Story Illustration – From Texts to Images

Yuexing Hao
Rutgers University
yh599@scarletmail.rutgers.edu

RUTGERS
Aresty Research Center for Undergraduates

Abstract

Nowadays, human beings came across many photos or pictures every day, from Instagram, Snapshots, Messages, etc. They are reading the information from images and imagining consequences of images from texts. For this project, I will transfer the story texts to consequences of connected visual images, which will be called as "story illustration". I will employ FLASK in Python as the web's framework, and build a webpage based on Hareesh Ravi's story model [1]. The webpage will get the users' texts inputs and send to the model in server to decode some keywords, and then match these words to the visual images in VIST storytelling dataset. Then it will display a consequence of visual images to the webpage.

The research summary will be about how to let the computer to display visual story illustrations from the consequences of texts.

Keywords: story illustration, connections, texts, visual images

Background

In this machine learning project, I will let the machine to illustrate stories from the pictures. Just like our parents did when I was young, they bought picture books and taught me words, logic, and grammar. I will employ Python as the basic language of my project and use deep learning packages like Pytorch, TensorFlow, numpy. By building a coherence model, I will first decode the pictures to derive some keywords, and then match these words to the texts.

For example, when people saw the pictures of a boy and a girl kiss, hug, play around, they may recognize these two people are a couple and enjoy their days for vacation. My machine will show one of the pictures, for instance, contains a boy, a girl, a kiss, a bunch of flowers. Then I will generate these words and encoded them in the new story model. The output of this story model will be a description of the pictures. Or when they saw pictures of a student wearing graduation attire and photting with family, having a party in the yard, it may be a story about a student graduates from some institutions and finishes up career.

Methods and Procedures

First, we are going to build a requests and responses function to allow users to input and let the database to get their inputs and send them to the server. In our server, we will insert the users' stories to the model, and run the model until it displays the matching five pictures in the VIST datasets. The pictures below illustrate the example of a model's input stories and outputs. It displays more than five pictures which are from CNSI model which wins the most votes from the AMT workers. [1] According to the matching level, the model will choose the best images for the stories.

Secondly, after getting the URLs of the model's output images from VIST datasets, we will send back the sequences of images back to the display page. The result pages are in below.
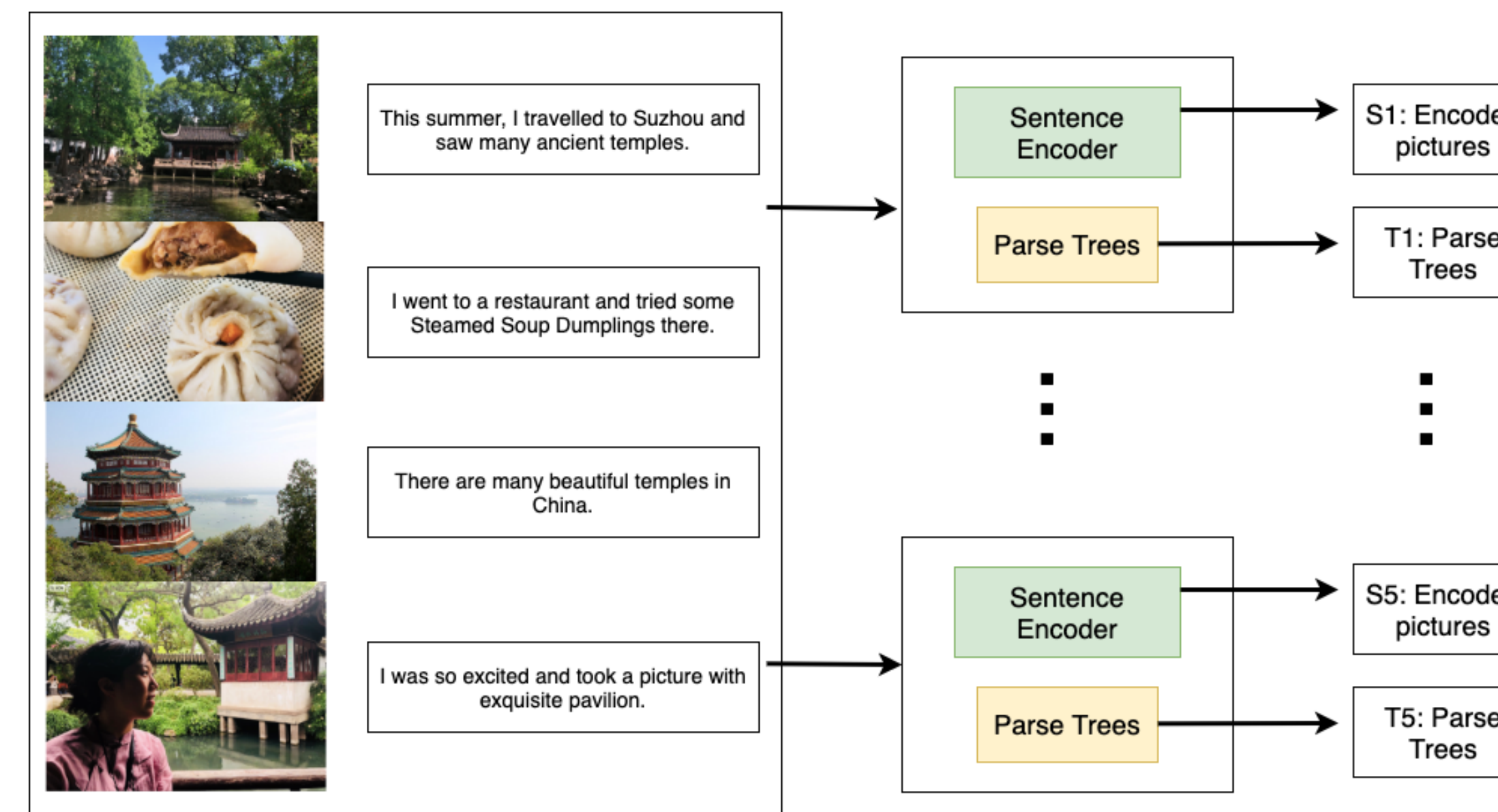
Results

In my website, users can type in their own stories (at most five sentences) to translate it to images. After they finish, they can press the green tab "Submit Text" and wait the server to "translate" their texts into images. The results will be evaluated based on the accuracy of the stories the webpage present. I will hold human evaluation tests on some sample stories, letting the respondents to select images in VIST datasets about the consequences of texts.

The display of the visual images which are transformed from the stories' users inserted. The images and sentences are one-to-one matched. There is also a blank space for users to enter their satisfaction level and suggestions.

Future Direction

With the rise of the Internet and the emergence of social media, more and more users employ visual images to record their life. The "translation" from texts to images can help people to record down the sequences from several pictures and combine them as a life story.

For further research, I will link the display images with users' local photos. For instance, when they text something like "my mother and I went to a party a few days ago", they will be informed that some photos are in their local photos. Users can choose to send the images rather than only sentences.

Society nowadays always considers machine learning as a negative or dangerous thing, but this research will show the understanding between humans and machines. The interactions between human beings and machines can even become friendly. Machines can not only take beautiful pictures for humans but also help them to write down the short but memorable stories. When people look back on the stories, they will have more warm visual memories than only texts.

Acknowledgements

References

1. Hareesh Ravi(2018). Show Me a Story: Towards Coherent Neural Story Illustration. In CVPR 2018.
2. Ryan, L. (2016). Visual storytelling with data. The Visual Imperative, 131–151.
3. Diakopoulos, N. (2018). Ethics in Data-Driven Visual Storytelling. Data-Driven Storytelling, 233–248. doi: 10.1201/9781315281575-10