

Image Caption Generator with Speech Using CNN and LSTM

Akshara Madhusoodanan¹ and Shyma Kareem²

Student, Department of Computer Applications¹

Assistant Professor, Department of Computer Applications²

Musaliar College of Engineering & Technology, Pathanamthitta, Kerala

***Abstract:** Image processing remains one of the most cutting-edge technologies employed by Google, the medical industry, and other sectors of the economy. Due to its free and open source tool, which every developer can afford, this technology has recently drawn many programmers and developers. Since it is currently used as a primary technique of gathering information from images, processing those images for various purposes, and performing various operations on those images, image processing also aids in learning a lot of information from a single image. The problem of creating voice-based image captions uses the idea of NLP (natural language processing) to comprehend the description of an image. The optimal method for this project is the combination of CNN and LSTM; the primary goal of the suggested research is to find the ideal caption for an image. The description will be translated into text after being obtained, and the text will then be given voice. For persons who are blind and cannot understand visuals, image descriptions can be obtained as a speech output using TTS engine.*

Keywords: CNN, LSTM, Deep learning, TTS engine

REFERENCES

- [1] Sumathi, T., and Hemalatha, M. presented "A combined hierarchical model for automatic image annotation and retrieval" at the 2011 International Conference on Advanced Computing (ICAC).
- [2] "Sentence Learning Deep Convolutional Neural Network for Image Caption Generation" was presented at the 13th International Conference on Ubiquitous Robots and Ambient Intelligence by Dong-Jin Kim, Donggeun Yoo, Bonggeun Sim, and In So Kweon.
- [3] Varsha Kesavan, Vaidehi Muley, and Megha Kolhekar presented "Deep Learning based Image Caption Generation" at the 2019 Global Conference for Advancement in Technology (GCAT).
- [4] "Visual Image Caption Generation for Industrial and Service Robotics Applications" by Yu-Ting Hsu, Yu-Cheng Wen, Huan-Jun Ye, and Ren C. Luo was published in IEEE-2019.
- [5] M.T. Yu and M.M. Sein presented their "Automatic image captioning system using integration of N cut and color-based segmentation method"