REFERENCE

[1]. Bazarevsky, V., Kartynnik, Y., Vakunov, A., Tkachenka, A., & Grundmann, M. (2019). BlazeFace: Sub-millisecond neural face detection on mobile GPUs. *arXiv*. https://doi.org/10.48550/arXiv.1907.05047

[2]. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

[3]. Chang, Y., Xu, T., He, Y., & Hu, X. (2021). A high-frame-rate eye-tracking framework for mobile devices. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1445–1449). IEEE. https://doi.org/10.1109/ICASSP39728.2021.9414624

[4]. Cheng, S., Sun, Z., Sun, L., Wang, K., Liu, Y., & Dong, L. (2022). EasyGaze: Hybrid eye tracking approach for handheld mobile devices. *Virtual Reality & Intelligent Hardware*, *4*(2), 173–188. https://doi.org/10.1016/j.vrih.2021.10.003

[5]. Fan, H., & Ling, H. (2017). Parallel tracking and verifying: A framework for real-time and high accuracy visual tracking. *arXiv*. https://doi.org/10.48550/arXiv.1708.00153

[6]. Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500. https://doi.org/10.1109/TPAMI.2009.30

[7]. Hart, N. J., Sui, X., Collins, L., & Wang, H. (2021). openEyeTrack: A high-speed multi-threaded eye tracker for head-fixed applications. *Journal of Neuroscience Methods*, *357*, 109134. https://doi.org/10.1016/j.jneumeth.2021.109134

[8]. Hinton, G. (2012). *Lecture 6e – RMSProp: Divide the gradient by a running average of its recent magnitude* [Lecture slides]. University of Toronto. http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

[9]. Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. (2019). Searching for MobileNetV3. *arXiv*. https://doi.org/10.48550/arXiv.1905.02244

[10]. Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

[11]. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. https://arxiv.org/abs/1412.6980

[12]. Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye tracking for everyone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2176–2184). IEEE. https://doi.org/10.1109/CVPR.2016.239.

Dataset available at https://gazecapture.csail.mit.edu/

[13]. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)* (pp. 689–696). PMLR.

[14]. Novák, J. Š., Masner, J., Benda, P., Šimek, P., & Merunka, V. (2024). Eye tracking, usability, and user experience: A systematic review. *International Journal of Human–Computer Interaction*, *40*(17), 4484–4500. https://doi.org/10.1080/10447318.2023.2221600

[15]. Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., Agber, V., Zhu, Y., Yu, J., Lu, Y., Yan, C., & Howard, A. (2024). MobileNetV4: Universal models for the mobile ecosystem. *arXiv*.

https://doi.org/10.48550/arXiv.2404.10518

[16]. Szeliski, R. (2022). *Computer vision: Algorithms and applications* (2nd ed.). Springer.

[17]. Wood, E., Baltrušaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. (2016). Learning an appearance-based gaze estimator from one million synthesised images. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (pp. 131–138). ACM. https://doi.org/10.1145/2857491.2857492