

Bangla Captioning Image Taken by Blind People

Extended Abstract:

Blind people are important part of our society. Due to technological revelation, it is easy enough to use smart phone and capturing image by voice or semi/fully guided command. It will be helpful to caption the captured image and describe the image to the blind people. Our target is solving this problem with computer vision, machine learning and image processing.

The idea originates from Gurari et al ^[1] who are responsible for creating a dataset with 39,000 images originating from people who are blind that are each paired with five English captions. We are going to build a similar dataset for Bangla language. For creating the dataset, we are considering two ways. One is using crowd workers and another is using translator API. At first, we are going to use translator API to build the pipeline for predicting captions from the images captured by blind people. In parallel, our crowd workers will caption the image in Bangla language. After building the dataset, we are likely to build our pipeline with pretrained models for captioning the image.

Currently, we are building our dataset pipeline. We are hoping to complete and generating a result within November, 2020.

References:

1. *Captioning Images Taken by People Who Are Blind*, Danna Gurari, Yinan Zhao, Meng Zhang, Nilavra Bhattacharya, EUROPEAN CONFERENCE ON COMPUTER VISION, 2020