IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

Visual Question Answering

Dr. Sai Madhavi D[1], Durga Shreya M[2], Manasa A[3], Pooja U Joshi[4],
Professor and HOD, Department of Computer Science and Engineering (AI & ML)[1]
Students, Department of Computer Science and Engineering[2,3,4]
Rao Bahadur Y Mahabaleshwarappa Engineering College, Ballari, Karnataka, India

Abstract: *We propose the task of free-form and open- ended Visual Question Answering (VQA). Given an image and a natural language question about the image, the task is to provide an accurate natural language answer. Mirroring real-world scenarios, such as helping the visually impaired, both the questions and answers are open-ended. Visual questions selectively target different areas of an image, including background details and underlying context. As a result, a system that succeeds at VQA typically needs a more detailed understanding of the image and complex reasoning than a system producing generic image captions. Moreover, VQA is amenable to automatic evaluation, since many open- ended answers contain only a few words or a closed set of answers that can be provided in a multiple choice format. We provide a dataset containing ~0.25M images, ~0.76M questions, and ~10M answers and discuss the information it provides. Numerous baseline for VQA are provided and compared with human performance. In this model, we have exclusively introduced a feature of voice to text using Speech recognition, Google Text-to-Speech and pygame module.*

Keywords: Visual Question Answering

REFERENCES

[1]. S. Antol, A. Agarwal, J. Lu, M. M. Mitchell, D. Batra, C. Lawrence Zitnick, and D.Parikh, "Vqa: Visual question answering," in Proceedings of the IEEE international conference on computer vision, 2015,pp. 2425–2433.

[2]. A. Agrawal, J. Lu, S. Antol, M. Mitchell, C. L. Zitnick, D. Parikh, and D. Batra, "Vqa: Visualquestion answering," International Journal ofComputer Vision, vol. 123,no. 1, pp. 4–31, May 2017.[Online].Available:https://doi.org/10.1007/s1 1263-016-0966-6.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

[3]. J. Lu, J. Yang, D. Batra, and D. Parikh, "Hierarchical question-image co-attention for visual question answering," in Advances In Neural Information Processing Systems, 2016, pp.289–297.

[4]. H. Xu and K. Saenko, "Ask, attend and answer: Exploring question guideds patial attention for visual question answering," in European Conference on ComputerVision.Springer,2016, pp.451–466.

[5]. K. Kafle and C. Kanan, "Analysis of visual question answering algorithms," in Proceedings of the IEEE International Conference on Computer Vision,2017,pp.1965– 1973.

[6]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition,2015,pp. 1–9