Image Caption Generator using Deep Learning

Farida Attar[1], Farzana Khan[2], Affan Ansari[3], Mujawar Saklen[4], Abubakr Shaikh[5], Danish Khan[6]

Assistant Professor, Department of Information Technology[1,2]
Students, Department of Information Technology[3,4,5,6]
M.H. Saboo Siddik College of Engineering, Byculla, Mumbai, India
farida.attar@mhssce.ac.in, farzana.khan@mhssce.ac.in, mohammedaffan.612008.it@mhssce.ac.in,
saklen.612037.it@mhssce.ac.in, abubakr.612047.it@mhssce.ac.in, danish.6119012.it@mhssce.ac.in

Abstract: *Image Caption Generation has always been a study of great interest to the researchers in the Artificial Intelligence department. Being able to program a machine to accurately describe an image or an environment like an average human has major applications in the field of robotic vision, business and many more. Automatic caption generation with attention mechanisms aims at generating more descriptive captions containing coarse to fine semantic contents in the image. This has been a challenging task in the field of artificial intelligence. In this paper, we present different image caption generating models based on deep neural networks, focusing on the various CNN techniques and analyzing their influence on the sentence generation. We have also generated captions for sample images and compared the different feature extraction and encoder models to analyse which model gives better accuracy and generates the desired results*

Keywords: CNN, RNN, LSTM , VGG, GRU, Encoder - Decoder, Image Captioning

REFERENCES

[1] CS771 Project Image Captioning by Ankit Gupta , Kartik Hira, Bajaj Dilip.

[2] "Every Picture Tells a Story: Generating Sentences from Images." Computer Vision ECCV (2016) by Farhadi, Ali, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hocken-maier, and David Forsyth

[3] Automatic Caption Generation for News Images by Yansong Feng, and Mirella Lapata, IEEE (2013).

[4] Image Caption Generator Based on Deep Neural Networks by Jianhui Chen, Wenqiang Dong and Minchen Li, ACM (2014)

[5] Show and Tell: A Neural Image Caption Generator by Oriol Vinyl, Alexander Toshev, Samy Bengio, Dumitru Erhan, IEEE (2015).

[6] Image2Text: A Multimodal Caption Generator by Chang Liu, Changhu Wang, Fuchun Sun, Yong Rui, ACM (2016).

[7] The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions by Sepp Hochreiter.

[8] Where to put the Image in an Image Caption Generator by Marc Tanti, Albert Gatt, Kenneth P. Camilleri.

[9] Sequence to sequence -video to text by Subhashini Venugopalan, Marcus Rohrbach, Jeffrey Donahue, Raymond Mooney, Trevor Darrell, and Kate Saenko.

[10] Learning phrase representations using RNN encoder-decoder for statistical machine translation by K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares,H. Schwenk, and Y. Bengi.

[11] TVPRNN for image caption generation .Liang Yang and Haifeng Hu.

[12] Image Captioning in the Wild: How People Caption Images on Flickr Philipp Blandfort, Tushar Karayil, Damian Borth, Andreas Dengel,German Institute for Artificial Intelligence, Kaiserslautern, Germany.

[13] Image Caption Generator Based On Deep Neural Networks Jianhui Chen ,Wenqiang Dong, Minchen Li ,CS Department. ACM 2014.

[14] BLEU: A method for automatic evaluation of machine translation. I