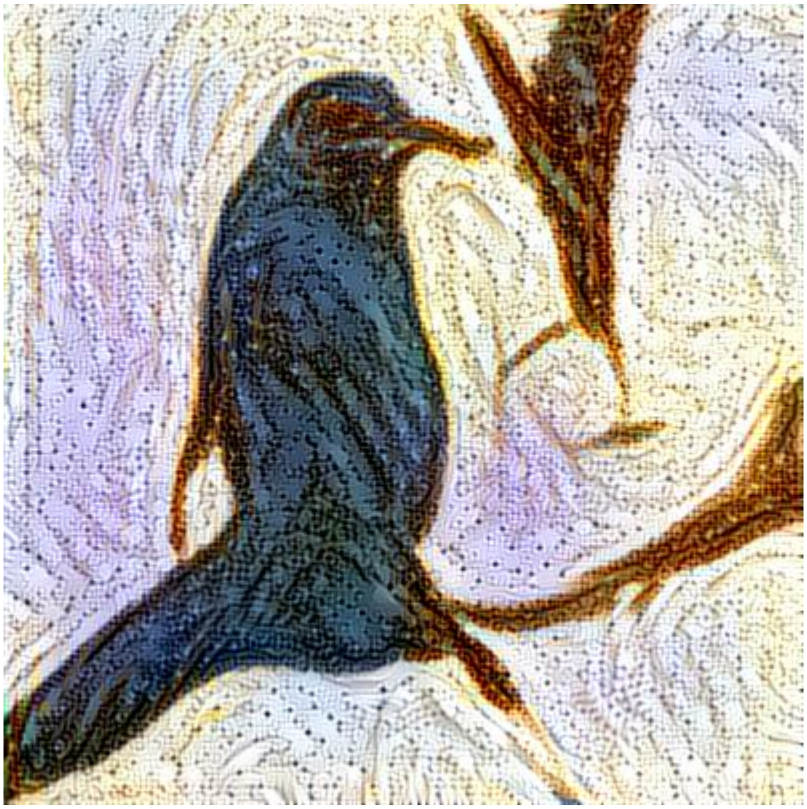
Machine Learning for the Arts
UCSD SPRING 2019
FINAL PROJECT

Generated Birds



(Sample final product)

Jeffrey Yeung
A12217277

Description

For my final project, I decided to extend and combine the applications of my work from project 1: text generation, and project 4: image generation. The overall project consists of three main components: generating a textual description of a bird (character-based RNN, generating an image of a bird given that textual description (AttnGAN), and finally adding artistic and aesthetic values with style transfer (Neural Style Transfer). The sample final product displayed in the title page is the result of starting with only a text file of descriptions and captions of scientific videos.

Concept

The concept developed as I was initially doing project 1: text generation. I had the idea to take it further by converting the textual output to a pictorial output. As the course advanced, I learned more about other architectures and techniques, and when we reached the later lectures in the course, I came across StackGAN and researching into that topic led me to AttnGAN, which is exactly what I needed for my idea. However, just having a plain image of a bird was dull and boring so I incorporated style transfer to add background to it by applying sceneries to my birds.

Technique

As introduced earlier, my final project consists of three main components that I will individually discuss.

Text generation with character-based RNN:

The idea behind character-based RNN is that given a sequence of characters from a training corpus, the model is able to predict the next character to appear following some input sequence. This technique is suitable for my needs because bird descriptions tend to be short and succinct, so each character is precious and there is no need to generate a large amount of text. This RNN works because it only not accounts for the most recent input character, but also the history of all the characters inputted.

AttnGAN:

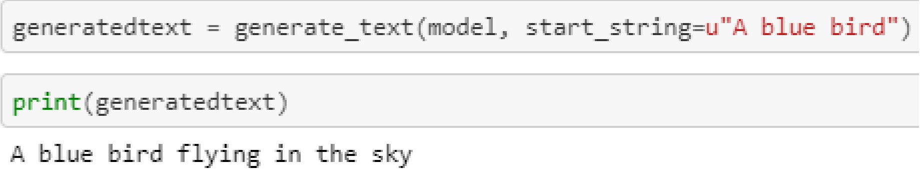
As its name states, AttnGAN is a Generative Adversarial Network that is attention-driven and consists of a multi-stage refinement process for text-to-image generation. AttnGAN works by starting with a simple, low-quality image generated using a GAN, then iteratively improving the image by adding details each step.

Neural Style Transfer:

This version of style transfer is based on using Convolutional Neural Networks to extract high level features from images to construct feature representations that can be generalized onto other images. These features fall under two categories, content representation and style representation. These representations are extracted separately using a CNN and then applied simultaneously to a new image, and the result is refined by minimizing the loss from the content and style from the white noise image.

Process

Starting with a text file of many video captions/descriptions of videos related to science and nature, I fed it into the character-based RNN and fed it a seed of “a blue bird” to ensure the output text is at least related to birds. I lowered the temperature to 0.5 from the default 1.0 to improve the stability of the output at the cost of novelty, which is actually bad for this project because the AttnGAN may have a difficult time understanding obscure descriptions. Even at a low temperature I had to generate many times before something simple and usable was generated. In the end, the best caption fitting my criteria was “a blue bird flying in the sky” – simple and straightforward.

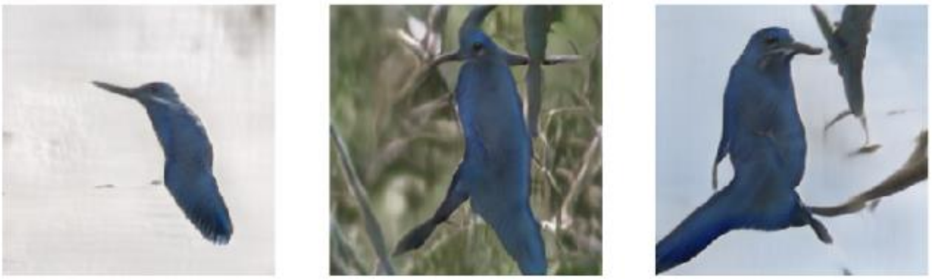


```
generatedtext = generate_text(model, start_string=u"A blue bird")
```

```
print(generatedtext)
```

```
A blue bird flying in the sky
```

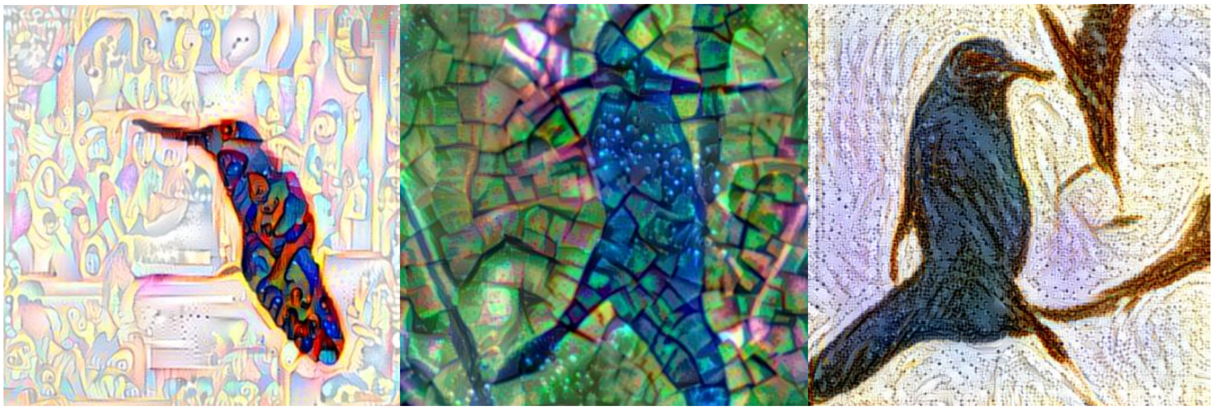
Once I have that caption generated, I fed it into a pretrained AttnGAN model that was trained on the Caltech-UCSD Birds dataset. Generating a normal-looking blue bird was much simpler and quicker than generating the caption because the dataset the model was trained on is much larger than the text file that I trained the RNN on. In the end, I chose the following three birds to continue my project; the middle bird is not as usual looking because I wanted to incorporate one that was strange.



After having the content images for my style transfer, I noticed that the backgrounds were very plain, and just having semi-realistic looking birds was boring to look at, so I wanted to create something more abstract by applying sceneries onto them. I applied a different style to each bird picture (strong white, gold, and green color schemes). The results can be seen in the next section.

Results

The final results turn out as well as I expected, the right and left images very obviously depict a bird, however, the middle image may not be as easy to tell because the content is closely blended with the background (color wise), and with the style patterns applied onto the bird, it makes it even harder to see since the beak and head are blended with the style.



Reflection

Overall, I am very satisfied with the project because the final product is as I envisioned. I learned many new techniques in completing this project and the only disappointment I have is that I tried to apply neural machine translation onto the caption and generate two separate branches of images to see how birds can be generated in different languages but I was unsuccessful with that process.

References

Andrej Karpathy, May 21, 2015
http://karpathy.github.io/2015/05/21/rnn-effectiveness/
Tensorflow Tutorials
https://www.tensorflow.org/tutorials/sequences/text_generation
Xu, Zhang, Huang, Zhang, Gan, Huang, He, Nov. 28, 2017
https://arxiv.org/pdf/1711.10485.pdf
Gatys, Ecker, Bethge, Sep. 2, 2015
https://arxiv.org/pdf/1508.06576v2.pdf

Github Code

https://github.com/ucsd-ml-arts/ml-art-final-jeffrey

Result Links

See Github submissions