2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

A Knowledge Distillation Architecture for Pressure Based In-Bed Human Body Reconstruction

Wentao Ni[1] Chen Lei[1] Liangjing Yang[1]

[1] Zhejiang University, China

Abstract. This paper addresses the critical challenge of supine human mesh reconstruction in clinical monitoring scenarios through an innovative knowledge distillation framework. Confronting the inherent limitations of pressure sensor data—including limb occlusion artifacts and limited 3D expressiveness—we propose a hierarchical teacher-student architecture that synergistically integrates cross-modal knowledge from visual domain expertise. Our method leverages a pre-trained CLIFF model as the teacher to guide pressure-map student networks (ResNet variants) in estimating SMPL body parameters. The framework achieved 2%~4% error reduction across key metrics. This work proposes a new solution to optimize pressure-based human body reconstruction and multimodal datasets utilization.

Keywords: Human Body Reconstruction, Knowledge Distillation, Multimodal Data Fusion.

1 Introduction

The estimation and reconstruction of supine human postures are progressively establishing itself as a pivotal research frontier within the domain of computer vision and biomechanical modeling domains, driving a paradigm shift profoundly aligned with the escalating global exigencies for optimized clinical care protocols. Such developments are necessitated by demographic aging trajectories and epidemiological transitions toward chronic disease predominance. Conventional optical modalities employing RGB-D cameras[20], infrared imaging arrays[21], or depth-sensing apparatus[3] confront intrinsic limitations stemming from textile-induced occlusions and photometric variability, compounded by non-trivial ethical considerations regarding psychological discomfort and confidentiality breaches inherent in continuous visual surveillance. Alternative monitoring paradigms leveraging wearable inertial measurement units, while operationally feasible, introduce iatrogenic risks including restricted mobility patterns and potential dermatological complications from prolonged device contact. These multifaceted constraints have precipitated scholarly investigations into unobtrusive pressure-sensitive sensor matrices as viable instrumentation for biomechanical monitoring, wherein distributed piezoresistive transducers quantify interface pressure topography to generate spatiotemporal pressure maps serving as foundational datasets for subsequent volumetric anatomical reconstructions.

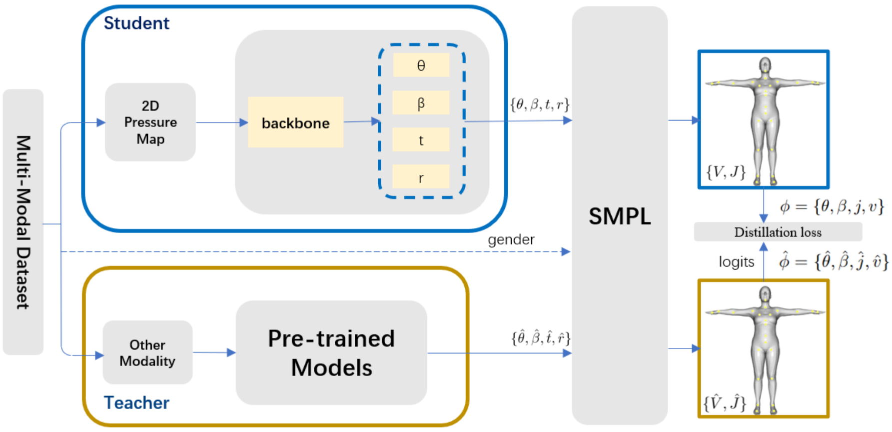

Fig. 1. Workflow of the knowledge distillation architecture

Despite the advancements, critical challenges persist regarding the inherent representational paucity of unidimensional pressure datasets. Singular pressure mappings merely encapsulate bidimensional contact force distributions at the mattress interface, inherently incapable of resolving anatomical occlusions from limb superposition or capturing three-dimensional musculoskeletal configurations. Monomodal training regimens confined to pressure datasets consequently engender model architectures deficient in holistic anatomical comprehension. Furthermore, the comparative scarcity of supine-oriented pressure repositories contrasts markedly with the maturing of multiview photogrammetric datasets within the computer vision research community. This dichotomy necessitates innovative cross-modal knowledge transfer methodologies, particularly given the complementary spatial perspectives afforded by ceiling-mounted optical systems (providing craniocaudal visual coverage) versus floor-embedded pressure grids (delivering ventrodorsal force measurements), thereby enabling synergistic data fusion to compensate for individual modality limitations.

To address these challenges, we propose a knowledge distillation architecture with modularized components specifically engineered for supine human mesh recovery from pressure topology inputs. The framework implements a dual-stage processing pipeline: initially transforming bidimensional pressure tensors into latent pose and shape descriptors through deep convolutional encoders, subsequently reconstructing differentiable human meshes via parametric body models. A hybrid teacher-student configuration is integrated to facilitate cross-modal feature distillation, where the teacher network, pretrained on other datasets, transfers geometrically enriched feature representations to the pressure-oriented student model through attention-guided regularization. This frame also integrates the SMPL (Skinned Multi-Person Linear) model, a differentiable humanoid parameterization that algebraically maps pose $\theta \in \mathbb{N}^{72}$ and shape $\beta \in \mathbb{N}^{10}$ coefficients to vertex-wise mesh coordinates $V \in \mathbb{N}^{6890\times3}$ with joints $J \in \mathbb{N}^{24\times3}$, and finally generate a mesh to represent human body.

During training stage, the teacher and student work on the same dataset to extract pose and shape features, teachers' output $\{\theta, \beta\}$ and its corresponding human mesh $\{V, J\}$form logits and is fed to the student as soft labels. The workflow is as shown in Fig.1, where global translation ($t \in \mathbb{N}^3$) and rotation ($r \in \mathbb{N}^3$) parameters are extracted for spatial calibration.

This architecture is promisingly instructive in the field of human body reconstruction in special scenarios. Our main contribution is to propose a knowledge distillation architecture to improve the performance of small and simple models by reducing error metrics up to 3.94% on MPJPE and 3.82% on V2V.

2 Related Works

2.1 Knowledge Distillation

The knowledge distillation method was first proposed by Hinton et al.[7] as a fundamental paradigm for model compression. The methodology introduces temperature-scaled soft targets from cumbersome teacher networks to guide lightweight student models, effectively transferring dark knowledge through softened probability distributions.

Gou et al.[4] developed a comprehensive survey systematically where they categorizes knowledge distillation approaches through four orthogonal dimensions: knowledge representation (response-based, feature-based and relation-based), training paradigms (offline, online and self-distillation), architectural configurations (multi-teacher ensembles, teaching assistant networks, and cross-modal distillation) and optimization strategies (adversarial distillation, curriculum-based scheduling, and quantization-aware training). The survey further identifies critical challenges including capacity mismatch and data scarcity, while outlining emerging directions such as data-free distillation through generative adversarial networks and neural architecture search-optimized student models.

2.2 Human Mesh Recovery

Human Body Models As a pioneering parametric human model, SCAPE[1] established the first data-driven framework (based on 3D scan data) to correlate body morphology with pose variations, laying the foundation for subsequent research. However, its triangle-based deformation approach suffered from computational inefficiency and joint distortion artifacts. SMPL[12] revolutionized the field by innovatively decoupling pose and shape parameters while adopting linear blend skinning, significantly improving real-time performance and geometric fidelity. Its modular design further spawned specialized variants like SMPL-II (hand-enhanced)[14] and SMPL-X (facial expression support)[13]. Building upon SMPL's parametric framework, SKEL[8] introduced a biomechanically accurate skeletal system with anatomical constraints to optimize joint kinematics, addressing SMPL's limitations in applications requiring physiological plausibility, such as medical rehabilitation and motion analysis. These parametric models collectively advanced human reconstruction from "geometrically plausible" to "physiologically valid", marking a paradigm shift in the field.

HMR Researches On the basis of SMPL and its derived models, researchers have advanced in HMR. CLIFF's[9] primary contribution in the field of 3D human reconstruction lies in resolving perspective ambiguity and scale indeterminacy in monocular image-based reconstruction. The method significantly enhances pose and shape parameter estimation accuracy under single-view conditions by normalizing 2D image features into the 3D camera coordinate system and integrating global contextual information through feature fusion. Particularly effective in scenarios with severe occlusions and unconventional viewpoints, its innovative architecture mitigates limb distortions caused by projection ambiguities in conventional approaches, providing enhanced geometric consistency for monocular vision applications. Addressing the scarcity of annotated pressure data, PressureNet[2] leverages physics-based simulations to generate synthetic pressure maps correlated with SMPL body configurations. A temporal encoder-decoder architecture then maps pressure sequences to 3D meshes, enabling pose estimation without visual privacy intrusions. To mitigate artifacts in single-view reconstruction, Zhang et al.[20] introduce a hierarchical feature fusion mechanism that progressively refines mesh predictions using multiscale contextual cues and propose PyMAF. Its iterative feedback loop between local feature pyramids and global mesh parameters effectively resolves ambiguities arising from self-occlusions and perspective distortions. Tandon et al.[15] proposed BodyMAP which addresses occlusion challenges in multiview reconstruction by integrating semantic segmentation masks with pose estimation. It employs adaptive feature alignment across viewpoints, dynamically weighting visual evidence based on per-joint visibility confidence, thereby improving reconstruction fidelity in cluttered environments. Wu et al.[16] pioneers a temporal pressure sensing framework for 3D human shape estimation in bedridden scenarios, addressing persistent occlusion challenges faced by visual modalities. By introducing a spatiotemporal convolutional network to model dynamic pressure sequences and integrating SMPL-based biomechanical constraints, the method infers anatomically plausible poses even under severe bedding occlusion. To bridge the real-synthetic domain gap, it employs physics-guided pressure simulation paired with SMPL annotations, enabling robust training with limited clinical data. The release of synthetic pressure datasets with biomechanical annotations further catalyzes research in tactile-based human reconstruction. Building upon SKEL, HSMR[17] incorporates biomechanical constraints into the optimization pipeline, including joint rotation limits and soft tissue collision avoidance. This hybrid data-driven and physics-based approach enables applications requiring physiological plausibility, such as rehabilitation monitoring and ergonomic assessment. Complementary works[5][18][19] explore the fusion of pressure data with visual modalities (RGB/depth) through knowledge distillation. By transferring geometric priors from vision-based teacher models to pressure-driven student networks, these methods alleviate information loss caused by limb occlusion in single-pressure-map observations.

3 Method

Our work proposes a knowledge distillation framework that enhances pressure map-based student models by assimilating expertise from teacher models trained on multi-modal data sources. The teacher-student architecture operates through coordinated parameter estimation pipelines: both models extract human pose and shape parameters from input data, which are subsequently fed into parametric human body models to generate corresponding 3D mesh representations.

3.1 Teacher-Student Architecture

Teacher Model Our framework employs a CLIFF model pre-trained on bedridden scenarios to process RGB inputs[10], extracting SMPL-compatible pose ($\theta \in \mathbb{N}^{72}$) and shape ($\beta \in \mathbb{N}^{10}$) parameters through its optimized regression head.

Student Model Constructed by established vision backbones (ResNet18/34/50[6], ConvNext[11]) with parallel feature extraction branches. These CNN-based architectures process pressure maps to produce intermediate tensors, which are then transformed through dedicated linear layers with ReLU activation functions into parameters matching SMPL's dimensional requirements and global translation ($t \in \mathbb{N}^3$) and rotation ($r \in \mathbb{N}^3$) parameters for spatial calibration.

3.2 Human Body Model

In this work we employ the SMPL model[12] for its established prevalence and simplicity. The SMPL pipeline operates through two independent parameter vectors: pose parameter ($\boldsymbol{\theta} \in \mathbb{N}^{72}$) governing skeletal rotations, and shape parameters ($\boldsymbol{\beta} \in \mathbb{N}^{10}$) encoding principal body morphology variations. These parameters are decoded from pressure map inputs through our network, subsequently driving the SMPL skinning function:

$$\mathcal{M}(\boldsymbol{\theta}, \boldsymbol{\beta}, g) = \mathrm{W}(T_P(\boldsymbol{\beta}, \boldsymbol{\theta}), J(\boldsymbol{\beta}), \boldsymbol{\theta}, \mathcal{W}, g) \tag{1}$$

where W denotes the skinning function, T_P the pose-corrective blend shapes, J joint locations, and $\mathcal{W}$ blend weights, $\boldsymbol{\theta}$ pose and $\boldsymbol{\beta}$ shape parameters, and g specify the gender.

3.3 Loss Design

During the training stage, with the features $\phi = \{\theta, \beta, j, v\}$ and $\hat{\phi} = \{\hat{\theta}, \hat{\beta}, \hat{j}, \hat{v}\}$ respectively from student and teacher we set the loss function $\mathcal{L}$ as follows:

$$\mathcal{L}(\phi, \hat{\phi}) = \lambda_\theta L_\theta + \lambda_\beta L_\beta + \lambda_j L_j + \lambda_v L_v \tag{2}$$

where L_θ, L_β, L_j are the MSE loss defined by

$$L_{i\in\{\theta,\beta,j\}} = \frac{1}{N_i}\sum(\mathrm{i} - \hat{\imath})^2 \tag{3}$$

and L_v is the MAE loss defined by

$$L_v = \frac{1}{N_v}\sum|v - \hat{v}| \tag{4}$$

The weights $\Lambda = [\lambda_\theta, \lambda_\beta, \lambda_j, \lambda_v]$ are [1.000,0.001,1.000,5.000].

4 Experiments

4.1 Datasets

We evaluate our framework on two benchmark datasets:
BodyPressureSD[3]: A synthetic pressure-map dataset with SMPL-annotated human poses.
SLP[10]: A real-world pressure sensing dataset exhibiting significant cross-modality discrepancies compared to synthetic counterparts.

Given the absence of SMPL ground truth in SLP, we leverage existing SMPL parameter annotations from BodyPressureSD through cross-dataset label alignment[3]. Specifically, we establish biomechanical correspondence between pressure distribution patterns and SMPL pose parameters via inverse kinematics optimization. Our protocol is designed as follows:
Training: 70 subjects from BodyPressureSD (synthetic domain).
Testing: 22 subjects from SLP (real-world domain)

4.2 Metrics

We adopt two pairs of metrics during evaluation stage: MPJPE (Mean Per Joint Position Error), V2V (Vertex to Vertex) and their Procrustes-Aligned version (e.g., PA-MPJPE and PA-V2V). These widely used metrics align with SMPL's parametric design and clinical requirements. MPJPE directly quantifies pose parameter accuracy by measuring joint position errors, while PA-MPJPE isolates geometric pose consistency after Procrustes alignment, critical for assessing biomechanical plausibility in variable patient orientations. V2V evaluates overall surface reconstruction quality influenced by both pose and shape parameters, capturing subtle anatomical details. PA-V2V further removes global transformations to focus on local anatomical fidelity, essential for pressure-based applications where tissue deformation patterns determine clinical outcomes. This dual-pair metric strategy comprehensively addresses SMPL's disentangled parameter space (pose vs. shape) while balancing global alignment and local geometric precision.

4.3 Evaluation Results

We devise two complementary experimental configurations to systematically evaluate the performance enhancement of our knowledge distillation framework across diverse student architectures. One group employs CLIFF as its teacher model and the other doesn't. The student model is selected from ResNet variants with different depths (ResNet18, ResNet34, ResNet50). We also evaluate ResNet34 with a Neutral SMPL model to investigate the interplay between anatomical priors and cross-modal knowledge transfer. In Table 1 we present the results.

Table 1. Evaluations of various student model and SMPL model settings w/ or w/o knowledge distillation, (*) indicates the decrement of the corresponding error metric

Teacher	Student	Metrics			
		MPJPE	V2V	PA-MPJPE	PA-V2V
None	ResNet18	97.29	121.65	88.83	111.67
	ResNet34	91.36	113.92	83.52	105.12
	ResNet50	83.88	105.53	76.01	97.22
	ResNet34 (Neutral SMPL)	74.57	95.14	68.48	88.96
CLIFF	ResNet18	93.46 (3.94%)	117.00 (3.82%)	85.08 (4.22%)	107.32 (3.90%)
	ResNet34	89.32 (2.23%)	110.87 (2.67%)	80.60 (2.67%)	101.47 (3.47%)
	ResNet50	81.48 (2.86%)	101.67 (3.66%)	73.41 (3.42%)	93.09 (4.25%)
	ResNet34 (Neutral SMPL)	73.83 (0.99%)	93.85 (1.36%)	66.26 (3.24%)	86.25 (3.05%)

We can observe that knowledge distillation enhance these models by decreasing the error metrics around 2%~4%.

5 Conclusion

Prior pressure-based human mesh recovery (HMR) methods have been constrained by inherent performance limitations, significant cross-device variability in sensor configurations, and prohibitive hardware costs associated with high-resolution pressure sensing systems. While multimodal approaches integrating pressure data with visual modalities (e.g., RGB/depth) attempt to mitigate these issues. Our work proposed an framework with knowledge distillation to enhance existing models' performance on pressure-based human body reconstruction tasks, restricted in in-bed scenarios. The knowledge distillation operation achieves 2%~4% decrement among selected metrics.

Given the rapid advancements in knowledge distillation and biomechanical modeling, the teacher-student architecture, and its integrated human body model demand critical updates to maintain state-of-the-art performance. Our future work aims to systematically investigate cross-component adaptation efficacy—particularly parameter space alignment between evolving teacher models (e.g., vision-language foundation models), next-generation student architectures (neural implicit representations), and anatomically informed human models (biomechanically constrained SMPL variants). This research trajectory works towards the establishment of an optimization protocol balancing model accuracy, computational efficiency, and clinical deploy ability, ultimately delivering an FDA-compliant, edge-computing-enabled product for real-world in-bed patient monitoring scenarios.

References

1. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Davis, J., Rodgers, J.: SCAPE: Shape completion and animation of people
2. Clever, H.M., Erickson, Z., Kapusta, A., Turk, G., Liu, C.K., Kemp, C.C.: Bodies at rest: 3d human pose and shape estimation from a pressure image using synthetic data. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp.6214–6223. IEEE. https://doi.org/10.1109/CVPR42600.2020.00625, https://ieeexplore.ieee.org/document/9157114/
3. Clever, H.M., Grady, P., Turk, G., Kemp, C.C.: BodyPressure–inferring body pose and contact pressure from a depth image. https://doi.org/10.48550/arXiv.2105.09936, http://arxiv.org/abs/2105.09936
4. Gou, J., Yu, B., Maybank, S.J., Tao, D.: Knowledge distillation: A survey 129(6), 1789–1819. https://doi.org/10.1007/s11263-021-01453-z, http://arxiv.org/abs/2006.05525
5. Gu, Z., Zhao, Z.Q., Ding, H., Shen, H., Zhang, Z., Huang, D.S.: Cross-domain knowledge distillation for low-resolution human pose estimation. https://doi.org/10.48550/arXiv.2405.11448, http://arxiv.org/abs/2405.11448
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015), https://arxiv.org/abs/1512.03385
7. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. https://doi.org/10.48550/arXiv.1503.02531, http://arxiv.org/abs/1503.02531
8. Keller, M., Werling, K., Shin, S., Delp, S., Pujades, S., Liu, C.K., Black, M.J.: From skin to skeleton: Towards biomechanically accurate 3d digital humans 42(6), 1–12. https://doi.org/10.1145/3618381, https://dl.acm.org/doi/10.1145/3618381
9. Li, Z., Liu, J., Zhang, Z., Xu, S., Yan, Y.: CLIFF: Carrying location information in full frames into human pose and shape estimation. https://doi.org/10.48550/arXiv.2208.00571, http://arxiv.org/abs/2208.00571
10. Liu, S., Huang, X., Fu, N., Li, C., Su, Z., Ostadabbas, S.: Simultaneously-collected multimodal lying pose dataset: Enabling in-bed human pose monitoring. IEEE Transactions on Pattern Analysis and Machine Intelligence 45(1), 1106–1118 (2023). https://doi.org/10.1109/TPAMI.2022.3155712
11. Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. CoRR abs/2201.03545 (2022), https://arxiv.org/abs/2201.0354512.
12. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: a skinned multi-person linear model 34(6), 1–16. https://doi.org/10.1145/2816795.2818013, https://dl.acm.org/doi/10.1145/2816795.2818013
13. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3D hands, face, and body from a single image. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 10975–10985 (2019)

14. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) 36(6) (Nov 2017)
15. Tandon, A., Goyal, A., Clever, H.M., Erickson, Z.: BodyMAP - jointly predicting body mesh and 3d applied pressure map for people in bed
16. Wu, Z., Xie, F., Fang, Y., Liang, Z., Wan, Q., Xiong, Y., Cai, X.: Seeing through the tactile: 3d human shape estimation from temporal in-bed pressure images 8(2), 1–39. https://doi.org/10.1145/3659612, https://dl.acm.org/doi/10.1145/3659612
17. Xia, Y., Zhou, X., Vouga, E., Huang, Q., Pavlakos, G.: Reconstructing humans with a biomechanically accurate skeleton. https://doi.org/10.48550/arXiv.2503.21751, http://arxiv.org/abs/2503.21751
18. Yang, Z., Zeng, A., Yuan, C., Li, Y.: Effective whole-body pose estimation with two-stages distillation. In: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp.4212–4222. IEEE. https://doi.org/10.1109/ICCVW60793.2023.00455, https://ieeexplore.ieee.org/document/10350450/
19. Zhang, F., Zhu, X., Ye, M.: Fast human pose estimation. https://doi.org/10.48550/arXiv.1811.05419, http://arxiv.org/abs/1811.05419
20. Zhang, H., Tian, Y., Zhou, X., Ouyang, W., Liu, Y., Wang, L., Sun, Z.: Py-MAF: 3d human pose and shape regression with pyramidal mesh alignment feedback loop. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 11426–11436. IEEE. https://doi.org/10.1109/ICCV48922.2021.01125, https://ieeexplore.ieee.org/document/9711286/
21. Zhu, Y., Xiao, M., Xie, Y., Xiao, Z., Jin, G., Shuai, L.: In-bed human pose estimation using multi-source information fusion for health monitoring in real-world scenarios 105, 102209. https://doi.org/10.1016/j.inffus.2023.102209, https://linkinghub.elsevier.com/retrieve/pii/S1566253523005250