

Object Permanence Emerges in a Random Walk along Memory



Pavel Tokmakov



Allan Jabri



Jie Li



Adrien Gaidon

Object Permanence



LA-CATER

[Shamsian et al., ECCV'20]

CenterTrack Framework



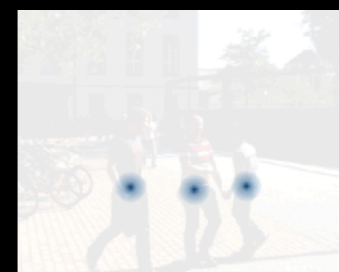
Frame t



Frame t-1



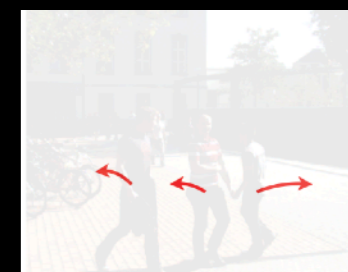
Tracks t-1



Detections t



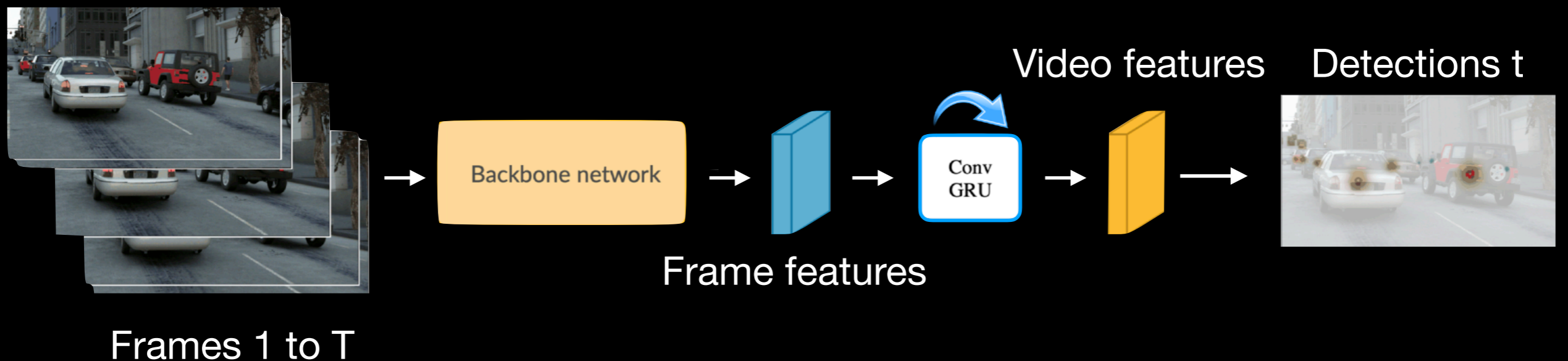
Boxes t



Offsets t -> t-1

[Zhou et al., ECCV'20]

Spatio-Temporal Object Representation



Supervising the Invisible in Sim



Annotations for a synthetic video

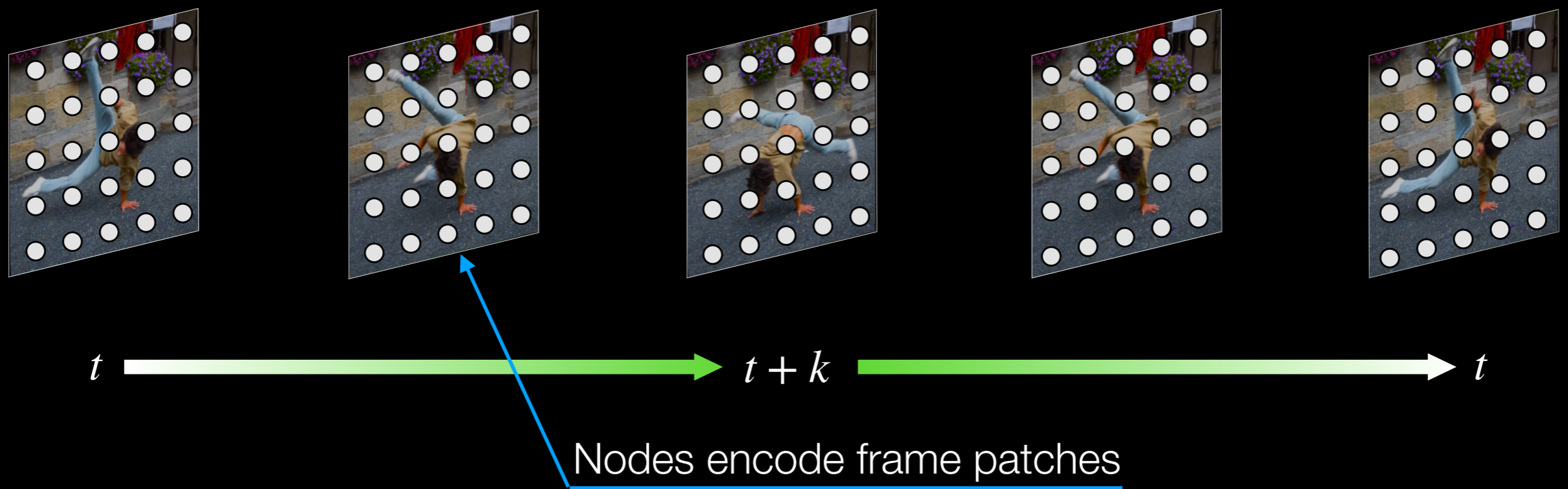
[Tokmakov et al., ICCV'21]

Videos as Random Graphs



[Jabri et al., NeurIPS'20]

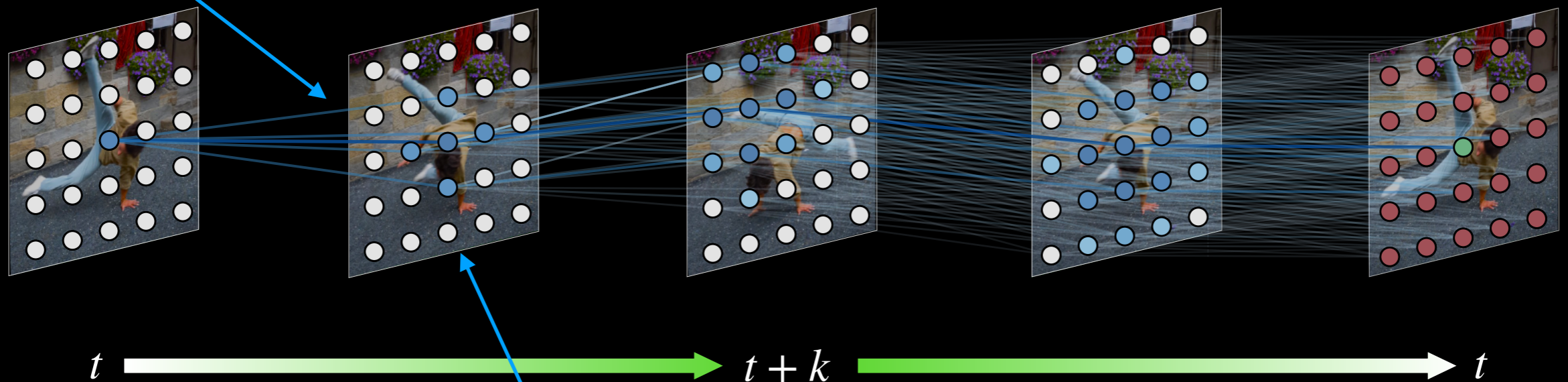
Videos as Random Graphs



[Jabri et al., NeurIPS'20]

Videos as Random Graphs

Tracking a point is a walk on the graph



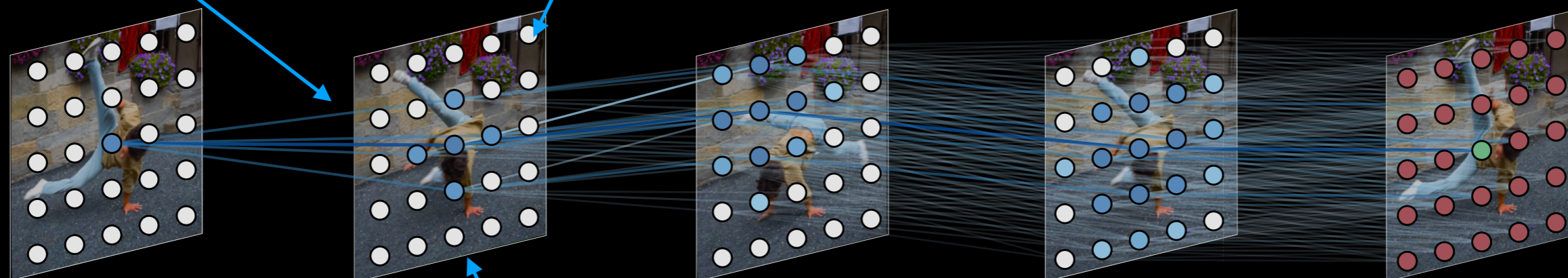
Nodes encode frame patches

[Jabri et al., NeurIPS'20]

Videos as Random Graphs

Tracking a point is a walk on the graph

Walker state represents belief about the target location



t  $t+k$  t

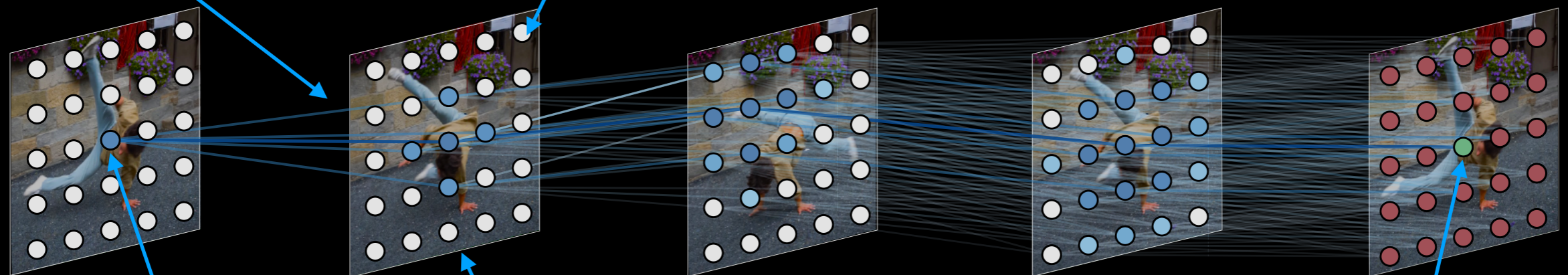
Nodes encode frame patches

[Jabri et al., NeurIPS'20]

Videos as Random Graphs

Tracking a point is a walk on the graph

Walker state represents belief about the target location



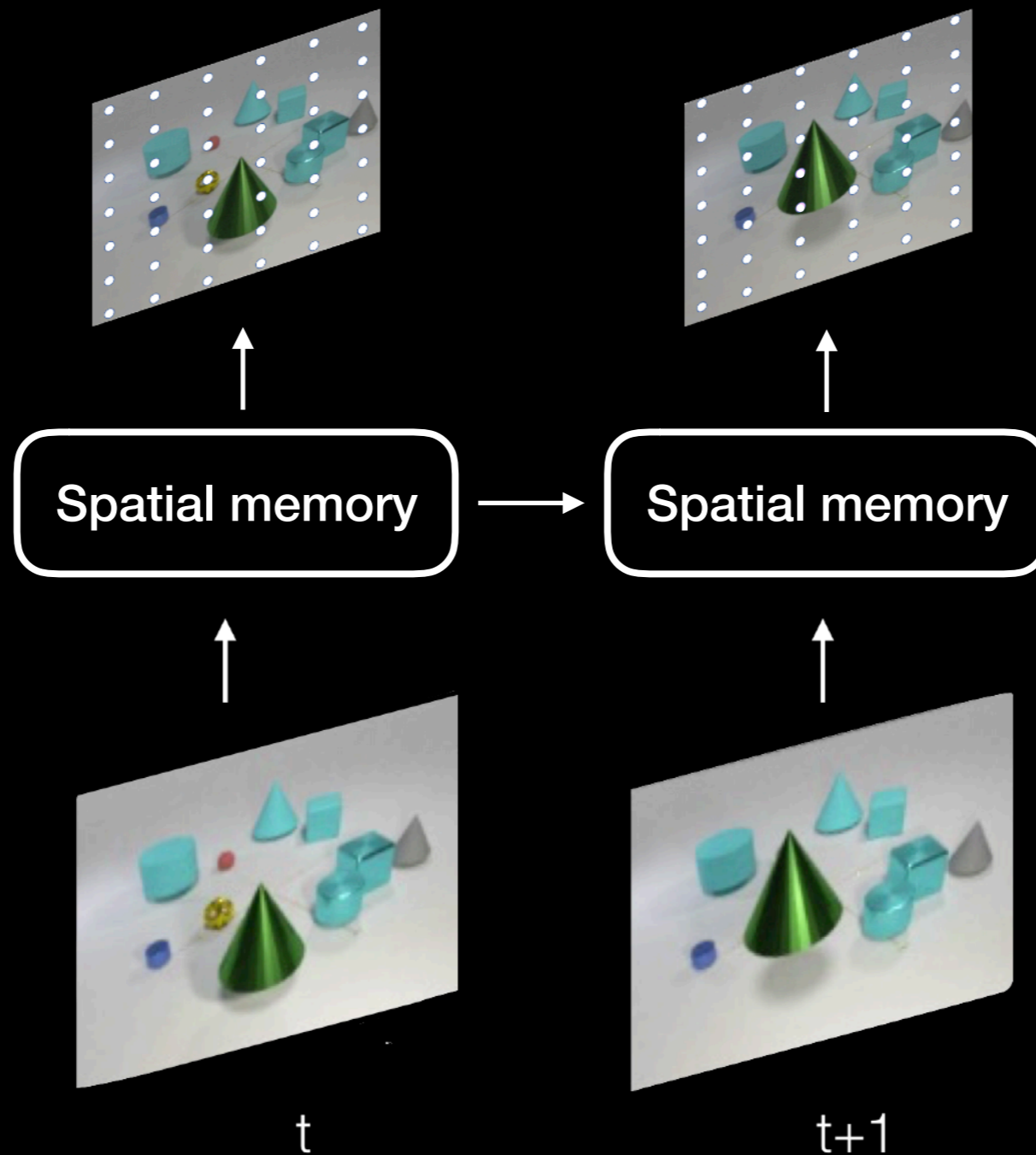
t $\xrightarrow{\text{green arrow}}$ $t+k$ $\xrightarrow{\text{green arrow}}$ t

Nodes encode frame patches

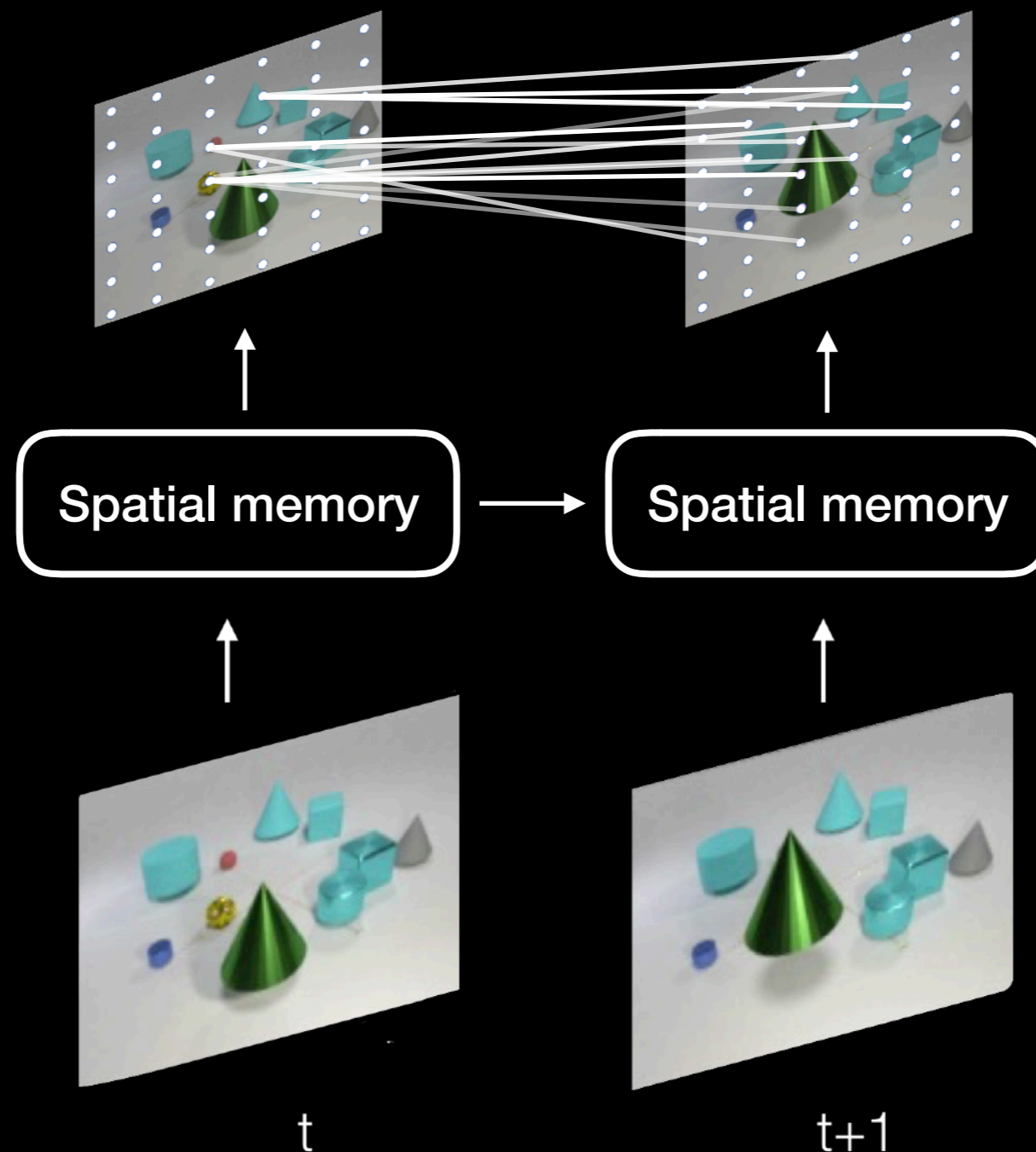
Only sparse supervision is required (start and end points)

[Jabri et al., NeurIPS'20]

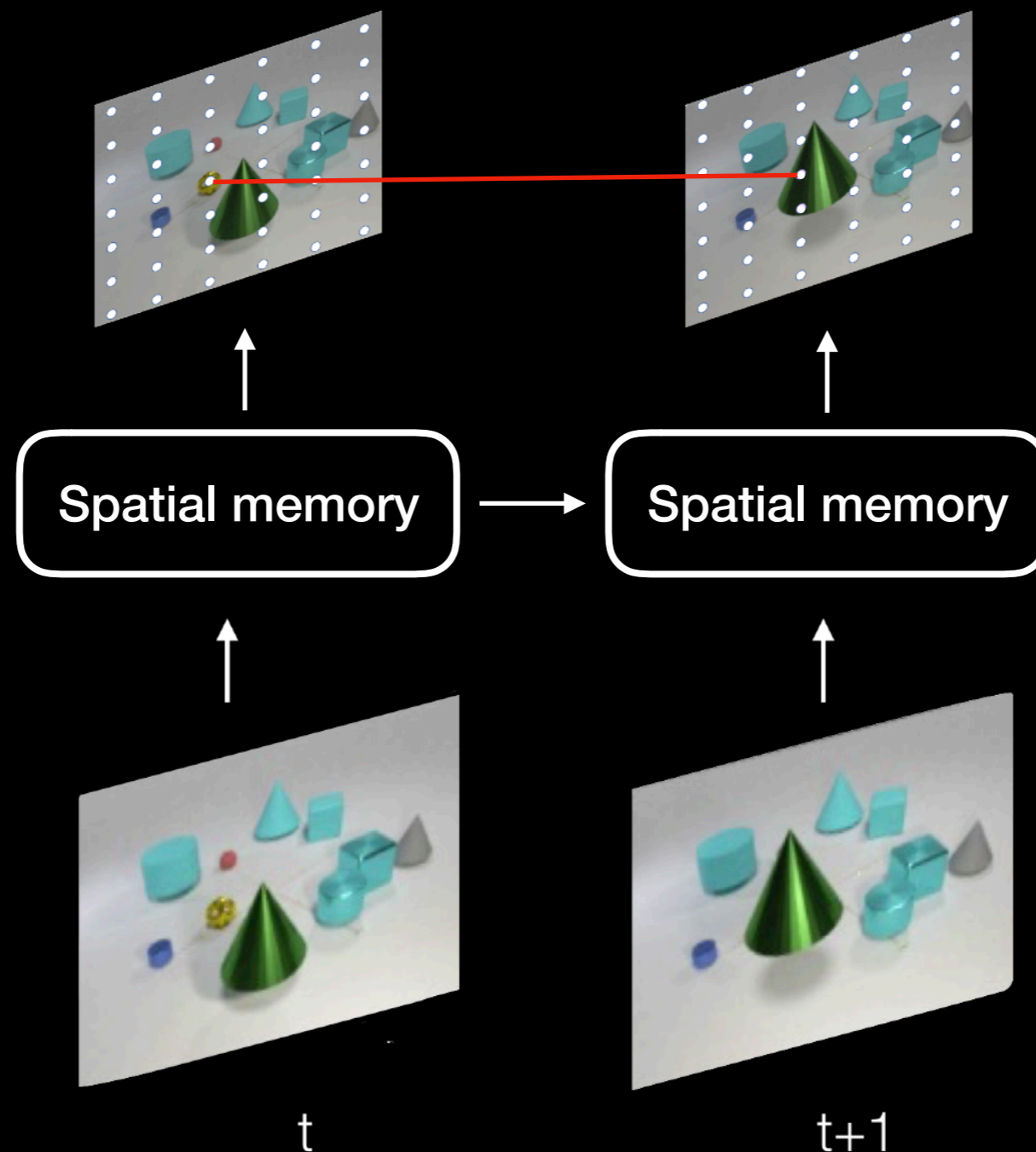
Random Walk Along Memory (RAM)



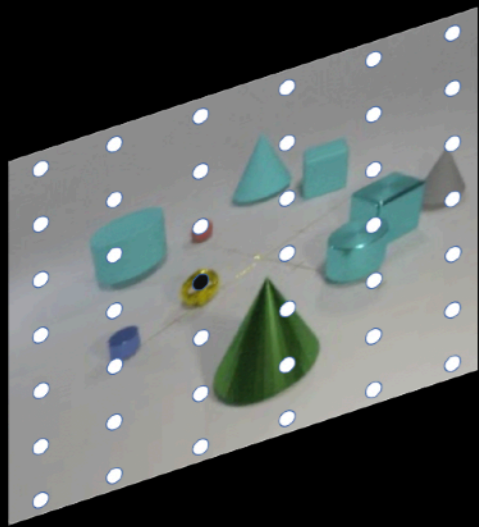
Random Walk Along Memory (RAM)



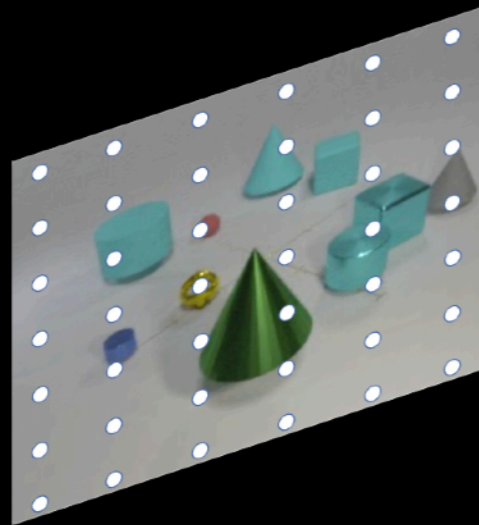
Random Walk Along Memory (RAM)



Fitting the Walk

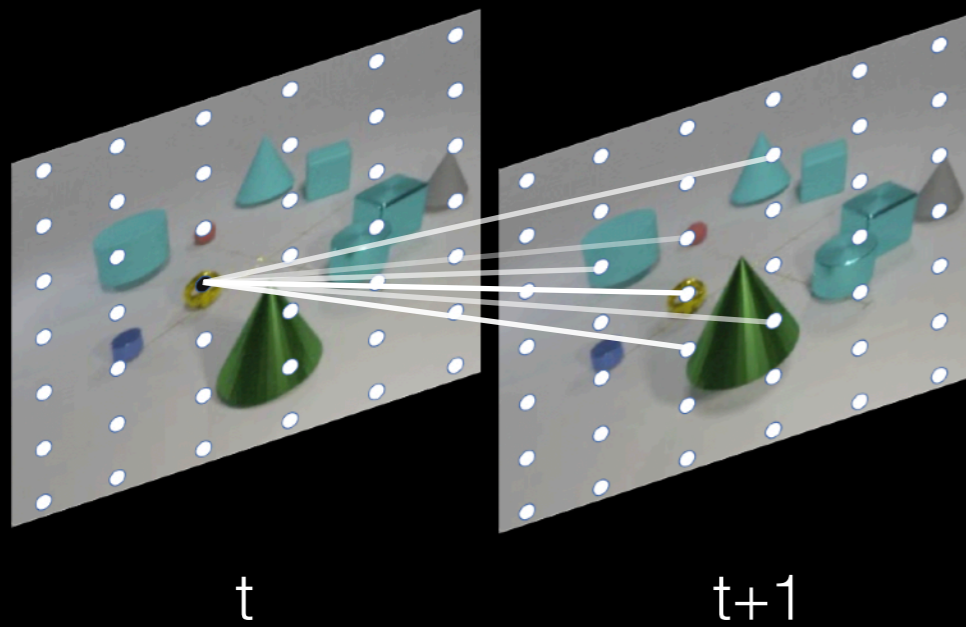


t

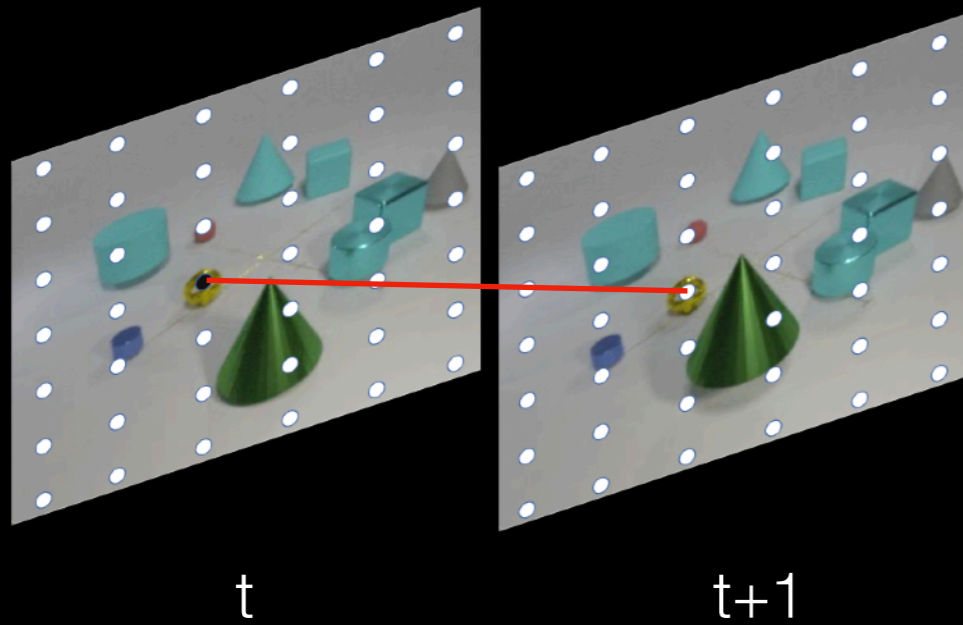


$t+1$

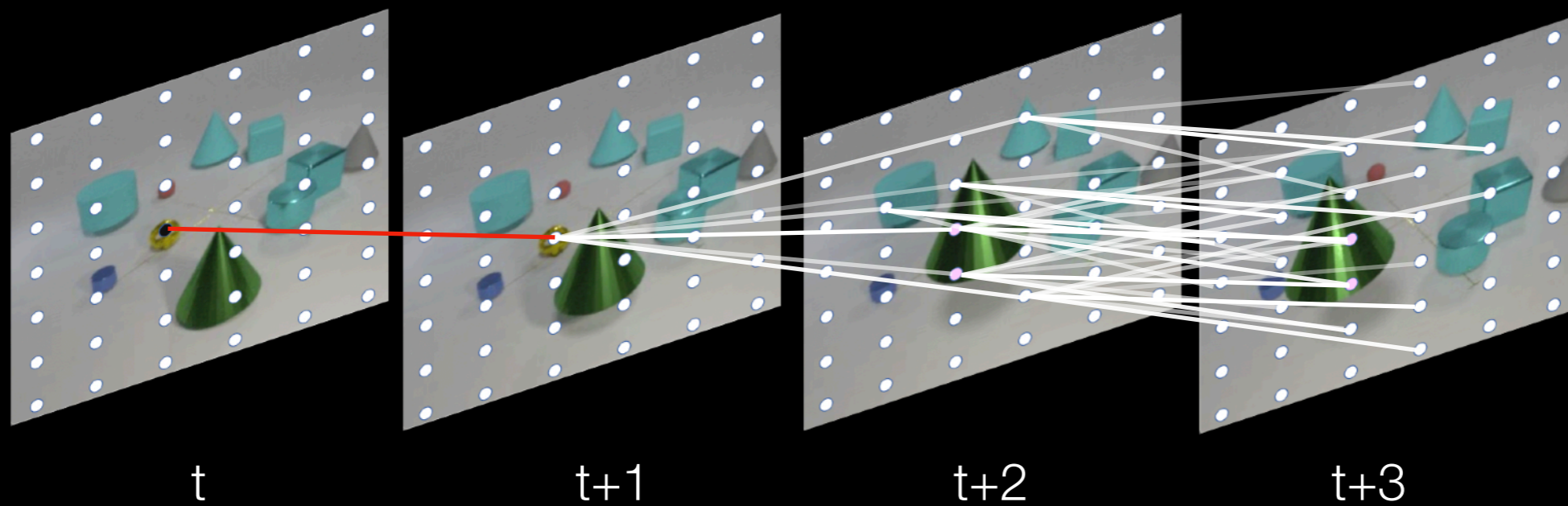
Fitting the Walk



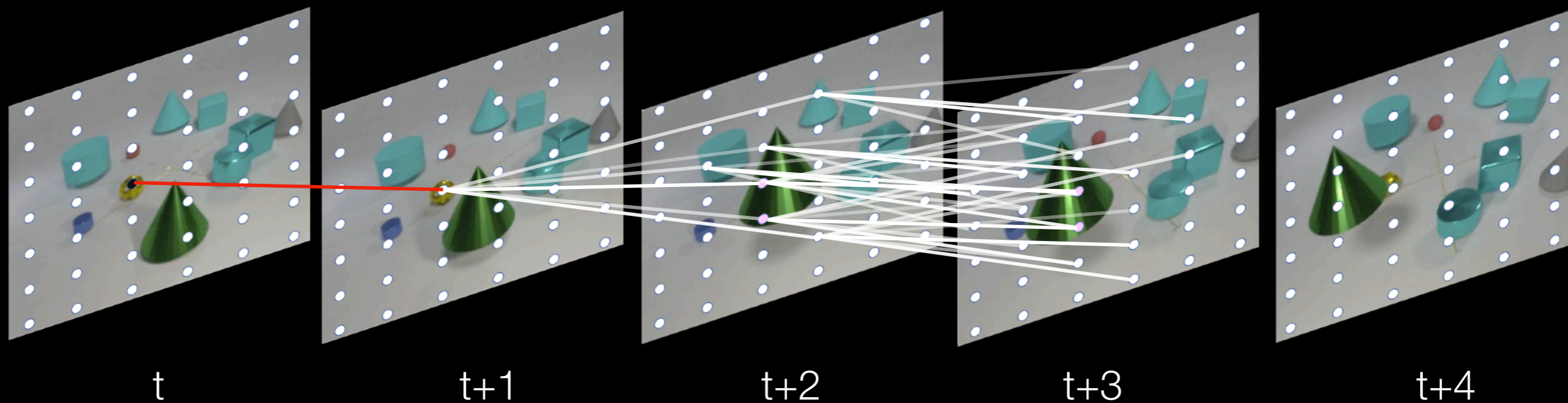
Fitting the Walk



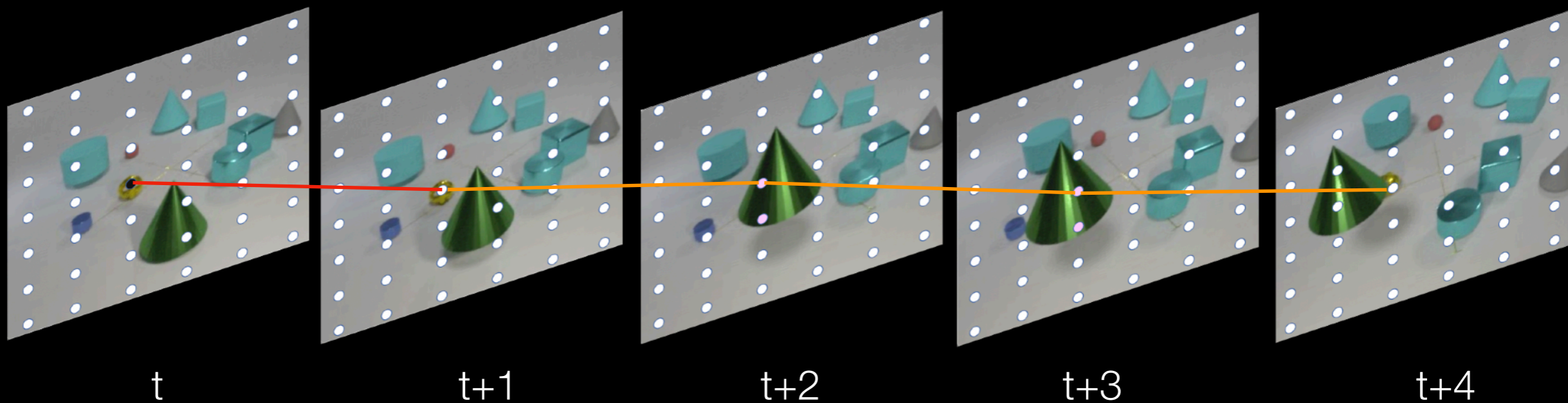
Fitting the Walk



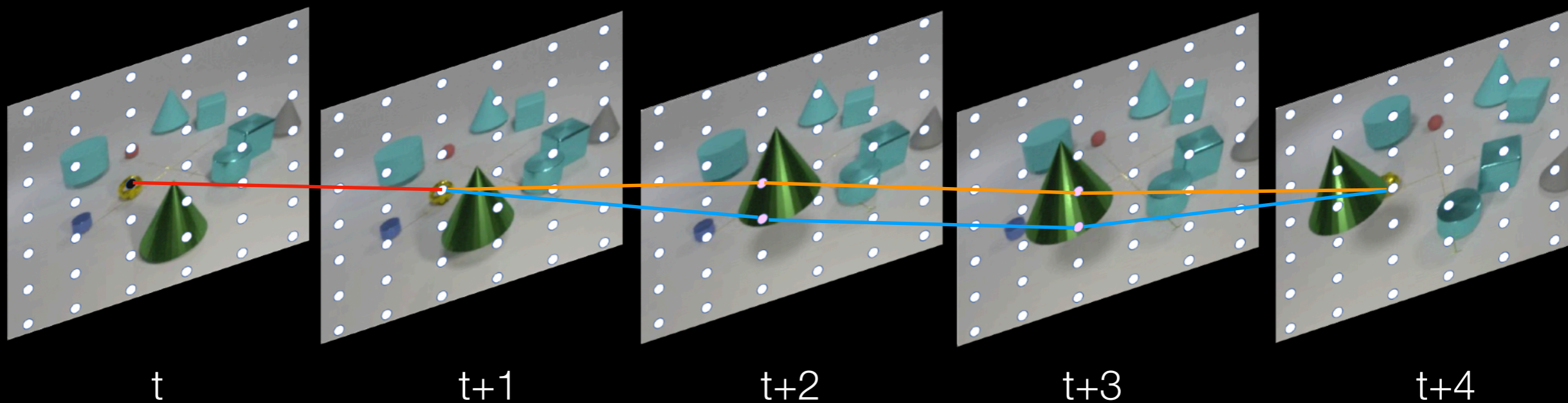
Fitting the Walk



Fitting the Walk



Fitting the Walk



Emerging Object Permanence



Object Permanence in the Real World

