CSCI 1430: *Attention Please!* Deep Dive into Image Captioning Model Efficacy

by Kevin Kang (kkang11)

Abstract

This project investigates the performance of transformer-based models in image captioning tasks, critically examining the limitations of generalized models like VGG19, ResNet, and InceptionX. We compare the performance of un-tuned models with a fine-tuned model on a specific dataset, and create a baseline model using TensorFlow for further comparison. Our twin objectives are to evaluate the state-of-the-art in image captioning and assess the adaptability of these models to specialized datasets, a crucial aspect for applications like medical imaging. This research aims to improve the granularity and specificity of AI image captioning.

Poster