

Deformable Pose Traversal Convolution for 3D Action and Gesture Recognition

Junwu Weng[†] [0000–0001–7983–1461], Mengyuan Liu[†] [0000–0003–4036–9315],
Xudong Jiang[†] [0000–0002–9104–2315], and Junsong Yuan[§] [0000–0002–7324–7034]

^YSchool of EEE, Nanyang Technological University

we0001wu@e.ntu.edu.sg Fliumengyuan, exdjiangG@ntu.edu.sg

^XDepartment of CSE, The State University of New York, Buffalo
jsyuan@buffalo.edu

Abstract. The representation of 3D pose plays a critical role for 3D action and gesture recognition. Rather than representing a 3D pose directly by its joint locations, in this paper, we propose a Deformable Pose Traversal Convolution Network that applies one-dimensional convolution to traverse the 3D pose for its representation. Instead of fixing the receptive field when performing traversal convolution, it optimizes the convolution kernel for each joint, by considering contextual joints with various weights. This deformable convolution better utilizes the contextual joints for action and gesture recognition and is more robust to noisy joints. Moreover, by feeding the learned pose feature to a LSTM, we perform end-to-end training that jointly optimizes 3D pose representation and temporal sequence recognition. Experiments on three benchmark datasets validate the competitive performance of our proposed method, as well as its efficiency and robustness to handle noisy joints of pose.

Keywords: Pose Traversal, Pose Convolution, Deformable Convolution, 3D Action and Gesture Recognition

1 Introduction

With the success of pose estimation methods [1, 2, 3] using depth sensor, 3D action and hand gesture recognition have drawn considerable attention. To recognize 3D action and gestures, each 3D pose is often characterized by its joints with 3D locations.

However, previous work [4, 5, 6] show that not every spatial joint is of equal importance to the recognition of actions, and human body movements exhibit spatial patterns among pose joints [7]. It is thus of great importance to identify those motion patterns and avoid the non-informative joints, via identifying the key combinations of joints that matter for the recognition. For instance, to recognize hand gesture “Okay”, the “approaching of index fingertip and thumb tip” as well as the “stretching of other three fingers apart from the palm” should be observed. The coordination of these five key fingertips is important to the recognition of gesture “Okay”.

