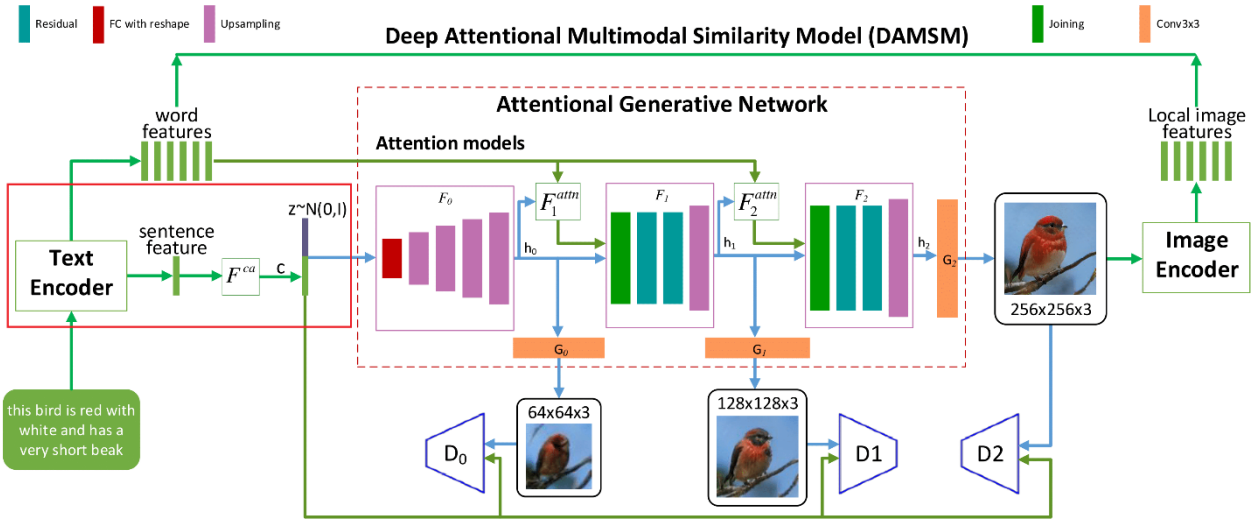
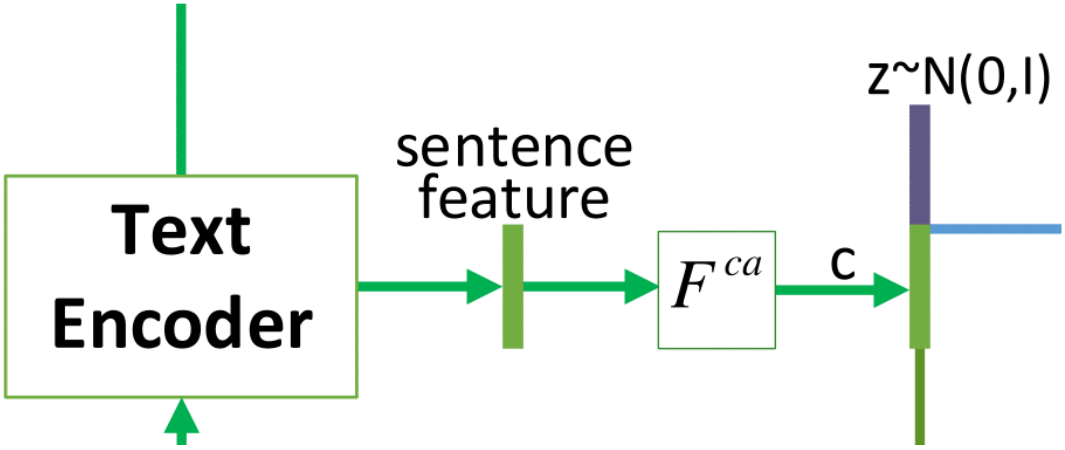
REPORT

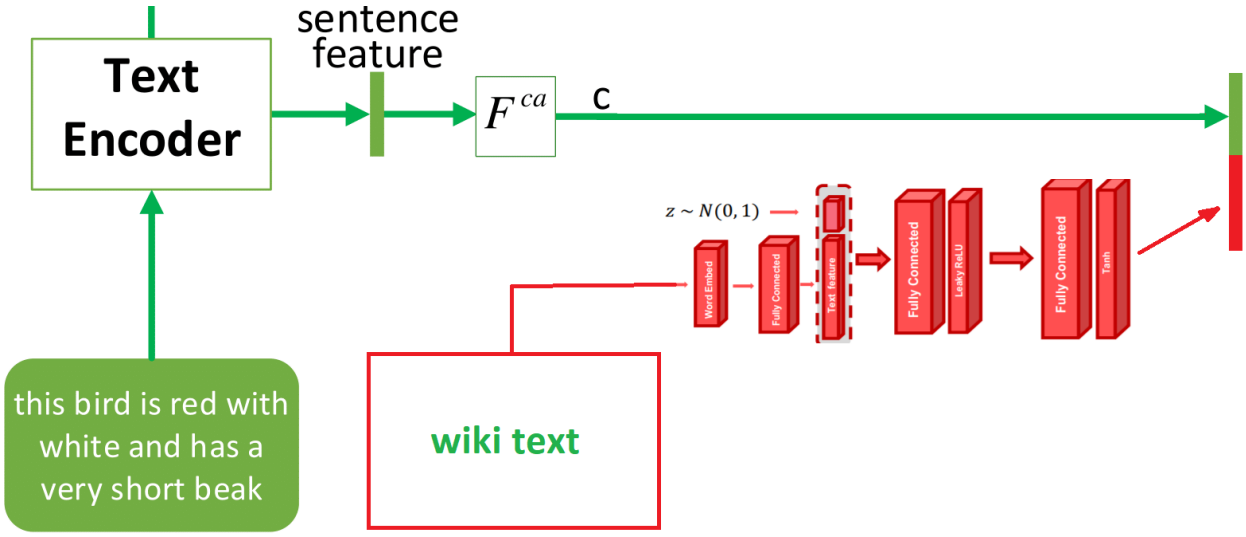
Old AttnGAN architecture:



Interesting part:



This part can be modified using module from ZSL_GAN (**final architecture**):



Motivation:

In AttnGAN only small phrases are used to extract features – this method is will not work with texts (where a lot of extra information can be seen). Second method works on big texts and it can filter provided information. To sum up: first approaches is not able to work with noise, while second is noise resistant.

We are not able to use second approach instead of first: second approach uses TF-IDF and it needs more text.

However, we can concatenate two approaches. The main disadvantage – we need dataset, which has Wiki links and small descriptions at the same time (for example, CUB).

Why it should work? Extracting features from phrases will give us a chance to find the most valuable parts for attention. Extracting features from text will give us more general information from descriptions – and so more details on an image.

About code:

It's not an interesting part, cause it's just about using AttnGAN code with some little changes from ZSL_GAN.

I've provided only parts with changes (as I know, I'm not able to publish all code in public github due to copyright rules). Unfortunately, CUB dataset wasn't available for some reasons and I couldn't debug this code. But idea can be seen quite clear.