Convolutional Neural Networks Architecture Analysis for Image Captioning

Authors : Jun Seung Woo, Shin Dong Ho

Abstract : The Image Captioning models with Attention technology have developed significantly compared to previous models, but it is still unsatisfactory in recognizing images. We perform an extensive search over seven interesting Convolutional Neural Networks(CNN) architectures to analyze the behavior of different models for image captioning. We compared seven different CNN Architectures, according to batch size, using on public benchmarks: MS-COCO datasets. In our experimental results, DenseNet and InceptionV3 got about 14% loss and about 160sec training time per epoch. It was the most satisfactory result among the seven CNN architectures after training 50 epochs on GPU.

Keywords : deep learning, image captioning, CNN architectures, densenet, inceptionV3

Conference Title : ICCV 2022 : International Conference on Computer Vision

Conference Location : Vancouver, Canada

Conference Dates : September 22-23, 2022