

Action
Detection

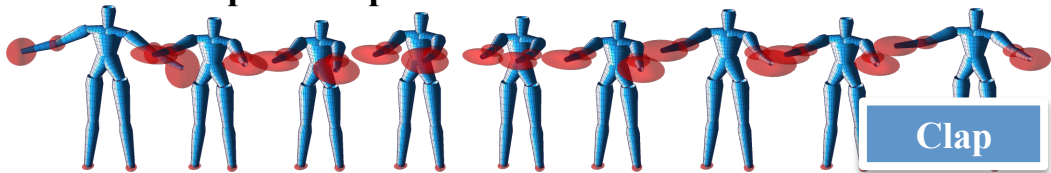


What?
(Action label)

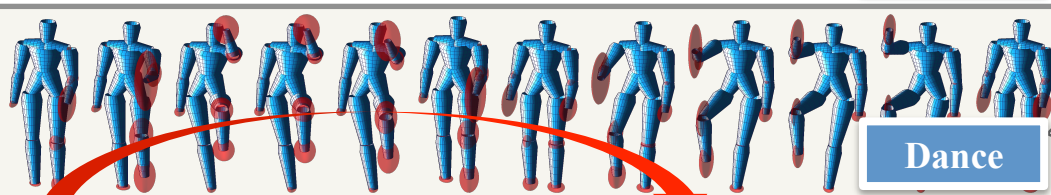


Where?
(Spatiotemporal Location)

Spatiotemporal Structure of Action Class



Clap



Dance

Time

$T - \Delta t$

The starting point of a
“dance” action is detected
at time = $T - \Delta t$.

T

At time T , the current pose is estimated from the
spatiotemporal structure of the “dance” action,
with respect to the relative time difference $-\Delta t$.

Action detection performs:

Action classification
determines the action class
and its spatiotemporal
structure.

Action localisation detects
the starting point of the
action.