[18] Yikang Li, Wanli Ouyang, Xiaogang Wang, et al. 2017. Vip-cnn: Visual phrase guided convolutional neural network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '17)*. 7244–7253.

[19] Xiaodan Liang, Lisa Lee, and Eric P Xing. 2017. Deep variation-structured reinforcement learning for visual relationship and attribute detection. In *IEEE Conference on Computer Vision and Pattern Recognitio (CVPR '17)*. IEEE, 4408–4417.

[20] Guang-Hai Liu, Jing-Yu Yang, and ZuoYong Li. 2015. Content-based image retrieval using computational visual attention model. *Pattern Recognition* 48, 8 (2015), 2554–2566.

[21] Yunfei Long, Lu Qin, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. A Cognition Based Attention Model for Sentiment Analysis. In *Conference on Empirical Methods in Natural Language Processing (EMNLP '17)*. 462–471.

[22] Cewu Lu, Ranjay Krishna, Michael Bernstein, and Li Fei-Fei. 2016. Visual Relationship Detection with Language Priors. In *European Conference on Computer Vision (ECCV '16)*.

[23] Pan Lu, Hongsheng Li, Wei Zhang, Jianyong Wang, and Xiaogang Wang. 2018. Co-attending Free-form Regions and Detections with Multi-modal Multiplicative Feature Embedding for Visual Question Answering.. In *The AAAI Conference on Artificial Intelligence (AAAI'18)*. 7218–7225.

[24] Lin Ma, Zhengdong Lu, and Hang Li. 2016. Learning to Answer Questions from Image Using Convolutional Neural Network.. In *The AAAI Conference on Artificial Intelligence (AAAI '16)*.

[25] Hyeonwoo Noh, Paul Hongsuck Seo, and Bohyung Han. 2016. Image question answering using convolutional neural network with dynamic parameter prediction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*.

[26] Mengye Ren, Ryan Kiros, and Richard Zemel. 2015. Exploring models and data for image question answering. In *Advances In Neural Information Processing Systems (NIPS '16)*.

[27] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[28] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems (NIPS '13)*. 926–934.

[29] Xuejian Wang, Lantao Yu, Kan Ren, Guanyu Tao, Weinan Zhang, Yong Yu, and Jun Wang. 2017. Dynamic attention deep model for article recommendation by learning human editors' demonstration. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD '17)*. ACM, 2051–2059.

[30] Qi Wu, Chunhua Shen, Lingqiao Liu, Anthony Dick, and Anton van den Hengel. 2016. What value do explicit high level concepts have in vision to language problems?. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*. 203–212.

[31] Qi Wu, Chunhua Shen, Peng Wang, Anthony Dick, and Anton van den Hengel. 2017. Image Captioning and Visual Question Answering Based on Attributes and External Knowledge. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).

[32] Qi Wu, Peng Wang, Chunhua Shen, Anthony Dick, and Anton van den Hengel. 2016. Ask me anything: Free-form visual question answering based on knowledge from external sources. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*. 4622–4630.

[33] Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, and Zheng Zhang. 2015. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*. 842–850.

[34] Caiming Xiong, Stephen Merity, and Richard Socher. 2016. Dynamic memory networks for visual and textual question answering. In *International Conference on Machine Learning (ICML '16)*.

[35] Huijuan Xu and Kate Saenko. 2016. Ask, attend and answer: Exploring question-guided spatial attention for visual question answering. In *European Conference on Computer Vision (ECCV '16)*. 451–466.

[36] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning (ICML '15)*. 2048–2057.

[37] Zhao Yan, Nan Duan, Junwei Bao, Peng Chen, Ming Zhou, Zhoujun Li, and Jianshe Zhou. 2016. Docchat: An information retrieval approach for chatbot engines using unstructured documents. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL '16)*, Vol. 1. 516–525.

[38] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola. 2016. Stacked attention networks for image question answering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*. 21–29.

[39] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. 2016. Hierarchical Attention Networks for Document Classification. In *The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL '16)*. 1480–1489.

[40] Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. 2016. Image captioning with semantic attention. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*. 4651–4659.

[41] Dongfei Yu, Jianlong Fu, Tao Mei, and Yong Rui. 2017. Multi-level attention networks for visual question answering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '17)*. 4709–4717.

[42] Shuangfei Zhai, Keng-hao Chang, Ruofei Zhang, and Zhongfei Mark Zhang. 2016. Deepintent: Learning attentions for online advertising with recurrent neural networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD '16)*. ACM, 1295–1304.

[43] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. 2018. User-guided Hierarchical Attention Network for Multi-modal Social Image Popularity Prediction. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. 1277–1286.