BodyMAP - Jointly Predicting Body Mesh and 3D Applied Pressure Map for People in Bed

Supplementary Material

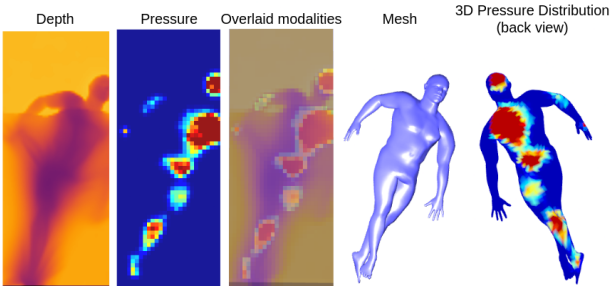

Figure 5. Depth and pressure image complement each other. Image of overlaid modalities (third column) depicts the enhanced context available to the model.

7. Input Modalities

BodyMAP uses as inputs the depth image and the corresponding pressure image for an individual in bed. It jointly predicts both the body mesh (3D pose & shape) and the pressure applied on the body, in the form of a 3D pressure map. As illustrated in Fig. 5, the depth image provides a top-to-bottom view, and the pressure image provides a 'bottom-to-up' view, capturing complementary features of the body. Using these two visual modalities enhances the model's context, allowing it to accurately predict the body mesh and 3D pressure map, even when the individual is heavily covered with blankets.

8. BodyMAP

8.1. BodyMAP-PointNet

BodyMAP-PointNet substantially surpasses prior methods by 25% on both body mesh and 3D pressure map prediction tasks, for in-bed, visually occluded people. Illustrated in Fig. 3, the model uses uses a ResNet18 [15] and MLP to predict the SMPL [29] parameters $\hat{\Psi}$. These parameters include body shape, joint angles, root-joint translation, and root-joint rotation: $\hat{\Psi} = [\hat{\beta}, \hat{\Theta}, \hat{s}, \hat{x}, \hat{y}]$. We use the SMPL embedding block [19] to obtain the SMPL [29] body mesh. The model then leverages the proposed Feature Indexing Module to form features for each vertex of the predicted mesh. Subsequently, it uses a PointNet [35] model to infer the 3D pressure map. This new method unifies the design for jointly predicting body mesh and 3D pressure map. The feature extractors used in our model, ResNet and PointNet, are easily replaceable with other modern feature extractors.

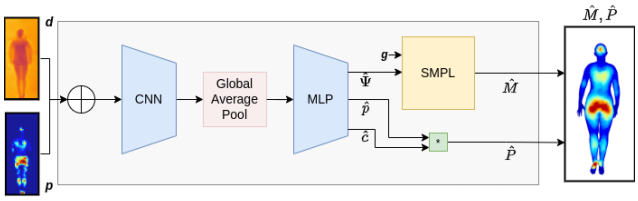

Figure 6. BodyMAP-Conv jointly predicts body mesh and 3D applied pressure map for an individual in bed. The model replaces the FIM and PointNet components of BodyMAP-PointNet, and instead predicts the 3D pressure map using an MLP.

8.2. BodyMAP-Conv

We develop a baseline method denoted as BodyMAP-Conv, depicted in Fig. 6. BodyMAP-Conv replaces FIM and PointNet with fully connected layers to predict the 3D pressure map jointly with the SMPL parameters. This model design allows us to evaluate the enhancements provided by FIM and PointNet for 3D pressure map prediction.

8.3. Training Strategy

We train BodyMAP-PointNet and BodyMAP-Conv to jointly predict the body mesh and 3D pressure map with the following loss function:

$$L = \mathcal{L}_{\text{SMPL}} + \lambda_1 \mathcal{L}_{\text{v2v}} + \lambda_2 \mathcal{L}_{\text{P3D}} + \lambda_3 \mathcal{L}_{\text{contact}} \quad (3)$$

Here, $\mathcal{L}_{\text{SMPL}}$ minimizes the absolute error on SMPL parameters $\hat{\Psi}$ and squared error on the 3D joint positions.

$$\mathcal{L}_{\text{SMPL}} = \frac{1}{N_\beta \sigma_\beta}\|\hat{\beta} - \beta\|_1 + \frac{1}{N_\Theta \sigma_\Theta}\|\hat{\Theta} - \Theta\|_1 + \frac{1}{6\sigma_y x}(\|\hat{x} - x\|_1 + \|\hat{y} - y\|_1) + \frac{1}{N_s \sigma_s}\sum_{j=1}^{N_s}\|\hat{s}_j - s_j\|_2 \quad (4)$$

where N_β = 10 body shape parameters, N_Θ = 69 joint angle parameters, x and y represent global rotation, and N_s = 24 joint positions in 3D with s_j representing each joint position. Each term is normalized by their respective standard deviations computed over the entire synthetic dataset. [9].

$\mathcal{L}_{\text{v2v}}$ (vertex-to-vertex loss) minimizes the Euclidean error between the predicted and ground truth vertex positions, normalized by the standard deviation σ_v calculated over the entire synthetic data, and is calculated as:

$$\mathcal{L}_{\text{v2v}} = \frac{1}{N_v \sigma_v} \sum_{j=1}^{N_v} ||\hat{v}_j - v_j||_2 \tag{5}$$

Here N_v represents the number of vertices for the body mesh which is equal to 6890 for SMPL [29] meshes.

$\mathcal{L}_{\text{P3D}}$ (3D pressure map loss) minimizes the squared error between predicted and ground truth per-vertex pressure value, normalized by the standard deviation of ground truth pressure map σ_p calculated over the entire synthetic data, and is modeled as:

$$\mathcal{L}_{\text{P3D}} = \frac{1}{N_v \sigma_p} \sum_{j=1}^{N_v} ||\hat{p}_j - p_j||_2 \tag{6}$$

$\mathcal{L}_{\text{contact}}$ is calculated as a cross-entropy loss between the predicted and ground truth per-vertex binary contact values, where the ground truth contact is obtained from all non-zero elements of ground truth 3D pressure maps.

$$\mathcal{L}_{\text{contact}} = \frac{1}{N_v \sigma_c} \sum_{j=1}^{N_v} -(y_j \log(p_j) + (1 - y_j) \log(1 - p_j)) \tag{7}$$

Here y_j represents the ground truth contact and p_j represents the predicted probability of contact value, calculated for each vertex. σ_c represents the standard deviation of ground truth contact calculated over the entire synthetic data [9].

Hyperparameters: We empirically set λ_1 = 0.25, λ_2 = 0.1, and λ_3 = 0.1 to train the models, BodyMAP-PointNet and BodyMAP-Conv, with the loss function defined in Eq. (3)

9. BodyMAP-WS

We introduce BodyMAP-WS, designed to implicitly learn 3D pressure maps without using direct supervision. BodyMAP-WS leverages a pre-trained body mesh regressor network and trains to predict a 3D pressure map. The mesh regressor network used in our work is depicted in Fig. 7(a). Using the mesh estimates, BodyMAP-WS leverages FIM and PointNet to predict the 3D pressure map. Unlike BodyMAP-PointNet which predicts the 3D pressure map as a product of the per-vertex pressure and binary contact value, BodyMAP-WS, depicted in Fig. 7(b), directly predicts the 3D pressure map modeled by the per-vertex pressure value.

For training, we form a differentiable 2D projection of the predicted 3D pressure map and align it with the input pressure image by minimizing loss between the formed projection and the input pressure image. This alignment allows the model to learn the correct 3D pressure map without any supervision with ground truth data. This training methodology allows the model to learn from unlabelled data, where collecting labels can be challenging and costly.

9.1. Training strategy

We pre-train the mesh regressor for BodyMAP-WS with the following loss function:

$$L = \mathcal{L}_{\text{SMPL}} + \lambda_1 \mathcal{L}_{\text{v2v}} \tag{8}$$

The mesh regressor model is then frozen and is used to train BodyMAP-WS with the loss function:

$$L = \mathcal{L}_{\text{P2D}} + \lambda_1 \mathcal{L}_{\text{Preg}} \tag{9}$$

Here, $\mathcal{L}_{\text{P2D}}$ (2D pressure loss) minimizes the squared error between the 2D projection of the predicted 3D pressure map and the input pressure image.

$$\mathcal{L}_{\text{P2D}} = \|\hat{p} - p\|_2 \tag{10}$$

where p is the input pressure image and $\hat{p}$ is the 2D projection of the predicted 3D pressure map.

We model $\mathcal{L}_{\text{Preg}}$ (pressure regularization loss) as a regularization loss to penalize pressure predicted on vertices that lie above the mattress plane, as these are not in contact with the mattress.

$$\mathcal{L}_{\text{Preg}} = \| \sum_{j=1}^{N_{\text{nocv}}} \hat{p_j} \|_2 \tag{11}$$

where p_j represents pressure on a vertex not in contact with the mattress, N_{nocv} represents the number of no-contact-vertices, i.e. vertices lying above the mattress plane ($Z = 0$) and is obtained as:

$$V_{\text{nocv}} = \{V | V_z > 0\} \tag{12}$$

$$N_{\text{nocv}} = |V_{\text{nocv}}| \tag{13}$$

Hyperparameters: For pre-training the mesh regressor we set λ_1 = 0.25 in Eq. (8). To train BodyMAP-WS, we set λ_1 = 500 in Eq. (9). These hyperparameter values are set empirically.

10. Pressure Metrics Details

To evaluate predicted 3D pressure maps, we have introduced two metrics: v2vP 1EA (one edge away) and v2vP 2EA. These metrics calculate the squared error between the predicted and ground truth 3D pressure at coarser levels. These coarser 3D pressure maps are formed by setting the pressure at a vertex as the average of its neighboring vertices. We illustrate the spread of neighboring vertices used for v2vP 1EA and v2vP 2EA in Fig. 8.

Network	Modalities	3D Pose Error - MPJPE (mm) ↓				3D Pressure Distribution Error - v2vP (kPa^2) ↓			
		Uncovered	Cover 1	Cover 2	Overall	Uncovered	Cover 1	Cover 2	Overall
Pyramid Fusion [45]	RGB-D-PI-IR	78.80	79.92	80.21	79.64	-	-	-	-
BPBnet [9]	D	70.16	76.99	76.49	74.54	2.87	2.88	2.87	2.87
BPWnet [9]	D	63.64	72.4	72.04	69.36	2.85	2.85	2.82	2.84
BodyMAP - Conv	D	55.7 ± 1.753	64.34 ± 2.099	63.82 ± 2.117	61.29 ± 1.978	2.68 ± 0.019	2.66 ± 0.02	2.63 ± 0.017	2.66 ± 0.018
BodyMAP - Conv	D - PI	49.48 ± 0.127	52.9 ± 0.576	53.0 ± 0.521	51.79 ± 0.379	2.59 ± 0.008	2.55 ± 0.007	2.53 ± 0.007	2.56 ± 0.007
BodyMAP - PointNet	D - PI	**48.77 ± 0.981**	**51.88 ± 1.142**	**52.36 ± 1.092**	**51.01 ± 1.071**	**2.16 ± 0.035**	**2.13 ± 0.029**	**2.12 ± 0.026**	**2.14 ± 0.030**

Table 4. Results for 3D pose error (MPJPE) & 3D pressure map error (v2vP) for the 22 test subjects from SLP dataset, over multiple blanket conditions - uncovered (no blanket), cover1 (light blanket), cover2 (heavy blanket) and overall (average case). The modalities used are: depth image (D), pressure image (PI), infrared images (IR) and RGB images. The metrics for BodyMAP models are shown with the mean and standard deviation computed over 3 random runs.

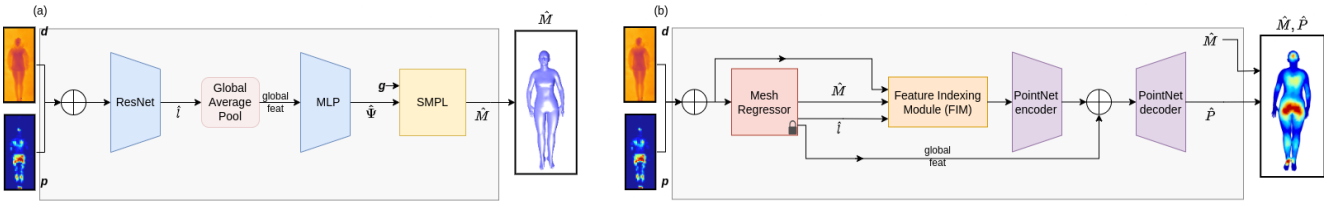

Figure 7. BodyMAP-WS implicitly learns to predict 3D pressure maps without using direct supervision. (a) The mesh regressor model is used as a frozen pre-trained model for BodyMAP-WS. (b) BodyMAP-WS leverages FIM and PointNet to directly predict the 3D pressure map (without predicting per-vertex binary contact like BodyMAP-PointNet).

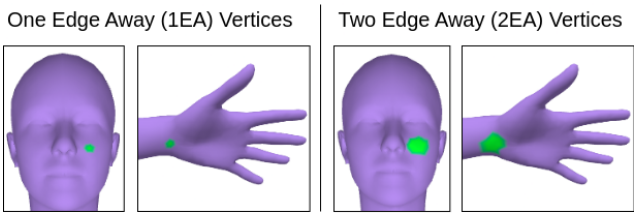

Figure 8. Smoothing region (shown in green color) of two vertices for one edge away (1EA) and two edge away (2EA) scenarios.

BodyMAP - Feature Indexing Module Ablation				3D Pressure Error (kPa^2) ↓		
Vertex Location	Input Image	ResNet Features	Global ResNet features	v2vP	v2vP 1EA	v2vP 2EA
-	✓	-	-	2.765	1.797	1.256
✓	-	-	-	2.478	1.552	1.043
-	✓	-	✓	2.455	1.526	1.033
-	✓	✓	-	2.369	1.419	0.921
✓	✓	-	-	2.334	1.442	0.969
✓	-	-	✓	2.339	1.429	0.945
✓	-	✓	-	2.205	1.311	0.855
✓	-	✓	✓	2.204	1.311	0.857
✓	✓	-	✓	2.195	1.315	0.862
✓	✓	✓	-	2.164	1.280	0.831
✓	✓	✓	✓	**2.130**	**1.260**	**0.822**

Table 5. Ablation study of Feature Indexing Module (FIM) and use of global ResNet features (formed after global average pooling) for 3D pressure map prediction. These results are evaluated for BodyMAP-PointNet model trained with depth and pressure images as input modalities, averaging over all blanket configurations.

11. Further Results

BodyMAP surpasses prior methods in both body mesh and 3D pressure map prediction. Tab. 4 presents detailed results over different blanket thickness configurations, evaluated over the 22 test subjects from the SLP dataset [25, 27].

While BPW [9], the previous state-of-the-art method, exhibits strong performance in body mesh prediction in scenarios where the individual is uncovered (uncovered case), its effectiveness notably diminishes when occlusions from blankets (cover1 and cover2) are introduced. In contrast, our methods, BodyMAP-PointNet and BodyMAP-Conv, when trained on multiple modalities, maintain consistent performance across all three blanket configurations, showcasing a significant improvement over BPW [9]. This evaluation reinforces the intuition that depth images capture limited details about the human body when individuals are covered with blankets. Instead, our approach of leveraging both modalities, with depth images providing the top-down view and 2D pressure images offering a 'bottom-to-up' view, yields robust performance across different blanket configurations.

Feature Indexing Module improves 3D pressure performance: Tab. 5 showcases the impact of different approaches to forming mesh features through the proposed Feature Indexing Module (FIM) and the use of global ResNet features (formed after global average pooling). The proposed fusing of vertex locations, accumulated input image and ResNet features (before global average pooling) through FIM, and global ResNet features, provides a comprehensive feature set for the model to learn from and result in accurate 3D pressure map predictions.

Ablating training components reduces performance: Ablating the $\mathcal{L}_{\text{v2v}}$ and $\mathcal{L}_{\text{contact}}$ loss functions from Eq. (3) in BodyMAP training results in a performance drop in both

body mesh and 3D pressure map prediction. Tab. 6 showcases the outcomes across the test set, where we observe a performance improvement from predicting 3D contact and minimizing loss for contact and vertex locations.

We also ablated training-time data augmentations and observed improvement in results from using random rotation and random erasing augmentations. Tab. 7 outlines these results.

12. Computational Analysis

We conducted a comparative analysis between BodyMAP-PointNet, BodyMAP-Conv, and the prior state-of-the-art method BPW [9], focusing on computational efficiency metrics including FLOP (total floating-point operations), parameter count, and inference time. The summarized results are detailed in Tab. 8.

Parameter count: BPW [9] employs two ResNet34 [15] models to predict SMPL [29] parameters for body mesh formation, alongside a white-box model and a compact convolutional network to estimate the 2D pressure image. This multi-model approach results in BPW [9] containing a total of 42.73 million parameters. In contrast, our models, BodyMAP-Conv and BodyMAP-PointNet, adopt simpler architectures with a reduced parameter count compared to BPW [9].

BodyMAP-Conv utilizes a ResNet18 [15] and fully connected layers to predict jointly the SMPL [29] mesh parameters and a pressure value for each vertex, leading to an output size of 6890 for the entire SMPL [29] mesh, resulting in a total parameter count of 34.03 million.

BodyMAP-PointNet also employs a ResNet18 [15] and fully connected layers for body mesh parameters and per-vertex pressure value prediction but optimizes for efficiency by sharing weights of the fully connected layers across vertices using the PointNet [35] architecture. This design choice results in a highly efficient model with a significantly reduced parameter count of only 13.56 million.

FLOPs: Although BPW [9] uses three separate neural-network models, it only a predicts 2D pressure image and not a per-vertex 3D pressure map through the models. This design choice allows BPW [9] to use a minimal number of fully connected layers, resulting in the lowest total of 1.25 giga FLOPs. In contrast, both BodyMAP-Conv and BodyMAP-PointNet predict a 3D pressure map, employing fully connected layers, resulting in higher total FLOPs.

BodyMAP-Conv utilizes a fully connected layer to predict a pressure value for each of the 6890 body mesh vertices, resulting in a total of 1.8 giga FLOPs.

BodyMAP-PointNet leverages a PointNet [35] model to predict the 3D pressure map. This requires the model to compute mesh features (features for each of the 6890 body mesh vertices) in many layers of the model, resulting in a total of 2.26 giga FLOPs. Despite the higher computational demand during inference, BodyMAP-PointNet's streamlined design results in lower memory requirements, facilitating easier scalability for jointly predicting body mesh and 3D pressure map.

BodyMAP-PointNet loss ablation	3D Pose Error	3D Pressure Map Error
Extra Losses - v2v, contact	MPJPE (mm) ↓	v2vP (kPa^2) ↓
-	55.34	2.182
✓	**52.63**	**2.130**

Table 6. Ablation study of extra losses, $\mathcal{L}_{\text{v2v}}$ (Eq. (5)) and $\mathcal{L}_{\text{contact}}$ (Eq. (7)), used for training BodyMAP-PointNet. Here the model is trained with depth and pressure image as input modalities, with evaluation conducted on the test set from the SLP dataset [25, 27], averaging over all blanket configuration (uncovered, cover1 & cover2)

BodyMAP-PointNet Augmentation ablation		3D Pose Error	3D Pressure Map Error
Modality	Data Augmentation	MPJPE (mm) ↓	v2vP (kPa^2) ↓
D	-	69.73	2.609
D	✓	59.44	2.473
D - PI	-	60.07	2.277
D - PI	✓	**52.63**	**2.130**

Table 7. Ablation study of data augmentations used for training BodyMAP-PointNet. The evaluation is conducted on the test set from the SLP dataset [25, 27], averaging over all blanket configurations (uncovered, cover1 & cover2)

Inference time: BPW [9] approximates the 3D pressure map by projecting pressure from the predicted 2D pressure image to the vertices of the predicted body mesh, necessitating vertex-wise iterations and resulting in an inference time of 3.5168 seconds. In contrast, BodyMAP-Conv and BodyMAP-PointNet predict the 3D pressure map directly, leading to significantly reduced inference times.

BodyMAP-Conv utilizes a fully connected layer to predict the 3D pressure map, resulting in the lowest inference time of 0.0042 seconds.

Employing the PointNet [35] architecture, BodyMAP-PointNet directly predicts the 3D pressure map in only 0.0046 seconds, demonstrating comparable inference time to BodyMAP-Conv.

Analysis: BodyMAP-PointNet demonstrates a substantial 68% reduction in the number of parameters and a 99% decrease in inference time compared to the prior state-of-the-art BPW [9]. These computational advantages translate to simplified scalability and real-time performance, crucial for implementation and deployment in hospitals and nursing facilities.

The joint prediction of body mesh and 3D pressure map for a person in bed by BodyMAP could empower caregivers with visualization of pressure distribution on the body. The low inference time required by BodyMAP allows for generating these visualizations in real-time. This eliminates the need for caregivers to manually assess pressure information, allowing them to prioritize providing care. These benefits could potentially enhance the pressure injury prevention process.

Network	FLOPs (G) ↓	Params (M) ↓	Inference Time (s) ↓
BPW [9]	**1.2469**	42.7316	3.516813
BodyMAP-Conv	1.8044	34.0252	**0.004224**
BodyMAP-PointNet	2.2619	**13.5580**	0.004650

Table 8. Results from model computation comparison. The evaluation is computed on a machine using Intel Core i9-13900KF-3 GHz-24 core CPU processor and NVIDIA GeForce RTX 4090 GPU processor.

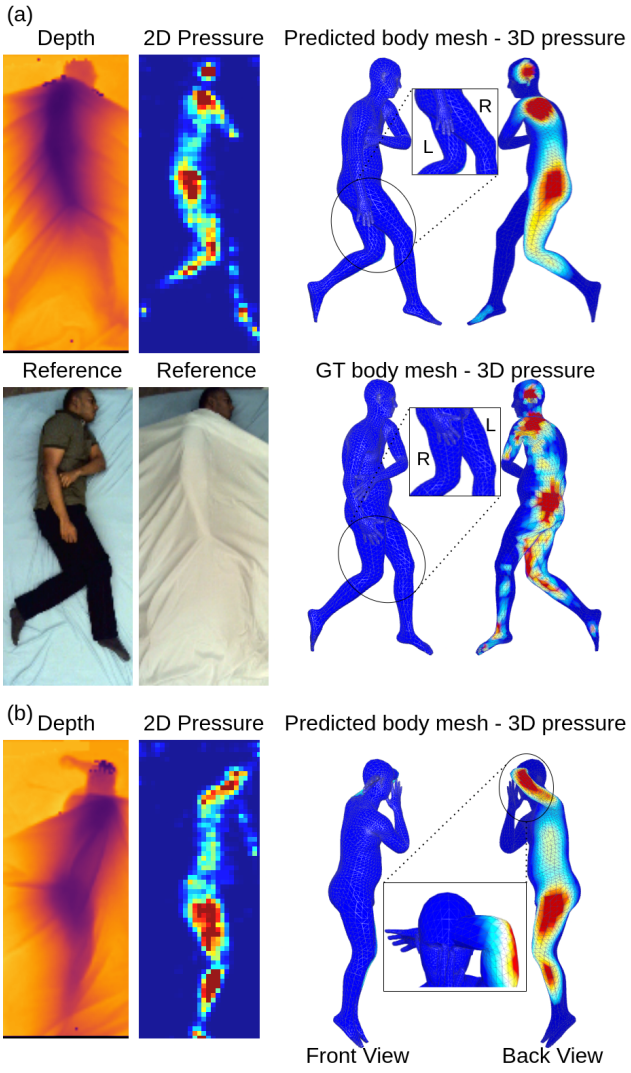

Figure 9. Failure cases of BodyMAP. (a) The Model predicts flipped orientation for legs stemming from inherent body symmetries. (b) Predicted mesh interpenetration failure example.

13. Future Work Opportunities

While BodyMAP demonstrates proficient estimation of body mesh and 3D applied pressure map, its translation into clinical efficacy requires validation through real-world studies. Clinical settings inherently differ from training scenarios, encompassing variations in mattresses, diverse body poses, partially elevated or tilted mattresses for actuated hospital beds, and the presence of additional objects like pillows and medical instruments. It is also crucial to quantify the benefits derived from the body mesh and 3D pressure map predictions & visualizations in improving pressure injury prevention in real clinical scenarios.

Additionally, our method encounters occasional false positives. For instance, in Fig. 9(a), the misalignment of leg orientation, attributed to inherent body symmetries, leads to an inaccurate 3D pressure map. Future enhancements may stem from expanded training datasets.

Furthermore, instances of mesh inter-penetrations can occur in our predictions. As depicted in Fig. 9(b), the left arm penetrates the head. Although incorporating loss terms to mitigate such occurrences is viable, our method currently leverages a fixed open-hand pose similar to BPW [9], contributing to occasional unnatural orientations.

Similar to BPW [9], our current approach fails to account for pressure from self-contact. Integrating methods that consider physics behind self-contact [10, 14, 32] could potentially augment our method's performance.

Our proposed BodyMAP-WS model employs an implicit learning strategy, effectively eliminating the need for ground truth 3D pressure map data. Although BodyMAP-WS requires a pre-trained mesh model, which traditionally necessitates 3D ground truth mesh data, our methodology paves the way for a fully self-supervised model. In this paradigm, both the body mesh and 3D applied pressure map could be learned without relying on any 3D labels. This advancement holds promise for the field, as it opens avenues to truly scale up to large-scale unlabeled data collected through real-world systems.

The optimized architecture of BodyMAP has significantly reduced inference time, enabling real-time analysis. This advancement creates opportunities to train models for predicting body mesh and 3D applied pressure maps on video data. This expanded capability would allow to capture richer insights and further improve the pressure ulcer prevention process.

14. Video Results

We release video results on our project website, visualizing the body mesh with a 3D pressure map to compare our method with ground truth and BPW [9]. The videos enable detailed visual analysis of the predicted pressure along with the predicted body mesh.

Additionally, we provide more visual results comparing our method with ground truth and BPW [9] in Fig. 10 & Fig. 11.

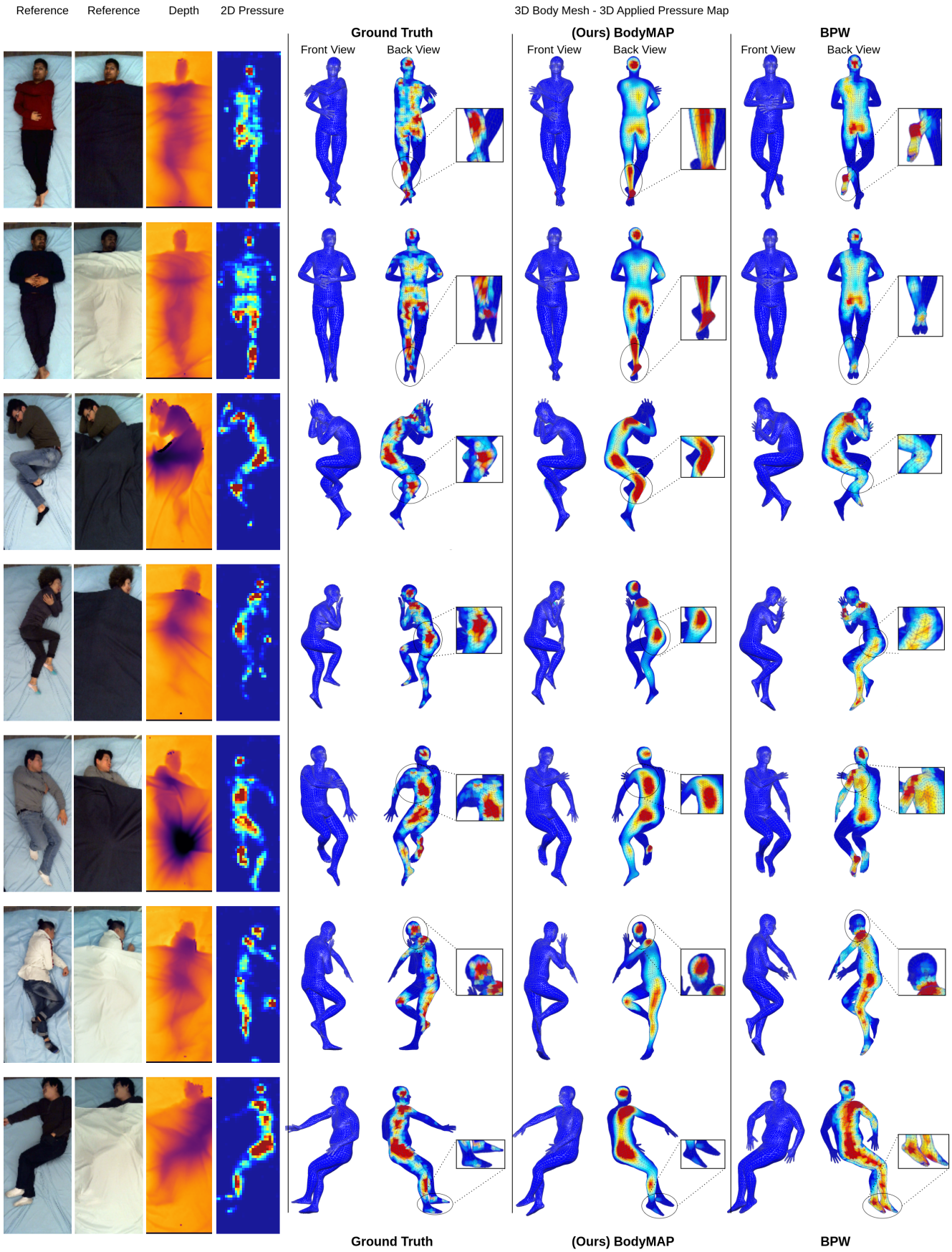

Figure 10. Results of inferring body mesh and 3D applied pressure map from examples in the SLP [25, 27] test set.

Figure 11. More results of inferring body mesh and 3D applied pressure map from examples in the SLP [25, 27] test set.