Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (n.d.). *Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks*. Retrieved July 20, 2024, from https://arxiv.org/pdf/1901.10323v3

The real-time recognition of dynamic hand gestures from video streams presents significant challenges, such as determining the start and end of gestures, ensuring gestures are recognized only once, and designing an architecture that meets memory and power constraints. To address these issues, a hierarchical structure using a sliding window approach is proposed, consisting of two models: a lightweight CNN for detecting gestures and a deep CNN (ResNeXt-101) for classifying them. Levenshtein distance is used as an evaluation metric to measure misclassifications, multiple detections, and missing detections. This approach achieves state-of-the-art offline classification accuracy of 94.04% and 83.82% on the EgoGesture and NVIDIA benchmarks, respectively. The novel two-model hierarchical architecture provides resource efficiency, early detections, and single-time activations, which are critical for real-time gesture recognition. The proposed approach is validated on two dynamic hand gesture datasets, achieving similar results and introducing the Levenshtein accuracy metric for real-time evaluation. Future work includes investigating statistical hypothesis testing for confidence measures of single-time activations and exploring different weighting approaches to further enhance performance.

—----------

Benitez-Garcia, G., Olivares-Mercado, J., Sanchez-Perez, G., & Yanai, K. (n.d.). *IPN Hand: A Video Dataset and Benchmark for Real-Time Continuous Hand Gesture Recognition*. Retrieved July 20, 2024, from https://arxiv.org/pdf/2005.02134v2

Continuous hand gesture recognition (HGR) is critical for human-computer interaction, with applications in sectors such as automotive, consumer electronics, and home automation. Despite the development of accurate deep learning models, existing publicly available datasets lack real-world elements necessary for building efficient HGR systems. To address this, a new benchmark dataset named IPN Hand is introduced, containing over 4,000 gesture samples and 800,000 RGB frames from 50 subjects, with 13 static and dynamic gestures designed for interaction with touchless screens. The dataset includes natural hand movements, diverse scenes, and varying background and illumination conditions. The performance of three 3D-CNN models is evaluated on isolated and continuous real-time HGR tasks, exploring the addition of modalities like optical flow and semantic segmentation to improve accuracy while maintaining real-time performance. Experimental results show that the state-of-the-art ResNeXt-101 model's accuracy decreases by about 30% with this real-world dataset, highlighting its potential as a benchmark for advancing continuous HGR. In conclusion, the IPN Hand dataset addresses real-world issues in continuous HGR, suggesting that RGB-seg is a suitable multi-modal alternative for real-time recognition, and can serve as a valuable benchmark for the research community.

—----------

Devineau, G., Moutarde, F., Xi, W., & Yang, J. (2018). Deep Learning for Hand Gesture Recognition on Skeletal Data. *2018 13th IEEE International Conference on Automatic*

Face & Gesture Recognition (FG 2018), 106–113.

https://doi.org/10.1109/FG.2018.00025

This paper presents a new 3D hand gesture recognition approach using a deep learning model that processes sequences of hand-skeletal joints' positions through parallel convolutions. The introduced Convolutional Neural Network (CNN) achieves state-of-the-art performance on the DHG dataset from the SHREC 2017 3D Shape Retrieval Contest, with classification accuracies of 91.28% for 14 gesture classes and 84.35% for 28 gesture classes. Key innovations include parallel processing branches for different time resolutions, residual connections for improved gradient backpropagation, and the use of dropout for regularization. This approach demonstrates that parallel processing with CNNs can compete with architectures like GRU and LSTM for sequence data, achieving significant improvements in classification accuracy.