

Title: Deep Learning and Visualization of Models for Image Captioning and Multimodal Translation

Author: Samuel Michalik

Institute: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Jindřich Helcl, Institute of Formal and Applied Linguistics

Abstract: In recent years, the machine learning paradigm known as deep learning has proven to be well suited for the exploitation of modern parallel hardware and large datasets, helping to advance the frontier of research in many fields of artificial intelligence and finding successful commercial applications. Deep learning allows end-to-end trainable systems to tackle difficult tasks by building complex hierarchical representations. However, these internal representations often avoid easy interpretation. We explore the possibilities of interpretable visualizations of attention components and beam search decoding at the task of image captioning and multimodal translation and build an application – Macaque, that can be run as an online service, to meet this end. Furthermore, we propose a novel attention function formulation, called scaled general attention. We experimentally evaluate scaled general attention along common attention functions on four different model architectures based on the encoder-decoder framework at the tasks of image captioning and multimodal machine translation. We utilise Macaque during qualitative analysis.

Keywords: deep learning visualisation image captioning multimodal translation