

Multi-view 3D Pose Estimation

20150342 Junkyu Park

20170317 Seungho Baek

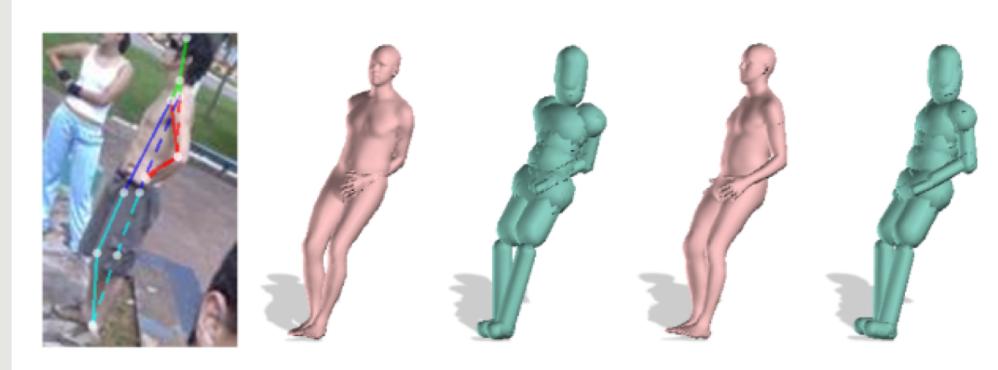
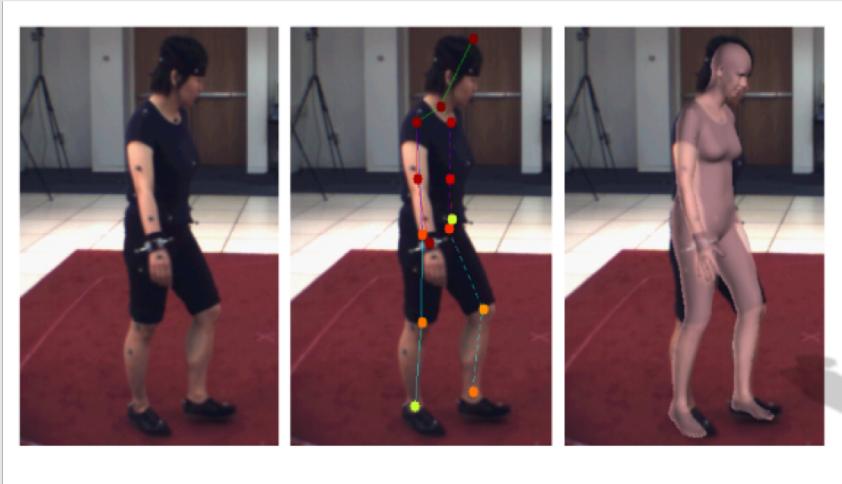
NCSOFT Vision AI Lab Human Pose Team Yeongeon Lee

Objective

**Improve 3D pose estimation accuracy
using multi-view images**

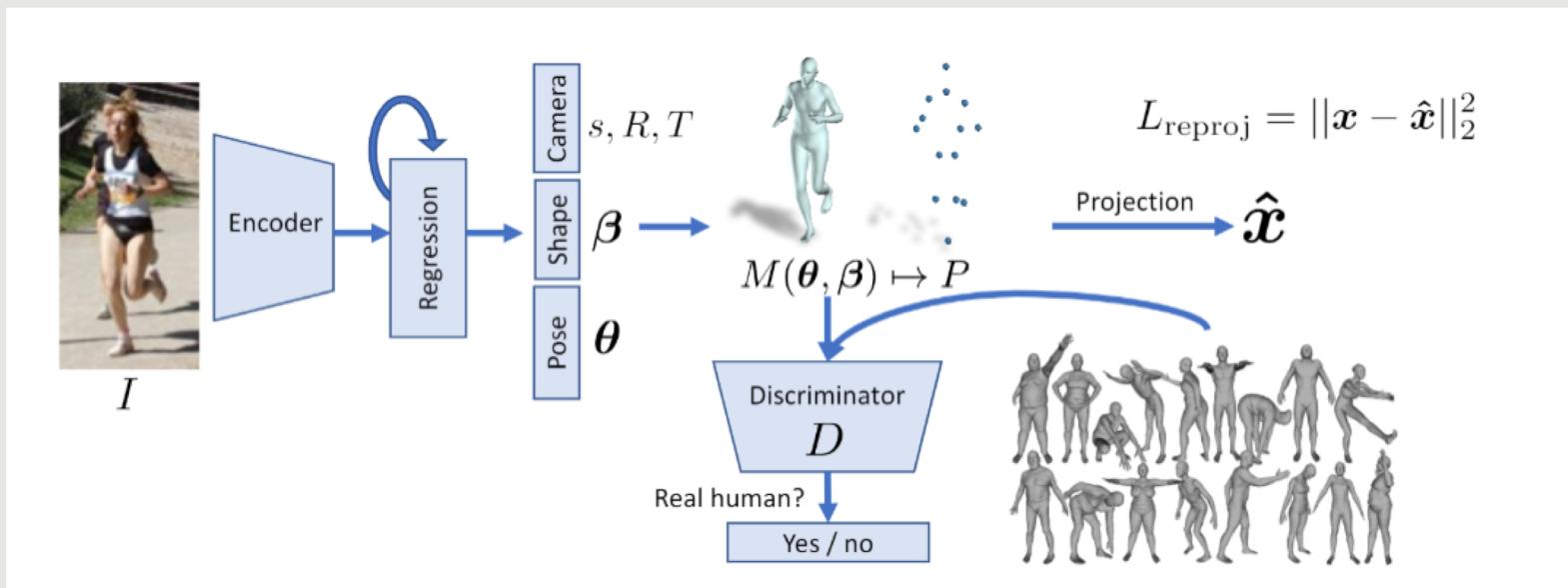
Related work - SMPLify

Image -> 2D annotation -> 3D annotation -> **OPTIMIZATION**

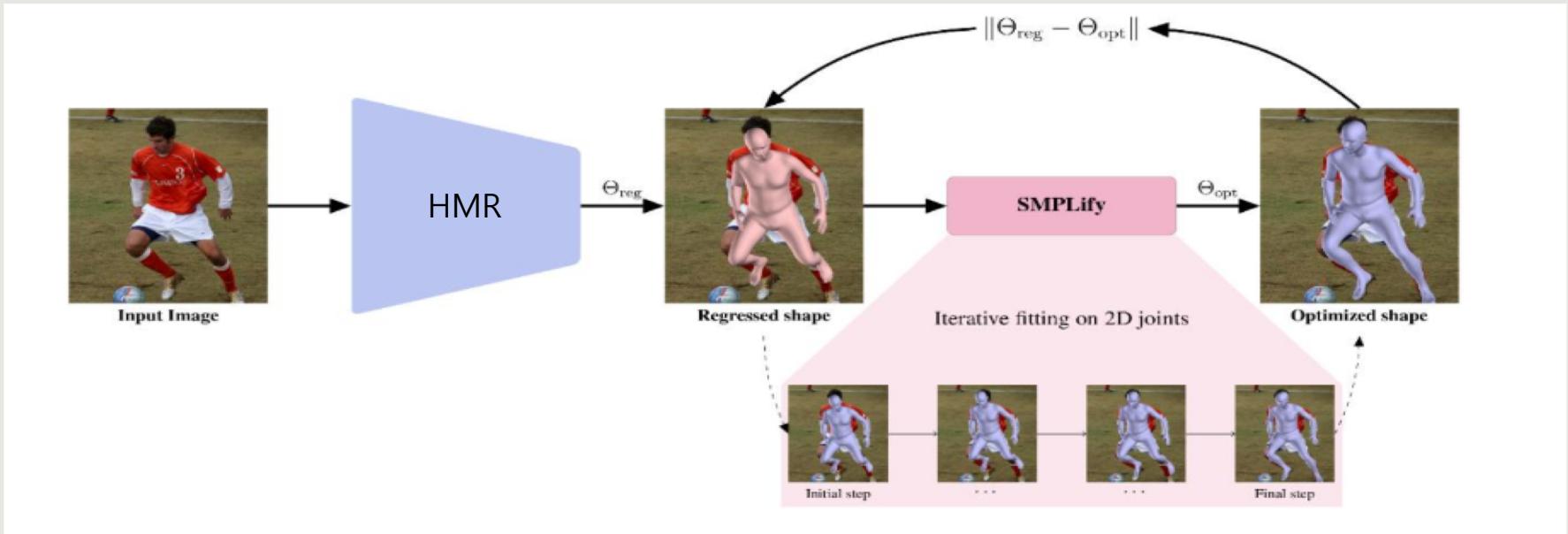


Related work - HMR

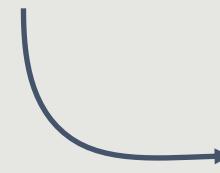
Image -> 3D annotation(using Discriminator)



Related work - SPIN



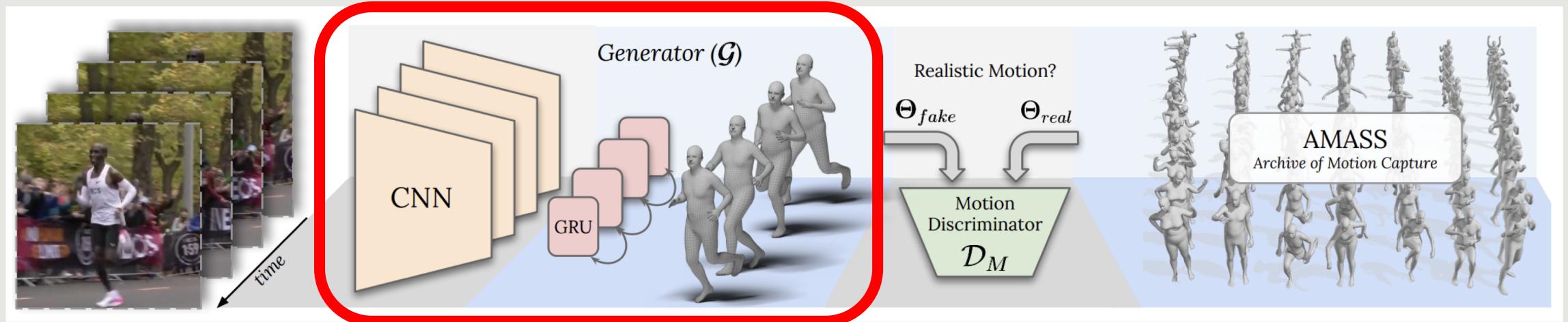
	Regression-based	Optimization-based
Strength	Can take all pixel values into consideration instead relying only on a sparse set of 2D location	Typically get a good fit
Weakness	A large amount of data is necessary to properly train the network	Slow and sensitive about choice of initialization



Use multi-view fitting
(Todo forward)

Related work - VIBE

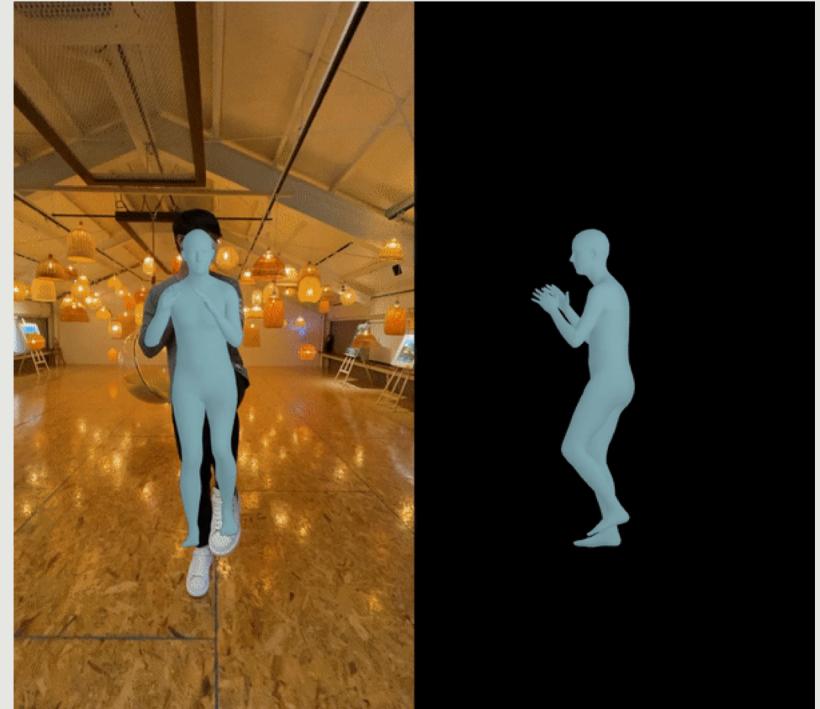
GRU outputs latent variables containing information incorporated from past and future frames.



Problem Statement

Low accuracy for unusual pose

Not enough multi-view datasets with Ground Truth

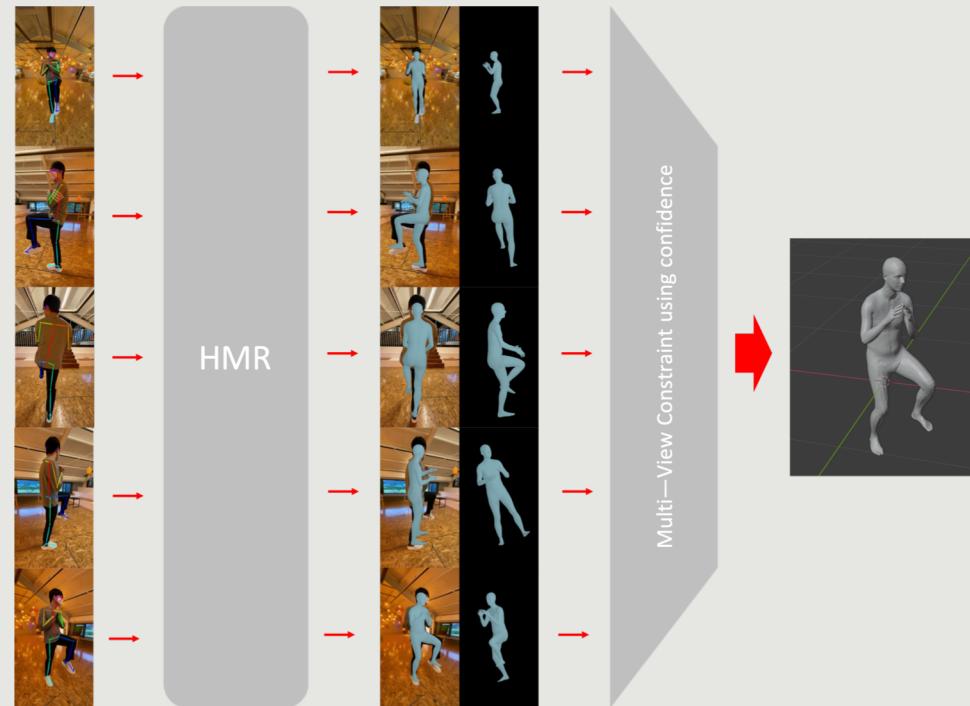


Solution

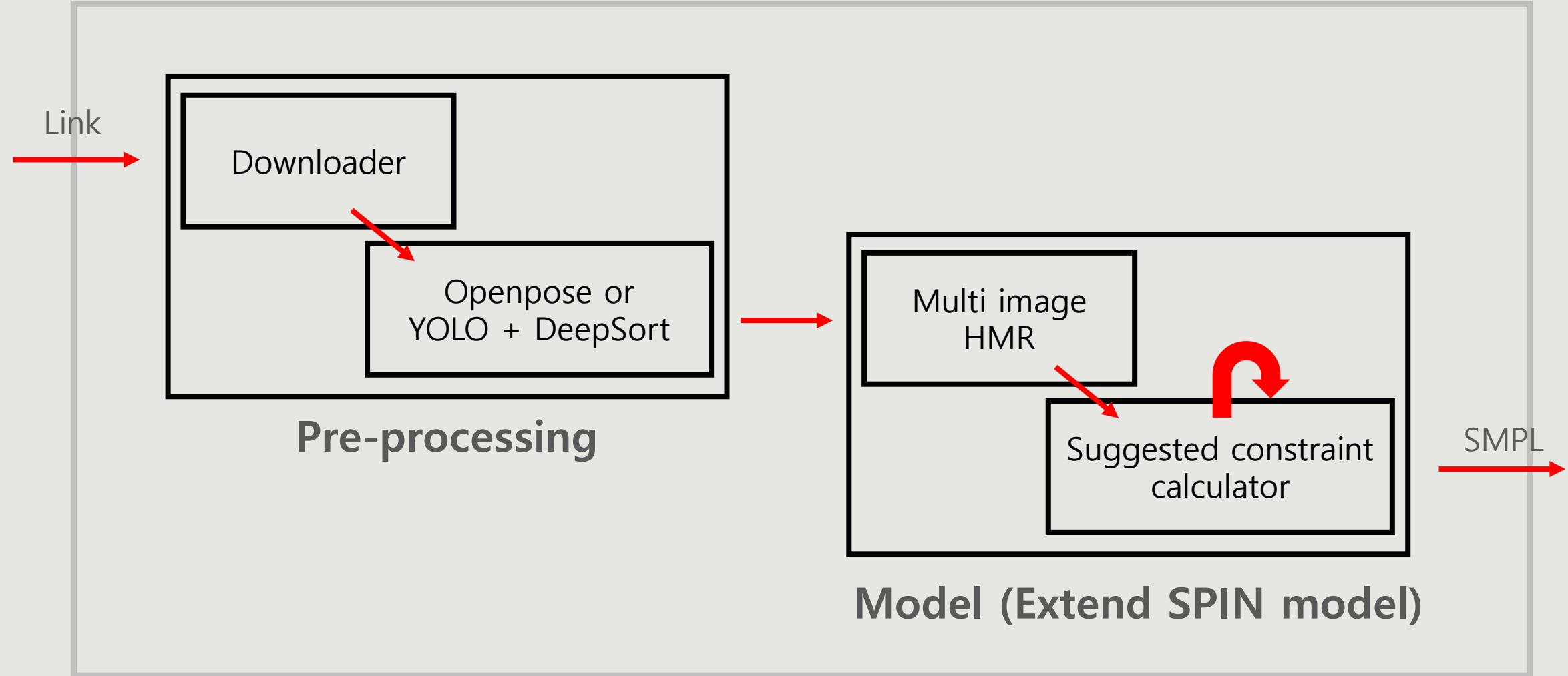
Find confidence map – SMPL relationship

Extend SPIN Model

$$E_J(S, \theta) = \sum_{v=1}^V s_v \{ E_J(\beta, \theta; K, J_{est}) + \lambda_\theta E_\theta(\theta) + \lambda_a E_a(\theta) + \lambda_\beta E_\beta(\beta) \}$$



Architecture



Requirement Analysis

Functional Requirement

- Design multi-view constraint and get better result
 - Modify the model to receive multi-view input.
 - Get confidence map from the openpose result.
 - Find relationship between confidence map and SMPL parameters.

	Absolute			Rigid Alignment		
	PCK	AUC	MPJPE	PCK	AUC	MPJPE
Mehta <i>et al.</i> [22]	75.7	39.3	117.6	-	-	-
VNect [23]	76.6	40.4	124.7	83.9	47.3	98.0
HMR [15]	72.9	36.5	124.2	86.3	47.8	89.8
Single-view	76.4	37.1	105.2	92.5	55.6	67.5

<For MPI-INF-3DHP>

	Rec. Error
NBF [24]	59.9
HMR [15]	56.8
Single-view	41.1

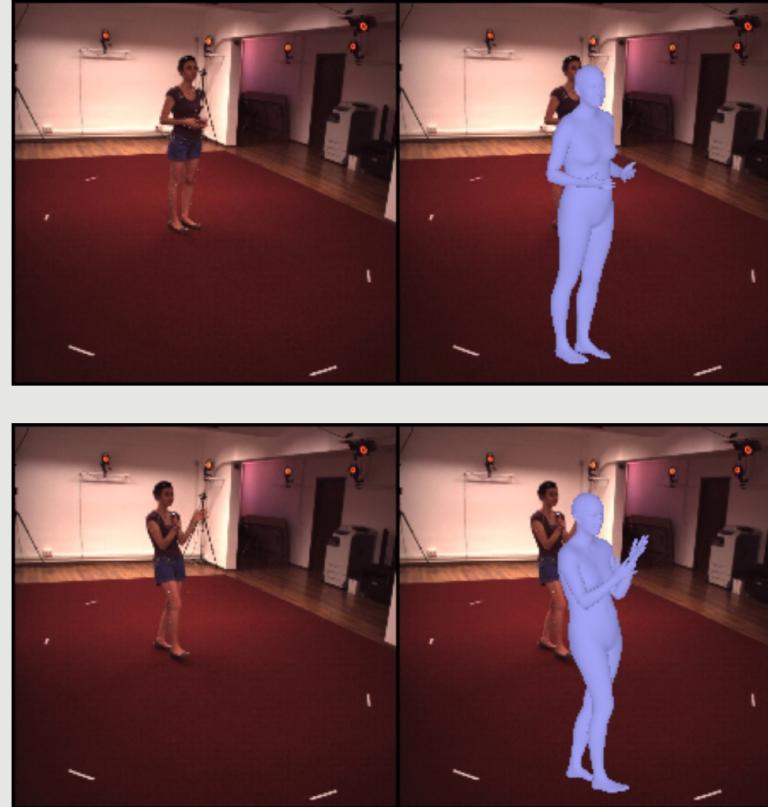
<For Human3.6m>

Non-functional Requirement

- Reconstruction error of multi-view 3D pose estimation should be better than existing models for human 3.6m
- PCK, AUC, MPJPE loss value should be lower than existing models for MPI-INF-3DHP dataset

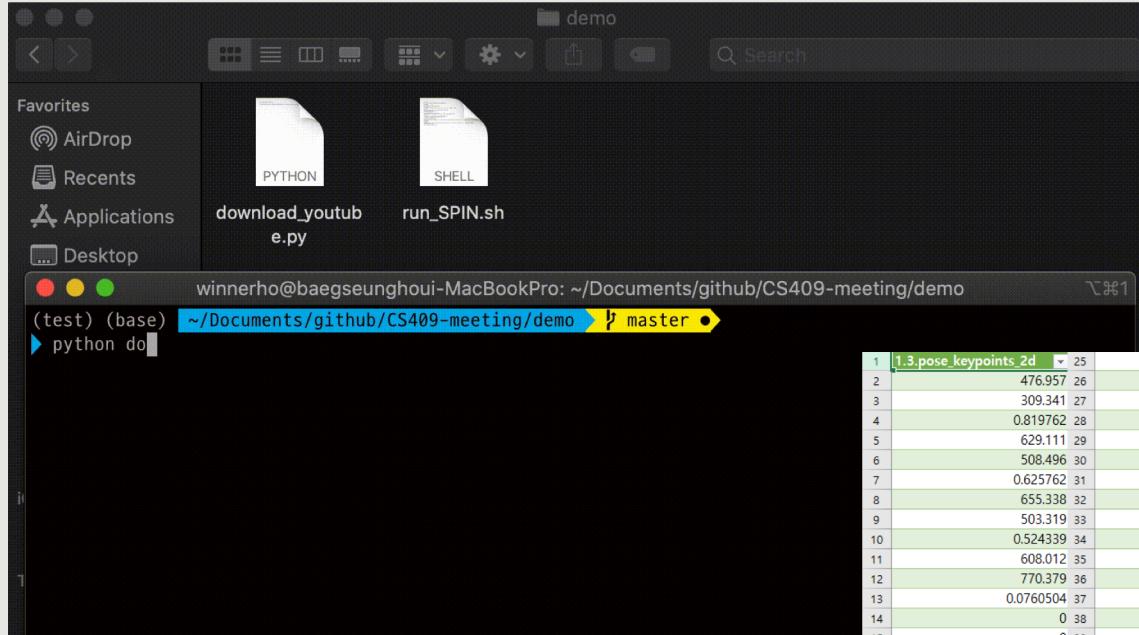
Acquired Datasets

```
✓ dataset_extras
  ≡ 3dpw_test.npz
  ≡ coco_2014_train.npz
  ≡ h36m_valid_protocol1.npz
  ≡ h36m_valid_protocol2.npz
  ≡ hr-lspet_train.npz
  ≡ lsp_dataset_original_train.npz
  ≡ lsp_dataset_test.npz
  ≡ mpi_inf_3dhp_train.npz
  ≡ mpi_inf_3dhp_valid.npz
  ≡ mpii_train.npz
```



Human3.6m(with SMPL), MPI(with SMPL), 3dpw, coco, etc

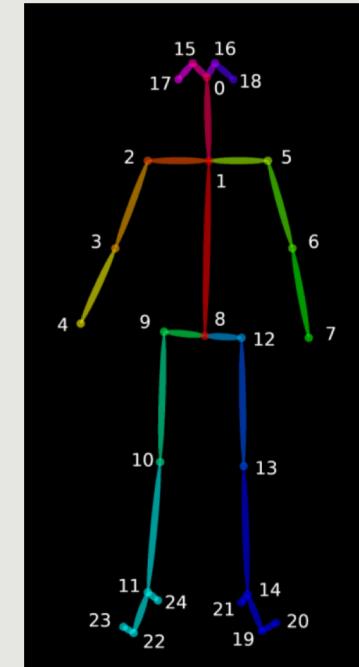
Detailed Pre-processing pipeline



The screenshot shows a Mac OS X desktop environment. In the foreground, a terminal window is open with the command `python do` entered. The terminal is located in a directory named `demo`. In the background, a file browser window titled "demo" is open, showing a folder structure with "PYTHON" and "SHELL" subfolders. Inside "PYTHON" are files "download_youtub e.py" and "run_SPIN.sh". The desktop icons include "AirDrop", "Recents", "Applications", and "Desktop".

1	1.3:pose_keypoints_2d	25	0.800121	50	513.729
2	476.957	26	560.786	51	278.21
3	309.341	27	1016.68	52	0.846813
4	0.819762	28	0.582026	53	0
5	629.111	29	602.85	54	0
6	508.496	30	1027.08	55	0
7	0.625762	31	0.526928	56	608.017
8	655.338	32	561.01	57	314.661
9	503.319	33	1336.11	58	0.956367
10	0.524339	34	0.772692	59	131.503
11	608.012	35	576.75	60	1419.91
12	770.379	36	1655.84	61	0.765518
13	0.0760504	37	0.735859	62	152.564
14	0	38	513.754	63	1430.5
15	0	39	1006.09	64	0.821748
16	0	40	0.563328	65	309.303
17	607.934	41	183.699	66	1378
18	508.479	42	974.662	67	0.731631
19	0.757514	43	0.849295	68	466.574
20	450.944	44	251.935	69	1640.01
21	770.585	45	1346.61	70	0.606601
22	0.838564	46	0.0705866	71	477.204
23	445.783	47	477.194	72	1634.7
24	613.498	48	272.657	73	0.573486
25	0.800121	49	0.252618	74	607.966
26	560.786	50	513.729	75	1702.89
27	1016.68	51	278.21	76	0.735467
28	0.582026	52	0.846813	77	
29	602.85	53	0	78	
30	1027.08	54	0	79	
31	0.526928	55	0	80	
32	561.01	56	608.017	81	
33	1336.11	57	314.661	82	
34	0.772692	58	0.956367	83	
35	576.75	59	131.503	84	

- For 25 body parts
 - $(3n-1)$ th data = x position of body part
 - $(3n)$ th data = y position of body part
 - $(3n + 1)$ th data = confidence of body part



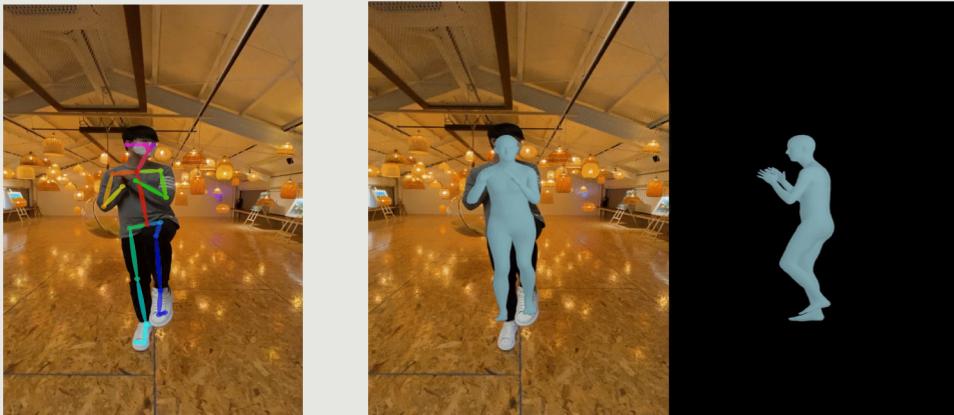
Visualizing

Youtube Datasets (Mannequin challenge)

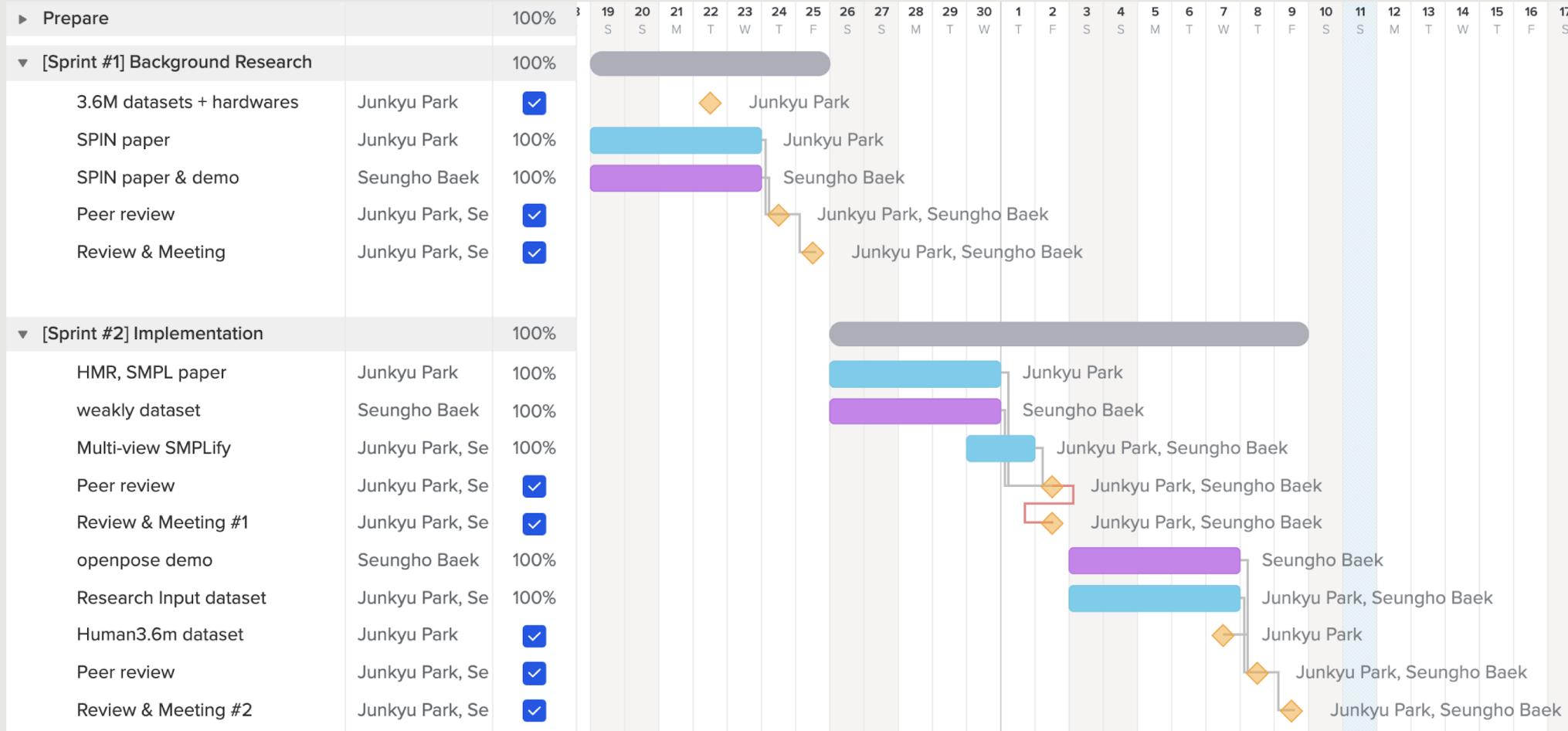


Our Datasets (Recorded by camera/phone)

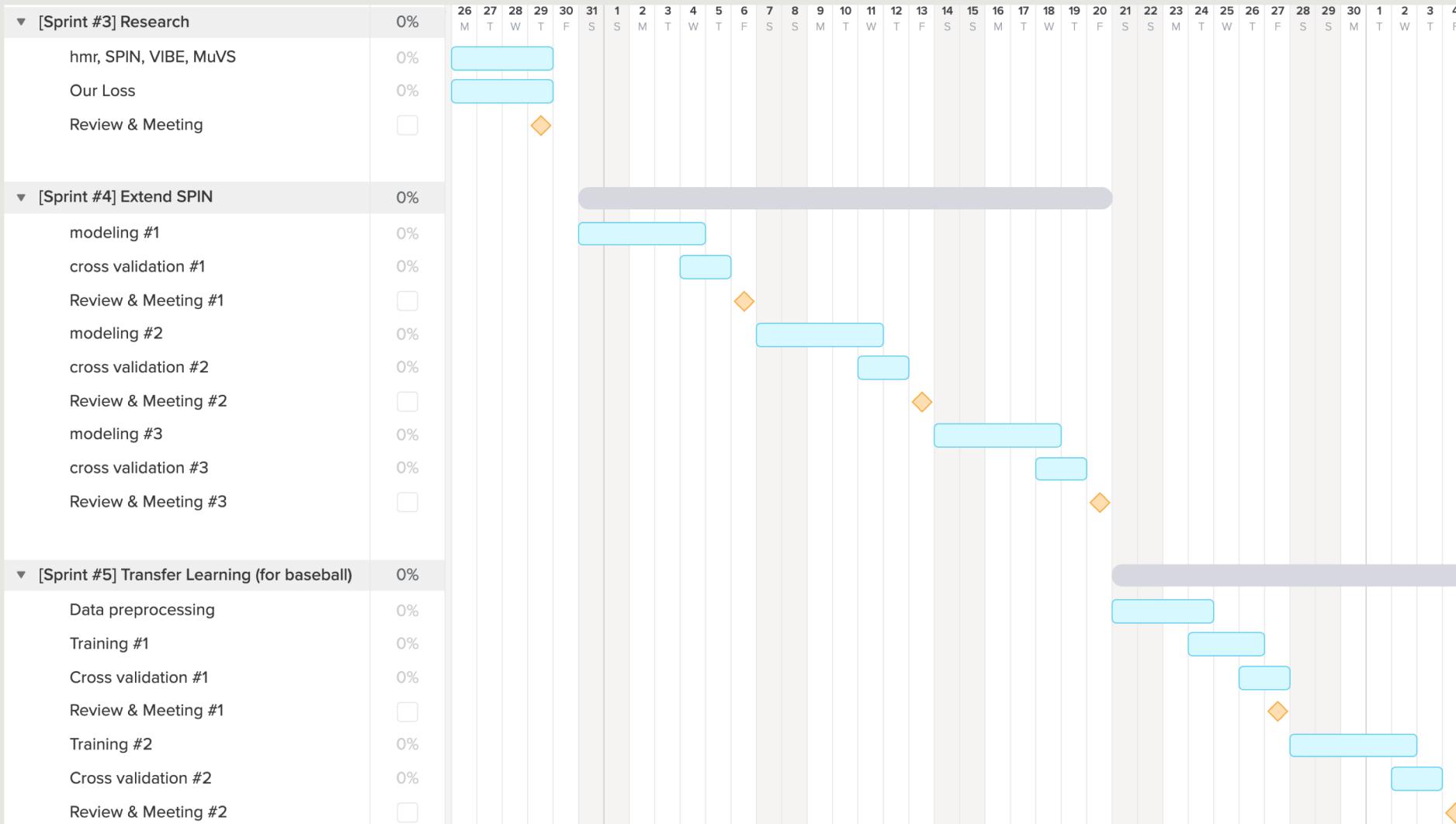
video_00000000299_keypoints.json
video_00000000298_keypoints.json
video_00000000297_keypoints.json
video_00000000296_keypoints.json
video_00000000295_keypoints.json
video_00000000294_keypoints.json
video_00000000293_keypoints.json
video_00000000292_keypoints.json
video_00000000291_keypoints.json
video_00000000290_keypoints.json
video_00000000289_keypoints.json
video_00000000288_keypoints.json
video_00000000287_keypoints.json



Schedule : before mid presentation



Schedule : after mid presentation



Role and Responsibility

Seungho Baek (Team Leader)

- Done
 - Run SPIN demo / Preprocess datasets for training / suggest constraint design
- To do
 - Make designed model

Junkyu Park

- Done
 - HMR, SMPLify, SPIN paper review / Get Human3.6m dataset / dataset labeling / environment settings
- To do
 - Acquire dataset and modeling

Yeongeon Lee (Mentor, NCSOFT)

- Done
 - Advice for project proceeding / Supply Human3.6m datasett
- To do
 - Feedback about our design and modeling / Supply needed datasets.