Image Captioning for Academic Figures and Graphs

Team Members: Bhaskar Bharat, Ekanki Agarwal, Prerna Mahajan, Viraj Kadam

Problem Statement

In academia we use figures and graphs to communicate rich,complex information. The captions of these figures are critical to convey effective messages. Through this project, we aim to build an end-to-end neural network model to automatically generate informative, high-quality captions for academic figures and graphs.

Scope for Modeling

1. To meet the objective, image features are extracted using pre-trained CNN architectures followed by using sequential modeling techniques like RNNs, transformers, etc.
2. Additionally, using attention techniques to visualize positions in the image corresponding to segments of words in the generated captions/description.
3. First we will be training on SCICAP dataset. SCICAP is a large-scale image captioning dataset that contains real-world scientific figures and captions. SCICAP was constructed using more than two million images from over 290,000 papers collected and released by arXiv.
4. Curation of captioned academic figures and graphs dataset based on Univ.AI course slides.
5. CNN + RNNs
6. CNN + Attention Models
7. Visualizing Attentions
8. Streamlit App

Datasets

1. SCICAP Dataset : SCICAP is a large-scale image captioning dataset that contains real-world scientific figures and captions. SCICAP was constructed using more than two million images from over 290,000 papers collected and released by arXiv.
2. Curation of captioned academic figures and graphs dataset based on Univ.AI course slides.

Potential Applications

1. Assistance for Visually Impaired
2. Image Search Applications
3. Visualizing Attentions can have applications in Medical Image Reading