

Broadening the Coverage of Scenes and Descriptions towards Versatile Image Captioning Systems

氏 名 本多右京

研 究 室 名 自然言語処理学

主指導教員名（論文博士の場合は推薦教員名） 渡辺太郎

内容梗概（1 ページ目に収めること）

Image captioning plays a fundamental role in vision-and-language research, which lies in the intersection of computer vision and natural language processing, by converting the information in images into natural language descriptions. To be versatile pipelines for downstream vision-and-language tasks, captioning systems should be able to describe various types of scenes with extensive information. However, current captioning systems are limited in their coverage of scenes and descriptions: they can handle limited types of scenes, and their output descriptions tend to be overly generic.

The goal of this dissertation is to broaden the limited coverage. The first half of this dissertation is devoted to addressing the limitation of describable scenes by introducing unsupervised image captioning methods. In the second half of this dissertation, we address the limitation of possible descriptions by enhancing the discriminability of current captioning models. Finally, we discuss the remaining problems and future directions for image captioning research.