Generating Sentences from Images

Qilei Zhang

Jun 30 2018

Abstract

Humans can prepare concise descriptions of pictures, focusing on what they find important.

1. Introduction

This paper describes a system that can calculate the image of a score connected to a sentence. This score can be used to attach descriptive sentences to a given image, or to get an image of a given sentence. For most pictures, humans can prepare a concise description in the form of a sentence relatively easily [1]. Such descriptions might identify the most interesting objects, what they are doing, and where this is happening. These descriptions are rich, because they are in sentence form. They are accurate, with good agreement between annotators [2].
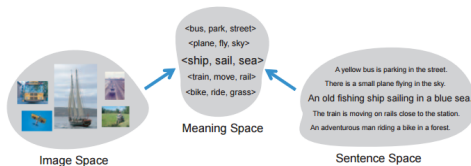

Figure 1. There is an intermediate space of meaning which has different projections to the space of images and sentences. Once we learn the projections we can generate sentences for images and find images best described by a given sentence.

2. Authors Suppressed Due to Excessive Length

Sentences can describe the same phenomena. Synecdoche and the general richness of vocabulary means that many different words can quite legitimately be used to describe the same picture [3].
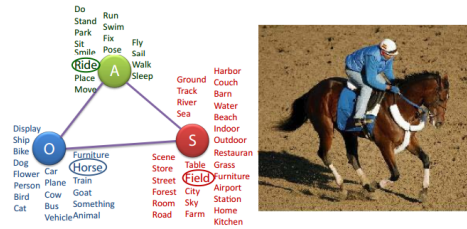

Figure 2. Multi-task Network Cascades for instance-aware semantic segmentation. At the top right corner is a simplified illustration.

3. Discussion and Future Work

Sentences are rich, compact and subtle information representation. Even so, it can predict good sentences of people's favorite images. The intermediate meaning representation is a key component of the model because it allows the benefit of distributed semantics. It is the right way to make deeper iteration in sentences and images [4].

References

[1] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi. Data mining for improving textbooks. *Acm Sigkdd Explorations Newsletter*, 13(2):7–19, 2012. 1

[2] K. Javorszky. Information processing in auto-regulated systems. *Entropy*, 5(2):161–192, 2003. 1

[3] P. Kakumanu, A. Esposito, O. N. Garcia, and R. Gutierrez-Osuna. A comparison of acoustic coding models for speech-driven facial animation. *Speech Communication*, 48(6):598–615, 2006. 1

[4] L. X. Zheng, Z. W. Xu, and H. M. Chen. Algorithm for generating short sentences from grammars based on branch coverage criterion. *Journal of Software*, 22(11):2564–2576, 2011. 1