Bibliography

[1] Khan, R., Islam, M. S., Kanwal, K., Iqbal, M., Hossain, Md. I., & Ye, Z. (2022, March 3). A Deep Neural Framework for Image Caption Generation Using GRU-Based Attention Mechanism. arXiv.Org. https://arxiv.org/abs/2203.01594/

[2] Herdade, S., Kappeler, A., Boakye, K., & Soares, J. (2019, June 14). Image Captioning: Transforming Objects into Words. arXiv.Org. https://arxiv.org/abs/1906.05963

[3] Olimov, F., Dubey, S., Shrestha, L., Tin, T. T., & Jeon, M. (2021, February 14). Image Captioning using Multiple Transformers for Self-Attention Mechanism. arXiv.Org. https://arxiv.org/abs/2103.05103

[4] Wang, C., Shen, Y., & Ji, L. (2021, October 1). Geometry Attention Transformer with Position-aware LSTMs for Image Captioning. arXiv.Org. https://arxiv.org/abs/2110.00335

[5] Wang, D., Liu, B., Zhou, Y., Liu, M., Liu, P., & Yao, R. (2022). Separate Syntax and Semantics: Part-of-Speech-Guided Transformer for Image Captioning. Applied Sciences, 12(23). https://doi.org/10.3390/app122311875

[6] Osolo, R. I., Yang, Z., & Long, J. (2021). An Analysis of the Use of Feed-Forward Sub-Modules for Transformer-Based Image Captioning Tasks. Applied Sciences, 11(24). https://doi.org/10.3390/app112411635

[7] Jasir, M.P. and Balakrishnan, K., 2022. Text-to-Speech Synthesis: Literature Review with an Emphasis on Malayalam Language. ACM TRANSACTIONS ON ASIAN AND LOW-RESOURCE LANGUAGE INFORMATION PROCESSING, 21(4).

[8] Yang, S. et al. (2017) 'Statistical parametric speech synthesis using generative adversarial networks under a multi-task learning framework', in 2017 IEEE automatic speech recognition and understanding workshop (ASRU). IEEE, pp. 685–691.

[9] Ning, Y., He, S., Wu, Z., Xing, C. and Zhang, L.J., 2019. A review of deep learning based speech synthesis. Applied Sciences, 9(19), p.4050.

[10] Kim, M. et al. (2021) 'Expressive text-to-speech using style tag', arXiv preprint arXiv:2104. 00436.

[11] Jain, R., Yiwere, M.Y., Bigioi, D., Corcoran, P. and Cucu, H., 2022. A Text-to-Speech Pipeline, Evaluation Methodology, and Initial Fine-Tuning Results for Child Speech Synthesis. IEEE Access, 10, pp.47628-47642.

[12] S. Wang and Y. Zhu, "A Novel Image Caption Model Based on Transformer Structure," 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE), Chengdu, China, 2021, pp. 144-148, doi: 10.1109/ICICSE52190.2021.9404124.

[13] Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T.-Y. (2020, June 8). FastSpeech2: Fast and high-quality end-to-end text to speech. arXiv.Org. https://arxiv.org/abs/2006.04558