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

Survey on Multimodal Image Captioning Approaches: Addressing Contextual Understanding, Cross-Dataset Generalization, and Multilingual Captioning

Mr. Nikhil Gopal Khodave[1] and Mr. Prathamesh S. Powar[2]
Student, Computer Science and Engineering [1]
Asst. Professor, Computer Science and Engineering[2]
Ashokrao Mane Group of Institution, Vathar, Kolhapur, India

Abstract: *This paper discusses new advances in multimodal image captioning, and the focus lies on improving access, contextual understanding, and generalization across many datasets. Current state-of-the-art uni-modal approaches are no longer good enough, whereas innovative multimodal techniques combine the visual and textual features to yield more accurate captions and richness of the captions produced. This includes the attention mechanism, scene graphs, and pre-trained transformer models through which more contextual descriptions are made. Further, the paper addresses challenges in cross-dataset generalization and multilingual captioning, pointing to the necessity of systems that adapt to real-world variability and support diversity in linguistic backgrounds. Through synthesizing the research conducted so far, this work outlines future directions for the creation of more inclusive, robust, and effective image captioning technologies, especially for applications in accessibility*

Keywords: Multimodal Image Captioning, Accessibility, Contextual Understanding, Cross-Dataset Generalization

DOI: 10.48175/IJARSCT-25104