Contribution ID: **44** Type: **not specified**

Better Text Understanding Through Image-To-Text Transfer

Generic text embeddings are successfully used in a variety of tasks. However, they are often learnt by capturing the co-occurrence structure from pure text corpora, resulting in limitations of their ability to generalize. In this paper, we explore models that incorporate visual information into the text representation. Based on comprehensive ablation studies, we propose a conceptually simple, yet well performing architecture. It outperforms previous multimodal approaches on a set of well established benchmarks. We also improve the state-of-the-art results for image-related text datasets, using orders of magnitude less data.

Author: Ms KAROL, Kurach (Google Brian)

Presenter: Ms KAROL, Kurach (Google Brian)