Department of Computer Engineering & Information Technology

Amirkabir University of Technology

# Research & Technical Presentation in Engineering

## Visual Question Answering

## Assignment 6

References Engineering

**Provided by**

Ali Gholami

**Student Number**

9531504

**Advisor**

Dr. Reza Safabakhsh

[1] Y. Li, N. Duan, B. Zhou, X. Chu, W. Ouyang, and X. Wang, “Visual Question Generation as Dual Task of Visual Question Answering.”

[2] A. V. Savchenko, “Maximum-likelihood approximate nearest neighbor method in real-time image recognition,” *Pattern Recognit.*, vol. 61, pp. 459–469, 2017.

[3] Y. Zhu, J. J. Lim, and L. Fei-Fei, “Knowledge Acquisition for Visual Question Answering via Iterative Querying,” *2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 6146–6155, 2017.

[4] D. Yu, J. Fu, T. Mei, and Y. Rui, “Multi-level Attention Networks for Visual Question Answering,” *2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4187–4195, 2017.

[5] A. Mahendru, V. Prabhu, A. Mohapatra, D. Batra, and S. Lee, “The Promise of Premise: Harnessing Question Premises in Visual Question Answering,” 2017.

[6] A. Das, H. Agrawal, L. Zitnick, D. Parikh, and D. Batra, “Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions?,” *Comput. Vis. Image Underst.*, 2017.

[7] J.-H. Huang, M. Alfadly, and B. Ghanem, “VQABQ: Visual Question Answering by Basic Questions,” 2017.

[8] Z. Yu, J. Yu, J. Fan, and D. Tao, “Multi-modal Factorized Bilinear Pooling with Co-attention Learning for Visual Question Answering,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, pp. 1839–1848, 2017.

[9] Y. Lin, Z. Pang, D. Wang, and Y. Zhuang, “Task-driven Visual Saliency and Attention-based Visual Question Answering,” 2017.

[10] P. Anderson *et al.*, “Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering,” 2017.

[11] Y. Jang, Y. Song, Y. Yu, Y. Kim, and G. Kim, “TGIF-QA: Toward Spatio-Temporal Reasoning in Visual Question Answering,” pp. 2758–2766, 2017.

[12] I. Calixto, Q. Liu, and N. Campbell, “Incorporating Global Visual Features into Attention-Based Neural Machine Translation,” 2017.

[13] D. Teney, P. Anderson, X. He, and A. van den Hengel, “Tips and Tricks for Visual Question Answering: Learnings from the 2017 Challenge,” 2017.

[14] A. Santoro *et al.*, “A simple neural network module for relational reasoning,” no. Nips, 2017.

[15] V. Kazemi and A. Elqursh, “Show, Ask, Attend, and Answer: A Strong Baseline For Visual Question Answering,” 2017.

[16] A. Das, S. Kottur, J. M. F. Moura, S. Lee, and D. Batra, “Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, pp. 2970–2979, 2017.

[17] K. Saito, A. Shin, Y. Ushiku, and T. Harada, “DualNet: Domain-invariant network for visual question answering,” *Proc. - IEEE Int. Conf. Multimed. Expo*, pp. 829–834, 2017.

[18] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[19] I. Schwartz, A. G. Schwing, and T. Hazan, “High-Order Attention Models for Visual Question Answering,” no. Nips, 2017.

[20] Q. Wu, D. Teney, P. Wang, C. Shen, A. Dick, and A. van den Hengel, “Visual question answering: A survey of methods and datasets,” *Comput. Vis. Image Underst.*, vol. 163, pp. 21–40, 2017.

[21] B. A. Plummer, A. Mallya, C. M. Cervantes, J. Hockenmaier, and S. Lazebnik, “Phrase Localization and Visual Relationship Detection with Comprehensive Image-Language Cues,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, pp. 1946–1955, 2017.

[22] K. Kafle and C. Kanan, “Visual question answering: Datasets, algorithms, and future challenges,” *Comput. Vis. Image Underst.*, vol. 163, pp. 3–20, 2017.

[23] A. Agrawal *et al.*, “VQA: Visual Question Answering: www.visualqa.org,” *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 4–31, 2017.

[24] H. Zhang, Z. Kyaw, S.-F. Chang, and T.-S. Chua, “Visual Translation Embedding Network for Visual Relation Detection,” pp. 5532–5540, 2017.

[25] P. Wang, Q. Wu, C. Shen, A. Dick, and A. van den Hengel, “FVQA: Fact-based Visual Question Answering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–16, 2017.

[26] R. Hu, J. Andreas, M. Rohrbach, T. Darrell, and K. Saenko, “Learning to Reason: End-to-End Module Networks for Visual Question Answering,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, no. Figure 1, pp. 804–813, 2017.

[27] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, and D. Parikh, “Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering,” pp. 6904–6913, 2016.

[28] H. Noh and B. Han, “Training Recurrent Answering Units with Joint Loss Minimization for VQA,” 2016.

[29] J.-H. Kim, K.-W. On, W. Lim, J. Kim, J.-W. Ha, and B.-T. Zhang, “Hadamard Product for Low-rank Bilinear Pooling,” pp. 1–14, 2016.

[30] D. Teney, L. Liu, and A. van den Hengel, “Graph-Structured Representations for Visual Question Answering,” pp. 1–9, 2016.

[31] H. Noh, P. H. Seo, and B. Han, “Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction,” *Cvpr*, pp. 30–38, 2016.

[32] C. Xiong, V. Zhong, and R. Socher, “Dynamic Coattention Networks For Question Answering,” pp. 1–14, 2016.

[33] M. Bolaños, Á. Peris, F. Casacuberta, and P. Radeva, “VIBIKNet: Visual Bidirectional Kernelized Network for Visual Question Answering,” 2016.

[34] T. Nguyen *et al.*, “MS MARCO: A human generated MAchine reading COmprehension dataset,” in *CEUR Workshop Proceedings*, 2016.

[35] D. Teney and A. van den Hengel, “Zero-Shot Visual Question Answering,” 2016.

[36] I. V. Serban *et al.*, “Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus,” pp. 2–3, 2016.

[37] R. Li and J. Jia, “Visual Question Answering with Question Representation Update (QRU),” *Nips*,

no. Nips, 2016.

[38] A. Fukui, D. H. Park, D. Yang, A. Rohrbach, T. Darrell, and M. Rohrbach, “Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding,” 2016.

[39] A. Agrawal, D. Batra, and D. Parikh, “Analyzing the Behavior of Visual Question Answering Models,” 2016.

[40] L. Ma, Z. Lu, and H. Li, “Learning to Answer Questions from Image Using Convolutional Neural Network,” *Aaai*, p. 7, 2016.

[41] G. Marcus, F. Rossi, and M. Veloso, “Beyond the Turing Test,” *AI Mag.*, pp. 3–4, 2016.

[42] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, “Bidirectional Attention Flow for Machine Comprehension,” pp. 1–13, 2016.

[43] A. Ray, G. Christie, M. Bansal, D. Batra, and D. Parikh, “Question Relevance in VQA: Identifying Non-Visual And False-Premise Questions,” 2016.

[44] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, 2016.

[45] A. Jabri, A. Joulin, and L. van Der Maaten, “Revisiting visual question answering baselines,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[46] L. Chen *et al.*, “SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning,” 2016.

[47] M. Malinowski, M. Rohrbach, and M. Fritz, “Ask Your Neurons: A Deep Learning Approach to Visual Question Answering,” 2016.

[48] H. Xu and K. Saenko, “Ask, attend and answer: Exploring question-guided spatial attention for visual question answering,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9911 LNCS, pp. 451–466, 2016.

[49] J. Lu, J. Yang, D. Batra, and D. Parikh, “Hierarchical Question-Image Co-Attention for Visual Question Answering,” no. c, pp. 1–9, 2016.

[50] A. Mallya and S. Lazebnik, “Learning models for actions and person-object interactions with transfer to question answering,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 414–428, 2016.

[51] K. J. Shih, S. Singh, and D. Hoiem, “Where To Look: Focus Regions for Visual Question Answering,” pp. 4613–4621, 2015.

[52] Q. Wu, P. Wang, C. Shen, A. Dick, and A. van den Hengel, “Ask Me Anything: Free-form Visual Question Answering Based on Knowledge from External Sources,” pp. 4622–4630, 2015.

[53] A. Jiang, F. Wang, F. Porikli, and Y. Li, “Compositional Memory for Visual Question Answering,” 2015.

[54] Y. Yang, Y. Li, C. Fermuller, and Y. Aloimonos, “Neural Self Talk: Image Understanding via Continuous Questioning and Answering,” 2015.

[55] K. Chen, J. Wang, L.-C. Chen, H. Gao, W. Xu, and R. Nevatia, “ABC-CNN: An Attention Based Convolutional Neural Network for Visual Question Answering,” 2015.

[56] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei, “Visual7W: Grounded Question Answering in Images,” 2015.

[57] J. Mao, J. Huang, A. Toshev, O. Camburu, A. Yuille, and K. Murphy, “Generation and Comprehension of Unambiguous Object Descriptions,” 2015.

[58] Z. Yang, X. He, J. Gao, L. Deng, and A. Smola, “Stacked Attention Networks for Image Question Answering,” *arXiv*, no. 1, pp. 21–29, 2015.

[59] S. Antol *et al.*, “VQA : Visual Question Answering,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2425–2433, 2015.