

Name/ID : ABDELKAREEM YOUSEF MAMDOH SOUBAR/19110022

Subject :IT Seminar

Assignment Title:VQA Paper

Submission Date :17/4/2023 on 12:00PM

VQA is a challenging task that requires the use of computer vision, natural language processing, and artificial intelligence., and artificial intelligence. The goal of VQA is to create computer programs that can answer questions about visual content like pictures and videos using both visual and textual information.

To do this, VQA algorithms first analyze the image or video using a deep neural network to identify objects and scenes. Afterwards, the question is subjected to natural language processing in order to generate a version of the question that can be comprehended by the computer.

The visual and textual representations are combined using a fusion model to generate a joint representation that captures the relevant information needed to answer the question. Finally, the computer program uses this joint representation to generate an answer to the question.

VQA has many practical applications, such as image and video captioning, autonomous driving, and assistive technologies for people with visual impairments. However, it also poses challenges such as handling ambiguity, variability, and complexity in visual and linguistic data, and addressing issues related to bias and fairness.

VQA algorithms use computer vision and natural language processing techniques to answer questions about visual content. The computer program takes information from the picture and question to come up with an answer, using different methods like categorization, guessing, or pattern recognition. Scientists continue to develop new ways to make VQA algorithms better and more accurate despite these challenges.

Here are a few more examples of VQA questions and answers:

Q: What is the person in the picture doing?

A: They are playing soccer.

Q: How many dogs are in the picture?

A: There are two dogs in the picture.

Q: What is the name of the landmark in the picture?

A: The landmark is the Eiffel Tower.

References:

* Yang, Z., He, X., Gao, J., Deng, L., & Smola, A. (2016). Stacked attention networks for image question answering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 21-29).
* Lu, J., Yang, J., Batra, D., & Parikh, D. (2016). Hierarchical question-image co-attention for visual question answering. In Advances in neural information processing systems (pp. 289-297).
* Kazemi, V., & Elqursh, A. (2017). Show, ask, attend, and answer: A strong baseline for visual question answering. arXiv preprint arXiv:1704.03162.
* Teney, D., Anderson, P., & He, X. (2018). Tips and tricks for visual question answering: Learnings from the 2017 challenge. In Proceedings of the European Conference on Computer Vision (pp. 413-429).
* Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6077-6086).