A Unified Deep Framework for Joint 3d Pose Estimation and Action Recognition from a Single Color Camera

Authors : Huy Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio Velastin

Abstract : We present a deep learning-based multitask framework for joint 3D human pose estimation and action recognition from color video sequences. Our approach proceeds along two stages. In the first, we run a real-time 2D pose detector to determine the precise pixel location of important key points of the body. A two-stream neural network is then designed and trained to map detected 2D keypoints into 3D poses. In the second, we deploy the Efficient Neural Architecture Search (ENAS) algorithm to find an optimal network architecture that is used for modeling the Spatio-temporal evolution of the estimated 3D poses via an image-based intermediate representation and performing action recognition. Experiments on Human3.6M, Microsoft Research Redmond (MSR) Action3D, and Stony Brook University (SBU) Kinect Interaction datasets verify the effectiveness of the proposed method on the targeted tasks. Moreover, we show that our method requires a low computational budget for training and inference.