Implicit 3D Human Mesh Recovery using Consistency with Pose and Shape from Unseen-view

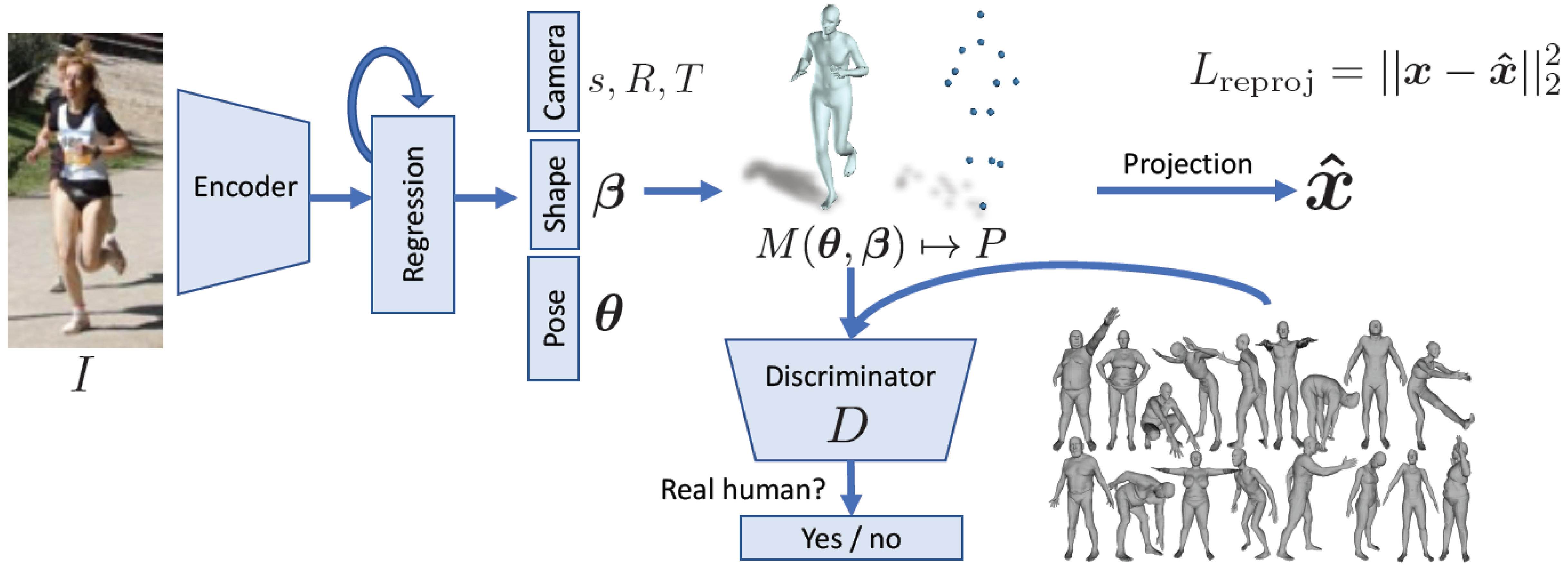
Hanbyel Cho, Yooshin Cho, Jaesung Ahn, Junmo Kim
School of Electrical Engineering, KAIST, South Korea

CVPR 2023 - THU-PM-052

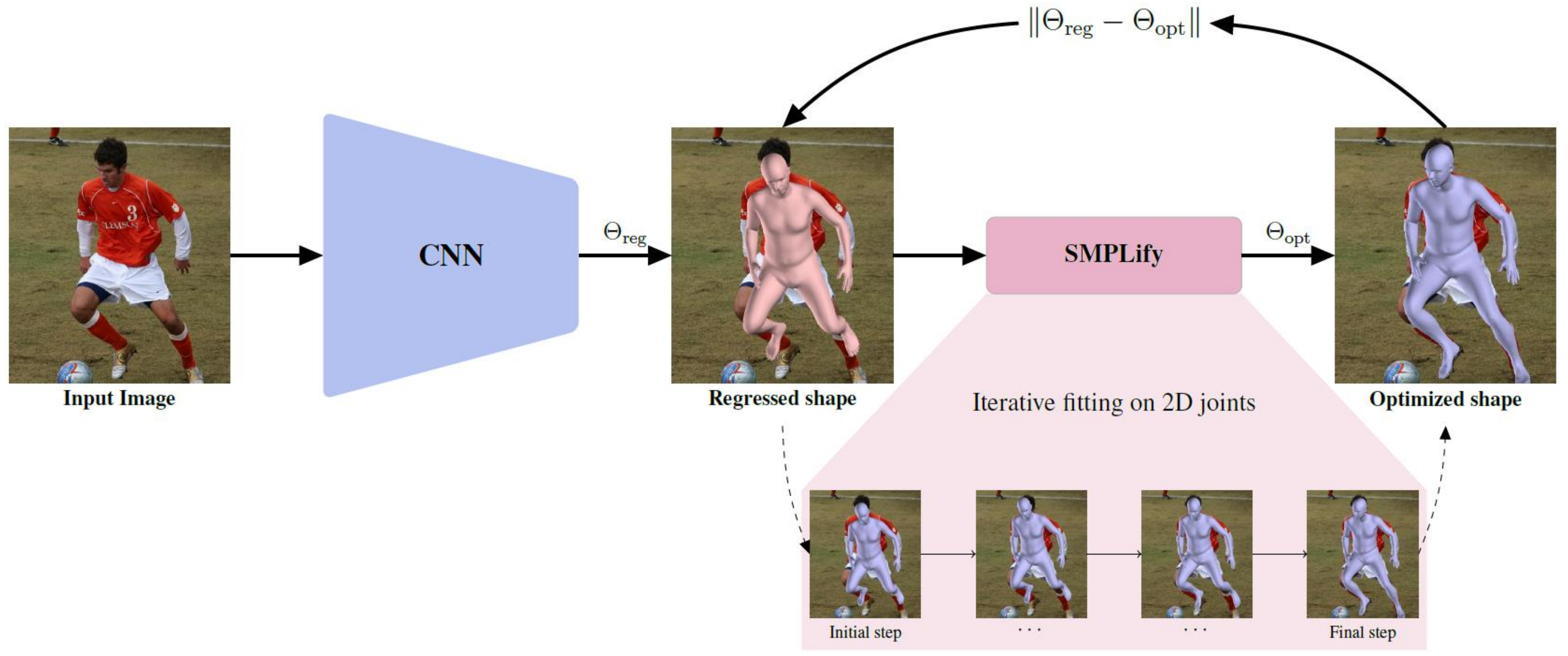
This video contains audio

Human Mesh Recovery

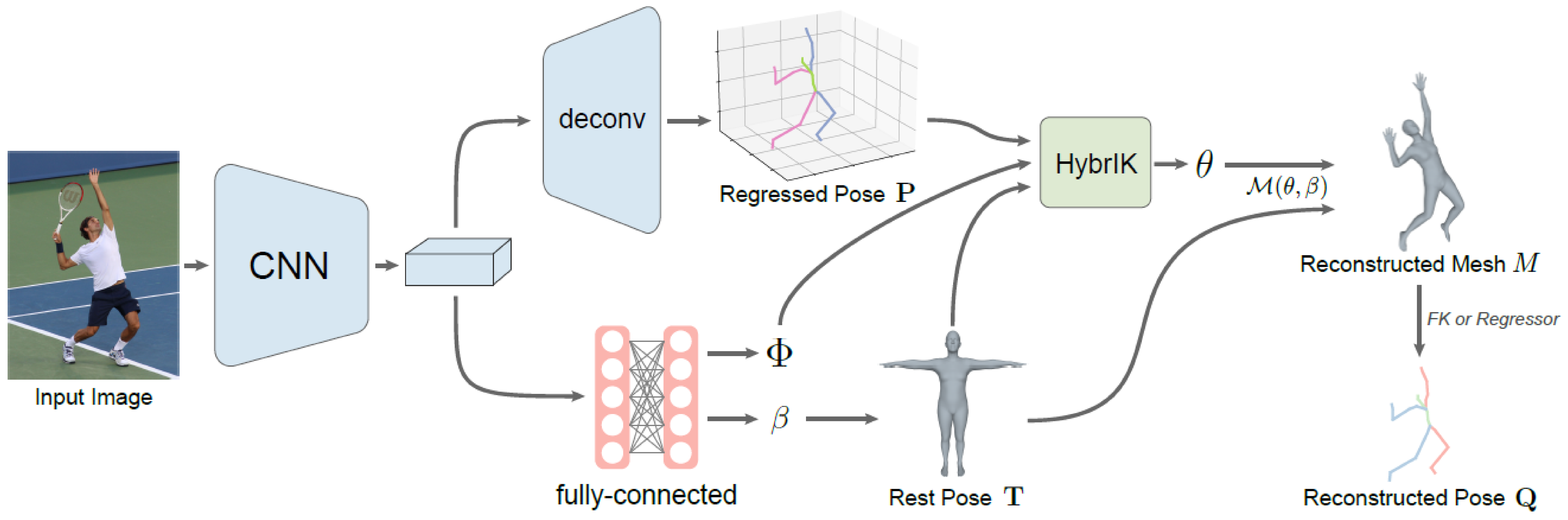
- HMR: task regresses 3D human body model (SMPL) parameters from RGB inputs



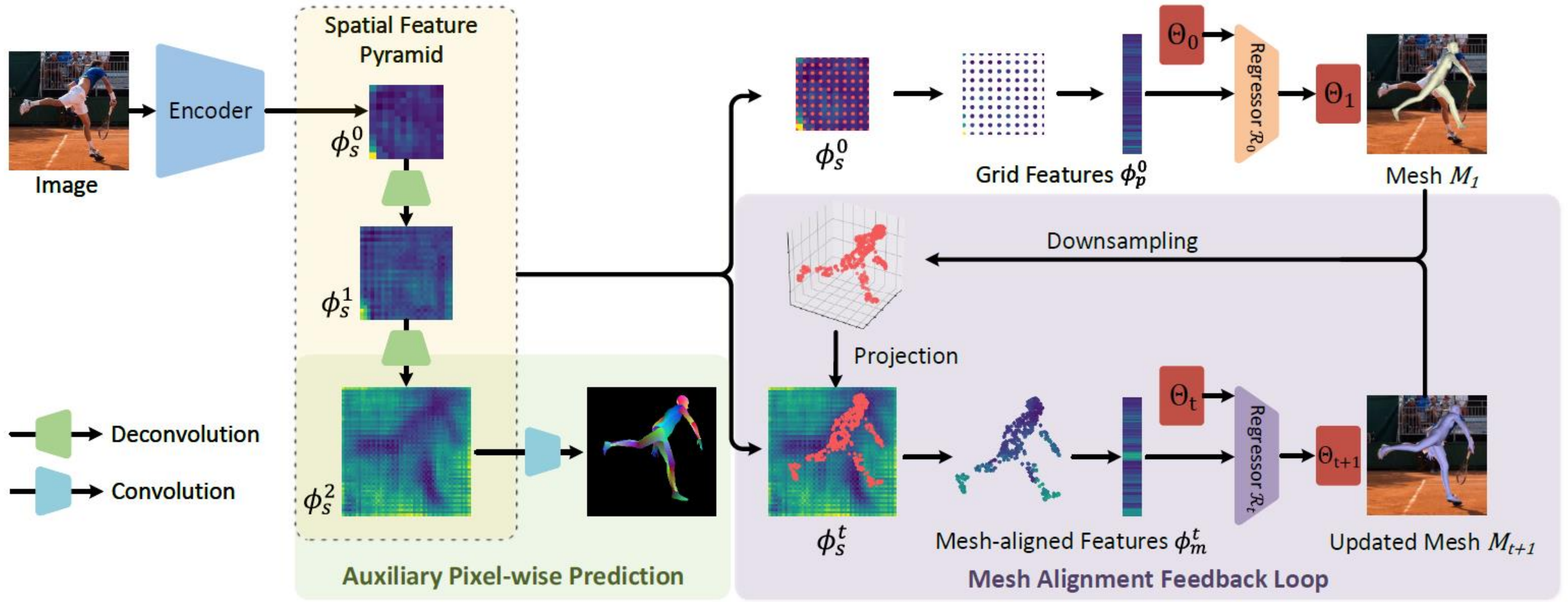
Kanazawa *et al.*, CVPR 2018 [R1]



Kolotouros *et al.*, ICCV 2019 [R2]



Li *et al.*, CVPR 2021 [R3]



Zhang *et al.*, ICCV 2021 [R4]

[R1] Kanazawa et al., End-to-End Recovery of Human Shape and Pose, CVPR 2018

[R2] Kolotouros et al., Learning to Reconstruct 3 D Human Pose and Shape via Model-fitting in the Loop, ICCV 2019

[R3] Li et al., HybrIK: A Hybrid Analytical-Neural Inverse Kinematics Solution for 3D Human Pose and Shape Estimation, CVPR 2021

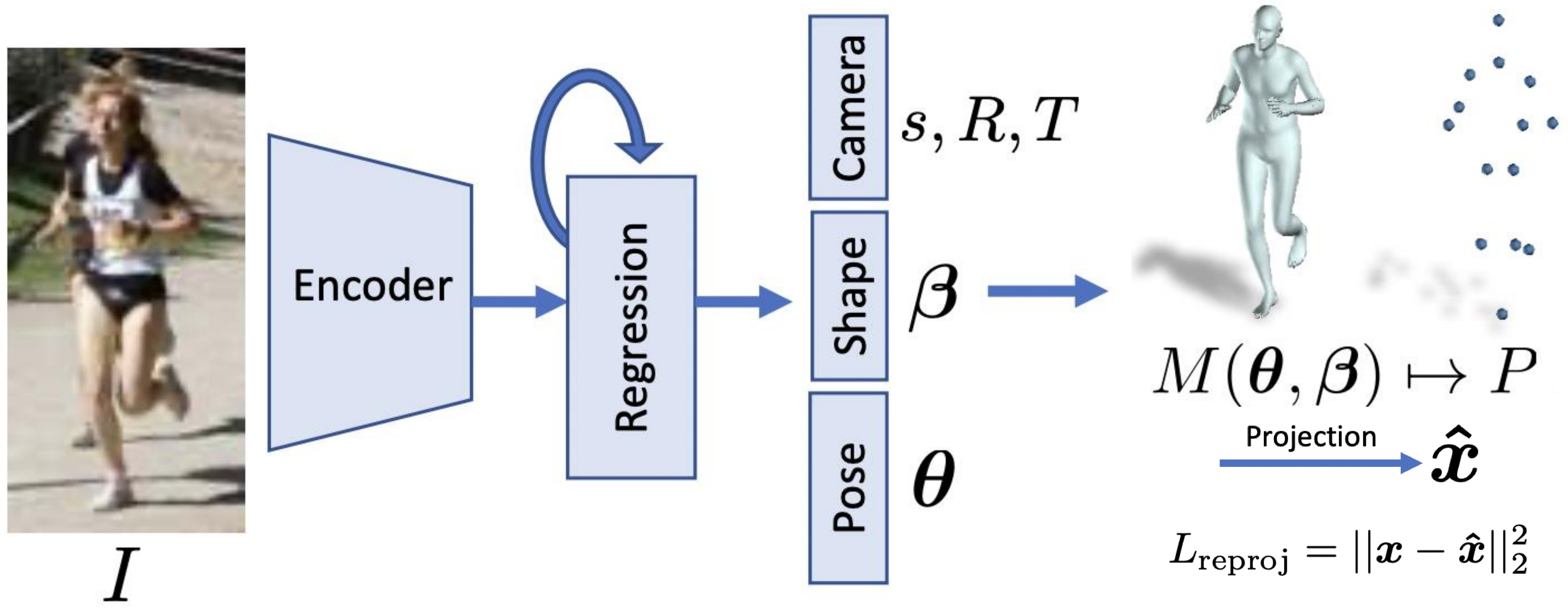
[R4] Zhang et al., PyMAF: 3D Human Pose and Shape Regression with Pyramidal Mesh Alignment Feedback Loop, ICCV 2021

Motivation

- Existing methods fail to regress SMPL when the ambiguity (*e.g.,* depth, occlusion) exists

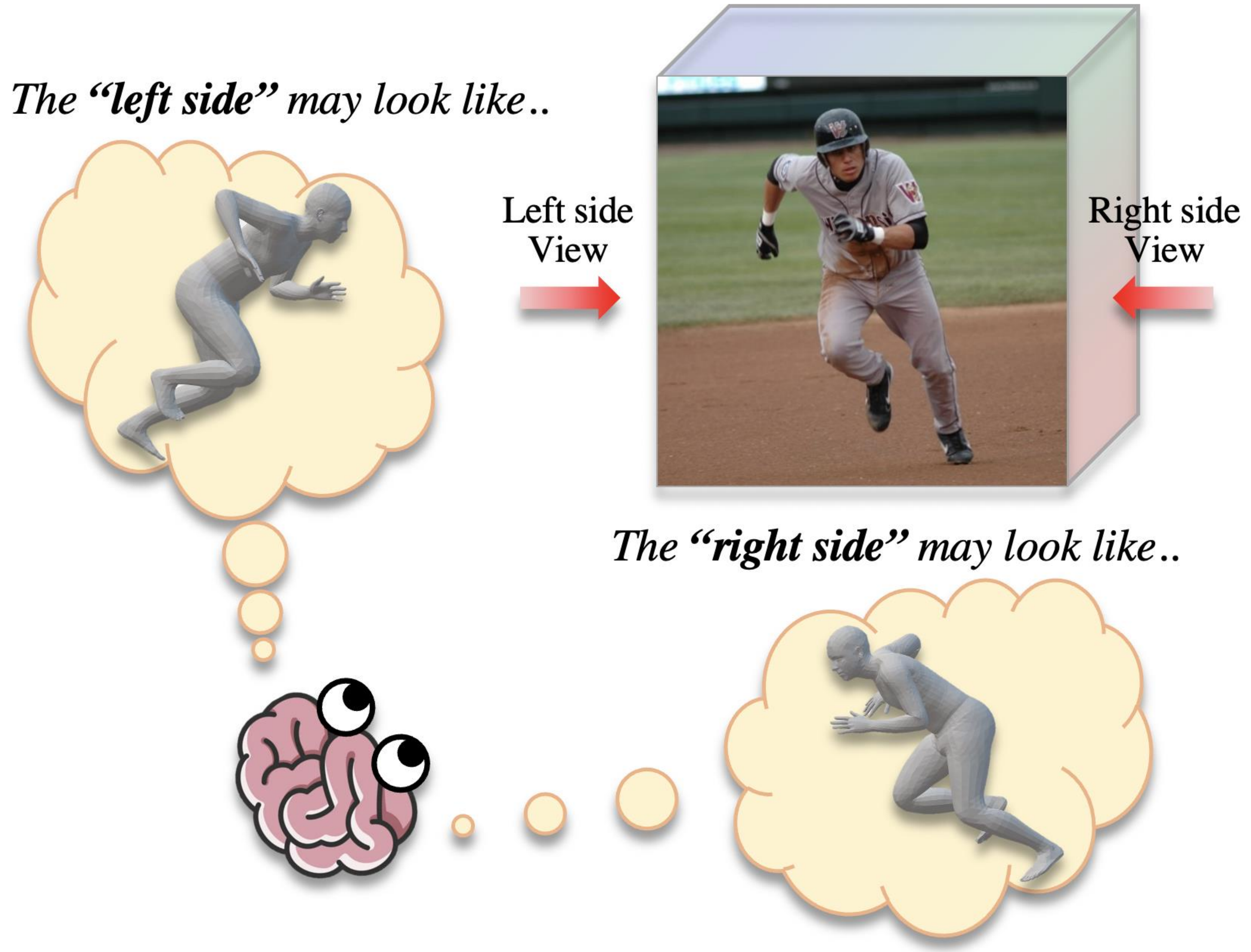
Existing methods

Only consider the direction in which the image was taken
Fail to reconstruct human body mesh if ambiguity exists (e.g., depth and occlusion)



Our motivation

Mimic the mental model of human
(1) Imagine a person at difference directions in 3D space
(2) Utilize consistency of pose and shape from those views



Goal & Method

- Make the model can imagine a person placed in a 3D space via neural feature fields
- **Training phase**: utilize the consistency of pose and shape by rotating viewing direction
- **Inference phase**: use results inferred from rendered feature in a canonical viewing dir.

Proposed model

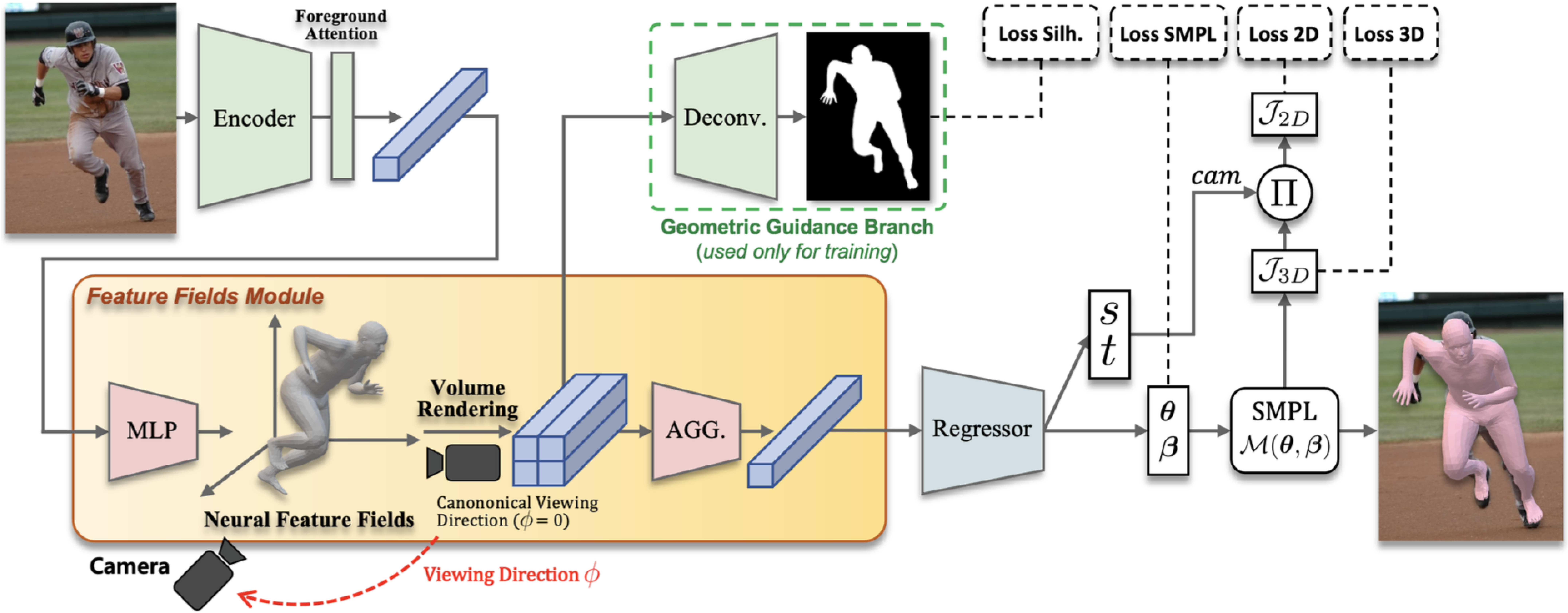


Figure 2. **Overview of ImpHMR architecture.** Given an image of a person, ImpHMR can implicitly imagine the person in 3D space and infer SMPL parameters viewed from an arbitrary viewing direction ϕ through *Feature Fields Module*. The model infers parameters from arbitrary directions during training to have a better 3D prior about person; consequently, regression performance in *Canonical Viewing Direction* is improved. For simplicity, we omit notation ϕ and write loss functions in Sec 3.4 abstractly according to the form of the output.

Method

- **Framework**: conventional HMR pipeline + Feature Fields Module + Geo. Guidance Branch
- **Objective**: canonical view regr. + appearance cons. + arbitrary view imagination loss

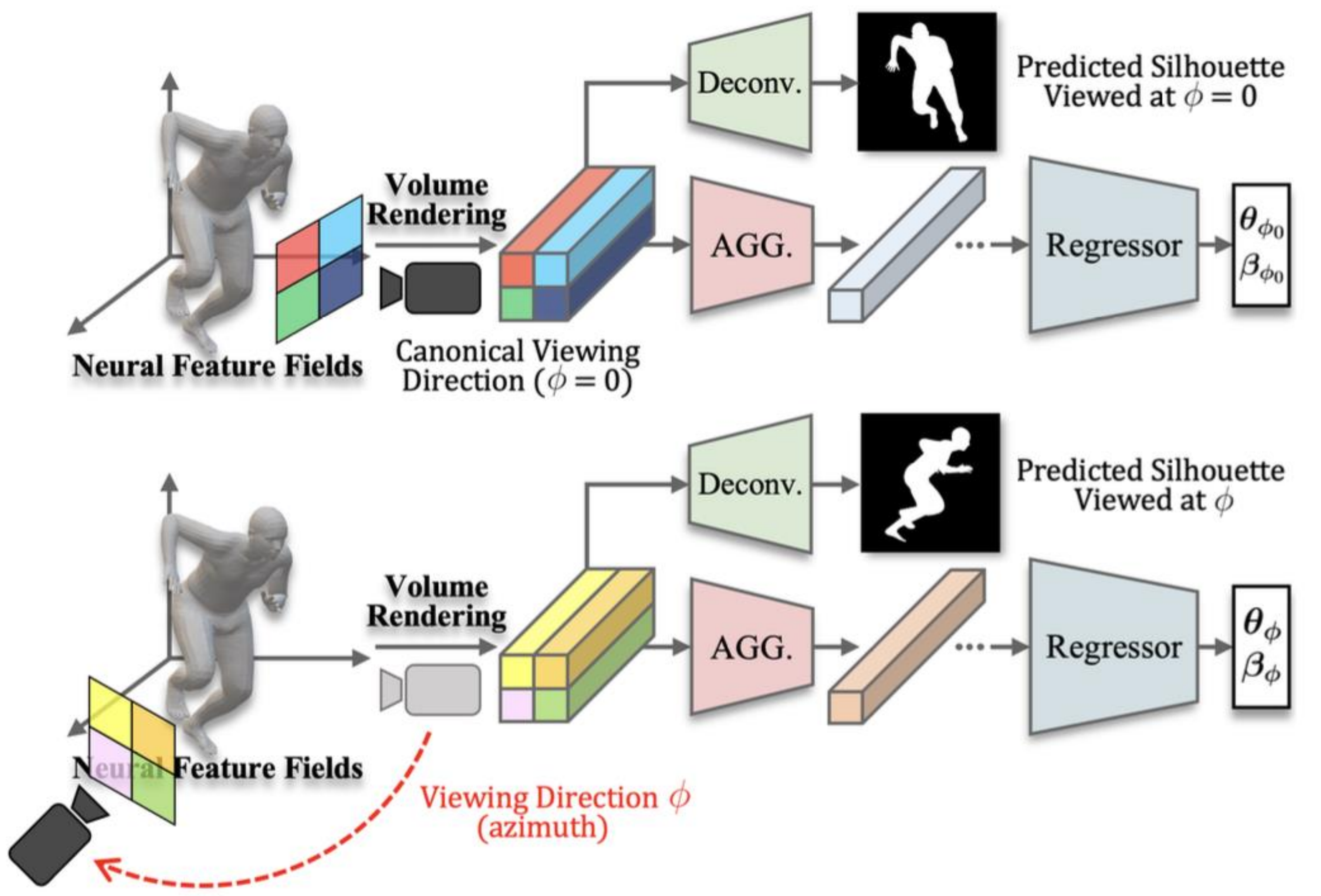
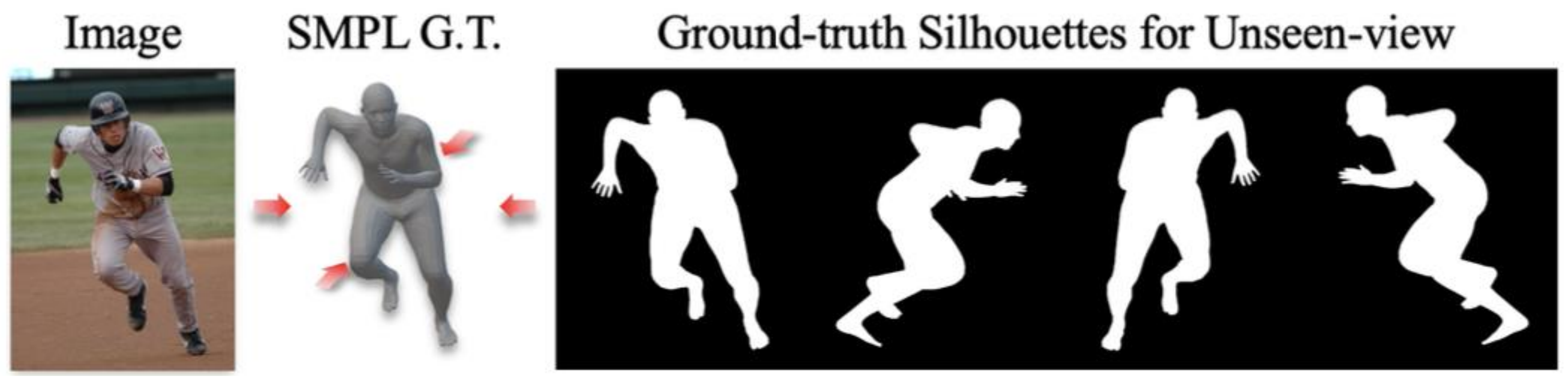


Figure 4. **SMPL parameter and silhouette regression with controlling camera viewing direction. Top:** regression from the *Canonical Viewing Direction* ($\phi = 0$), as in conventional methods. **Bottom:** regression from an arbitrary viewing direction.



(1) Canonical view regression loss

- Constraint for inference from the canonical viewing direction like conventional HMR methods

$$\mathcal{L}_{reg} = \lambda_{2d}||K_{\phi_0} - \hat{K}|| + \lambda_{3d}||J_{\phi_0} - \hat{J}|| + \lambda_{pose}||\boldsymbol{\theta}_{\phi_0} - \hat{\boldsymbol{\theta}}|| + \lambda_{shape}||\boldsymbol{\beta}_{\phi_0} - \hat{\boldsymbol{\beta}}||,$$

Method

- **Framework**: conventional HMR pipeline + Feature Fields Module + Geo. Guidance Branch
- **Objective**: canonical view regr. + appearance cons. + arbitrary view imagination loss

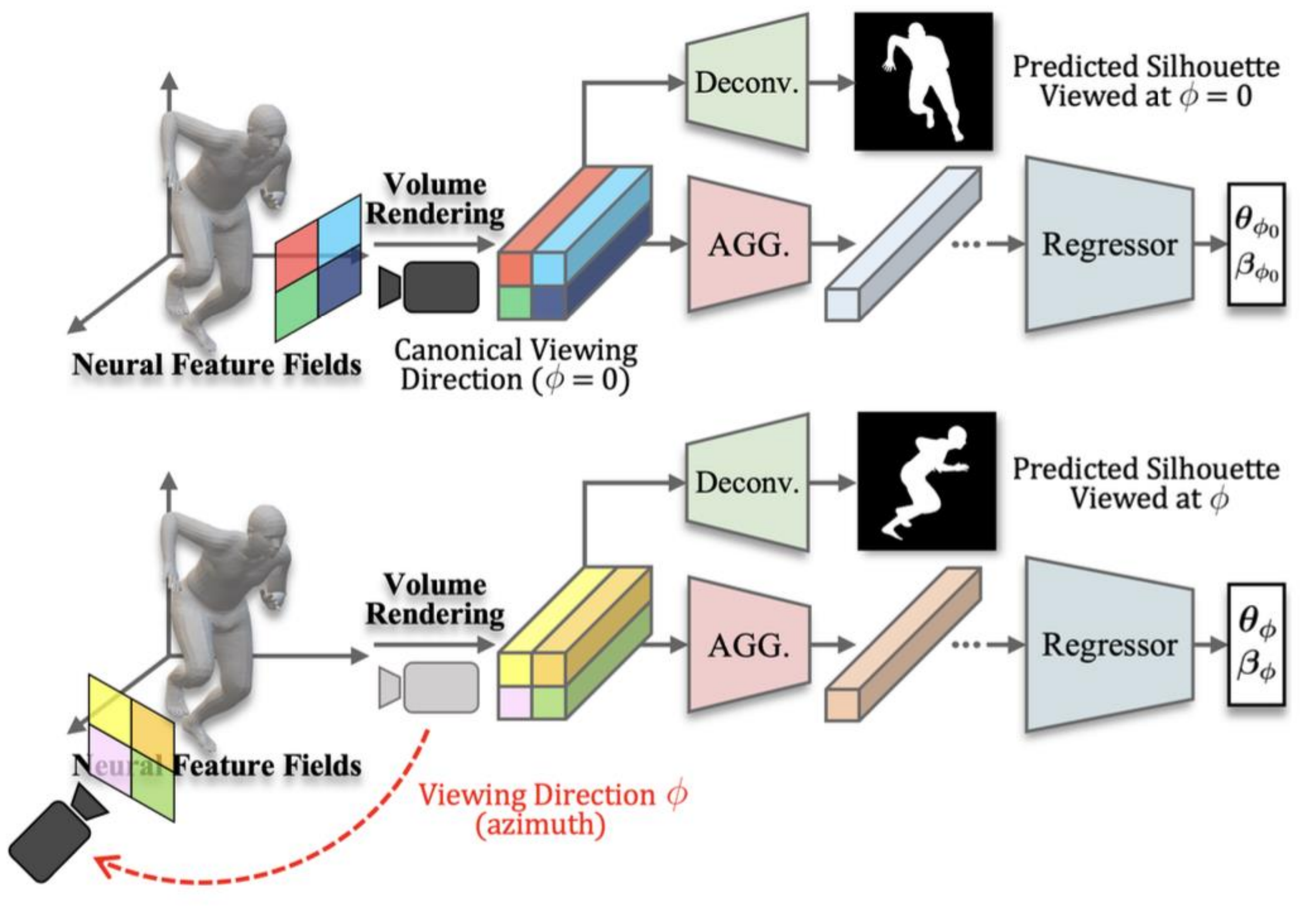
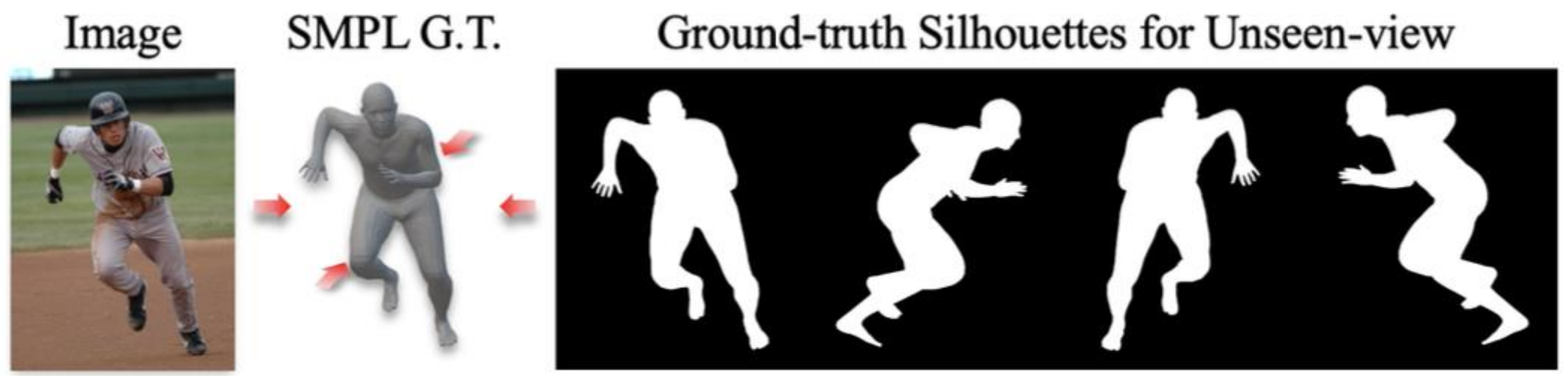


Figure 4. **SMPL parameter and silhouette regression with controlling camera viewing direction. Top:** regression from the *Canonical Viewing Direction* ($\phi = 0$), as in conventional methods. **Bottom:** regression from an arbitrary viewing direction.



(1) Canonical view regression loss

- Constraint for inference from the canonical viewing direction like conventional HMR methods

$$\mathcal{L}_{reg} = \lambda_{2d}||K_{\phi_0} - \hat{K}|| + \lambda_{3d}||J_{\phi_0} - \hat{J}|| \\ +\lambda_{pose}||\boldsymbol{\theta}_{\boxed{\phi_0}} - \hat{\boldsymbol{\theta}}|| + \lambda_{shape}||\boldsymbol{\beta}_{\phi_0} - \hat{\boldsymbol{\beta}}||,$$

(2) Arbitrary view imagination loss

- Constraint that the predicted results (including silh.) from an arbitrary viewing direction should be equal to the rotated G.T.

$$\mathcal{L}_{imag} = \mathbb{E}_{\phi \sim p_{cam}}[\lambda_{3d}||J_{\phi} - \hat{J}_{-\phi}|| + \lambda_{silh.}||\boxed{S_{\phi} - \hat{S}_{-\phi}}|| \\ +\lambda_{pose}||\boldsymbol{\theta}_{\boxed{\phi}} - \hat{\boldsymbol{\theta}}_{\boxed{-\phi}}|| + \lambda_{shape}||\boldsymbol{\beta}_{\phi} - \hat{\boldsymbol{\beta}}||],$$

Method

- **Framework**: conventional HMR pipeline + Feature Fields Module + Geo. Guidance Branch
- **Objective**: canonical view regr. + appearance cons. + arbitrary view imagination loss

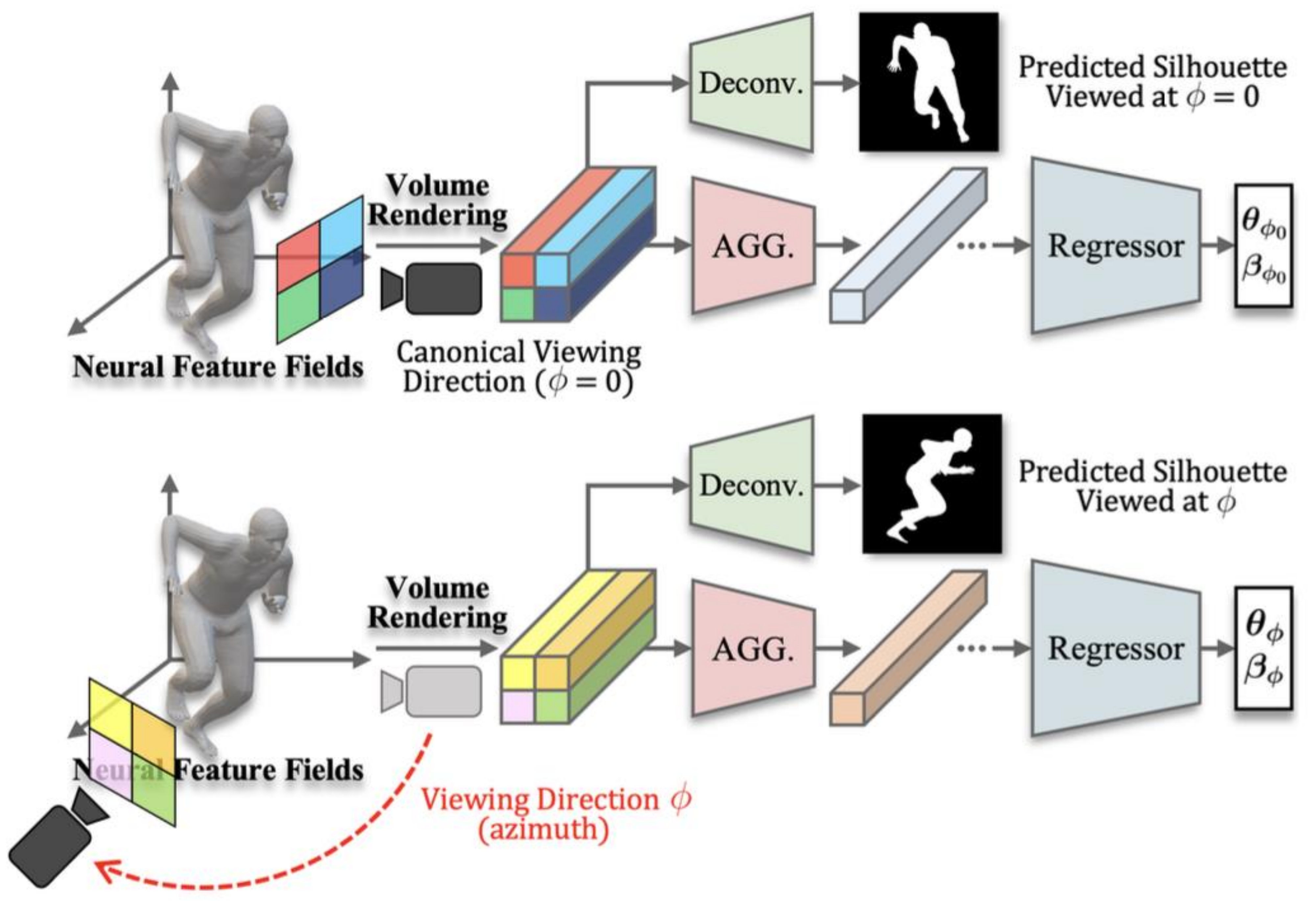
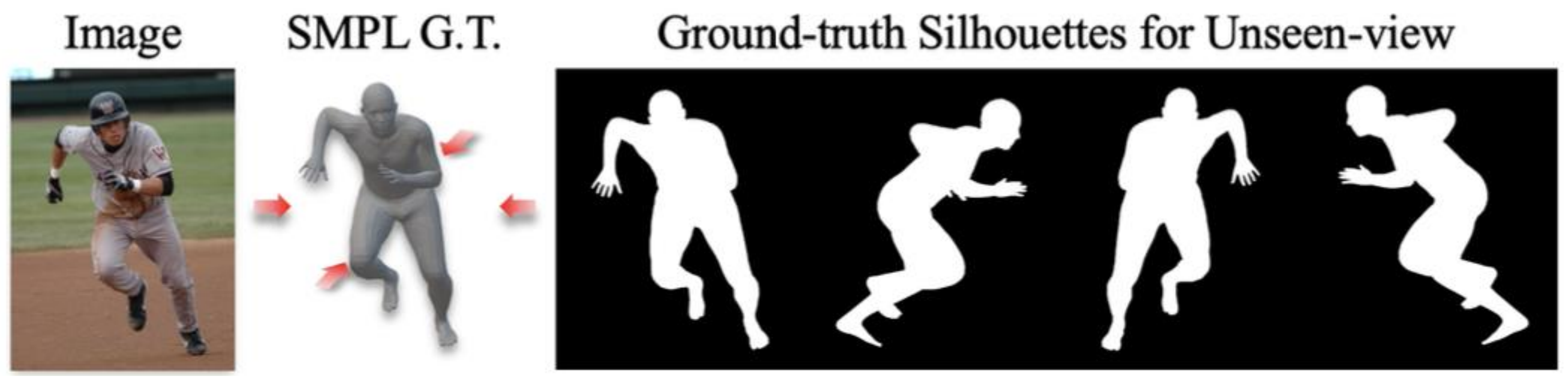


Figure 4. **SMPL parameter and silhouette regression with controlling camera viewing direction. Top:** regression from the *Canonical Viewing Direction* ($\phi = 0$), as in conventional methods. **Bottom:** regression from an arbitrary viewing direction.



(1) Canonical view regression loss

- Constraint for inference from the canonical viewing direction like conventional HMR methods

$$\mathcal{L}_{reg} = \lambda_{2d}||K_{\phi_0} - \hat{K}|| + \lambda_{3d}||J_{\phi_0} - \hat{J}|| \\ +\lambda_{pose}||\boldsymbol{\theta}_{\boxed{\phi_0}} - \hat{\boldsymbol{\theta}}|| + \lambda_{shape}||\boldsymbol{\beta}_{\phi_0} - \hat{\boldsymbol{\beta}}||,$$

(2) Arbitrary view imagination loss

- Constraint that the predicted results (including silh.) from an arbitrary viewing direction should be equal to the rotated G.T.

$$\mathcal{L}_{imag} = \mathbb{E}_{\phi \sim p_{cam}}[\lambda_{3d}||J_{\phi} - \hat{J}_{-\phi}|| + \lambda_{silh.}||\boxed{S_{\phi} - \hat{S}_{-\phi}}|| \\ +\lambda_{pose}||\boldsymbol{\theta}_{\boxed{\phi}} - \hat{\boldsymbol{\theta}}_{\boxed{-\phi}}|| + \lambda_{shape}||\boldsymbol{\beta}_{\phi} - \hat{\boldsymbol{\beta}}||],$$

(3) Appearance consistency loss

- Constraint that the pose and shape parameters inferred from different directions should be the same

$$\mathcal{L}_{cons} = \mathbb{E}_{\phi_1, \phi_2 \sim p_{cam}}[\lambda_{pose}||\boldsymbol{\theta}'_{\boxed{\phi_1}} - \boldsymbol{\theta}_{\boxed{\phi_2}}|| \\ +\lambda_{shape}||\boldsymbol{\beta}_{\phi_1} - \boldsymbol{\beta}_{\phi_2}||],$$

Results

- Quantitative results (3DPW)
  - 8.1% improv. in PA-MPJPE

		3PDW		
	Method	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Temporal	HMMR [20]	116.5	72.6	139.3
	DSD [50]	-	69.5	-
	Arnab *et al.* [2]	-	72.2	-
	Doersch *et al.* [11]	-	74.7	-
	VIBE [23]	93.5	56.5	113.4
	TCMR [8]	95.0	55.8	111.3
	MPS-Net [55]	91.6	54.0	109.6
Frame-based	HMR [19]	130.0	76.7	-
	GraphCMR [26]	-	70.2	-
	SPIN [25]	96.9	59.2	116.4
	PyMAF [62]	92.8	58.9	110.1
	I2L-MeshNet [39]	100.0	60.0	-
	ROMP [49]	89.3	53.5	105.6
	HMR-EFT [17]	-	54.2	-
	PARE [24]	<u>82.9</u>	<u>52.3</u>	<u>99.7</u>
	ImpHMR (Ours)	**81.8**	**49.8**	**96.4**
	ImpHMR (Ours) w. 3DPW	74.3	45.4	87.1

Results

- Quantitative results (3DPW)
  - 8.1% improv. in PA-MPJPE

	Method	3PDW MPJPE ↓	PA-MPJPE ↓	PVE ↓
Temporal	HMMR [20]	116.5	72.6	139.3
	DSD [50]	-	69.5	-
	Arnab *et al.* [2]	-	72.2	-
	Doersch *et al.* [11]	-	74.7	-
	VIBE [23]	93.5	56.5	113.4
	TCMR [8]	95.0	55.8	111.3
	MPS-Net [55]	91.6	54.0	109.6
Frame-based	HMR [19]	130.0	76.7	-
	GraphCMR [26]	-	70.2	-
	SPIN [25]	96.9	59.2	116.4
	PyMAF [62]	92.8	58.9	110.1
	I2L-MeshNet [39]	100.0	60.0	-
	ROMP [49]	89.3	53.5	105.6
	HMR-EFT [17]	-	54.2	-
	PARE [24]	82.9	52.3	99.7
	ImpHMR (Ours)	**81.8**	**49.8**	**96.4**
	ImpHMR (Ours) w. 3DPW	74.3	45.4	87.1

- Quantitative results (3DPW-OCC)

Method	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Zhang *et al.* [63]	-	72.2	-
HMR-EFT [17]	94.4	60.9	111.3
PARE [24]	90.5	56.6	107.9
ImpHMR (Ours)	**86.5**	**54.4**	**104.7**

Results

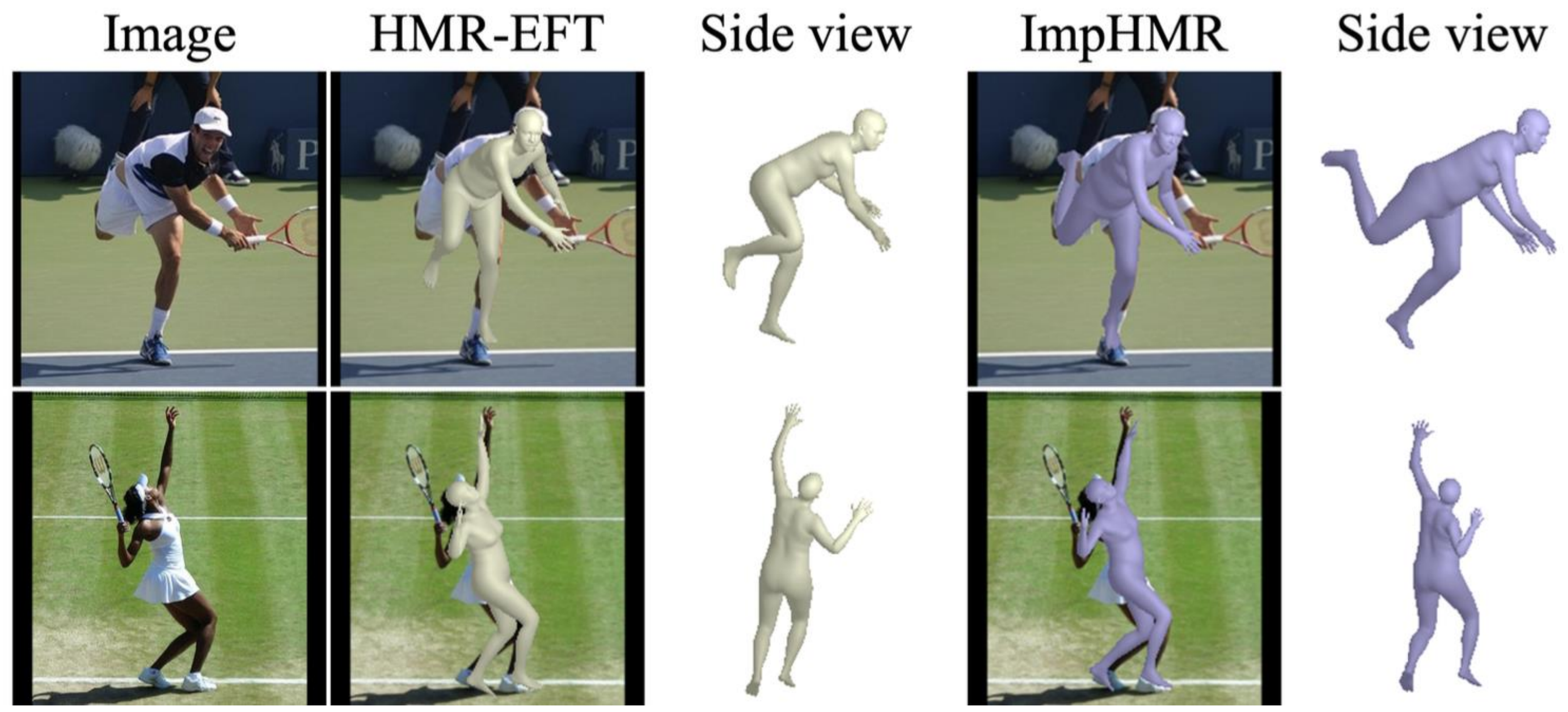
- Quantitative results (3DPW)
  - 8.1% improv. in PA-MPJPE

	Method	3PDW MPJPE ↓	PA-MPJPE ↓	PVE ↓
Temporal	HMMR [20]	116.5	72.6	139.3
	DSD [50]	-	69.5	-
	Arnab *et al.* [2]	-	72.2	-
	Doersch *et al.* [11]	-	74.7	-
	VIBE [23]	93.5	56.5	113.4
	TCMR [8]	95.0	55.8	111.3
	MPS-Net [55]	91.6	54.0	109.6
Frame-based	HMR [19]	130.0	76.7	-
	GraphCMR [26]	-	70.2	-
	SPIN [25]	96.9	59.2	116.4
	PyMAF [62]	92.8	58.9	110.1
	I2L-MeshNet [39]	100.0	60.0	-
	ROMP [49]	89.3	53.5	105.6
	HMR-EFT [17]	-	54.2	-
	PARE [24]	82.9	52.3	99.7
	ImpHMR (Ours)	**81.8**	**49.8**	**96.4**
	ImpHMR (Ours) w. 3DPW	74.3	45.4	87.1

- Quantitative results (3DPW-OCC)

Method	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Zhang *et al.* [63]	-	72.2	-
HMR-EFT [17]	94.4	60.9	111.3
PARE [24]	90.5	56.6	107.9
ImpHMR (Ours)	**86.5**	**54.4**	**104.7**

- Qualitative results



Results

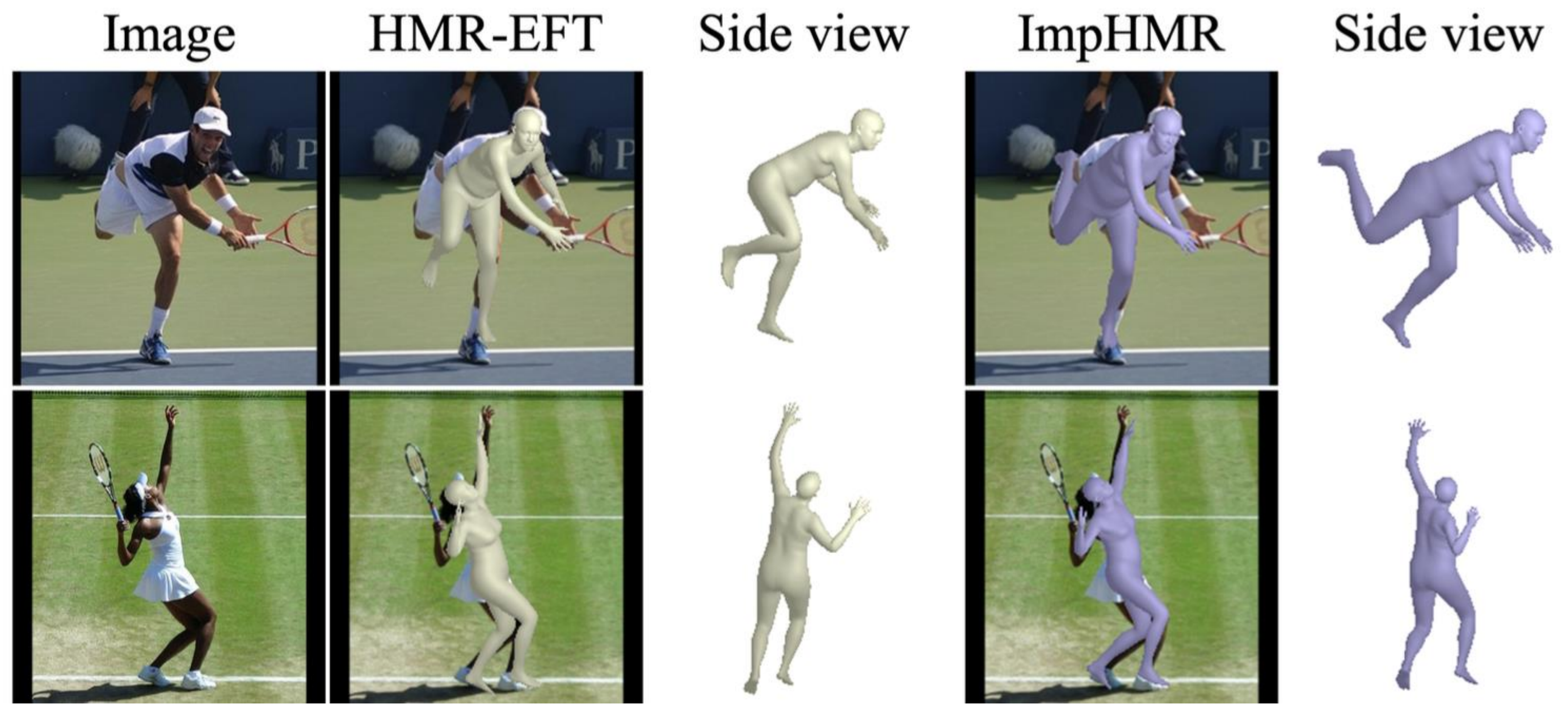
- Quantitative results (3DPW)
  - 8.1% improv. in PA-MPJPE

	Method	3PDW MPJPE ↓	PA-MPJPE ↓	PVE ↓
Temporal	HMMR [20]	116.5	72.6	139.3
	DSD [50]	-	69.5	-
	Arnab *et al.* [2]	-	72.2	-
	Doersch *et al.* [11]	-	74.7	-
	VIBE [23]	93.5	56.5	113.4
	TCMR [8]	95.0	55.8	111.3
	MPS-Net [55]	91.6	54.0	109.6
Frame-based	HMR [19]	130.0	76.7	-
	GraphCMR [26]	-	70.2	-
	SPIN [25]	96.9	59.2	116.4
	PyMAF [62]	92.8	58.9	110.1
	I2L-MeshNet [39]	100.0	60.0	-
	ROMP [49]	89.3	53.5	105.6
	HMR-EFT [17]	-	54.2	-
	PARE [24]	82.9	52.3	99.7
	ImpHMR (Ours)	**81.8**	**49.8**	**96.4**
	ImpHMR (Ours) w. 3DPW	74.3	45.4	87.1

- Quantitative results (3DPW-OCC)

Method	MPJPE ↓	PA-MPJPE ↓	PVE ↓
Zhang *et al.* [63]	-	72.2	-
HMR-EFT [17]	94.4	60.9	111.3
PARE [24]	90.5	56.6	107.9
ImpHMR (Ours)	**86.5**	**54.4**	**104.7**

- Qualitative results



- Results from different views

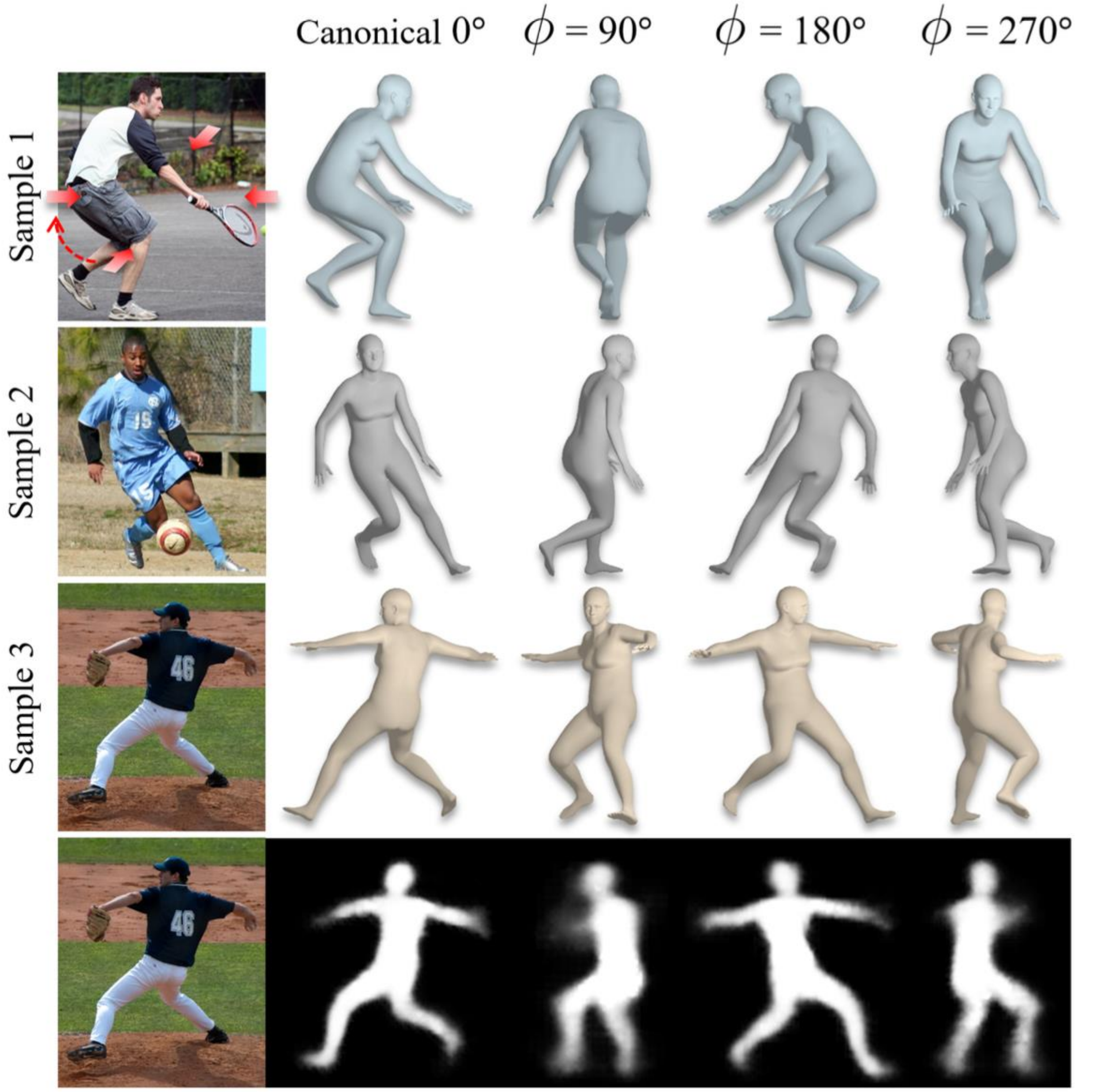


Figure 7. **Inferred SMPL mesh and silhouettes viewed from different viewing directions.** Results inferred *by changing the viewing direction* clockwise by 90° from canonical viewing direction. Note that the inference results are *not by rotating the mesh* inferred from the canonical viewing direction, but *directly inferring a person viewed from different directions in 3D space*.

Thank you!