

Performance Evaluation of CNN Frameworks and Image Transformer on Hindi Image Captioning

Aparna Paliwal

Department of Computer Science
Banasthali Vidyapith, Aliyabad, Rajasthan, India

Abstract: *In this paper, we have shown the development of a multimodal machine translation system that provides the textual description of an image. We have shown how Image captions in Hindi can be generated using an English Image captioning dataset. The Flickr dataset was used for training the model for image captioning and English-Hindi parallel corpus was used across domains for training the Neural Based Encoder-Decoder model. For our study, we developed a model based on sequence-to-sequence architecture. We trained two image captioning models based on VGG16 and RESNET50 architectures which generated English captions and were then translated into Hindi using seq2seq NMT model. The systems were evaluated using the standard BLEU MT evaluation metric. It was found the RESNET50 based model produced better score than VGG16 based model.*

Keywords: Multimodal Data, Image Captioning, Machine Translation, Encoder-Decoder Model