

# Attention GAN Training Using ZSL\_GAN Text Encoder

Ammar Ali<sup>1</sup> <sup>1</sup>ITMO University, St. Petersburg, Russia

## 0.1 Basic Solution

AttnGAN training consists of two stages. The first stage is to train the DAMSM model to provide Useful features from the text context to generate the image. The second step includes full training of the conditional-GAN that uses the first stage weights to get the text encoding and image encoding to train the Generator and Discriminator see Fig(1). On the other hand, ZSL\_GAN can handle noisy

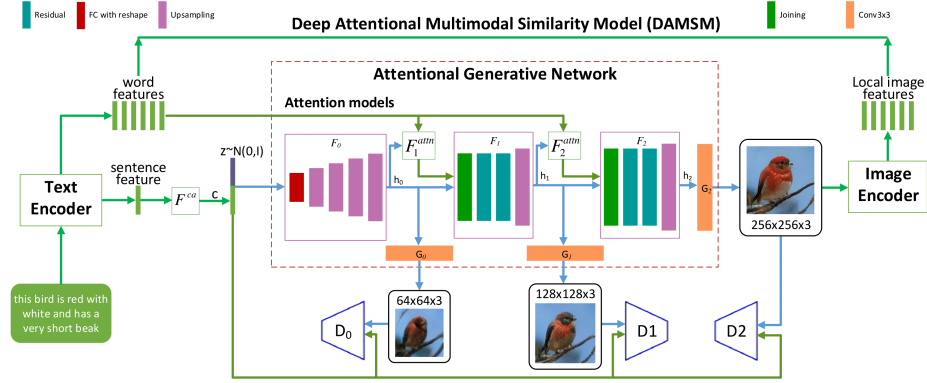


Figure 1: Attention GAN structure

inputs. and the goal is to use ZSL\_GAN text encoder in the Attention GAN. So let us take a look at ZSL\_GAN architecture Fig(2). A simple way to do the

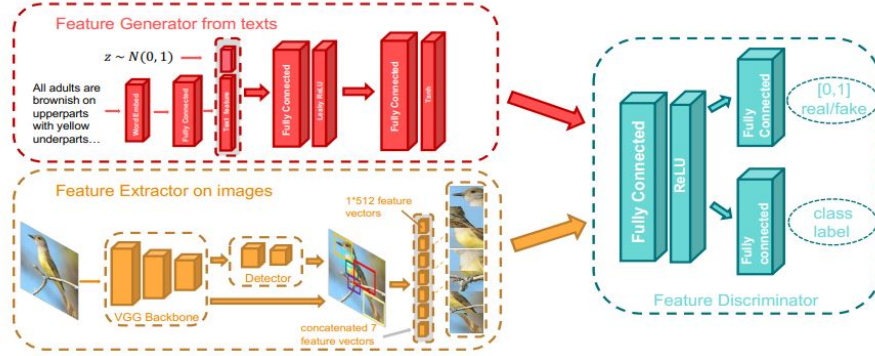


Figure 2: ZSL\_GAN structure

merge is to split the train into 3-stages. The first stage is by training ZSL\_GAN in the standard way. In the second stage, Take the text encoder use it instead of embedding layer in the Attention GAN and train the DAMSM model. The final stage, train the Attention GAN. See Fig(3).

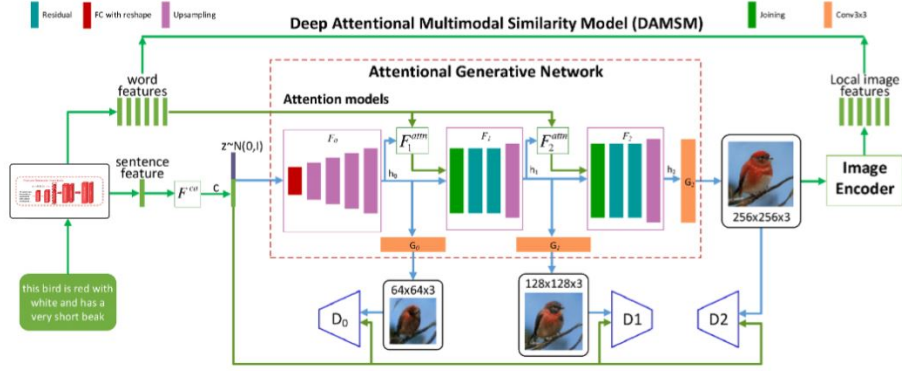


Figure 3: Basic solution

This solution has been implemented in this repo: <https://github.com/ammaraali32/AttnGAN-with-ZSL-Text-Encoder>

## 0.2 Future Development

There multiple directions to improve the current model. Unfortunately because of the task description that was about merging these two solutions. I haven't done any experiments for these suggestions.

1. Using Bert encoder to encode the text could improve the results.
2. The feature Extractor in Attention GAN is Inception. (multiple models have been released after changing it could also improve the results"
3. Trick that in most cases works for me that after training. Generate different images then train a classifier to classify between the fake images and the real ones. freeze the model and use its features As a second image encoder.
4. The State of Art now is DF\_gan <https://github.com/tobran/DF-GAN>. It is similar to AttnGAN in the Text Encoding. I think it is better to apply ZSL\_GAN text encoding on it.