

Recently, the state-of-the-art models for image captioning have overtaken human performance based on the most popular metrics, such as BLEU, METEOR, R