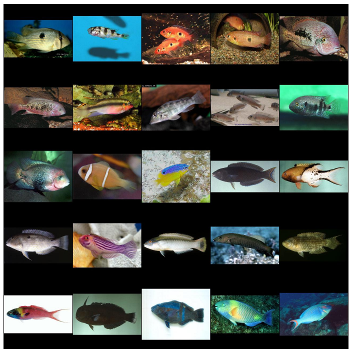

Fig. 2. 25 images from the subset chosen for training.

I was aiming to get the highest-resolution, highest-fidelity images I could, which I measured, for the most part, qualitatively.

V. Experiments and Results

A. Lightweight StyleGAN2

The first model I trained was ”lightweight” StyleGAN2 [3], for 150 epochs at 128x128 output resolution. This was a small, proof-of-concept model that was easy to train. The results were encouraging, but they weren’t as good as I hoped they would be. I used `@lucidrains`’s ”lightweight” StyleGAN2 implementation.

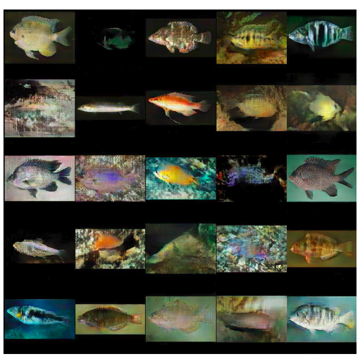

Fig. 3. 128x128 images generated by ”lighweight” StyleGAN2 after 150 epochs

B. StyleGAN2

I could see that I was onto something. With the data I had collected and a simpler model (”lightweight” StyleGAN2), I had gotten pretty decent results. The fish looked lifelike. However, they weren’t as realistic as I wanted them to be, so I turned to a more complex model, StyleGAN2. Again, I used `@lucidrain`’s StyleGAN2 implementation.

As another proof of concept, I trained StyleGAN2 with 128x128 output images (this took far less training time than 256x256 outputs). I continued training for 131 epochs (at which point I stopped seeing meaningful improvement in generated images).

I was quite happy with the 128x128 images. When ensembled, the model produced lifelike fish images more often than not, and the fish generated were pretty diverse in their shape, color, and backgrounds. However, when I viewed the 128x128 images on their own, I was disappointed by their small size. So I went bigger!

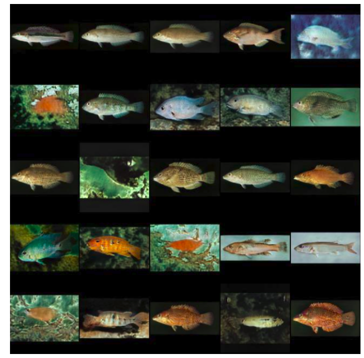

Fig. 4. 128x128 images generated by StyleGAN2 after 131 epochs

I trained the StyleGAN2 model to produce 256x256 output images. This took a LOT longer than the 128x128 model, so I wasn’t able to run it for as long. Still, after only 48 epochs, the ensembled model was creating lifelike fish pretty often (and I was out of money).

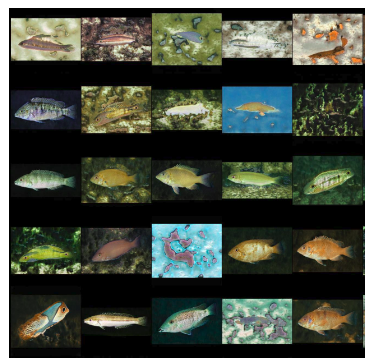

Fig. 5. 256x256 images generated by StyleGAN2 after 48 epochs

Although the images aren’t as lifelike as the 128x128 images produced by the 128x128 model that trained for longer, many are still realistic enough that I would believe they are real fish. Additionally, they are bigger images that have an impressive amount of detail.

VI. This Fish Does Not Exist

I though the results of my model were pretty cool, so I decided to put some of them online. Unfortunately, because of the dual limitations a) not all the images produced by the 256x256 StyleGAN2 model looked like fish and b) I couldn’t afford to keep a GPU running for my website, I hand-picked realistic-looking generated fish to display on my website.

The super simple site is at This Fish Does Not Exist. You can refresh the page to see another never-before-seen fish generated by StyleGAN2. Check it out!

Acknowledgment

Many thanks to the users of FishBase for taking such great photos of fish and the TAs and instructors of CS236G, especially Sharon Zhou, for showcasing the beauty of GANs.

References

[1] Karras, Tero; Laine, Samuli; Aila, Timo. ”A Style-Based Generator Architecture for Generative Adversarial Networks.” 2019. https://arxiv.org/pdf/1812.04948.pdf.

[2] Karras, Tero et al. ”Analyzing and Improving the Image Quality of StyleGAN.” 2019. https://arxiv.org/pdf/1912.04958

[3] Liu, Bingchen; Zhu, Yizhe; Song, Kunpeng; Elgammal, Ahmed. ”Towards Faster and Stabilized GAN Training for High-Fidelity Few-Shot Image Synthesis.” ICLR 2021. https://openreview.net/pdf?id=1Fqg133qRaI