







**Method - Architecture**



→ process videos per-frame, using guidance from the previous frame,

as in MaskTrack [Khoreva et al., CVPR'17]

→ it can be trained using static images only

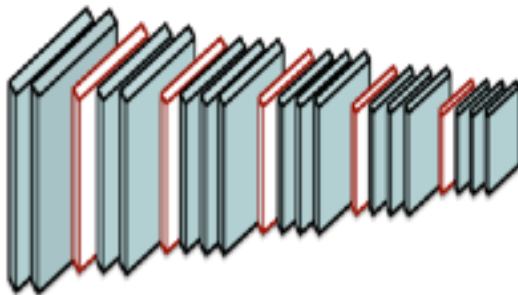
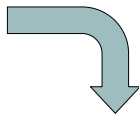
DeeperCut

ResNet-101

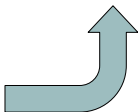
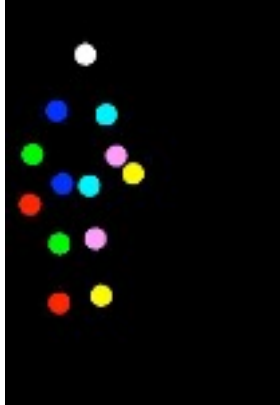




Frame  $t$   
input



Frame  $t$   
output pose



Frame ***t-1***  
predicted pose

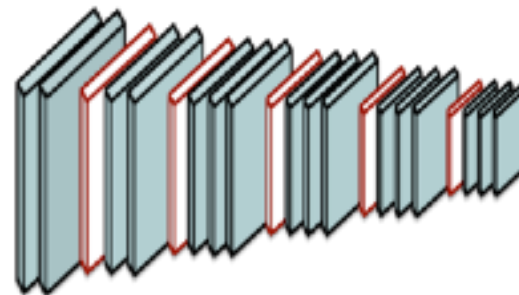
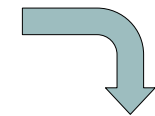
# Method - Architecture



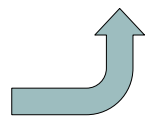
Frame  $t$   
input



Frame  $t-1$   
predicted pose



*HumanPoseTracker*



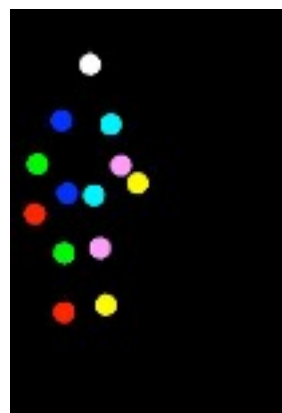
Frame  $t$   
output pose

# Method - Architecture

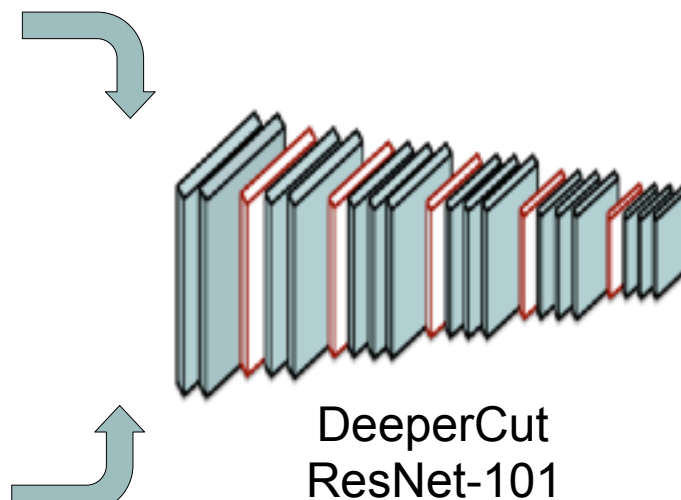
→ process videos per-frame, using guidance from the previous frame,  
as in MaskTrack [Khoreva et al., CVPR'17]



Frame  $t$   
input



Frame  $t-1$   
predicted pose



Frame  $t$   
output pose

→ it can be trained using **static images** only