A neural network based algorithm for classification of sets of human body keypoints

I. Kilbas, D. Gribanov, R. Paringer

Introduction

Human action recognition has attracted a lot of attention due to rapid development of surveillance systems and intelligent medical applications. There are multiple approaches to action recognition, such as processing images as a whole or performing analysis of a specific set of features. One of the prominent approaches is to use pose estimation to perform posture analysis. In this paper we consider the latter and present a neural network based algorithm for classification of sets of human body keypoints. The key feature of the algorithm is data preprocessing approach, which allows to increase generalization of the neural network performing classification.

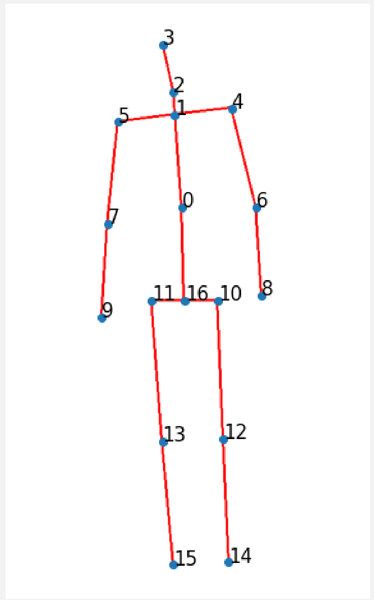

Figure 1. Example of a keypoints set (a pose)

Methodology

The algorithm consists of two parts: a preprocessing module and a neural network. The preprocessing module is responsible for extracting features from the input set of keypoints. The neural network (multi-layer perceptron) classifies the vector of extracted features.

The preprocessing module extracts two types of features: normalized vectors representing directions of limbs and normalized distances from hands to knees and feet. The latter is required to retain high accuracy on some of the poses (the forward tilt of the torso, later poses of type S). The normalized distances computed as:

$$d_{norm} = ||p_{knee\_left} - p_{foot\_left}|| + ||p_{knee\_right} - p_{foot\_right}||$$

$$D(d) = exp\,(-\sigma * d / d_{norm}),$$

where d is a distance between hand and knee as well as hand and foot; σ is a positive constant which value is set empirically.

To illustrate the importance of the developed preprocessing module, we do a comparative study of different versions of the algorithm: the original algorithm - version A, without the preprocessing – version B.

Table 1. Accuracy of different versions of the algorithm

Algorithm version	*Average accuracy*	*Accuracy on poses of type S*
A	**89,8**%	**78,0**%
B	79,9%	51,2%

Results

Figure 2. Confusion matrix for the version A

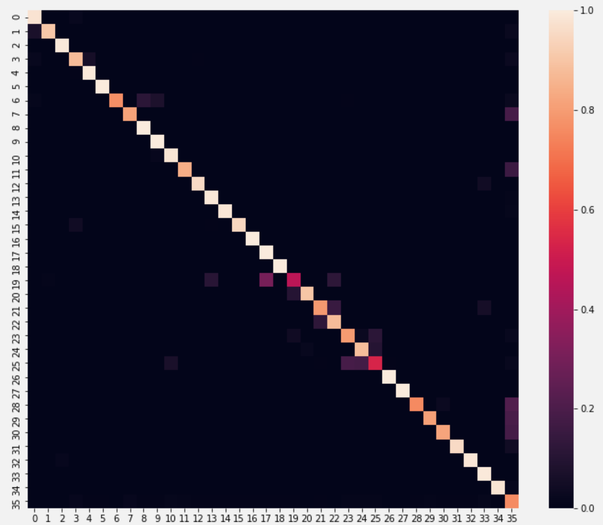

Figure 3. Confusion matrix for the version B