sketch assistant : Human in the loop application to drawing sketches

Kumar Bhargav Srinivasan
Department of Computer Science
University of Colorado Boulder
Boulder, CO 80309
kusr7198@colorado.edu

Abstract

Auto generation of images from text has several applications in image manipulation, computer gaming etc. It could be used as a drawing aid in the perceptive of people with Parkinson's disease, (A disorder of the central nervous system that affects movement) who usually face problems like shakiness, tremors and difficulty writing. The proposal here is to create an application which is based on Human in the loop machine learning and will take description of the image as input and generate an image, If the user isn't satisfied with image generated, he/she provides feedback via natural language instruction. This feedback is fed into network to improvise on the image generated. Plan is to use doodle dataset Quick draw, open sourced by Google,a collection of 50 million drawings across 345 categories.

1 Introduction

1.1 Motivation

With the recent successes of Generative Adversial Networks (GANs) and Variational Autoencoders (VAEs), there has been a lot of work and interest in the research community on image generation from text input [1-2-3]. Some Systems also help Machines learn to caption Images via human feedback[8]. Microsoft created a system using Attentional Generative Adversial Network[1] can synthesize details at different subregions of the image by paying attentions to the relevant words in the natural language description, But these systems lack a feedback loop which could help in refining the image generated by understanding the features extracted from the feedback. These state-of-the-art models[1] for image generation are capable of generating images on geometrical datasets, but don't do very well on datasets like Quickdraw where subjects are not always centered in the image and not perfect in terms of shape as they are handwritten. Also description of the image might not contain all the relevant details which adds on to the complexity and limit the amount of image-text data pairs. It would be hard to understand which object belongs to the text when multiple objects are present drawn out for the same text description in multiple ways.

The Quick Draw Dataset is a collection of 50 million drawings across 345 categories, contributed by players of the game Quick, Draw!.[4] The drawings were captured as timestamped vectors, tagged with metadata including what the player was asked to draw and in which country the player was located. There are multiple images associated with the category, some are complete and where as others are incomplete. Initial step would be to annotate/create caption to the images using crowd-sourcing and add update information to dataset by provide an interface to edit the images and description and capture the changes as a state machine. Once this data is obtained we could use the GAN models to generate images using description.

2 Proposal

Sketches are typically drawn based on a back and forth conversation/discussion about the attributes of the final sketch in mind, A typical artist would start with bare skeleton the sketch and add on to it by knowing the details of intended sketch based on feedback on the current state of the sketch. Here I propose a similar approach to build a assistant/chat bot which would generate an initial sketch based on the description of the image and User can modify the generated image by providing verbal feedback.

2.1 Dataset Preparation

Data provided by Google just has category of the image, We need more information about the image to improvise on the content which could be generated.Following steps are formulated to synthesize data for the experiment.

- **Crowd-sourcing to update the description for the image**
  Create an interface where user can provide better caption of the objects present in the image and indicate how well image could identify the intended object.
  For example, these are set of the flamingo sketches in different ways. Each of them lack some attributes which need to be annotated.

Figure 1. shows set of flamingos drawn in Quickdraw

- **Capture the sequence of edits to the image based on the feedback**
  Collect the feedback via chat interface pairing two workers on Amazon Mechanical Turk (AMT). One of them provides feedback on current image to add or remove the elements from the image, other creates an updated sketch based on the feedback. Following is an example of how we collect the feedback:

Figure 2. shows flamingo after feedback to add beak and legs.

3 Approach

Attention in Neural Networks is modeled by considering How Humans focus on a particular subset of their sensory input, and tune-out the rest[6]. It is employed in applications where you have a collection of data points, all of which may not be pertinent to the task at hand. In our case we want to give attention over user feedback and generate the image.

I would propose to use a modified attention Generative adversarial network which could take in feedback and text description along with memory state as inputs and produce image as output. A Generative adversarial network consist of a generator network G and a discriminator network D. Given training data x, G takes input from a random noise z and tries to generate data that has the

similar distribution as x. Discriminator network D takes input from both training data x and generated data from G, it estimates the probability of a sample came from training data rather than G[3].

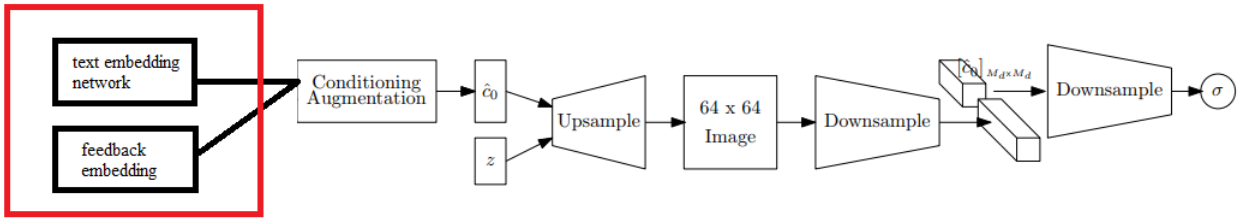

Figure 2. shows the modified GAN with Feedback and conditional augmentation.

Approach would be to build the model in following order:
Generate text embedding by encoding the captions with a skipthoughts pretrained model and for feedback embeddings use a Recurrent encoder to generate Skip-Thought vectors. concatenate the text and feedback embeddings and pass to Conditioning Augmentation module to produce latent variable inputs for the generator. These embedding are fed into GAN to generate a low resolution image. This image can be enhanced by stacking GAN. GAN's generator consists of fully connected layer followed by convolutional network for upsampling. Discriminator would involve deep convolutional layers with text embedding and generated image as input.

3.1 Evaluation

Evaluating the quality of these systems is a hard problem. Salimans propose to use an image classifier to evaluate the quality of generated images called Inception score. The intuition behind this that we want out model to generate meaningful objects. We could use top-r metric for selecting r relevant images out of all the images generated by the network.

3.2 Resources needed

For dataset preparation, server need to be hosted to collect better text description for image and Interface involving two people to provide feedback and draw updates to image.

4 Future work

a sketch assistant could be integrated with software tools used by painters and interior designers to auto-generate images. Also further extend it to take voice input to refine photos.

5 Dataset

Following are the potential datasets which could be used for solving the problem.

- Quickdraw dataset with 345 categories of images
- COCO dataset with Visual Dialog Dataset (VisDial)
  Common objects in context, is a large-scale object detection, segmentation, and captioning dataset with around 330K images and 1.5 million object instances and 80 categories.
- CODraw - clipart dataset annotated with updates on the images with update description.
  This is not publicly released but can potentially resolve the dataset preparation problem discussed earlier.

6 References

[1] Tao Xu, Pengchuan Zhang, Qiuyuan Huang,Han Zhang, Zhe Gan, Xiaolei Huang, Xiaodong He, Nov 2017, AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks

[2] Han Zhang , Tao Xu , Hongsheng Li, Shaoting Zhang , Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, Aug 2017, StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks

[3] Jiale Zhi, 2017, PixelBrush: Art Generation from text with GANs

[4] Quick draw dataset https://github.com/googlecreativelab/quickdraw-dataset

[5] Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, Yoshua Bengio, Feb 2018, ChatPainter: Improving Text to Image Generation using Dialogue.

[6] Understanding AttnGAN: Text-to-Image convertor - https://codeburst.io/understanding-attngan-text-to-image-convertor-a79f415a4e89

[7] Seitaro Shinagawa, Koichiro Yoshino, Sakriani Sakti, Yu Suzuki, Satoshi Nakamura, Feb 2018, Interactive Image Manipulation with Natural Language Instruction Commands

[8] Huan Ling, Sanja Fidler, June 2017, Teaching Machines to Describe Images via Natural Language Feedback.