CMU RI 16-995: Independent Study:
Listening Generative Models

Advisor: Jean Oh
The Robotics Institute, Carnegie Mellon University
Fall 2018

1 Course Description

Objective: The goal of this independent study is to develop an algorithm for generating and interactively modifying an image based on an input sequence of language descriptions. The idea is related to Interactive Generative Adversarial Networks (i-GANs) [20] where a user can modify an image interactively using a graphical user interface (GUI). Instead of using a GUI to interact with the system, we propose a system that listens to natural language descriptions to interactively (re-)generate images.

Background research: This study requires solid understanding of basic deep learning approaches including feed-forward and recurrent neural networks that can be reviewed in textbooks such as [4] (Part I and II). As background research for this study, a literature survey will be conducted on recent progress on image synthesis, in particular Generative Adversarial Networks (GANs) models [5, 10, 18, 20, 7, 21, 13, 14, 22, 16, 6, 15, 17, 11, 8, 1, 19], Variational Autoencoders (VAEs) [9], and Flow-based models [2].

Datasets: For this study, we plan to use sketches rather than photo-realistic imagery. This study will utilize publicly available datasets including the human sketch dataset [3] and Google Quick! Draw dataset 1. The student is

1https://github.com/googlecreativelab/quickdraw-dataset

Figure 1: An motivational example of incremental hand drawing [12]
expected to do further research on additional datasets as needed. Additionally, the student will be responsible for collecting language description data over the chosen sketch dataset using Amazon Mechanical Turk. The cost of this new data collection will be paid by the course advisor.

**Experiments:** On one set of experiments, we will evaluate the image synthesis on the final image only as follows. The algorithm will be evaluated on the dual-learning manner. We first train a multi-class classifier that maps a sketch to a label using the sketch dataset. Next, we use this classifier to classify the generated sketch from our system. Finally, we compare the accuracy of the classifier on the generated images against that of human drawings, e.g., if the accuracy is close to that on the test dataset that includes human drawings then it indicates comparable performance.

On the second set of experiments, we will create an online/offline game version of our system and collect user statistics, e.g., how long each participants engage in the game and their reactions, for future study.

**Evaluation:** The student and the advisor will co-author a technical paper that includes a formal problem definition, related work, detailed technical approach, experiments and results, and conclusion and future directions. The report will be written incrementally we we keep track of the progress.

**Reading list**


