

Developing a deep learning model for real-time human pose estimation and action recognition using multi-modal data sources.

The goal of this research project is to develop a deep learning model that can perform real-time human pose estimation and action recognition by using multi-modal data sources. The methodology involves collecting data from different modalities, such as video, audio, and motion capture data, and preprocessing it to remove noise, outliers, and artifacts. Next, features will be extracted using deep learning techniques, such as CNNs and RNNs, and a multi-modal fusion method will be developed to combine the information from different modalities. Based on the fused features, a deep learning model will be developed for joint human pose estimation and action recognition, which will incorporate attention mechanisms, transfer learning, and optimization from end-to-end. The developed model will be implemented in real-time on a suitable system, such as a GPU-based server or an embedded system, and evaluated on benchmark datasets. Finally, the developed model will be applied to real-world problems in domains like human-robot interaction, virtual reality, and gaming.

References

- [1] Luvizon, Diogo C., David Picard, and Hedi Tabia. "2d/3d pose estimation and action recognition using multitask deep learning." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [2] Luvizon, Diogo C., David Picard, and Hedi Tabia. "Multi-task deep learning for real-time 3D human pose estimation and action recognition." *IEEE transactions on pattern analysis and machine intelligence* 43.8 (2020): 2752-2764.
- [3] Diogo Carbonera Luvizon, Hedi Tabia, and David Picard. "Learning features combination for human action recognition from skeleton sequences." In: *Pattern Recognition Letters* 99 (2017), pp. 13–20.
- [4] Diogo C. Luvizon, David Picard, and Hedi Tabia. "Consensus-Based Optimization for 3D Human Pose Estimation in Camera Coordinates." In: *Int. J. Comput. Vis.* 130.3 (2022), pp. 869–882.
- [5] Diogo C Luvizon, Hedi Tabia, and David Picard. "Human pose regression by combining indirect part detection and contextual information." In: *Computers & Graphics* 85 (2019), pp. 15–22.
- [6] Luvizon, Diogo Carbonera, Hedi Tabia, and David Picard. "Ssp-net: Scalable sequential pyramid networks for real-time 3d human pose regression." *Pattern Recognition* (2023): 109714.
- [7] Akremi, Mohamed, Rim Slama, and Hedi Tabia. "SPD Siamese Neural Network for Skeleton-based Hand Gesture Recognition." *17th International Conference on Computer Vision Theory and Applications VISAPP 2022*. SCITEPRESS-Science and Technology Publications, 2022.
- [8] Ben Charrada, Tarek, et al. "TopoNet: Topology Learning for 3D Reconstruction of Objects of Arbitrary Genus." *Computer Graphics Forum*. Vol. 41. No. 6. 2022.