

Audio and Caption Generation using ML

R P Rajeshwari¹, Thammineni Pushyamitra², Madhuri G³, B Pavani Kumari⁴.

Assistant Professor, Department of Computer Science and Engineering¹

Students, Department of Computer Science and Engineering^{2,3,4}

Rao Bahadur Y. Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract: *Using deep learning and computer vision, image caption generation identifies the context of an image and comments on it with the most relevant captions. It entails an image's epithet combined with English keywords, as well as datasets provided during model training. We produce a vectorized representation of an image using a deep Convolutional neural network, which is then fed into a long-short-term memory (LSTM) network, which provides captions. LSTM (Long short-term memory) is also one of the RNN (Recurrent Neural Network) constructs, mainly utilized in the deep learning discipline. The vivid applications of image captioning are natural language processing applications, recommendations in editing applications, helpful for virtual assistants, for indexing the image, for the blind or partially-sighted people, for the internet community, etc.*

Keywords: Deep Learning and Computer Vision

REFERENCES

- [1]. Liu, Shuang & Bai, Liang & Hu, Yanli & Wang, Haoran. (2018). Image Captioning Based on Deep Neural Networks. MATEC Web of Conferences. 232. 01052. 10.1051/mateconf/201823201052
- [2]. A. Hani, N. Tagougui and M. Kherallah, "Image Caption Generation Using A Deep Architecture," 2019 International Arab Conference on Information Technology (ACIT), 2019, pp. 246-251, doi: 10.1109/ACIT47987.2019.8990998.
- [3]. S. Das, L. Jain and A. Das, "Deep Learning for Military Image Captioning," 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 2165-2171, doi: 10.23919/ICIF.2018.8455321.
- [4]. Learning CNN- LSTM Architectures For Image Caption Generation, Moses Soh, 2016.
- [5]. G Geetha, T. Kirthigadev, G GODWIN Ponsam, T. Karthik, M. Safa, "Image Captioning Using Deep Convolutional Neural Networks (CNNs)" Published under license by IOP Publishing Ltd in Journal of Physics :Conference Series ,Volume 1712, International Conference On Computational Physics in Emerging Technologies (ICCPET) 2020 August 2020, Mangalore India in 2015.
- [6]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [7]. Donahue, Jeffrey, et al. "Long-term recurrent convolutional networks for visual recognition and description." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [8]. Lu, Jiasen, et al. "Knowing when to look: Adaptive attention via a visual sentinel for image captioning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Vol. 6. 2017.
- [9]. Ordonez, Vicente, Girish Kulkarni, and Tamara L. Berg. "Im2text: Describing images using 1 million captioned photographs." Advances in neural information processing systems. 2011.
- [10]. Chen, Xinlei, and C. Lawrence Zitnick. "Mind's eye: A recurrent visual representation for image caption generation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.S

