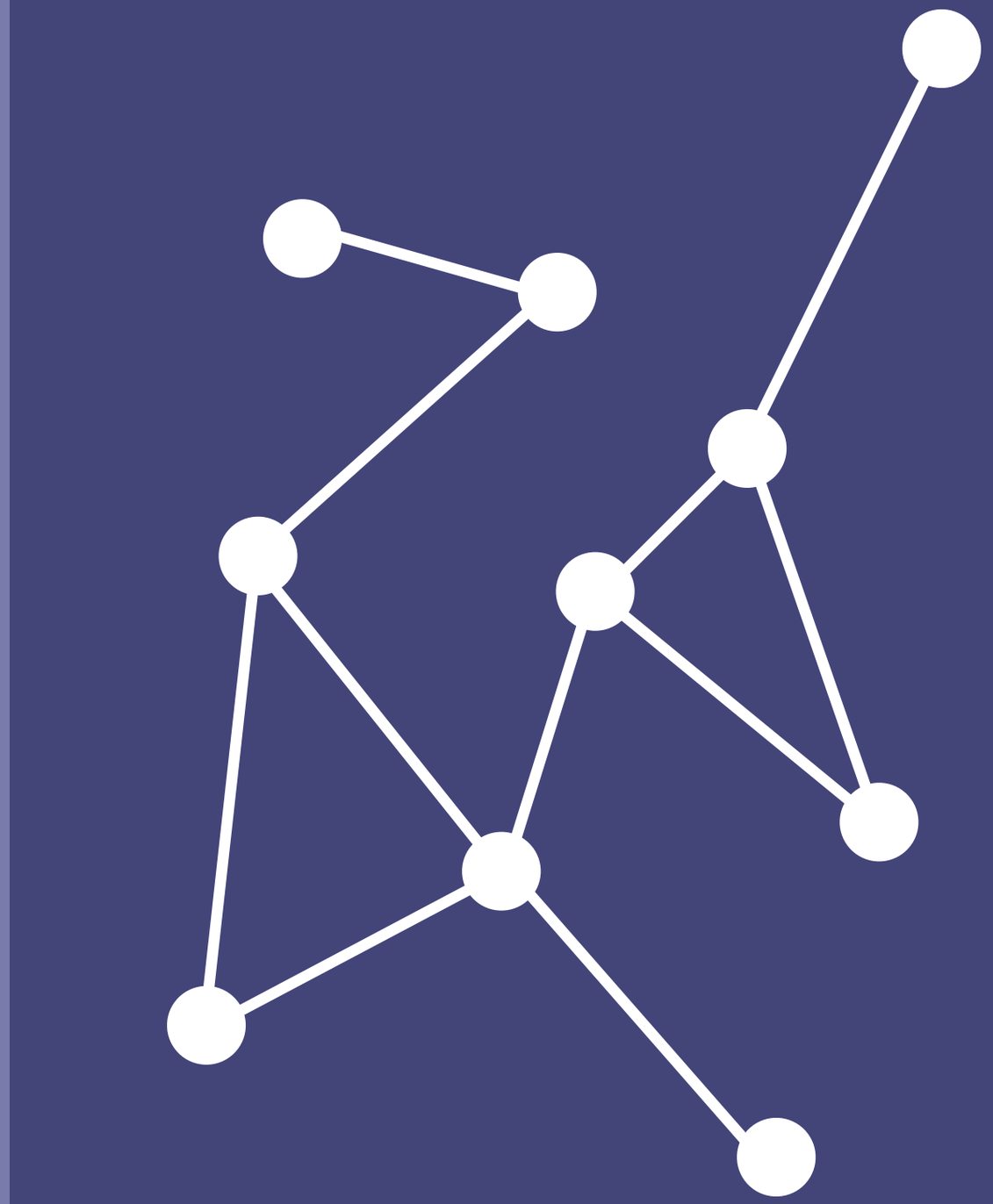
Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition

A presentation by Faizaan Sakib

Spatial Temporal Graph Convolutional Network (STGCN)

Graph Convolutional Network



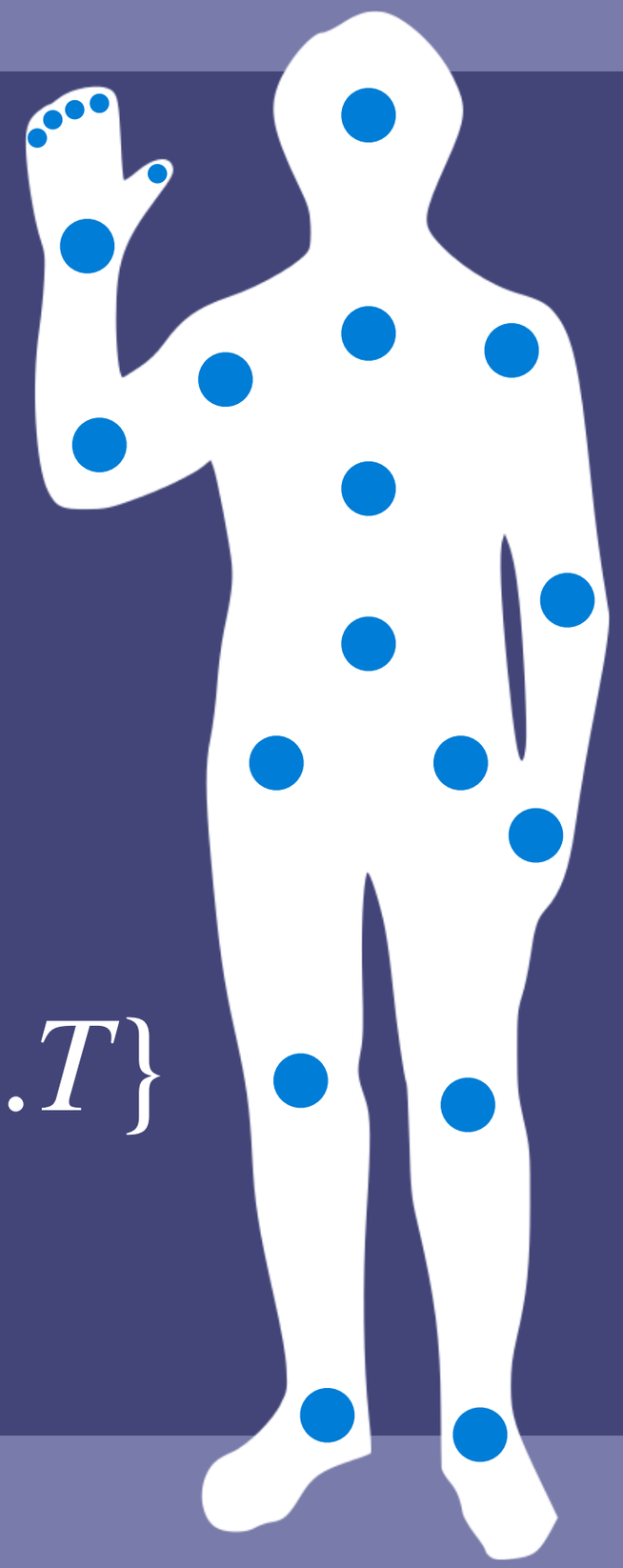
$$G = (V, E)$$

$$G = (V, E)$$

Graph Construction

$$G = (V, E)$$

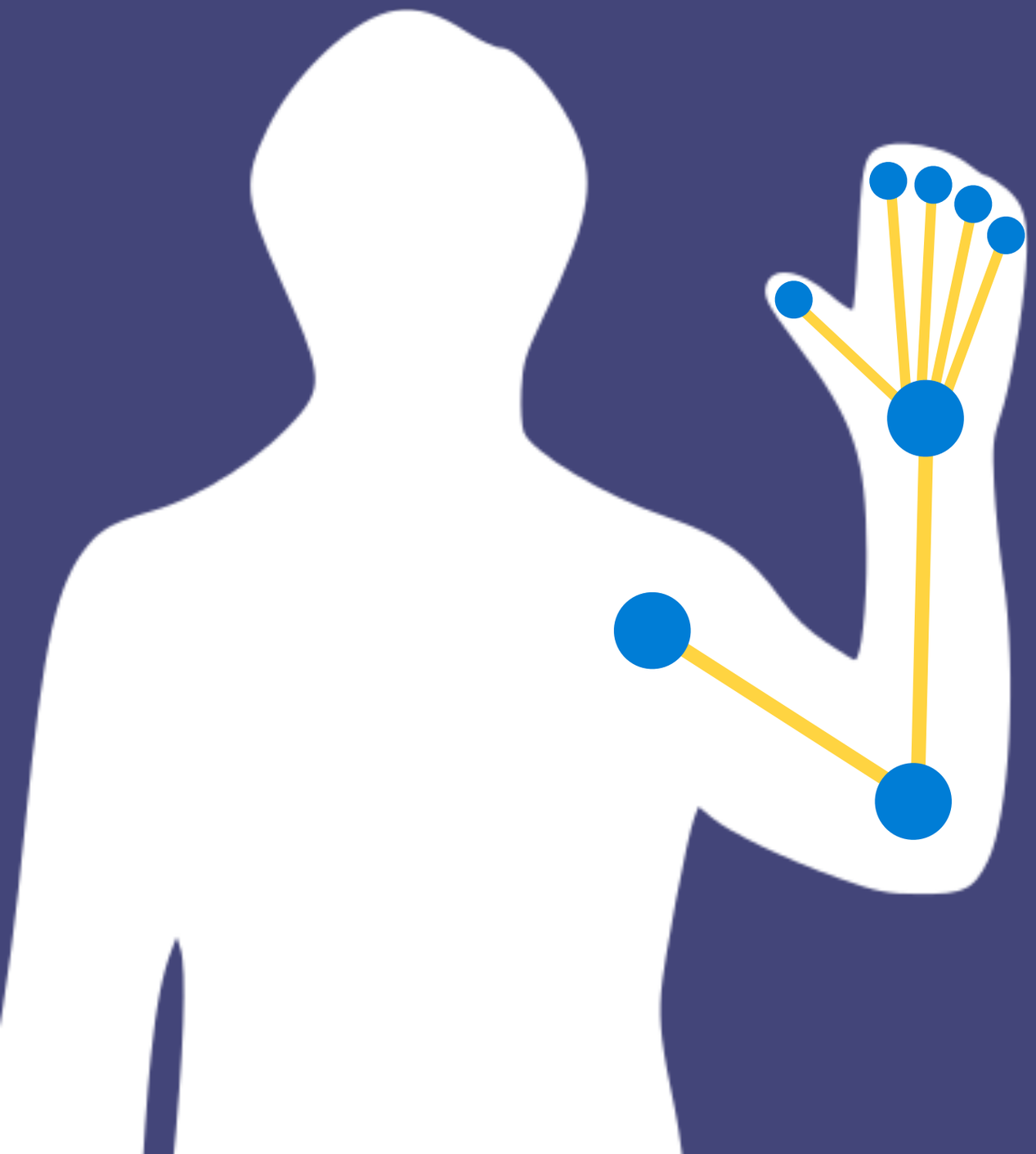
$$V = \{v_{it} \mid i = 1 \ldots N, t = 1 \ldots T\}$$



$$G = (V, E)$$

“intra-skeleton”

$$E_S = \{v_{it}v_{jt} \mid (i,j) \in H\}$$

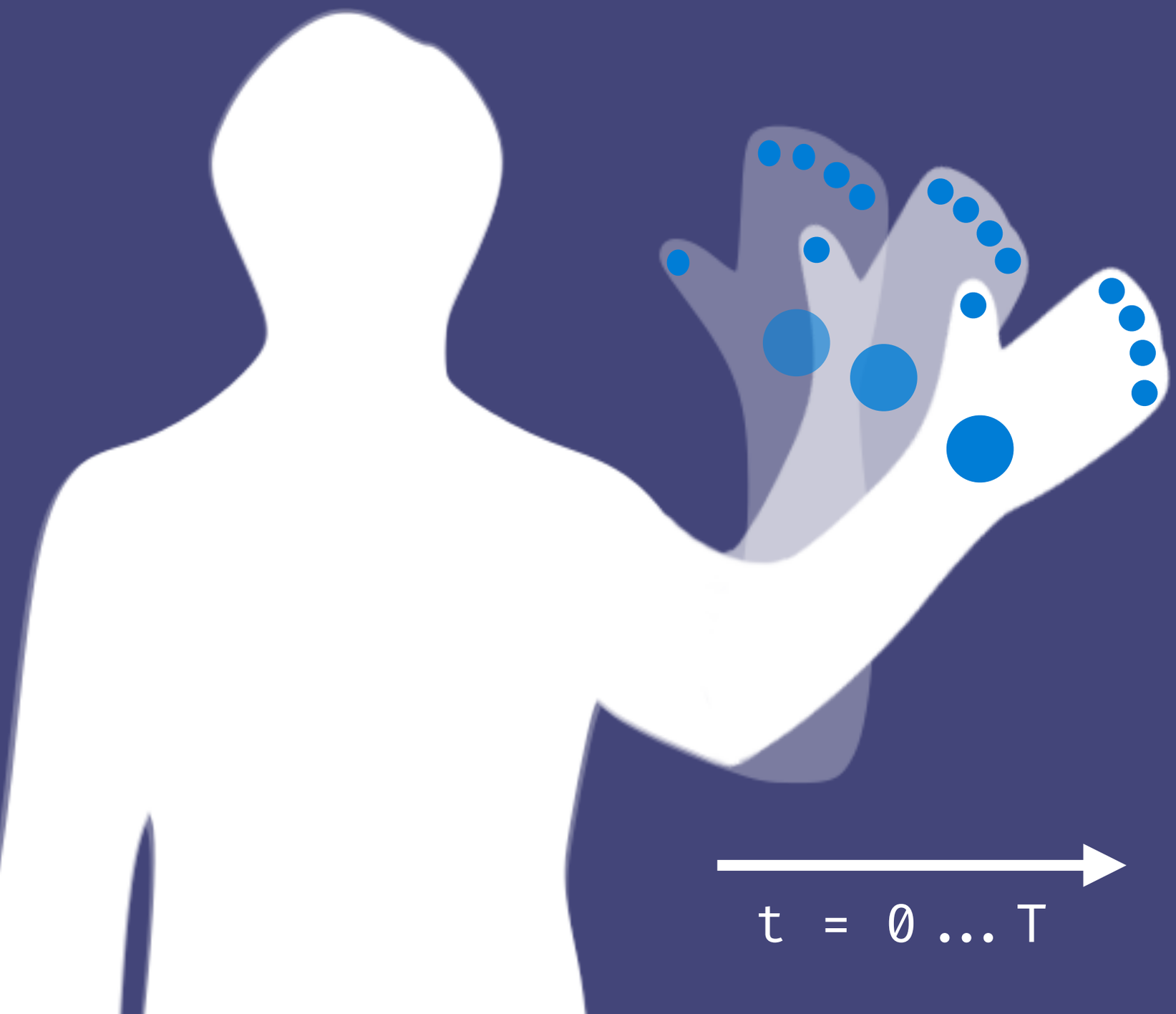


Spatial

Temporal

"intra-frame"

$$E_F = \{v_{it} v_{i(t+1)}\}$$

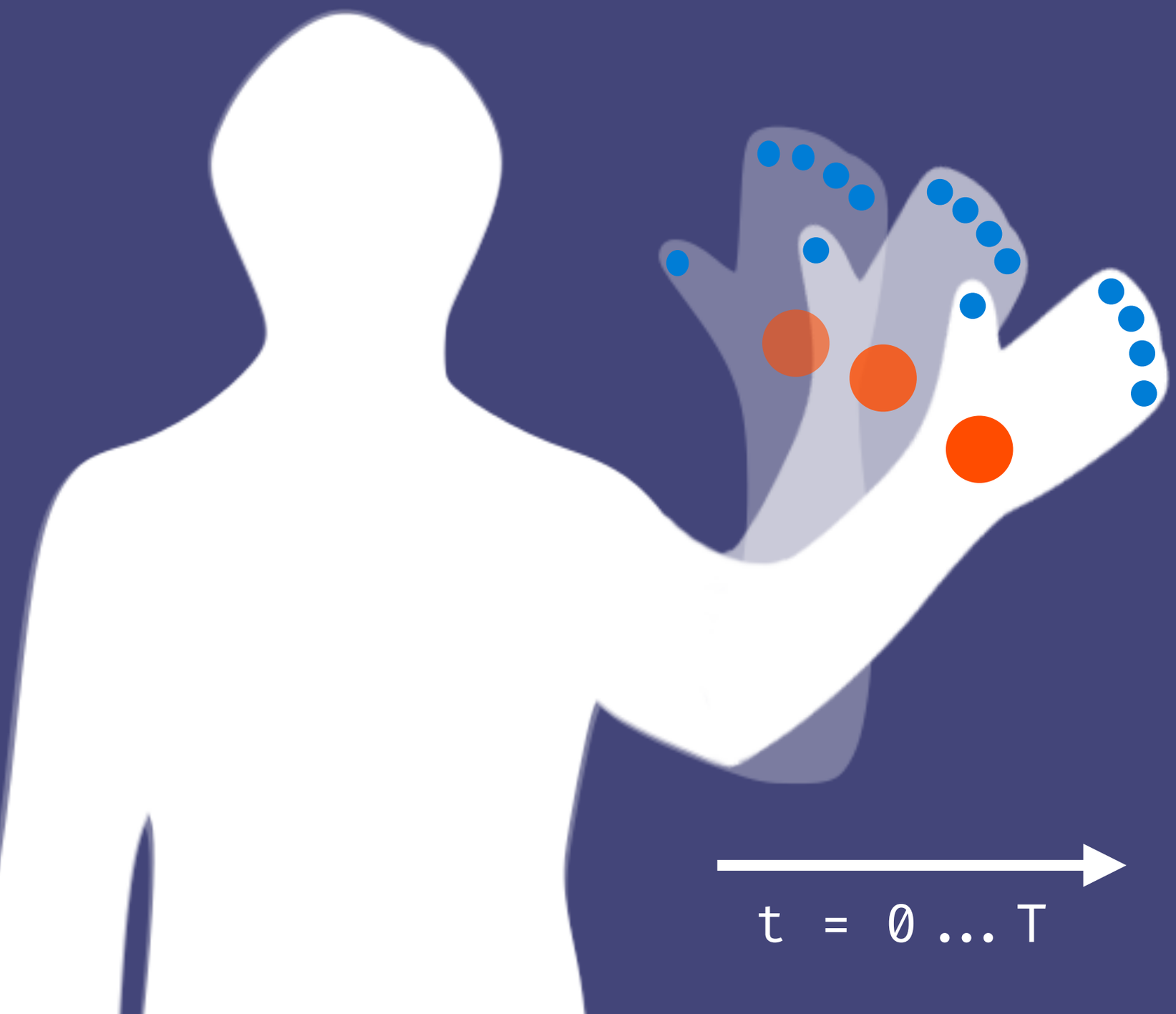


Spatial

Temporal

"intra-frame"

$$E_F = \{v_{it} v_{i(t+1)}\}$$

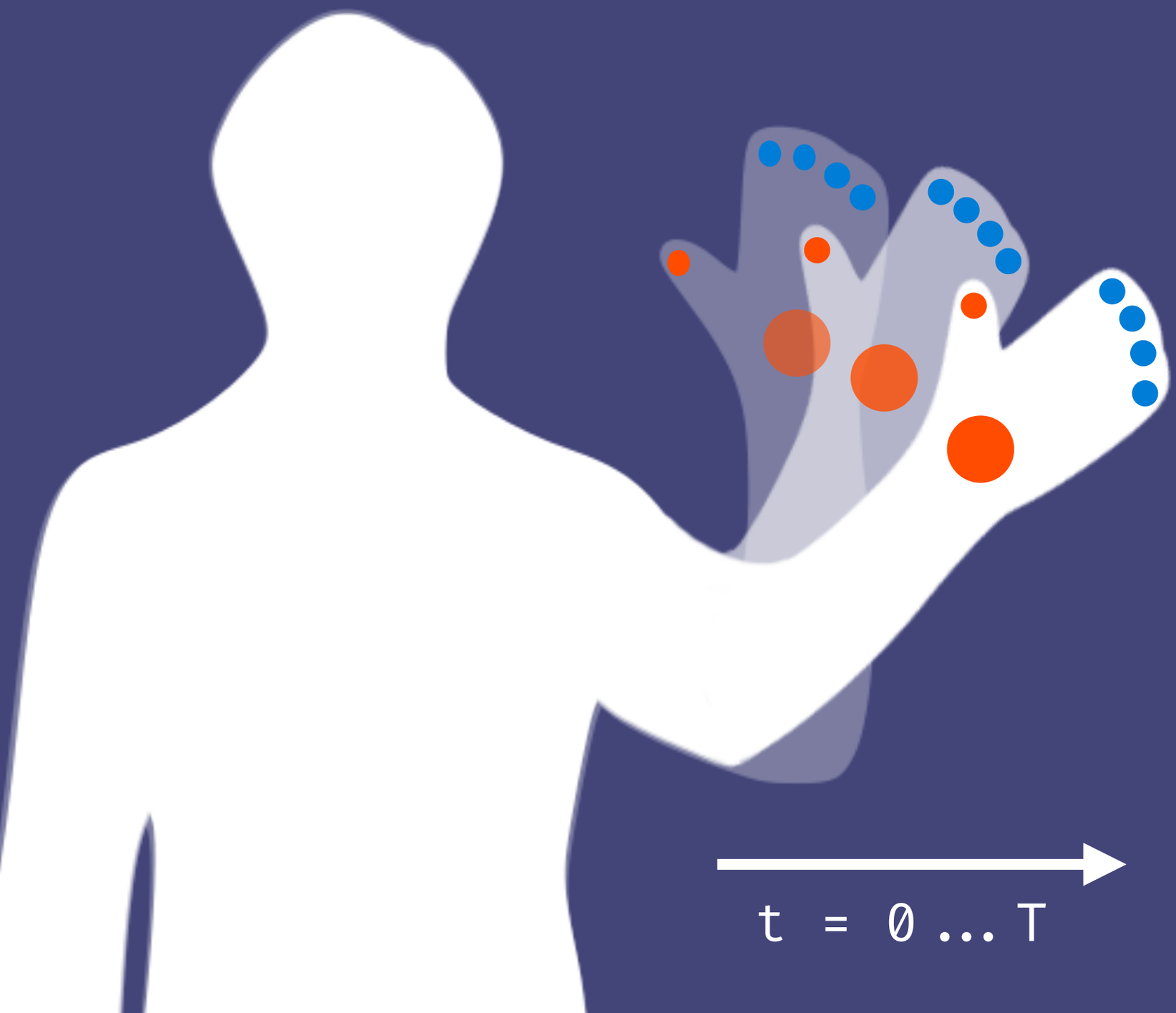


Spatial

Temporal

"intra-frame"

$$E_F = \{v_{it} v_{i(t+1)}\}$$

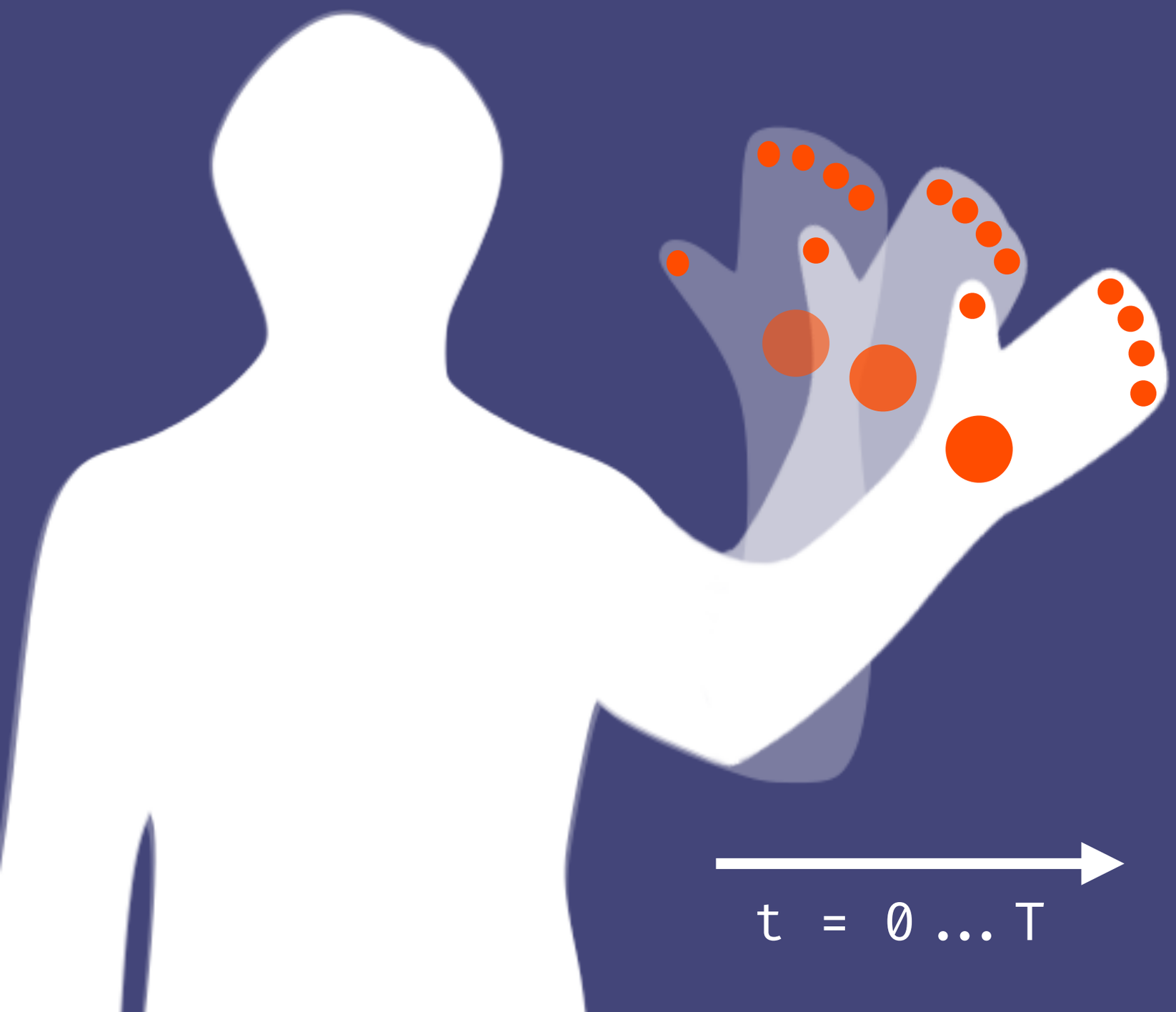


Spatial

Temporal

"intra-frame"

$$E_F = \{v_{it} v_{i(t+1)}\}$$

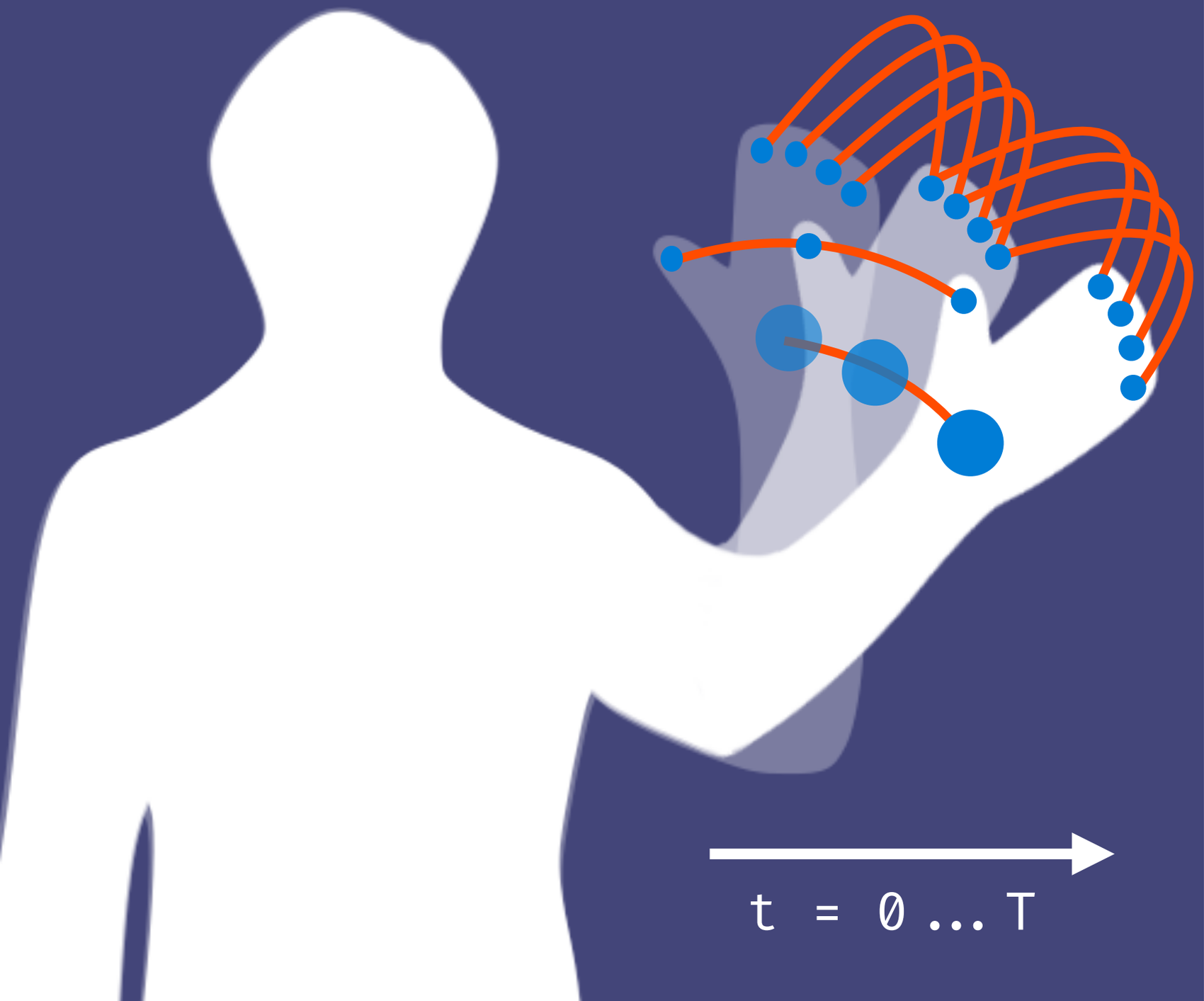


Spatial

Temporal

"intra-frame"

$$E_F = \{v_{it} v_{i(t+1)}\}$$



Spatial

Temporal

Before STGCNs

class = waving

softmax

CNN

singular RGB frame

CNN

consecutive multi-frame of optical flow

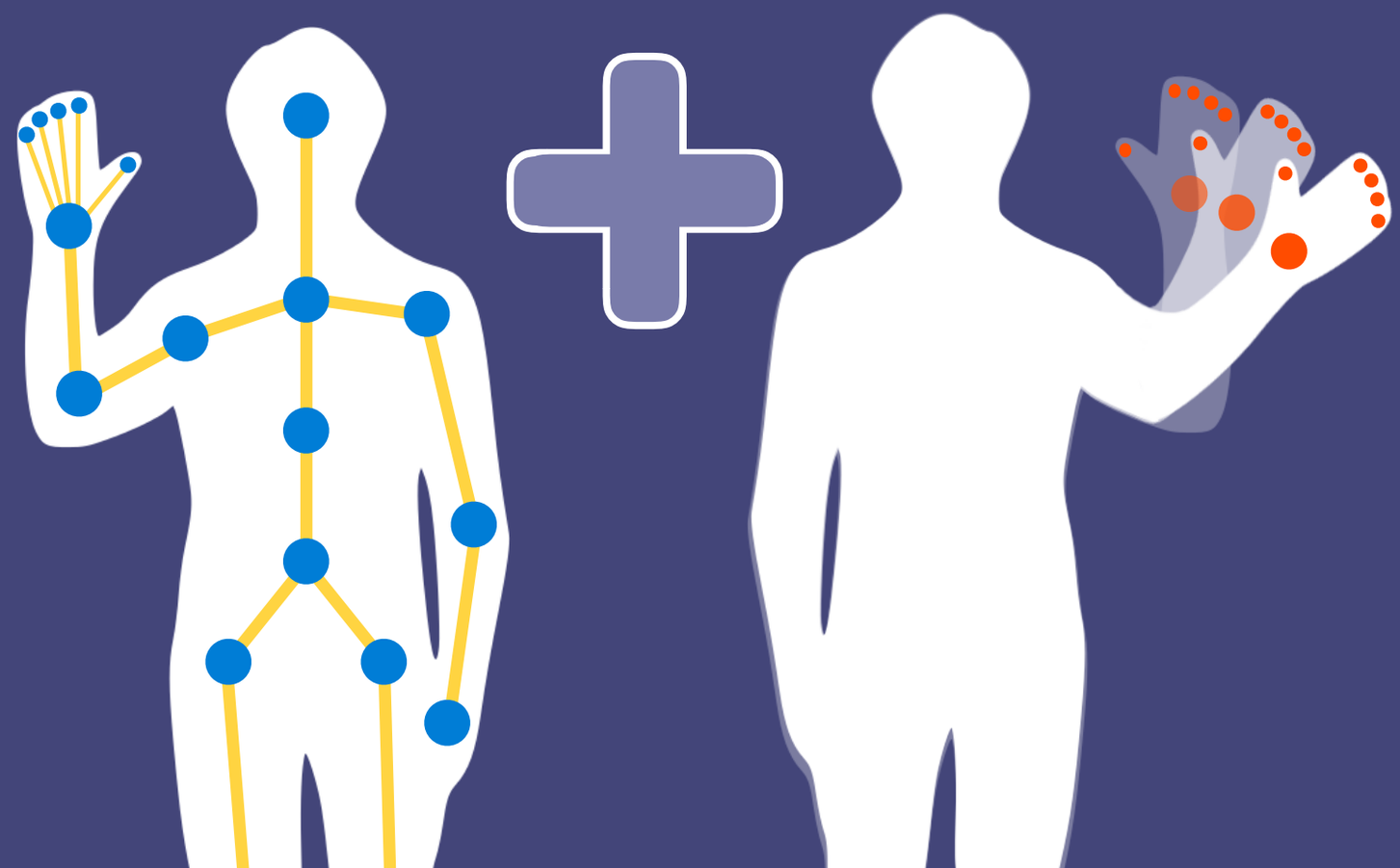
input video

Pose
estimation
(e.g. OpenPose)

class = waving

softmax

STGCN ~9 layers



Graph Convolution

image

kernel

2	1	2	0	2
1	3	2	2	1
1	2	1	0	3
2	0	0	1	0
0	1	1	3	1

2	1	1
0	0	2
1	2	0

kernel

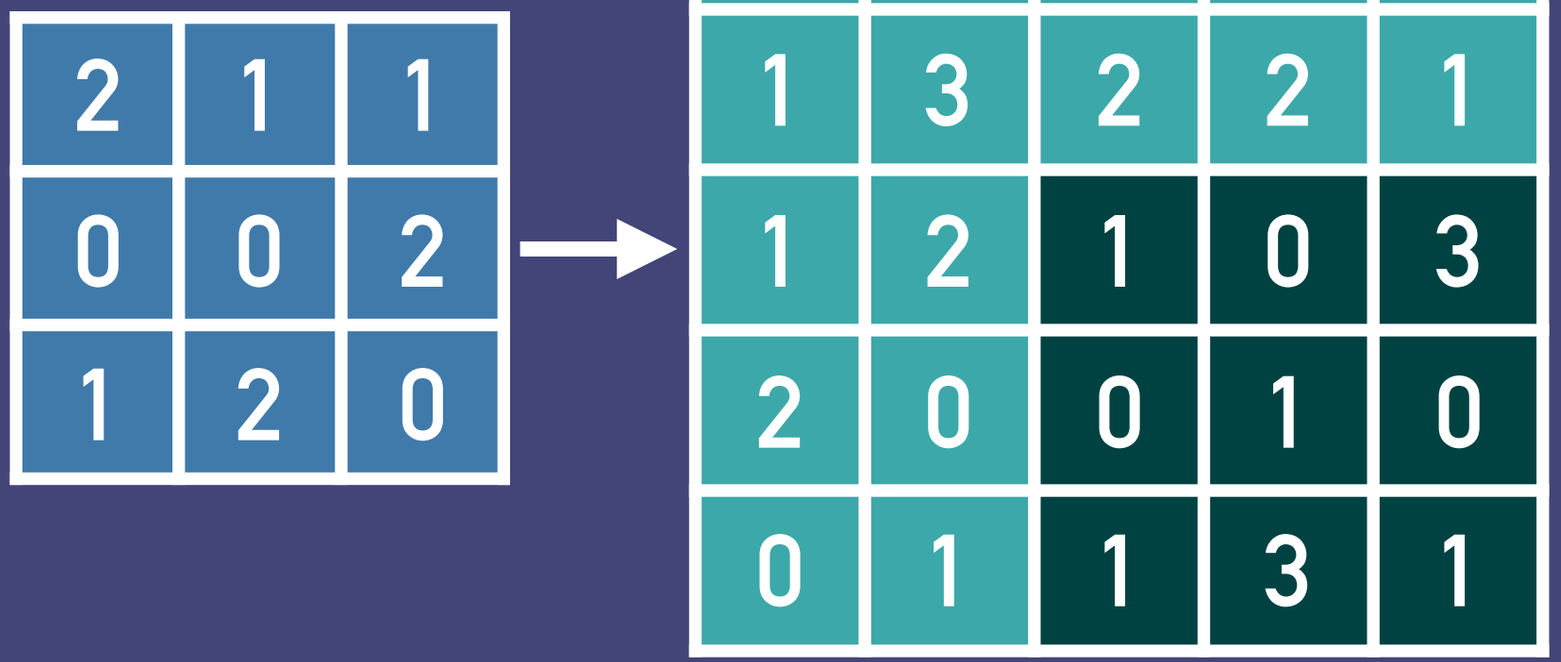
2	1	1
0	0	2
1	2	0

image

2	1	2	0	2
1	3	2	2	1
1	2	1	0	3
2	0	0	1	0
0	1	1	3	1

image

kernel



image

kernel

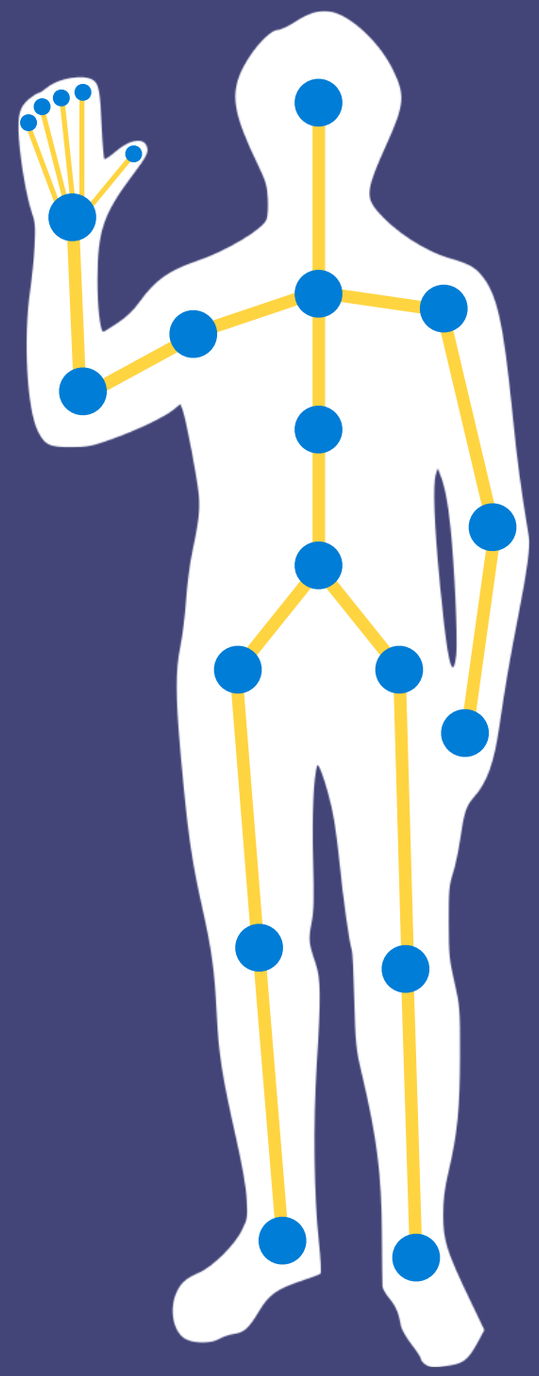
2	1	1
0	0	2
1	2	0

2	1	2	0	2
1	3	2	2	1
1	2	1	0	3
2	0	0	1	0
0	1	1	3	1

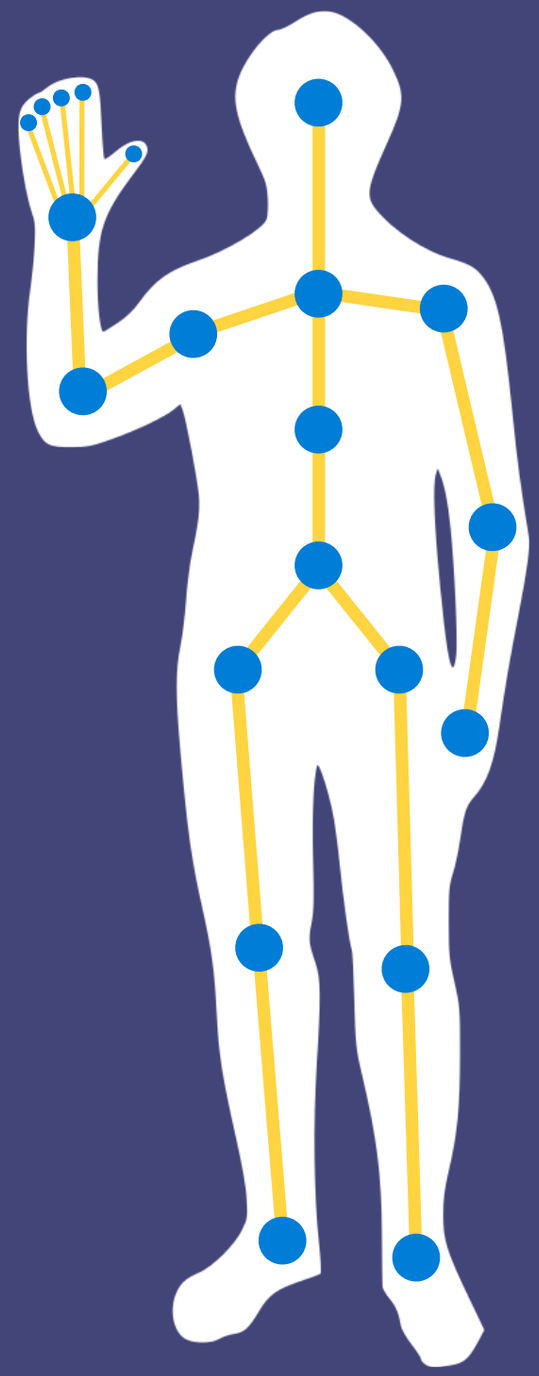
output

2	1	1
0	0	2
1	2	0

2	1	1
0	0	2
1	2	0

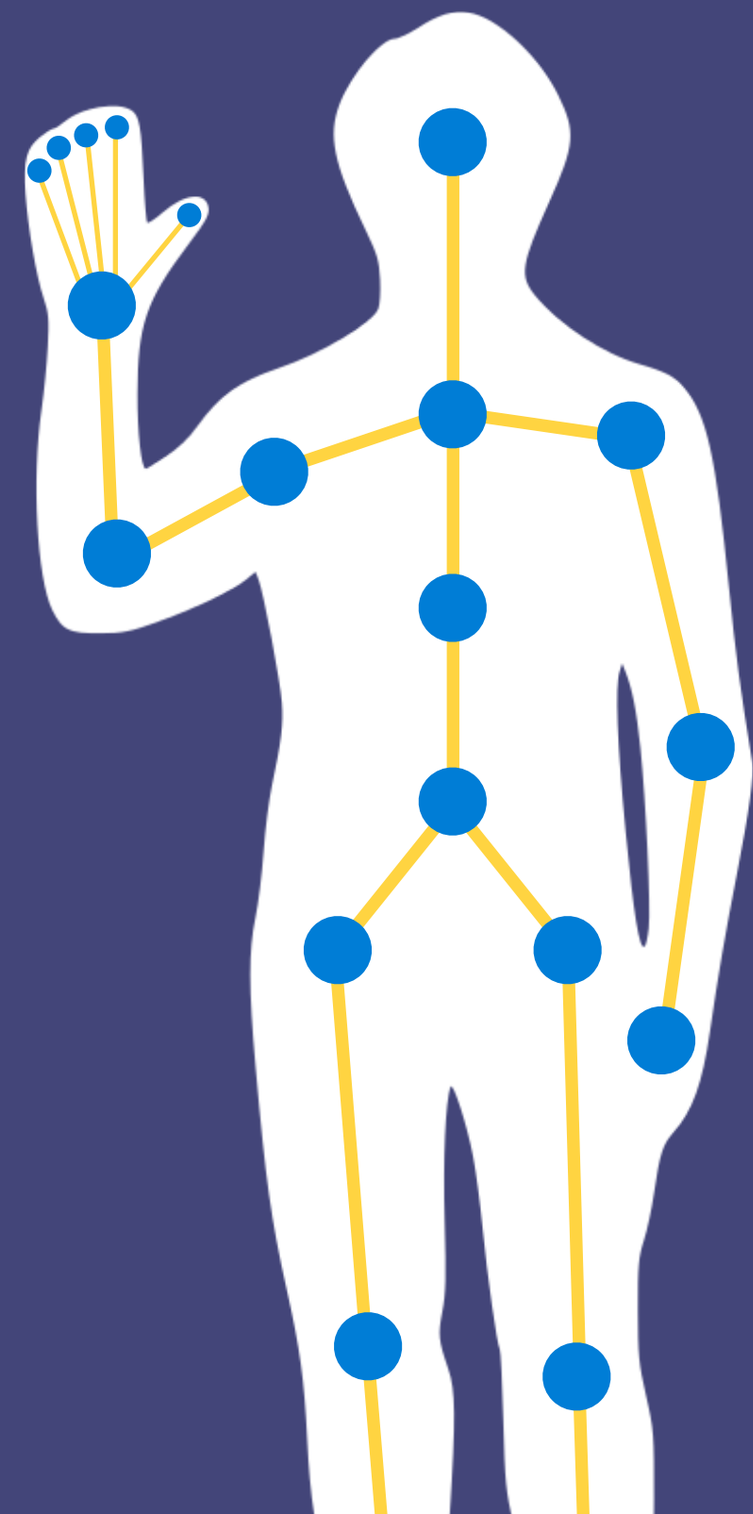


2	1	1
0	0	2
1	2	0



?

Spatial Convolution

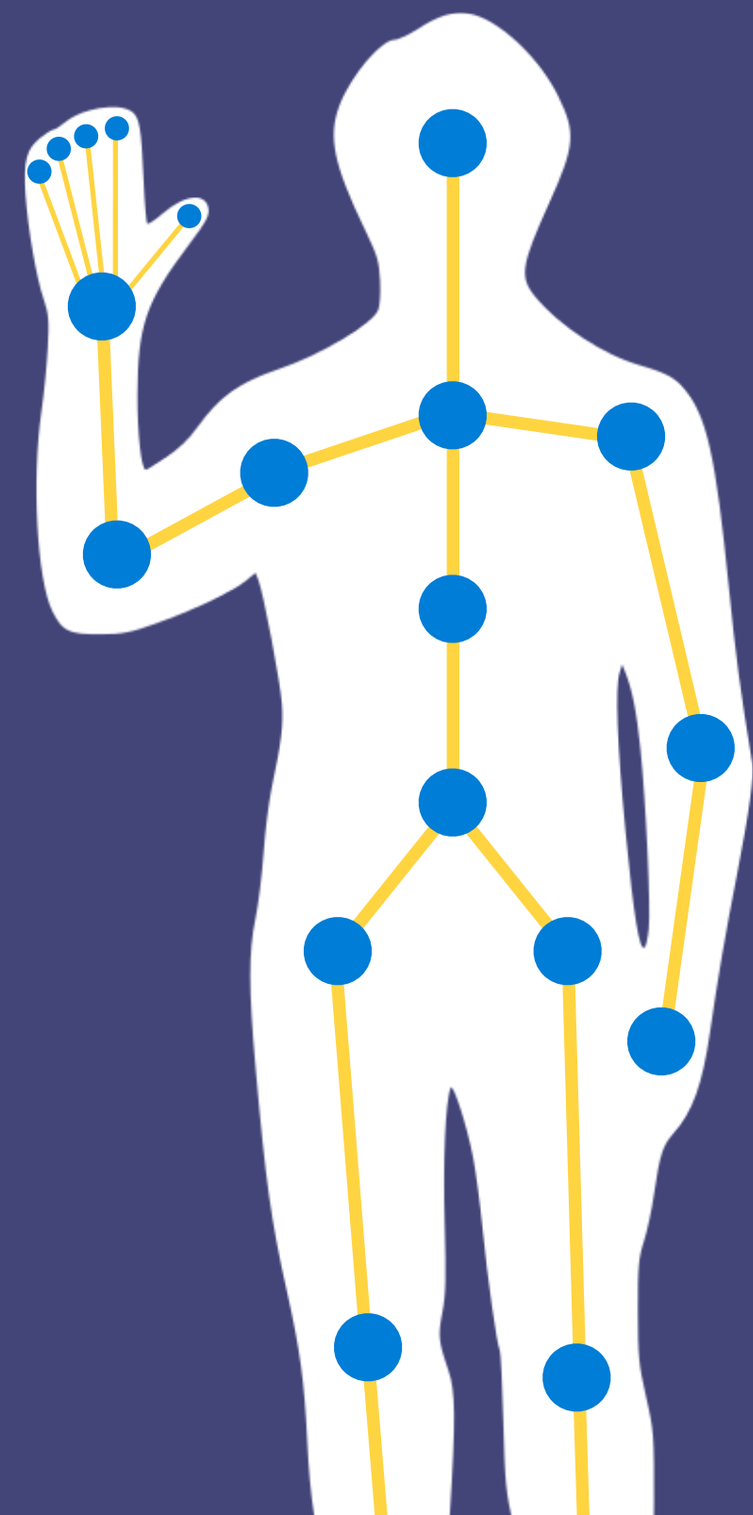


Spatial Configuration Partitioning

Root Node

Centripetal

Centrifugal

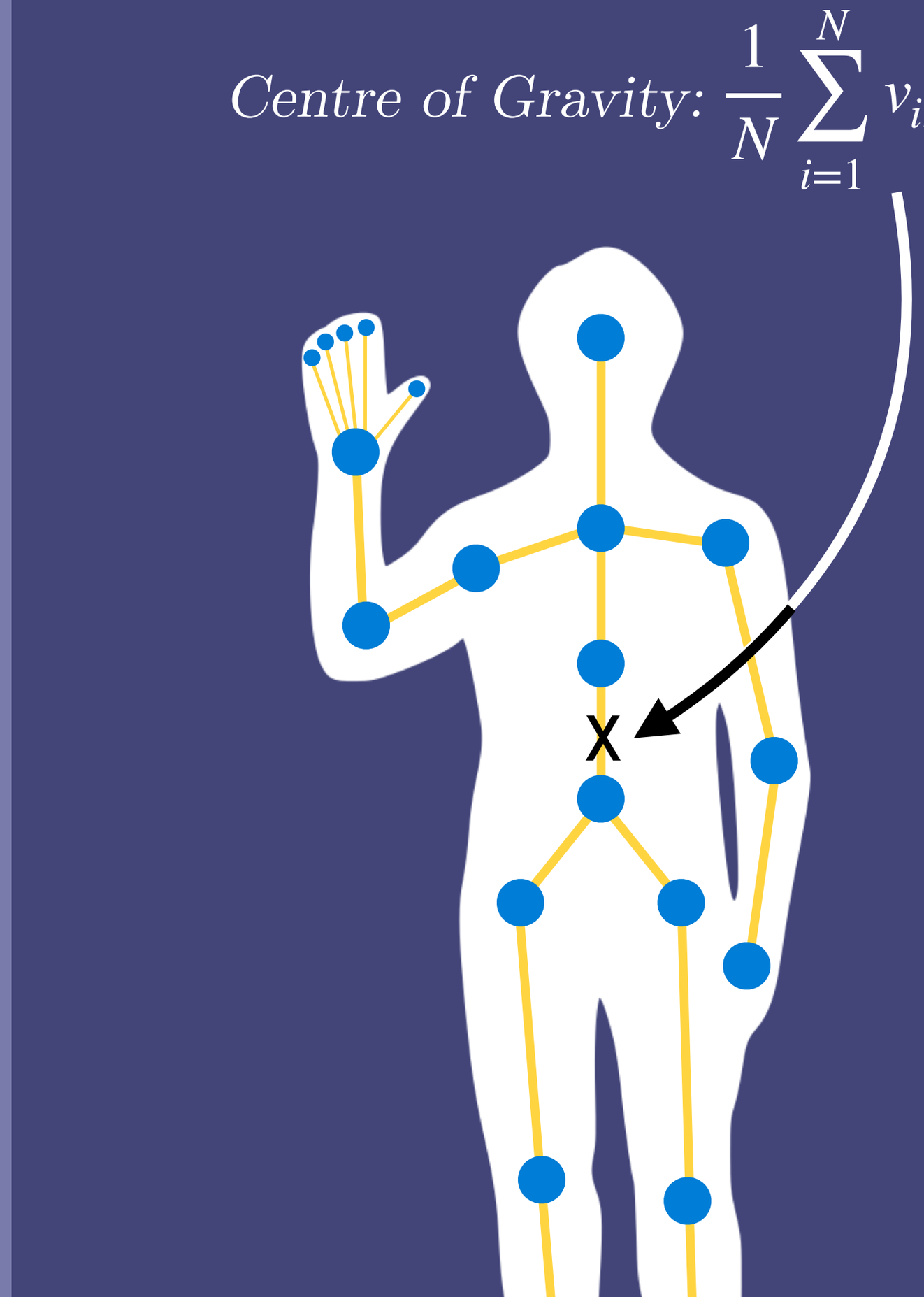


Spatial Configuration Partitioning

Root Node

Centripetal

Centrifugal



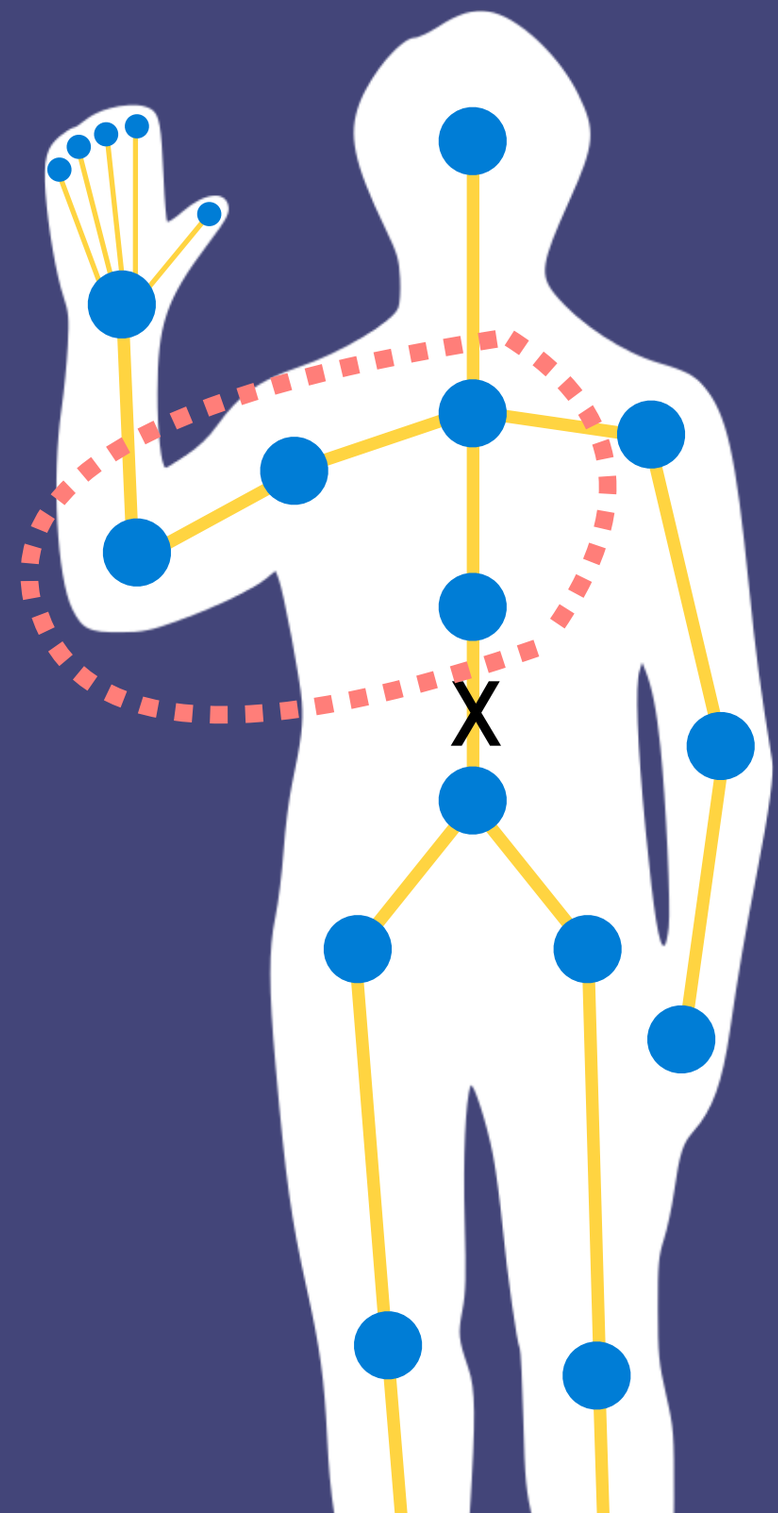
Spatial Configuration Partitioning

Root Node

Centripetal

Centrifugal

$$\textit{Centre of Gravity: } \frac{1}{N}\sum_{i=1}^{N} v_{it}$$

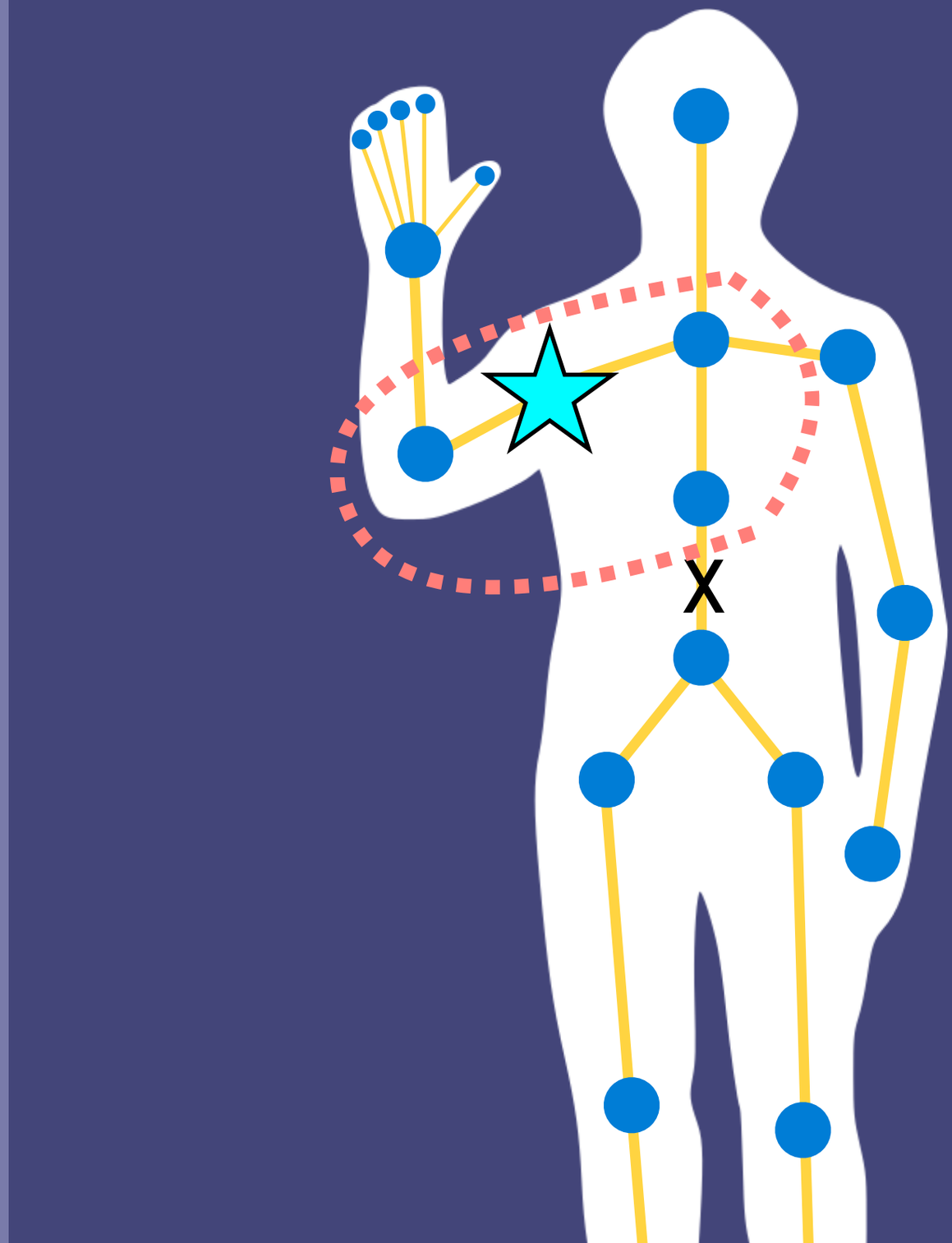


Spatial Configuration Partitioning

Root Node

Centripetal

Centrifugal

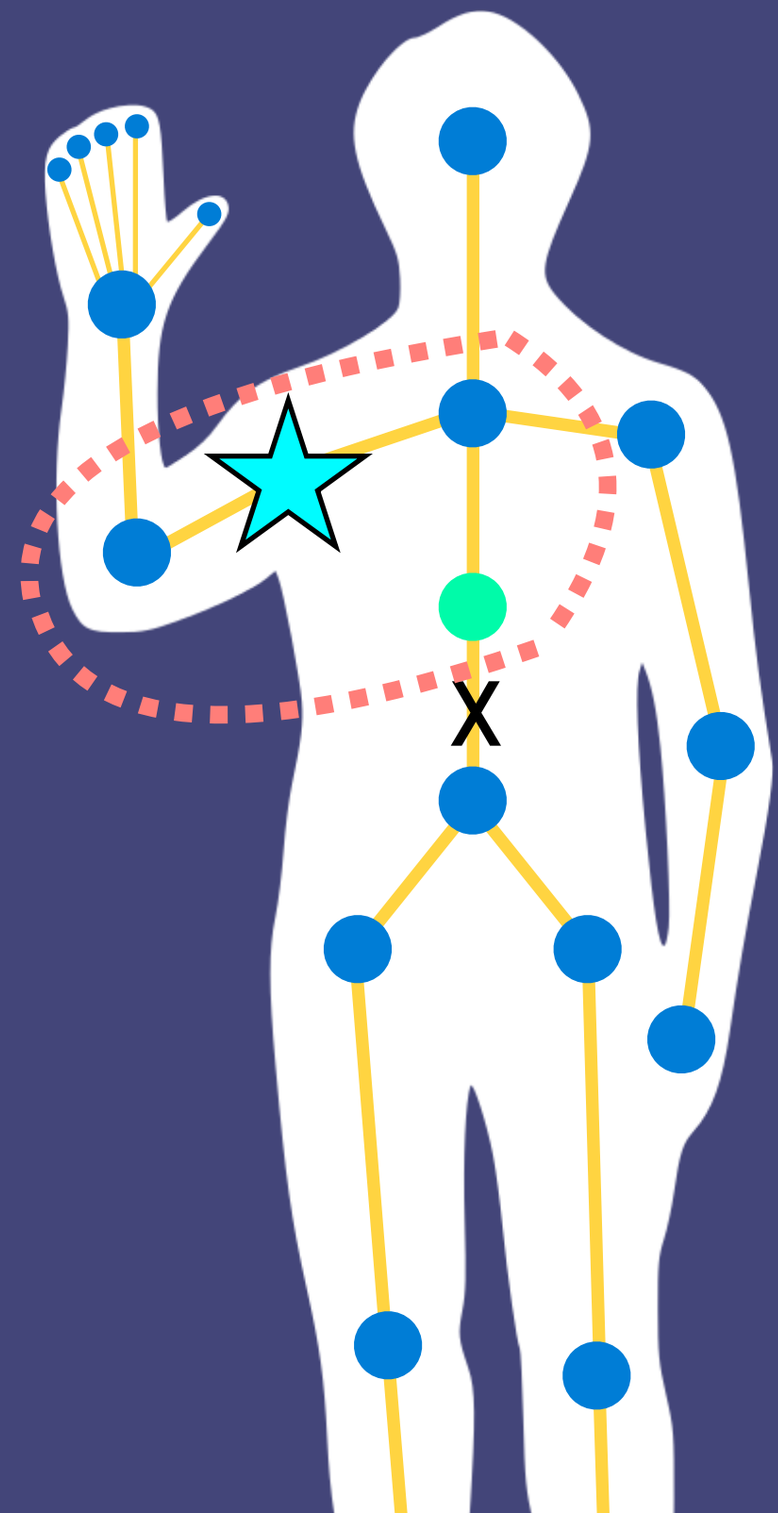


Spatial Configuration Partitioning

Root Node

Centripetal

Centrifugal

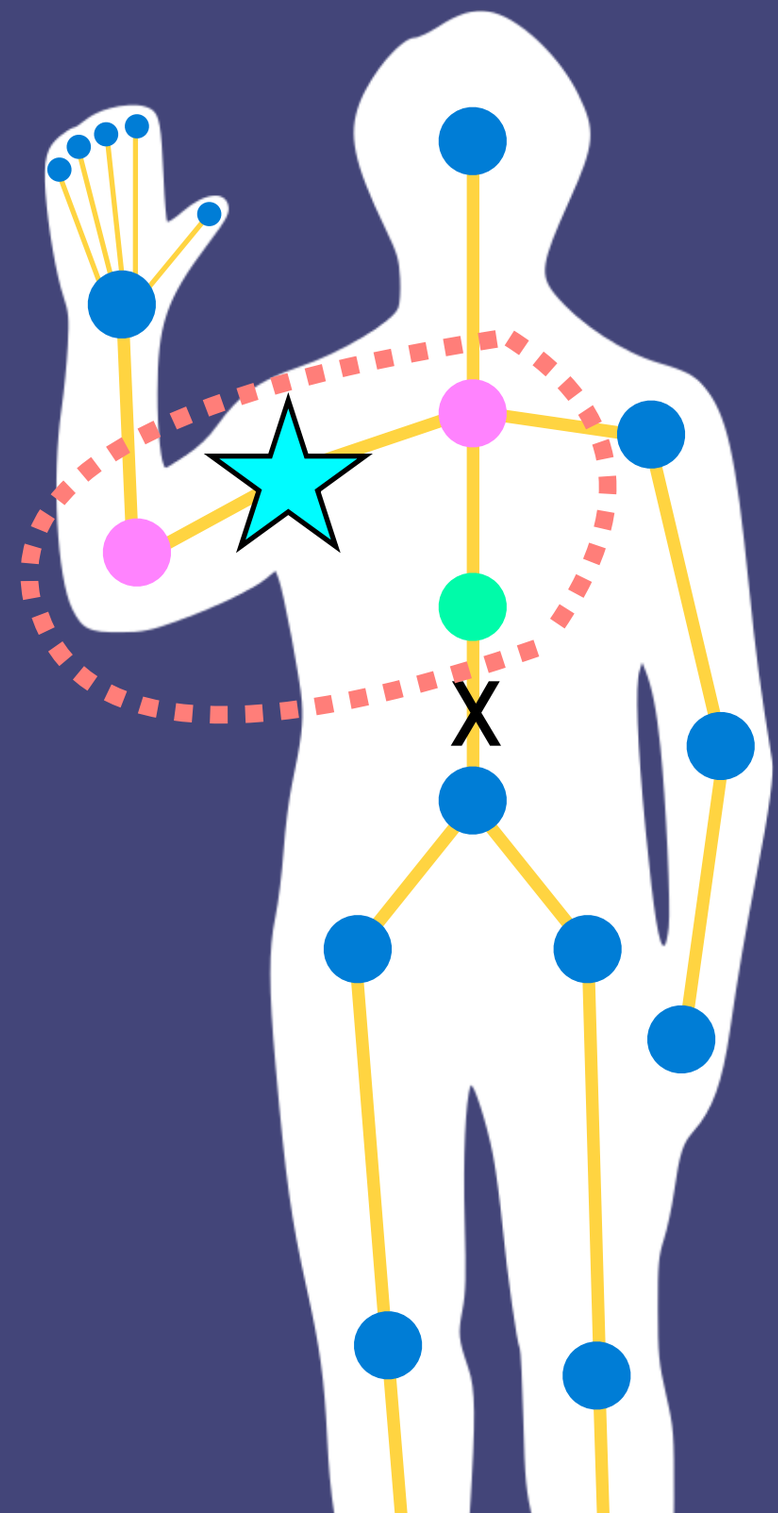


Spatial Configuration Partitioning

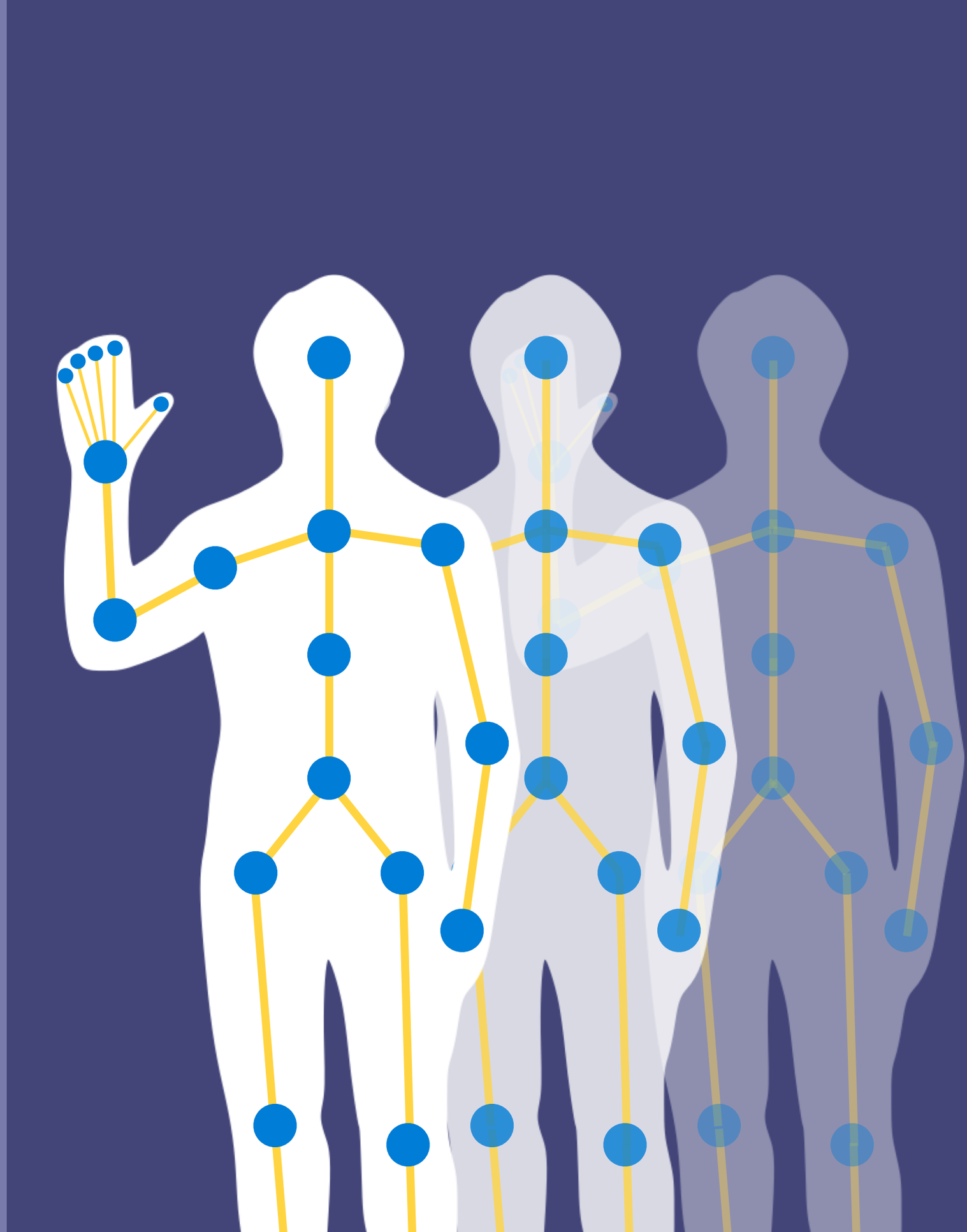
Root Node

Centripetal

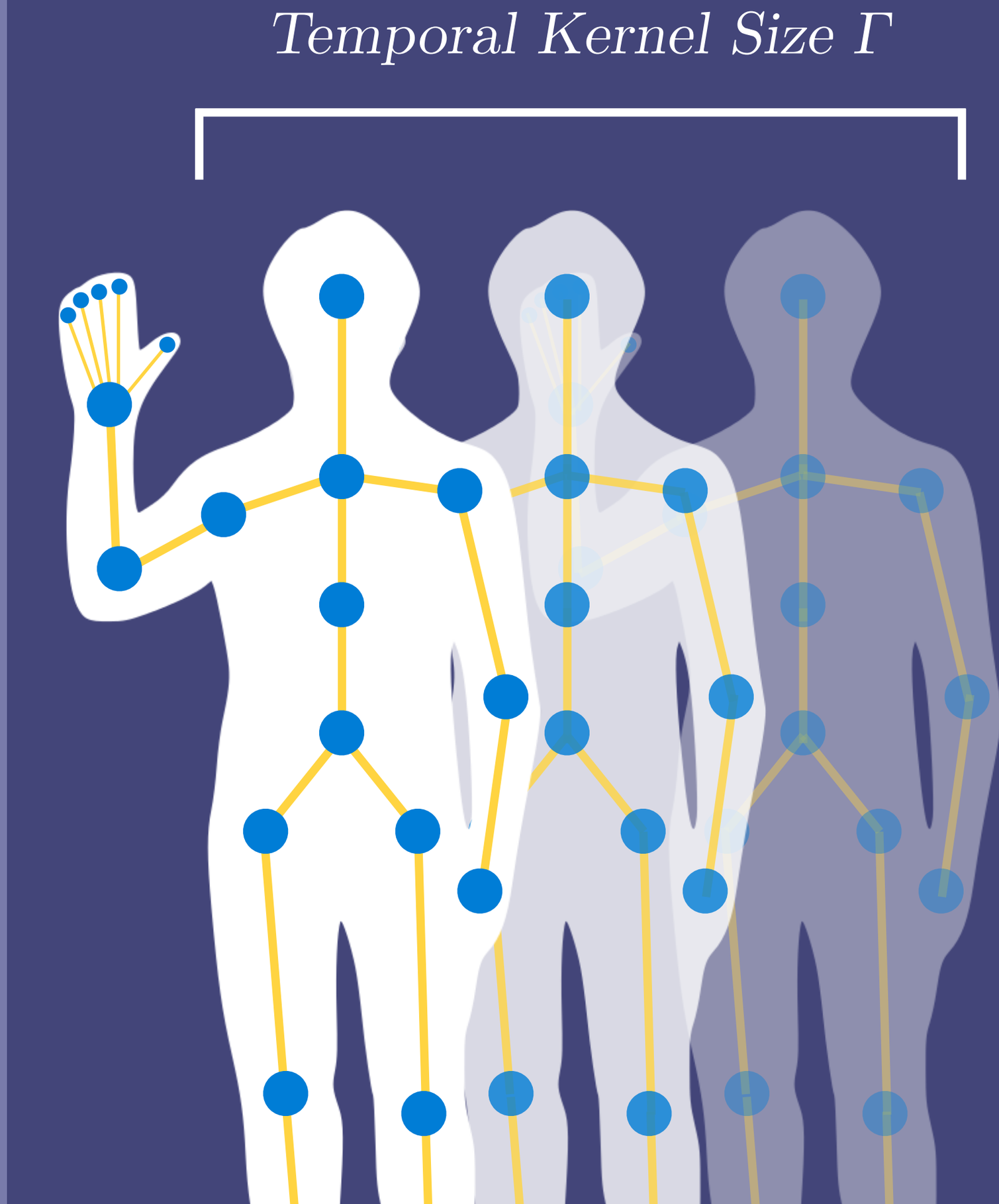
Centrifugal



Temporal Convolution



Temporal Convolution



Performance

NTU-RGB+D

Kinetics

ST-GCN

C-CNN + MLTN

Temporal Conv

ST-LSTM + TS

PA-LSTM

Deep-LSTM

H-RNN

Lie Group

NTU-RGB+D

Kinetics

Two-Stream: RGB + Flow

ST-GCN

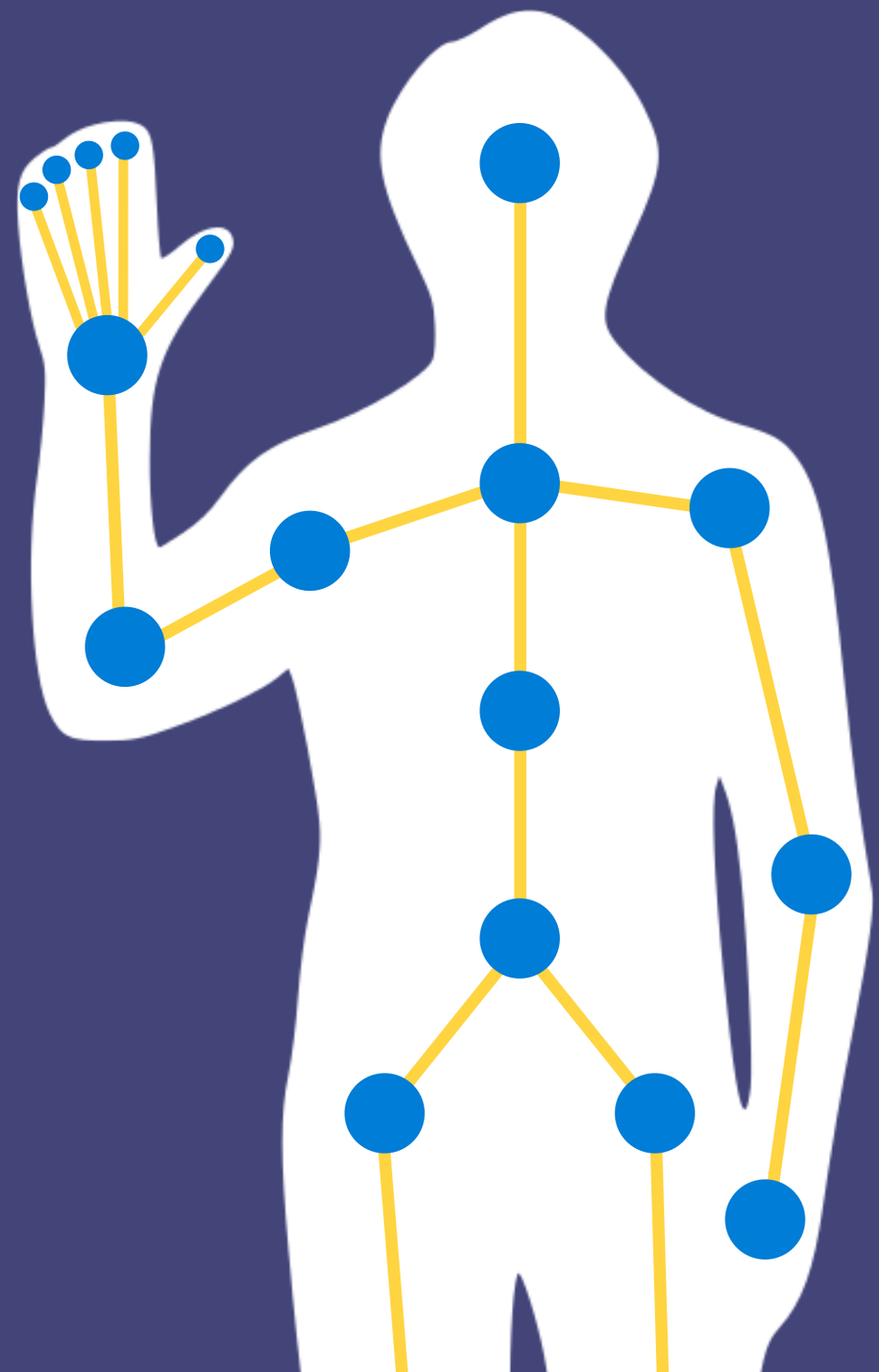
Temporal Conv

Deep LTSM

VideoDarwin

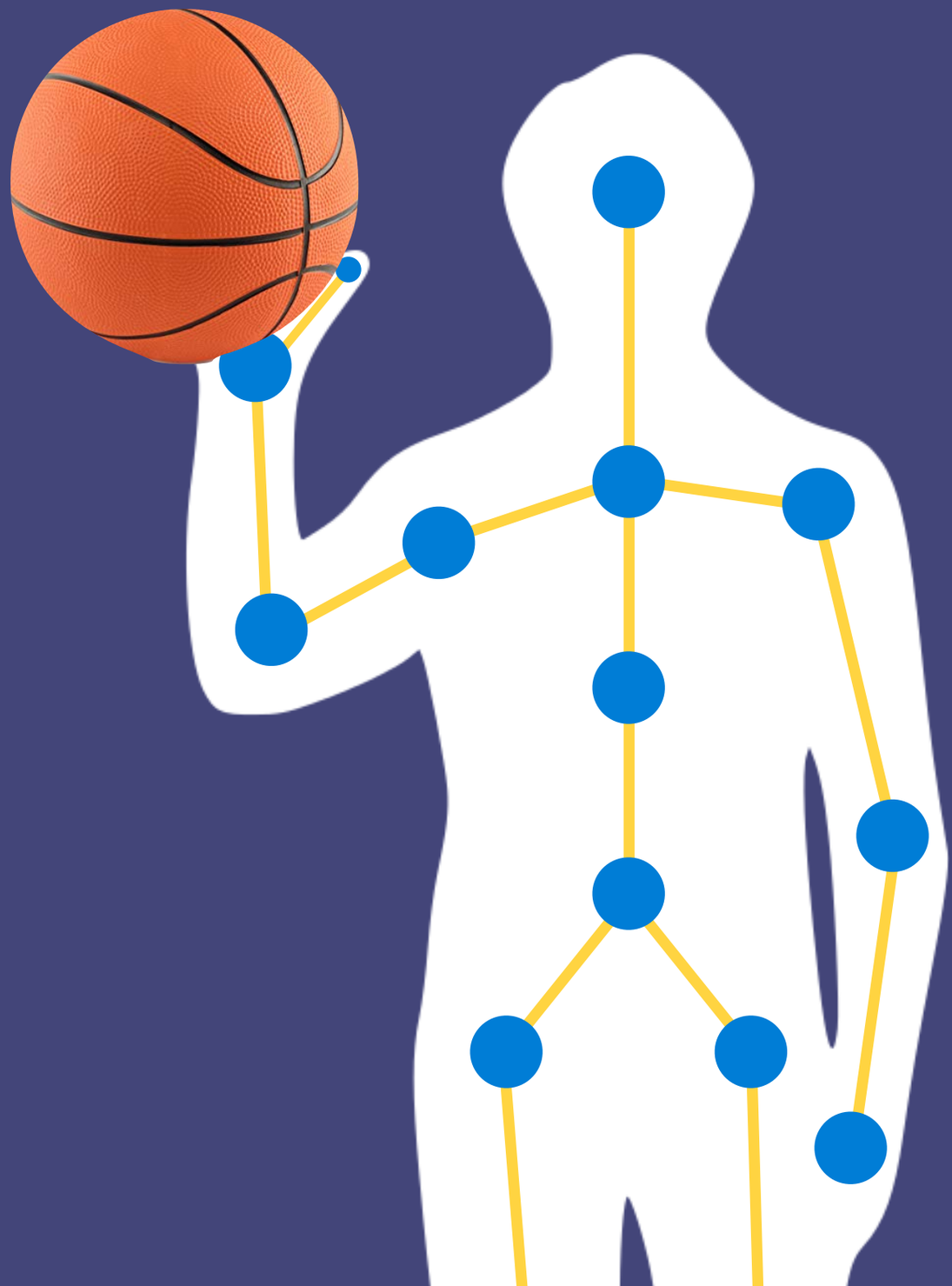
Heterogeneous Graphs

Depth of Model



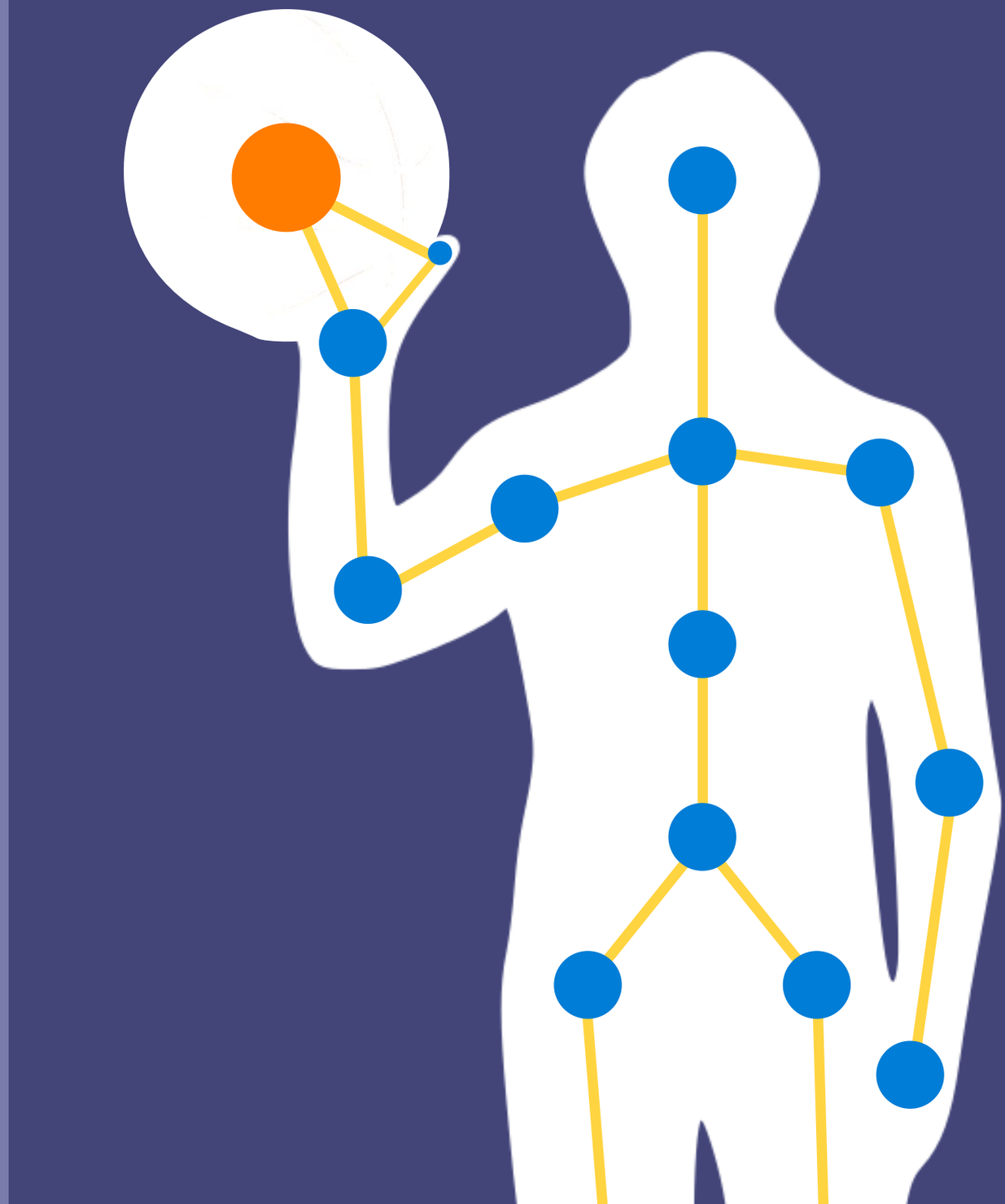
Heterogeneous Graphs

Depth of Model



Heterogeneous Graphs

Depth of Model



Heterogeneous Graphs

Depth of Model

Thank you