



THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

Room 1 ,Ziskind Building
on Thursday, Dec 28, 2017
at 12:15

Greg Shakhnarovich
TTI-Chicago

Discriminability loss for learning to generate descriptive image captions

Abstract:

Image captioning -- automatic production of text describing a visual scene -- has received a lot of attention recently. However, the objective of captioning, evaluation metrics, and the training protocol remain somewhat unsettled. The general goal seems to be for machines to describe visual signal like humans do. We pursue this goal by incorporating a discriminability loss in training caption generators. This loss is explicitly "aware" of the need for the caption to convey information, rather than appear fluent or reflect word distribution in the human captions. Specifically, the loss in our work is tied to discriminative tasks: producing a referring expression (text that allows a recipient to identify a region in the given image) or producing a discriminative caption which allows the recipient to identify an image within a set of images. In both projects, use of the discriminability loss does not require any additional human annotations, and relies on collaborative training between the caption generator and a comprehension model, which is a proxy for a human recipient. In experiments on standard benchmarks, we show that adding discriminability objectives not only improves the discriminative quality of the generated image captions, but, perhaps surprisingly, also makes the captions better under a variety of traditional metrics.

Joint work with Ruotian Luo (TTIC), Brian Price and Scott Cohen (Adobe).