OVERVIEW

1. Project Background and Description

The last few years have seen amazing advancements in image classification and text-to-image generative models, from the highly malleable generative model put forth in “Generating images from captions with attention.” (Mansimov, Elman, et al. 2015) to the highly realistic generated images put forth in “Generative adversarial text to image synthesis.” (Reed, Scott, et al. 2016). As these generative models get more powerful, it is interesting to explore new contexts in which they might be utilized.

The focus of my project will be to analyze data from the New Yorker’s cartoon captioning contest, and see whether I can use these generative image models or Image search tools such as Bing Image Search and Google Custom Search to generate New Yorker cartoon style images. I will first attempt this using a set of hand-written descriptions that has been generated for the dataset. Should this endeavor be successful, I will move on to attempt this generation using the humorous captions collected from the New Yorker image captioning contest. In both cases, I hope to apply style transfer to the image that I generate or query to bring it more in line with the New Yorker’s specific comic aesthetic.
The first stage of this project will be developing a reliable pipeline for generating images given a few key descriptors, and then applying style transfer to convert those images to the required style. From there, I hope to explore a mapping between the humorous captions that were given in the captioning contest and the physical descriptions generated by hand. Should a mapping be found, I hope to add this conversion to the pipeline, converting from humorous description to physical descriptions, and then from physical descriptions to New Yorker style images, which can be then modified to match the cartoon aesthetic.

2. Project Scope

By the end of the project, I hope to:

- Establish a pipeline for converting image feature lists to New Yorker style images
- Identify a mapping between jokes commonly made about an image and that image’s most important features
- Utilize that knowledge to build a system which can take in a set of suggested captions, and generate a list of probable image features
- Integrate this feature recognition system into the photo generation pipeline, allowing the system to take in a set of captions, and generate a photo that is hopefully representative if the content discussed in the captions

I do not intend to:

- Develop a competitive text-to-image generative model
• Generate photos that properly arrange features in the humorous juxtaposition that makes the comics interesting
• Generate photos which closely resemble New Yorker cartoons

3. Deliverables

By the end of the project, I hope to have:

• An implementation of some text-to-image algorithm which can be used to generate new-yorker-style cartoons given image descriptions
• Either code or analysis, exploring the relationship between common captions and features within a cartoon
• Documentation describing how the system works, what decisions were made in its building, and the underlying logic behind those decisions
• A moving-forward plan for the system, detailing interesting directions to take the work in the future.

4. Referenced papers:
