Captions for Accessibility

Team name: Captioning

Members: Christian Martin, Awais Naeem

Project Goal: Model that can create image captions with longer, more descriptive captions.

Benefits: This should improve the value of the captions for accessibility use cases by providing more information about the image.

Dataset:

Smithsonian Open Access[1][2], filtered to images with a description of decent length (exact length TBD during analysis). The dataset includes newline-delimited JSON files that include captions and media metadata.

Proposed Methods:

The Cooper Hewitt Museum has been working on creating better guidelines for image description[3], and has run at least one event[4] to create better captions for its images. Based on the blog post from the event, there are >800 images with improved descriptions.

Download the JSON part of the dataset, search the metadata for captions of reasonable length, and use the caption+image pairs to fine-tune the Salesforce BLIP model.

Second, we'll attempt to train our own custom model (CNN/RNN) to compare with. We may also use the Flickr or LAION image+caption datasets for more training data and then fine-tune the model built from those (similar to the process for the existing BLIP model).

Milestones:

- Search the SI Open Access metadata
- Determine what a reasonable length of caption is
- Download the captions+image pairs meeting the criteria for being of a reasonable length
- Train-test split the pairs
- Fine-tune the BLIP model

- Create a custom model
- Use the testing pairs using text analysis evaluation metrics

Evaluation Criteria: use text analysis similarity metrics, e.g. BLEU/ROUGE/METEOR. We may need to test multiple metrics to see what works well for this use case.

Deliverables:

- Colab/Jupyter (ipynb) notebook for training and testing the model
- Pairs of captions and file names (image files won't be included due to size)
- Report
- Presentation

References:

[1] https://www.si.edu/openaccess

[2] https://registry.opendata.aws/smithsonian-open-access/

[3] https://www.cooperhewitt.org/cooper-hewitt-guidelines-for-image-description/

[4] https://www.cooperhewitt.org/2020/05/11/forming-meaning-through-our-most-personal-sense-copy/