Team name	Stevenwudi
Team leader name	Stevenwudi
Team leader address, phone number and email	17 Marlborough Road, Sheffield, U.K. +447411671105 stevenwudi@gmail.com
Rest of team members	
Team website URL (if any)	http://vision.group.shef.ac.uk/DiWu.html

Title of the contribution	Deep Learning with Graphical Models (ergodic state HMM)
General method description	1) SkeletonModule: Deep Belief Networks for modeling singleton factors for Hidden Markov Model. 2) Depth Module: 3D convolutional Neural Networks for modeling singleton factors for Hidden Markov Model by stacking 4 frames together. 3) Viterbi path decoding for simultanesouly segmenting and classifying gestures. 4) The overall performance should be around 0.8 in terms of Jaccard Index. If the performance of the submission is vastly different, please inform me of the issue (maybe the wrongly submitted file?)
References	[1]Playing Atari with Deep Reinforcement Learning, Volodymyr Mnih et al. [2]Leveraging Hierarchical Parametric Networks for Skeletal Joints Based Action Segmentation and Recognition, Di Wu, Ling Shao, CVPR 2014

Describe data preprocessing techniques applied (if any)	**(1) Skeleton: used_joints = ['ElbowLeft', 'WristLeft', 'ShoulderLeft', 'HandLeft', 'ElbowRight', 'WristRight', 'ShoulderRight', 'HandRight', 'Head', 'Spine', 'HipCenter'] preprocessing as in Ref[2]** **(2) Depth data:** **1)taking the foreground mask,** **2)template matching** **i)actor's chest to image centre→ find SHIFT,** **ii)normalize the depth image by its centre median depth→ find scale** **iii) Median filter by 5*5 (not sure whether it's necessary, but done it anyway)** **iv) Resize centre cbuoid of 320*320 to 90*90 (simply because. Stacking 4 frames together as in Ref[1]** **Note that (i) – (iv) as basic image processing takes a loooong time on CPU (though still 2 times faster than real time, approx. 30 fps)**
Describe features used or data representation model (if any)	Skeleton→ Deep Belief Network Depth→ 3D Convolutional Neural Network
Data modalities used, i.e.	Depth (with user segmentation), Skeleton,

Temporal clustering approach (if any)	Nope
Temporal segmentation approach (if any)	Ergodic state HMM
Gesture representation approach (if any)	?
Classifier used (if any)	Viterbi path decoding
Large scale strategy (if any)	Deep Learning Model (?)

Transfer learning strategy (if any)	Can Deep Belief Network pre-training count?
Temporal coherence and/or tracking approach considered (if any)	Almost no tracking (apart From depth normalization according to template matching, but I didn't do it on a frame to frame base, maybe too computational intensive.)
Other technique/strategy used not included in previous items (if any)	Combining multiple different initialization Nets always helps better estimation
Method complexity analysis	Though learning the network using stochastic gradient descent is tediously lengthy, once the model finishes training, with low inference cost, our framework can perform in realtime action segmentation/recognition. Specifically, a single feed forward neural network incurs trivial computation time, linearly in O(T) and the complexity of Viterbi algorithm is O(T ∗|S|^2) with number of frames T and state number S.

Qualitative advantages of the proposed solution	**(1) Simultaneouly gesture recognition and segmentation** **(2) Real time performance (in theory)** **(3) The larger the dataset, the better for the Deep Learning model to learn**
Results of the comparison to other approaches (if any)	Dunno
Novelty degree of the solution and if is has been previously published	(1)Skeleton method has been published at CVPR2014 (its performance I reckon should around 0.8 by Jarccard Index) (2) Depth method, dunno

Language and implementation details (including platform, memory, parallelization requirements)	**Mainly: Python (C++ back end)** **Memory requirement: almot no requirement** **Parralleization: Depth Image classification requires GPU, Skeleton Deep Belief Network can be run on GPU (the very powerful and handy Theano internally handles that)**
Human effort required for implementation, training and validation?	My PhD study is sort about it….. The idea was intensively matured and trialed during the last week. The software packages handles' the validation set, therefore reduce the human effort for implementing validation.
Training/testing expended time?	Training Skeleton: 10 hours ish Training Depth: preprocessing 10+ hours, training 3D-CNN 10 hours
General comments and impressions of the challenge	Glad the organizors provide us with such a large scale gesture recognition dataset, allowing some of my ideas work comparatively well wrt small dataset.