

Standard image captioning tasks such as COCO and Flickr30k are factual, neutral in tone and (to a human) state the obvious (e.g., "a man playing a guitar"). W