ستاره باباجانی – 99521109 – خلاصه این مقاله

In “The Visual Question Answering: A Survey of Methods and Datasets”, the authors provide a comprehensive overview of the field of Visual Question Answering (VQA) by examining the state of the art and reviewing available datasets for training and evaluating VQA systems. The paper compares modern approaches to VQA, classifying methods by their mechanism to connect the visual and textual modalities. It discusses the common approach of combining convolutional and recurrent neural networks to map images and questions to a common feature space, as well as memory-augmented and modular architectures that interface with structured knowledge bases. Promising future directions for the field and the use of natural language processing models were also explored. The paper also discusses a variety of datasets for training and evaluating VQA systems, including DAQUAR, COCO-QA, VQA-real, Visual Genome, FM-IQA, and more. Additionally, the article includes discussions of specific datasets designed to evaluate the performance of VQA systems that make use of external knowledge bases, such as the KB-VQA and FVQA datasets.

The authors provide a detailed analysis of the evaluation measures including metrics such as accuracy with respect to ground truth, the Wu-Palmer similarity, and the Visual Turing Test using human judges. Furthermore, the paper reports the results of existing methods on major datasets for VQA, providing insights into the performance and limitations of these methods. The paper also highlights the challenges in VQA and the varying protocols for collecting data, impacting the complexity, and external knowledge required.

Overall, the article provides a comprehensive and thorough overview of the state of the art in Visual Question Answering, including the current methods, datasets, and future directions for the field.