

Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image

Denis Tomè

Chris Russell

Lourdes Agapito

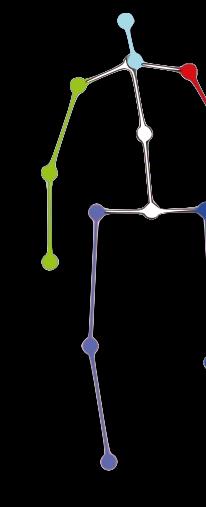
We introduce a novel approach to solve the problem of

3D human pose estimation

from a single RGB image



Input image

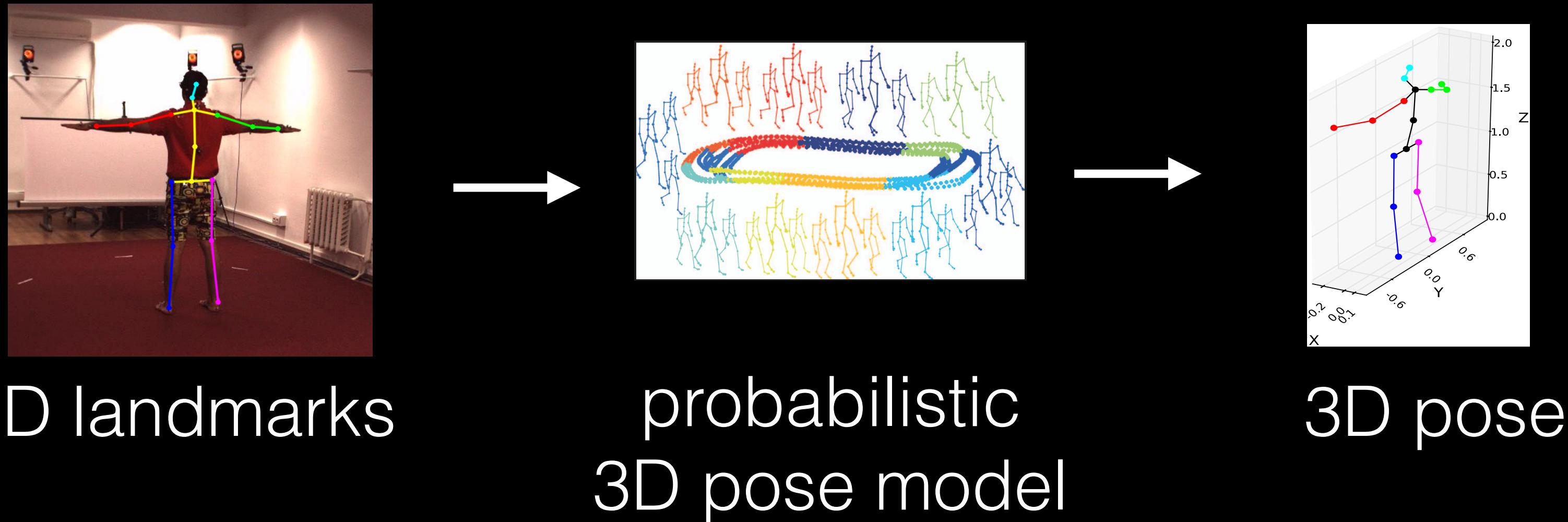


Output 3D Pose

Our method reasons **jointly** about 2D joint estimation
and 3D pose reconstruction to **improve both tasks.**

Our approach

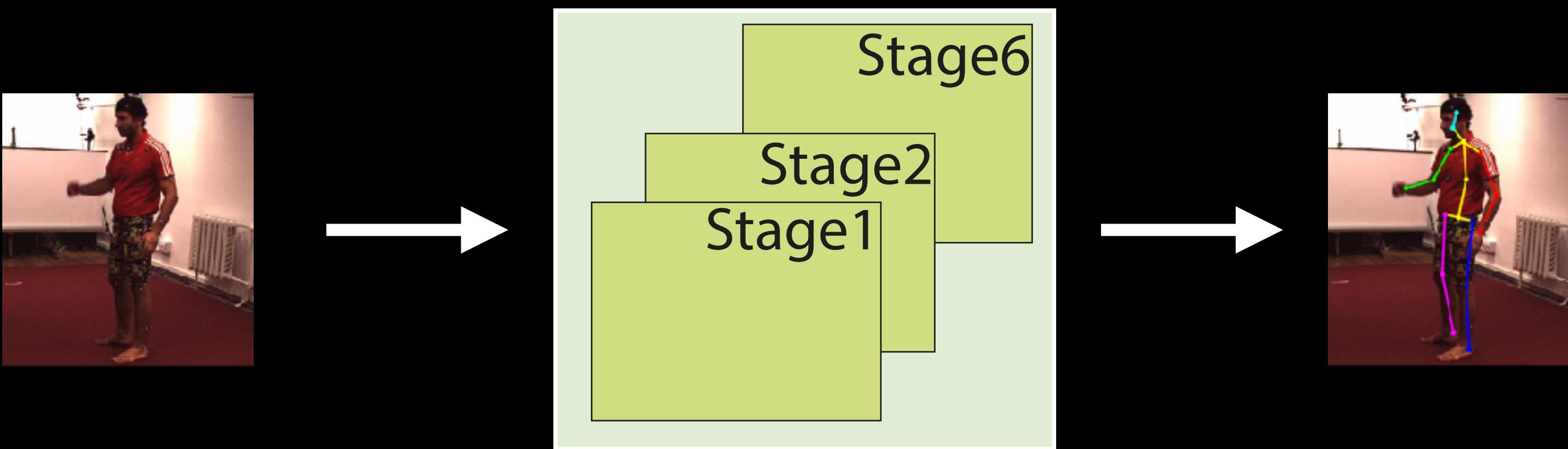
- First, we learn a **probabilistic model** of 3D human pose from 3D mocap data



This model lifts 2D joint positions (**landmarks**) into 3D

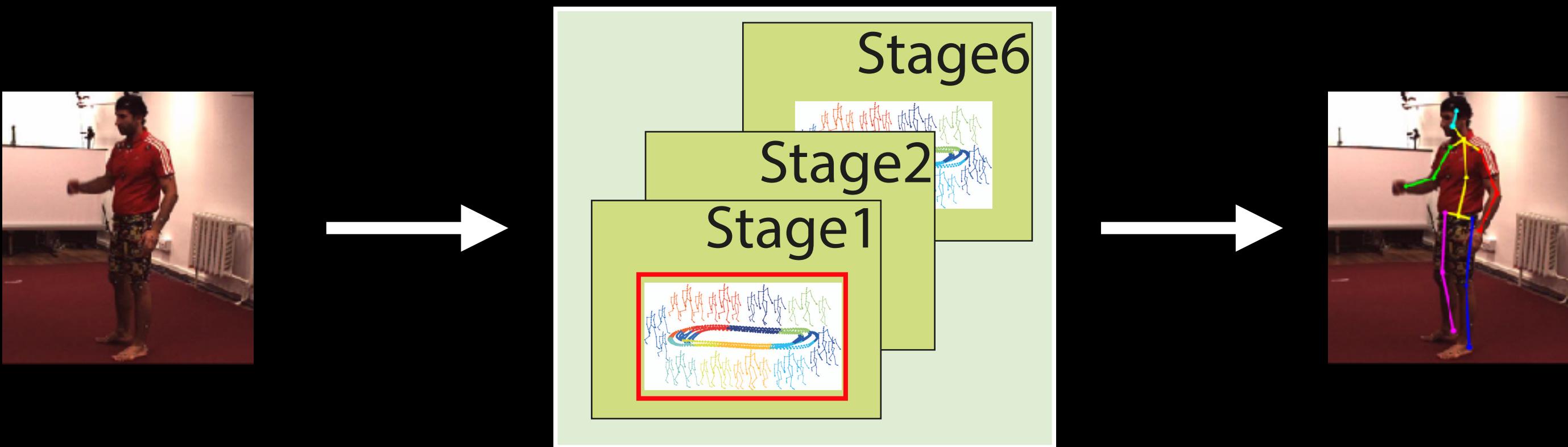
Our approach

- Next, we train a novel end-to-end multi-stage CNN for 2D landmark estimation



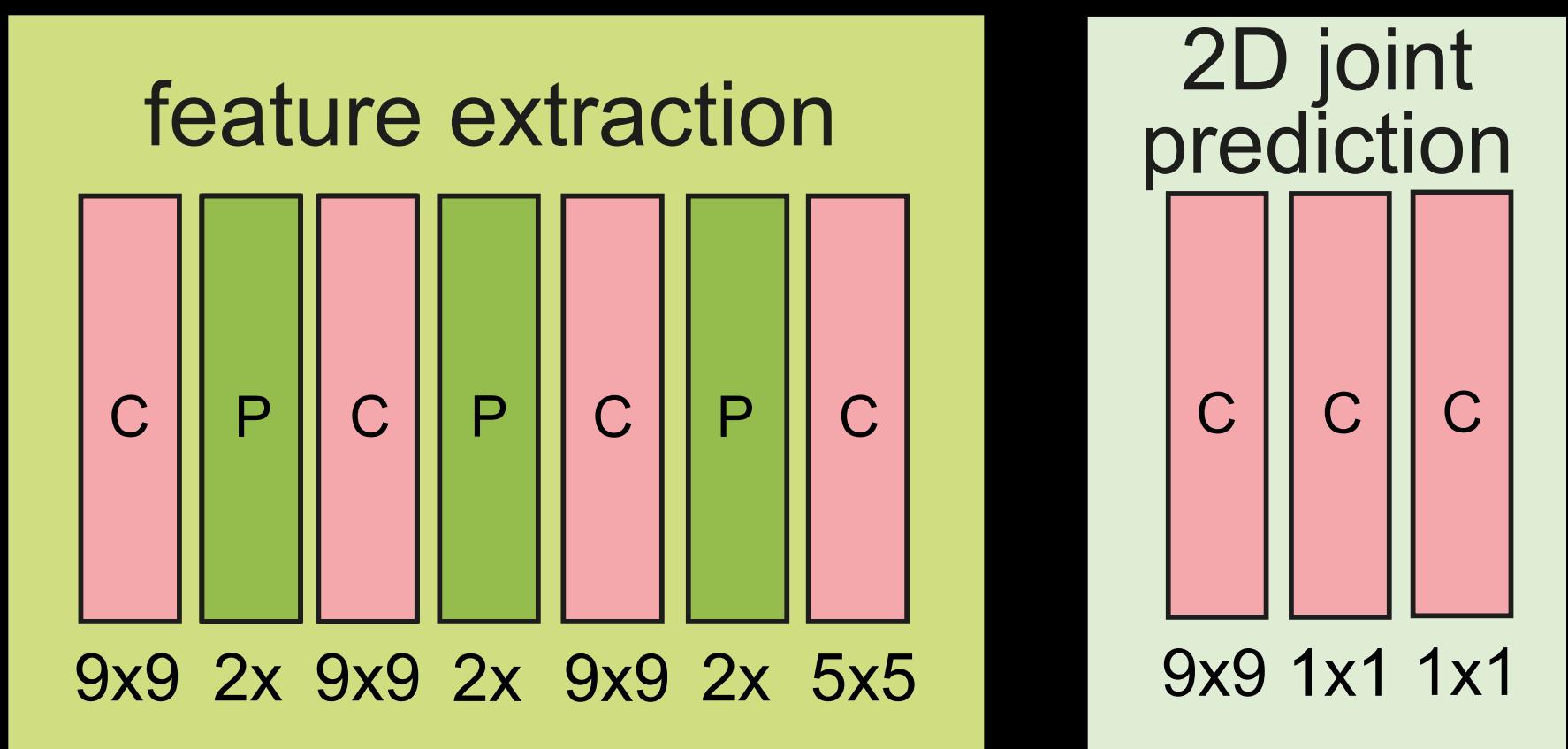
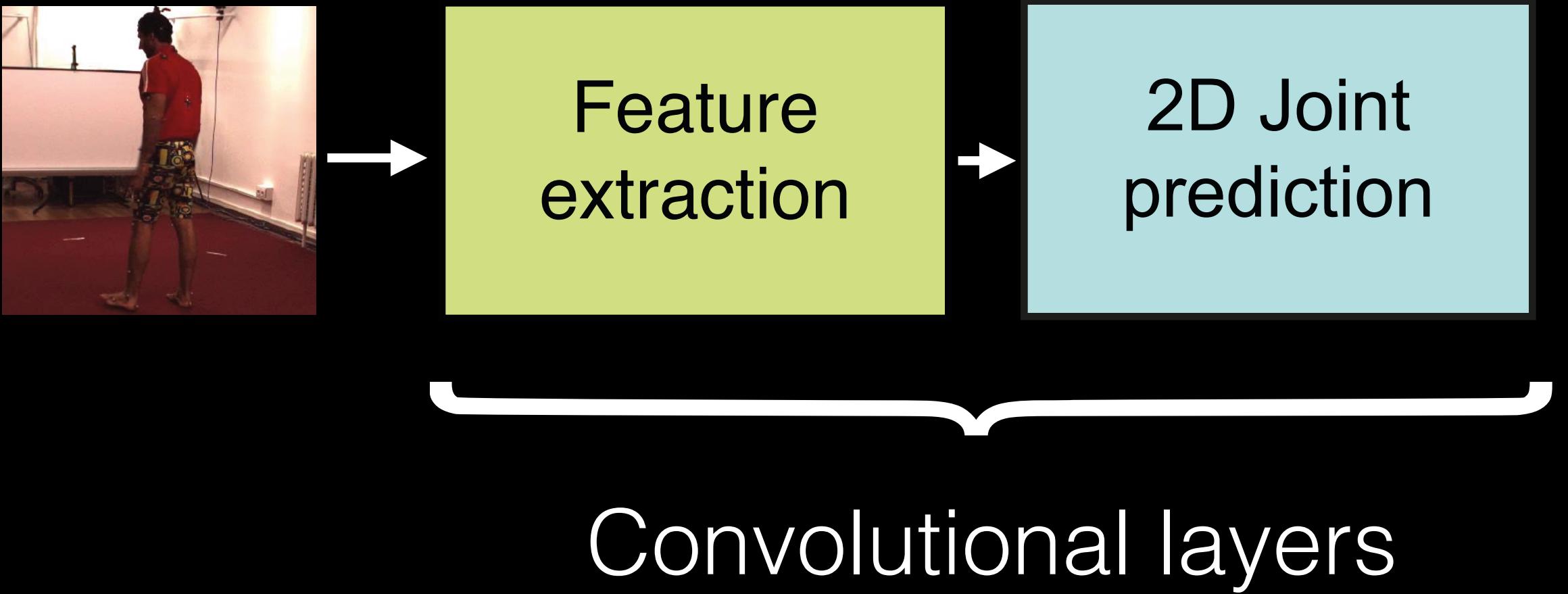
Our approach

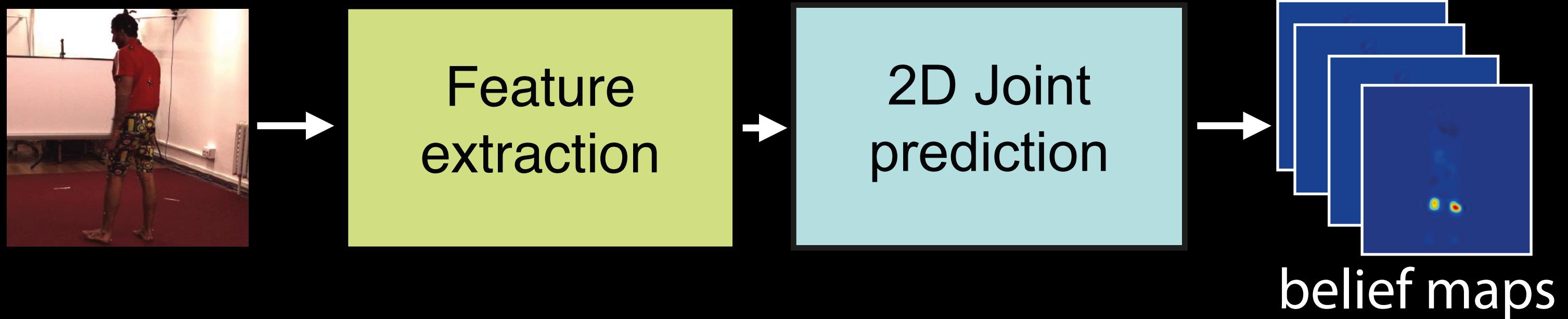
- Next, we train a novel end-to-end multi-stage CNN for 2D landmark estimation



- Each stage includes a new layer based on our **probabilistic 3D pose model** of human poses to enforce 3D pose constraints

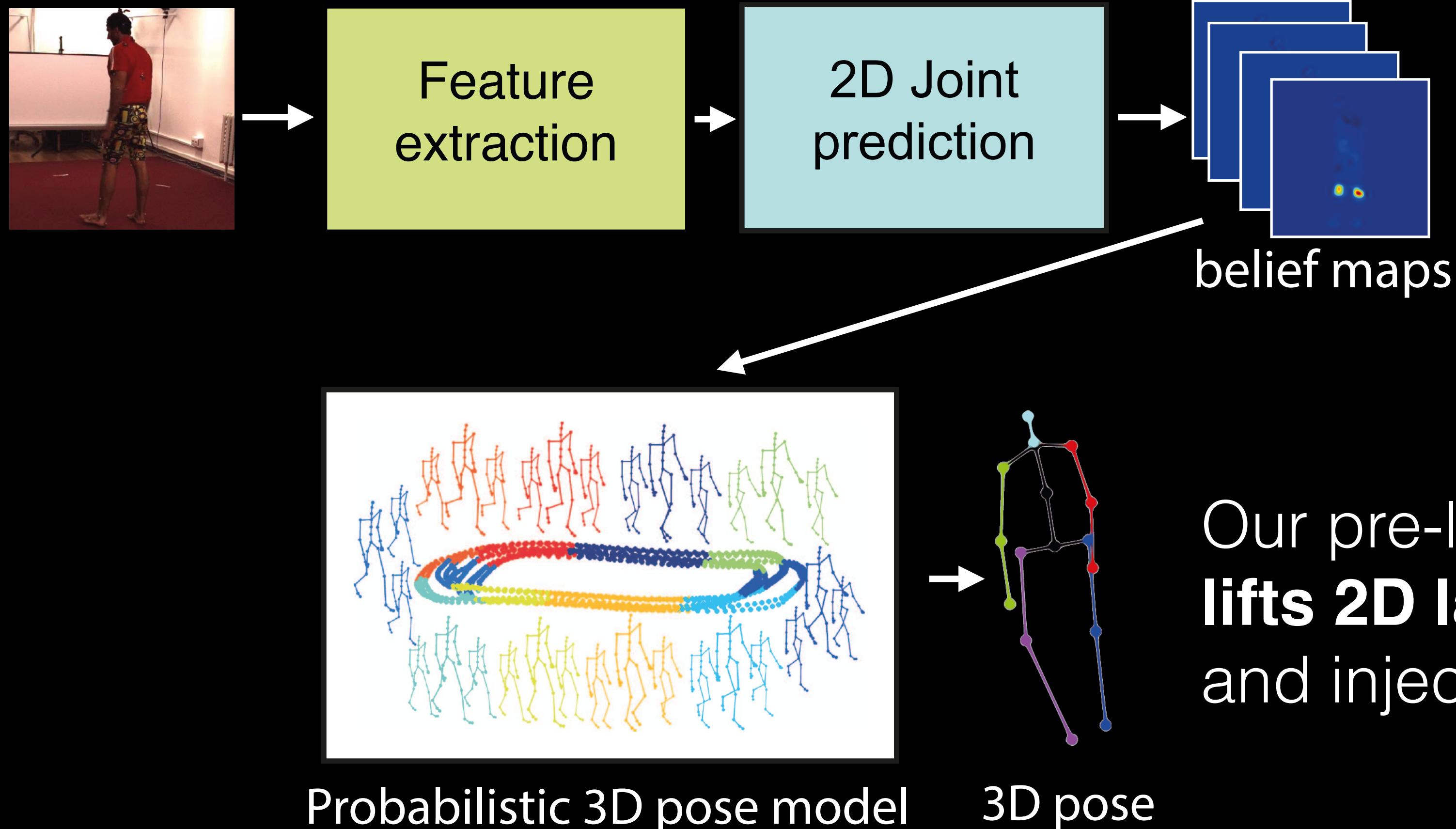
Detailed architecture



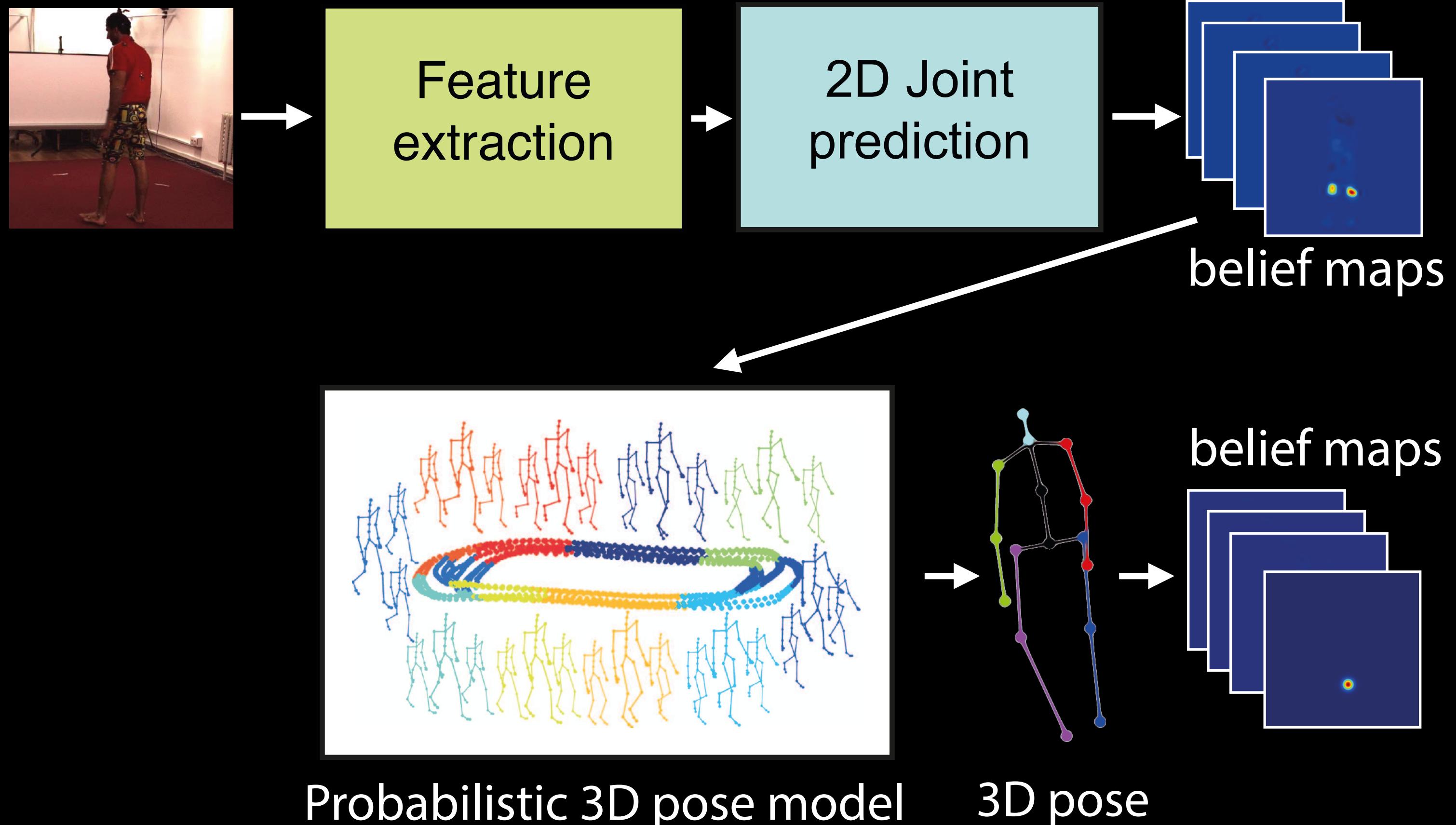


For each landmark, a **2D belief map** is generated

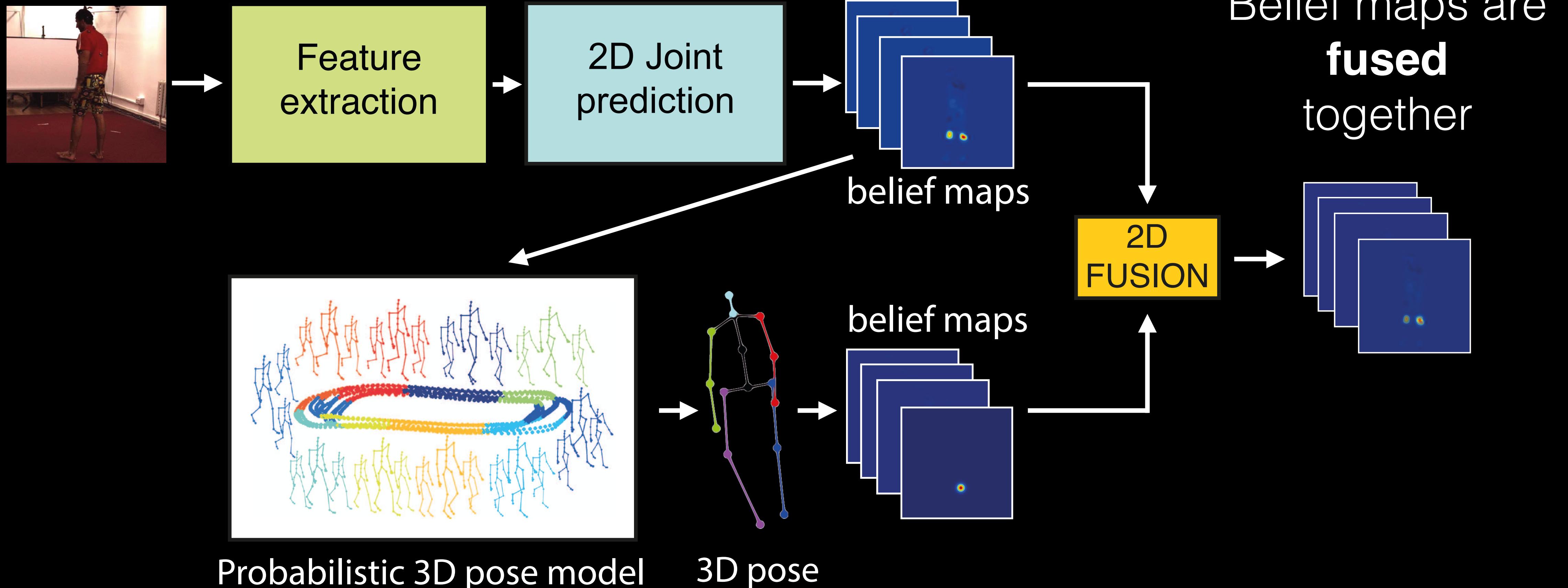
This defines how confident the architecture is that a specific landmark occurs at any given pixel (u, v) of the input image

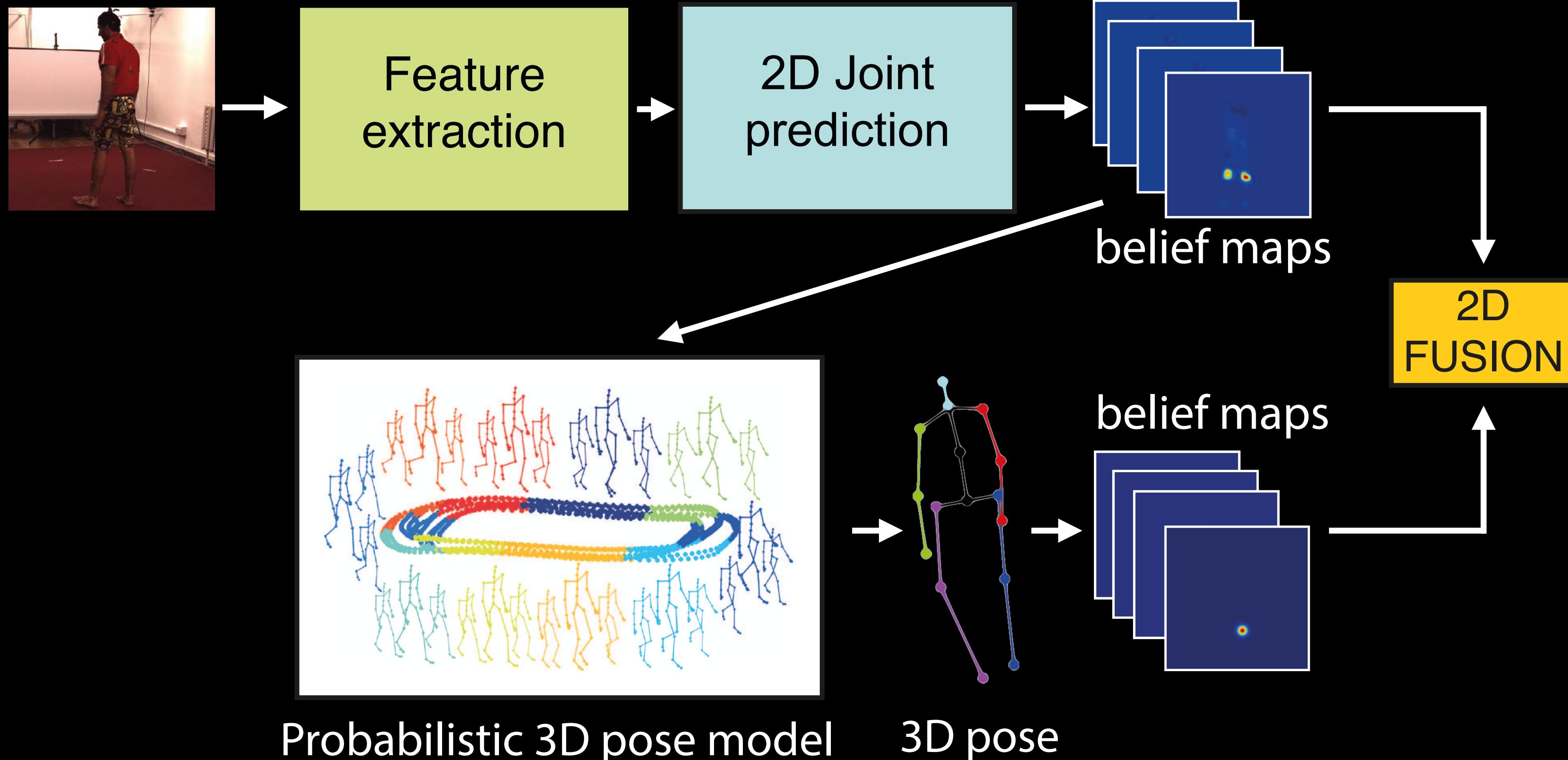


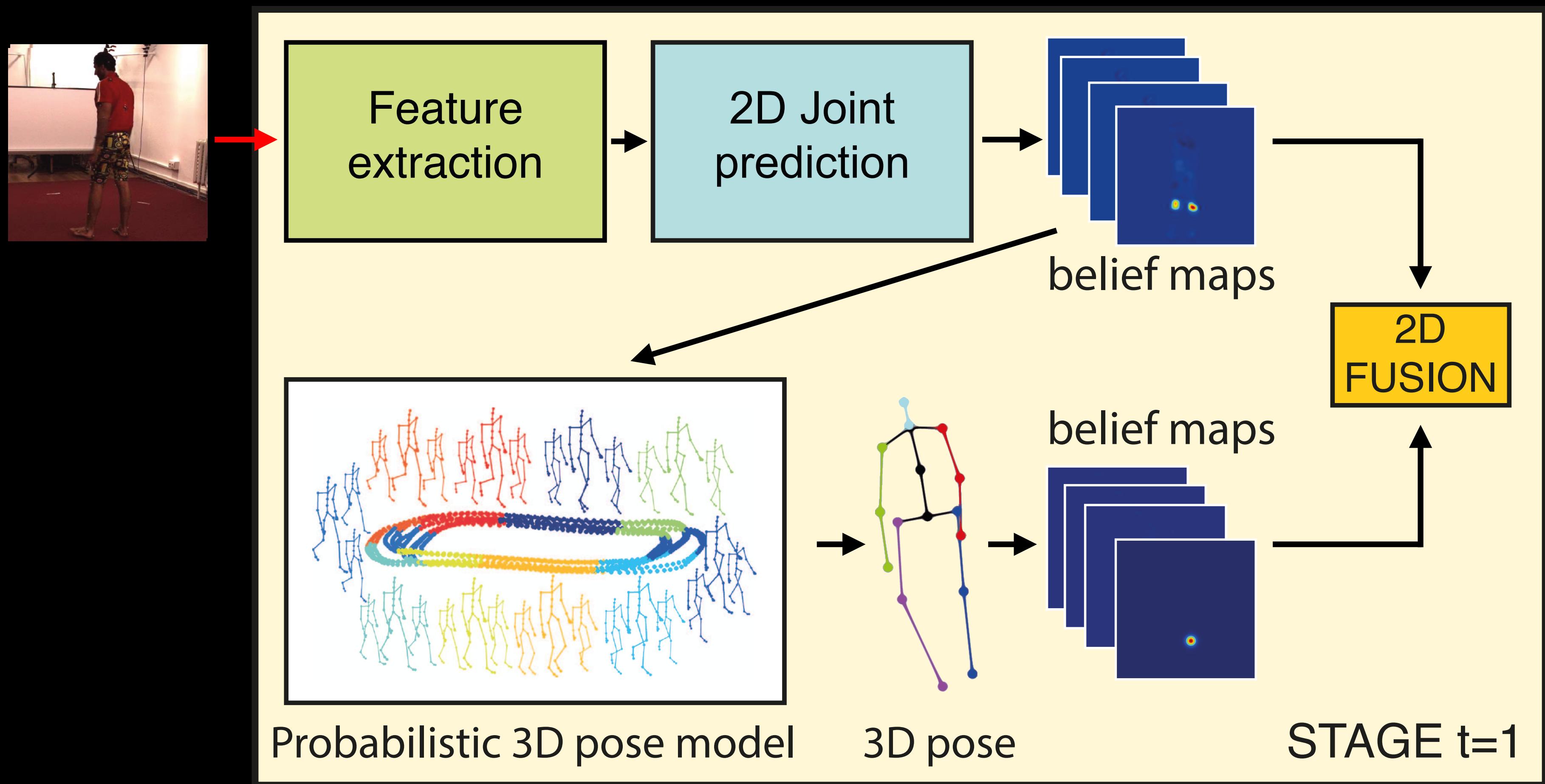
Our pre-learned probabilistic model
lifts 2D landmarks into 3D
and injects 3D pose information



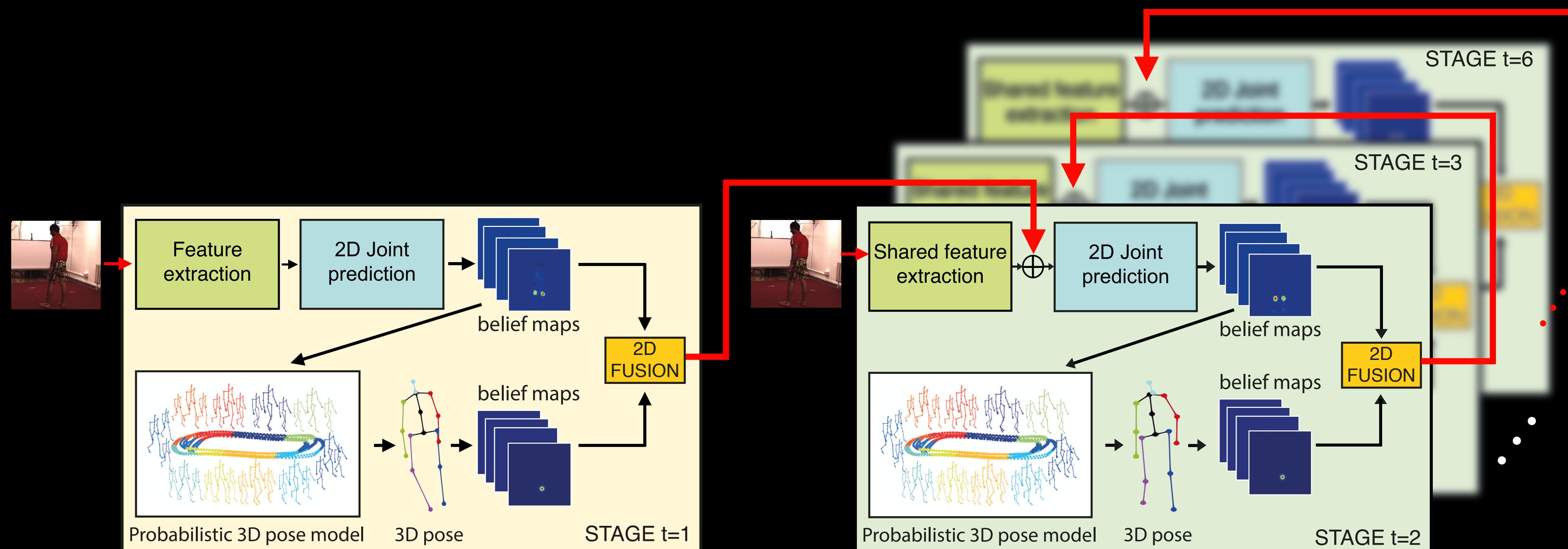
The 3D pose is used to generate a new set of **2D belief maps**



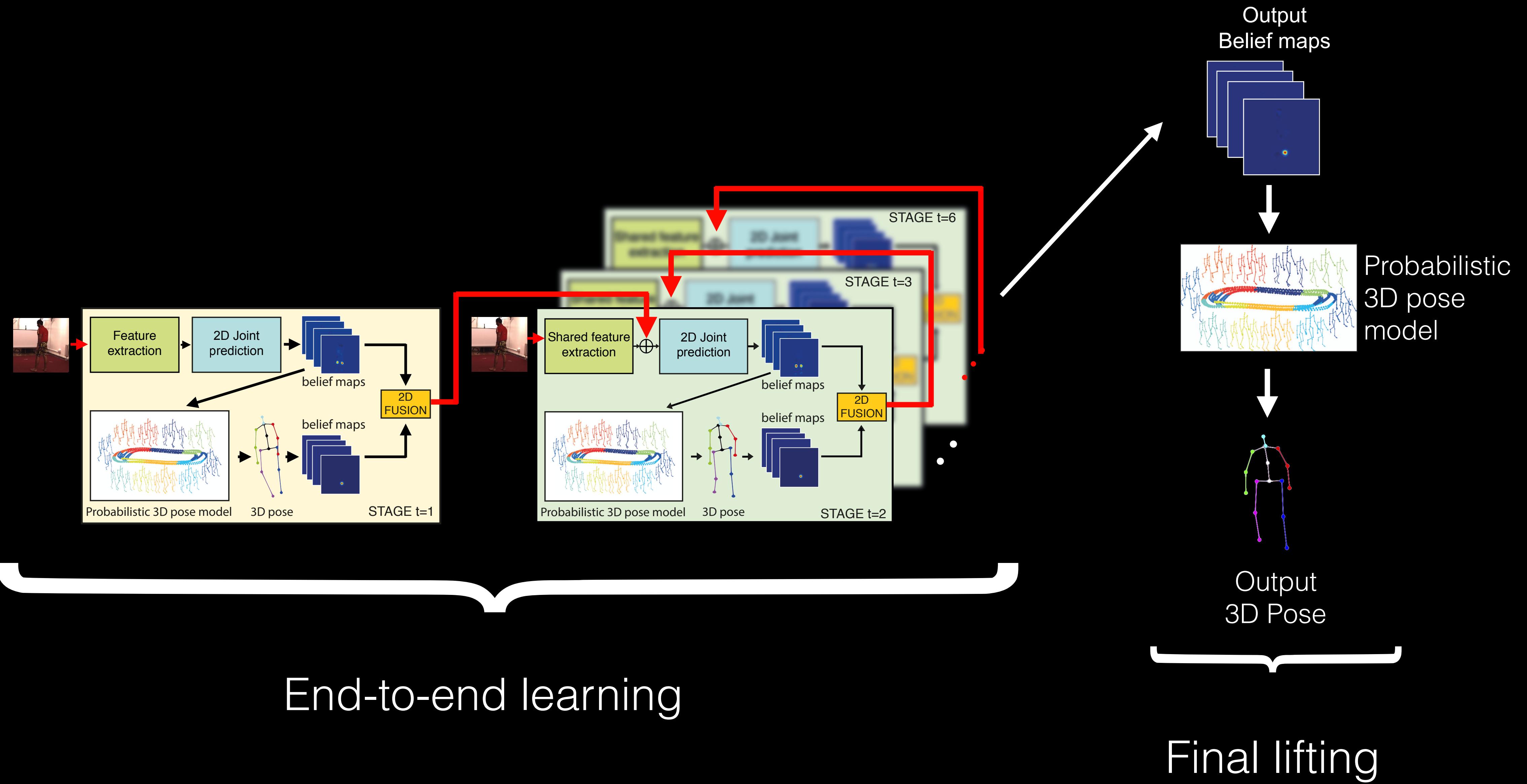




The **2D belief maps** from each stage are used as input to the next **stage**



The accuracy of the belief maps **increases progressively** through the stages



Our approach achieves state-of-the-art results on the Human3.6M dataset

	Directions	Discussion	Eating	Greeting	Phoning	Photo	Posing	Purchases
LinKDE [14]	132.71	183.55	132.37	164.39	162.12	205.94	150.61	171.31
Li <i>et al.</i> [19]	-	136.88	96.94	124.74	-	168.68	-	-
Tekin <i>et al.</i> [32]	102.39	158.52	87.95	126.83	118.37	185.02	114.69	107.61
Tekin <i>et al.</i> [31]	-	129.06	91.43	121.68	-	162.17	-	-
Zhou <i>et al.</i> [44]	87.36	109.31	87.05	103.16	116.18	143.32	106.88	99.78
Sanzari <i>et al.</i> [27]	48.82	56.31	95.98	84.78	96.47	105.58	66.30	107.41
Ours - Single PPCA Model	68.55	78.27	77.22	89.05	91.63	110.05	74.92	83.71
Ours - Mixture PPCA Model	64.98	73.47	76.82	86.43	86.28	110.67	68.93	74.79
	Sitting	Sitting Down	Smoking	Waiting	Walk Dog	Walking	Walk Together	Average
LinKDE [14]	151.57	243.03	162.14	170.69	177.13	96.60	127.88	162.14
Li <i>et al.</i> [19]	-	-	-	-	132.17	69.97	-	-
Tekin <i>et al.</i> [32]	136.15	205.65	118.21	146.66	128.11	65.86	77.21	125.28
Tekin <i>et al.</i> [31]	-	-	-	-	130.53	65.75	-	-
Zhou <i>et al.</i> [44]	124.52	199.23	107.42	118.09	114.23	79.39	97.70	113.01
Sanzari <i>et al.</i> [27]	116.89	129.63	97.84	65.94	130.46	92.58	102.21	93.15
Ours - Single PPCA Model	115.94	185.72	88.25	88.73	92.37	76.48	77.95	92.96
Ours - Mixture PPCA Model	110.19	173.91	84.95	85.78	86.26	71.36	73.14	88.39

Example results on the Human3.6M dataset

