IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 4, February 2025

Comparing CNN Architectures and Image Transformer for Hindi Image Captioning

Aparna Paliwal
Department of Computer Science
Banasthali Vidyapith, Banasthali, Rajasthan, India

Abstract: *In this paper, we have shown the development of some image captioning 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 that Image Transformers performed better then RESNET50 and VGG16 based models.*

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