|  |  |  |  |
| --- | --- | --- | --- |
| Title | Comparison of Different CNN Model Used as Encoders for Image Captioning | | |
| Author(s) Name | Md Sahrial Alam, Md Sayedur Rahman, Md Ikbal Hosen, Khairul Anam Mubin, Sharif Hossen, MF Mridha | | |
| Contact Email(s) | firoz.mridha@aiub.edu | | |
| Published Journal Name | International Conference on Data Analytics for Business and Industry (ICDABI) | | |
| Type of Publication | Conference | | |
| Volume |  | Issue |  |
| Publisher | IEEE | | |
| Publication Date | 2021/10/25 | | |
| ISSN |  | | |
| DOI | 10.1109/ICDABI53623.2021.9655846. | | |
| URL | https://ieeexplore.ieee.org/abstract/document/9655846 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Image captioning is the process of automatically describing the content of an image which connects computer vision and natural language processing. In this paper, we compare five popular convolutional neural networks architecture. They are Vgg16, InceptionV3, Resnet50, Densenet201 and Xception Model. By using these preprocessing model for same image captioning model. Encoder-Decoder model is a part of recurrent neural networks for sequence-to-sequence prediction problems. Encoders are usually used pre-trained convolutional neural networks for large datasets. There are many different types of Encoder-Decoder architecture used for generating caption. But it is very complicated to evaluate the performance of the architecture. In this paper, we used categorical-crossentropy for loss function, RMSprop for optimizer in Vgg16, Resnet50, InceptionV3, Densenet201 and Xception model. | |