

Image Captioning with External Knowledge

In modern automatic image captioning, generating straightforward visual descriptions of images is largely a solved problem. One of the biggest challenges that still remains is incorporating information that cannot be inferred from the image alone: its context and related real world knowledge. In this dissertation, we tackle this challenge by developing a new method of enriching an otherwise standard captioning pipeline with contextually relevant image-external knowledge.

Our method starts with identifying the subset of data from external sources that is relevant to a given image. The retrieved data is integrated into the caption generation process, aiming to influence the resulting caption and extend it beyond a purely visual description. Based on this general method, we develop three neural image captioning models. The first model addresses a specific problem of generating references to the geographic context of the image. The second model expands to broad encyclopedic knowledge about the depicted geographic entities. Finally, the third model generalizes beyond the geographic domain and applies our method to diverse images from newspaper articles. The evaluation of the models shows that our method is indeed effective for producing contextualized and informative captions with factually accurate references to relevant external knowledge.

ISBN 978-94-6093-426-1

DOI <https://dx.medra.org/10.48273/LOT0641>



Netherlands Graduate School of Linguistics
Landelijke Onderzoekschool Taalwetenschap



Utrecht
University

Institute for
Language Sciences

LOT
641

Sofia Nikiforova

Image Captioning with External Knowledge

Sofia Nikiforova

Image Captioning with External Knowledge