**Bibliography:**

[1] J. Gu, G. Wang, J. Cai, and T. Chen, “An Empirical Study of Language CNN for Image Captioning,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-October, pp. 1231–1240, 2017, doi: 10.1109/ICCV.2017.138.

[2] J. Aneja, A. Deshpande, and A. G. Schwing, “Convolutional Image Captioning,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 5561–5570, 2018, doi: 10.1109/CVPR.2018.00583.

[3] K. Xu et al., “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.” Available: <http://proceedings.mlr.press/v37/xuc15>.

[4] K. Xu, H. Wang, and P. Tang, “Image Captioning With Deep Lstm Based On Sequential Residual Department of Computer Science and Technology , Tongji University , Shanghai , P . R . China Key Laboratory of Embedded System and Service Computing , Ministry of Education ,” no. July, pp. 361–366, 2017.

[5] S. Liu, L. Bai, Y. Hu, and H. Wang, “Image Captioning Based on Deep Neural Networks,” MATEC Web Conf., vol. 232, pp. 1–7, 2018, doi: 10.1051/matecconf/201823201052.

[6] R. Subash, R. Jebakumar, Y. Kamdar, and N. Bhatt, “Automatic image captioning using convolution neural networks and LSTM,” J. Phys. Conf. Ser., vol. 1362, no. 1, 2019, doi: 10.1088/1742- 6596/1362/1/012096.

[7] C. Wang, H. Yang, and C. Meinel, “Image Captioning with Deep Bidirectional LSTMs and Multi-Task Learning,” ACM Trans. Multimed. Comput. Commun. Appl., vol. 14, no. 2s, 2018, doi: 10.1145/3115432.

[8] M. Han, W. Chen, and A. D. Moges, “Fast image captioning using LSTM,” Cluster Comput., vol. 22, pp. 6143–6155, May 2019, doi: 10.1007/s10586-018-1885-9.

[9] H. Dong, J. Zhang, D. Mcilwraith, and Y. Guo, “I2T2I: Learning Text To Image Synthesis With Textual Data Augmentation.”

[10] Y. Xian and Y. Tian, “Self-Guiding Multimodal LSTM - When We Do Not Have a Perfect Training Dataset for Image Captioning,” IEEE Trans. Image Process., vol. 28, no. 11, pp. 5241–5252, 2019, doi: 10.1109/TIP.2019.2917229.