Supplementary Material for "S2R: Illumination-Aware Synthetic-to-Real Adaptation for Single-Image Human Material Estimation"

Yu Jiang
jiangyu1181@whu.edu.cn
School of Computer Science
Wuhan University
Wuhan, China

Jiahao Xia
jiahao.xia-1@uts.edu.au
Faculty of Engineering and IT
University of Technology Sydney
Ultimo, NSW, Australia

Jianchi Sun
sunjc0306@whu.edu.cn
School of Computer Science
Wuhan University
Wuhan, China

Jiongming Qin
2022102110003@whu.edu.cn
School of Computer Science
Wuhan University
Wuhan, China

Tuo Cao
tcao@iaii.ac.cn
Institute of AI for Industries
Chinese Academy of Sciences
Wuhan, China

Yusen Wang
wangyusen@whu.edu.cn
DFRD Institute
Wuhan, China

Chunxia Xiao
cxxiao@whu.edu.cn
School of Computer Science
Wuhan University
Wuhan, China

1 Implementation Details

Rendering Function. For rendering supervision, we use a physics-based shader to synthesize appearance from the estimated material maps and illumination:

$$L(x) = \int_{\Omega} L_i \, f_r(\mathbf{N}, \mathbf{A}_d, \mathbf{R}, \mathbf{A}_s, \mathbf{S}, w_i, w_o) \, (w_i \cdot \mathbf{N}) \, dw_i. \quad (1)$$

Here, $L(x)$ denotes the radiance reflected at surface point x, L_i is the incident illumination, and f_r denotes the BSDF. We instantiate f_r with the Disney BSDF model [1], which provides a practical and expressive parameterization for skin reflectance, particularly for modeling subsurface scattering effects. The material parameters are represented by $\mathbf{N}$, $\mathbf{A}_d$, $\mathbf{R}$, $\mathbf{A}_s$, and $\mathbf{S}$, corresponding to the surface normal, diffuse albedo, roughness, specular albedo, and subsurface scattering, respectively, while w_i and w_o denote the incident and outgoing directions.

Evaluation Metrics. We evaluate both material estimation and relighting quality across varying lighting conditions. For normals, we use mean angular error (degrees), RMSE, and PSNR. For diffuse albedo, subsurface scattering, and relighting results, we report PSNR, SSIM, and LPIPS to capture both numerical accuracy and perceptual quality. Because the roughness and specular-albedo annotations in OpenHumanBRDF [5] are category-based, while our objective is pixel-based estimation, we interpret their direct pixel-wise metrics with caution and complement them with relighting evaluation in Section 4.2.

Training Details. Our model is implemented in PyTorch and trained on four NVIDIA RTX 3090 GPUs (24 GB memory each) with a batch size of 1. The training procedure consists of two stages. We use OpenHumanBRDF [5] for the first stage because it is constructed from scan-based data, which is of higher quality and closer to real-world distributions than purely CG datasets. We first train the model on OpenHumanBRDF for 200 epochs with a learning rate of 1×10^{-4}, which takes about four days. We then fine-tune the model on the real dataset CosmicMan-HQ 1.0 [9] for 100 epochs using a learning rate of 5×10^{-5}, which requires approximately three additional days. In both stages, we use the AdamW [10] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 0.01.

2 Datasets

OpenHumanBRDF. For supervised training and benchmarking, we employ OpenHumanBRDF, which consists of 147 full-body human subjects. The dataset is divided into 127 identities for training and 20 identities for evaluation. For each subject, high-resolution rendered images of size 1024×1024 are provided from multiple viewpoints, including 100 views for each training identity and 10 views for each test identity. Each image is further associated with a foreground mask and a complete collection of PBR material properties: normal, diffuse albedo, roughness, specular albedo, and subsurface scattering maps.

Synthetic CG Dataset. We also leverage the synthetic dataset released by AFHR [2]. Relative to OpenHumanBRDF, this dataset offers a much larger scale, comprising 2,500 subjects in total, including 2,400 for training and 100 for testing, all rendered at 1024×1024 resolution. Each sample is associated with a foreground mask, normal, diffuse-albedo, roughness, and specular maps, and ground-truth relighting results under HDR environments. Although no explicit subsurface-scattering maps are provided, the included skin mask enables us to approximate subsurface-scattering values in the same manner as HM [5].

CosmicMan-HQ 1.0. For real-data adaptation, we further leverage CosmicMan-HQ 1.0, a large-scale real-world human-image dataset introduced in CosmicMan [9], and construct a filtered full-body subset for fine-tuning to reduce the domain gap between

Table 1: Appearance reconstruction on CosmicMan-HQ 1.0.

Method	PSNR↑	SSIM↑	LPIPS↓
Full w/o Stage II	21.36	0.8668	0.1562
Full	**24.58**	**0.8989**	**0.0704**

Figure 1: Appearance-reconstruction results on CosmicMan-HQ 1.0.

synthetic and real data. The original dataset contains 6M high-resolution single-person images with a mean resolution of 1488 × 1255, covering full-body, half-body, and headshot views. Since our fine-tuning stage focuses on complete full-body humans, we retain 10,000 images with intact full-body subjects, including 9,000 for training and 1,000 for testing.

3 Baselines

We compare S2R against seven recent methods in three categories. Single-image material-estimation methods include RADN [3] and HATSNet [4]. Human-specific material and relighting approaches include SL [7], HM [5], and AFHR [2]. Relighting-driven baselines include FBHR [8] and TR [11]. Except for SL, which is only accessible through its online demo, all competing methods are retrained on both evaluation datasets following their official protocols. For compact qualitative comparisons, we present HATSNet and TR from the first and third categories, respectively, since they deliver the strongest results within their groups.

4 Additional Quantitative Evaluations

4.1 Effect of Stage II on Real Images

CosmicMan-HQ 1.0 does not provide ground-truth PBR material maps or relighting images, so direct material-space evaluation is unavailable. We therefore use an appearance-reconstruction proxy: the estimated materials and illumination are used to re-render the input appearance, and PSNR, SSIM, and LPIPS are computed against the input image within the foreground mask. As shown in Table 1 and Fig. 1, Stage II substantially improves all three metrics. This result supports that real-image adaptation improves the consistency among estimated materials, illumination, and observed appearance, thereby reducing the synthetic-to-real gap. The reconstruction score is a consistency measure rather than direct ground-truth material accuracy.

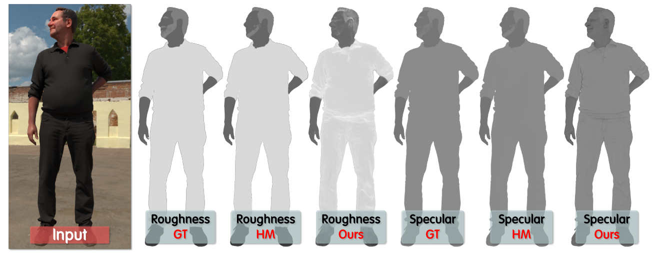

Figure 2: Qualitative comparison of roughness and specular albedo.

4.2 Roughness, Specular Albedo, and Relighting

The ground-truth roughness and specular albedo in OpenHumanBRDF are defined using five material categories, whereas S2R predicts spatially varying, pixel-based maps (Fig. 2). Their representations and optimization objectives therefore differ, and direct pixel-wise metrics do not fully reflect perceptual or physical quality. We nevertheless report these metrics for completeness in Tables 2 and 3. Without direct roughness or specular losses, S2R remains competitive on both channels and obtains the best real- and synthetic-lighting relighting scores among the compared methods. The relighting gains indicate that the predicted material components work coherently in image formation even when individual channel annotations are category-based.

4.3 Lighting Correction

We additionally evaluate lighting correction through appearance reconstruction on OpenHumanBRDF. The variant without correction directly uses the illumination estimated by AFHR, whereas the full variant uses the illumination corrected by our method. Table 4 shows consistent improvements in PSNR, SSIM, and LPIPS after correction. Because this test jointly evaluates estimated material and illumination through rendering, it is evidence of improved image-formation consistency rather than a direct measurement against ground-truth lighting parameters.

4.4 Sensitivity to Noisy Priors

Our method treats predictions from Sapiens, AFHR, and HM as noisy but complementary cues rather than copying them directly. The cues are integrated through adaptive feature fusion and physics-inspired constraints, allowing the network to retain useful information while correcting prior errors. To evaluate robustness, we corrupt the depth, segmentation, normal, and albedo priors by alpha-blending each with random noise using $\alpha = 0.5$. Tables 5–7 show that corrupted priors reduce performance, as expected, but the complete Stage-I model consistently performs best. Together with the "w/o IAFI" variant, these results show that the proposed model benefits from informative priors without simply inheriting their predictions.

5 Comparison with Priors

We exploit multi-source priors for material estimation. Specifically, we adopt Sapiens [6] to predict surface-normal priors. For diffuse albedo, we further integrate prior maps from AFHR [2] and HM [5].

Table 2: Roughness and specular-albedo evaluation on OpenHumanBRDF. Category-based annotations are compared with our pixel-based predictions.

Method	Roughness			Specular Albedo		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
AFHR	0.955	22.92	0.0201	0.958	38.93	0.0211
HM	**0.963**	**23.99**	**0.0160**	**0.989**	**39.78**	**0.0087**
S2R	0.959	23.58	0.0183	0.985	39.42	0.0092

Table 3: Relighting evaluation on OpenHumanBRDF.

Method	Real Relighting			Synthetic Relighting		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
AFHR	0.786	22.27	0.0364	0.811	23.94	0.0399
HM	0.813	22.42	0.0378	0.832	23.97	0.0323
S2R	**0.821**	**23.14**	**0.0335**	**0.844**	**24.17**	**0.0243**

Table 4: Lighting correction evaluated by appearance reconstruction on OpenHumanBRDF.

Method	PSNR↑	SSIM↑	LPIPS↓
S2R w/o Correct	24.82	0.916	0.0391
S2R w/ Correct	**26.33**	**0.943**	**0.0374**

By retaining reliable prior cues while correcting their errors, our method produces more accurate normal and diffuse-albedo estimates than the raw priors, as demonstrated quantitatively in the main paper. We also provide qualitative comparisons with Sapiens on real-world images in Fig. 3. While Sapiens tends to produce over-smoothed normals with missing fine details, our results preserve richer geometric structures, especially on facial regions.

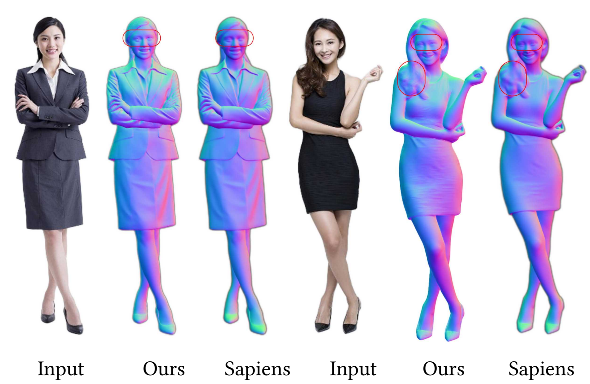

Figure 3: Comparison with Sapiens on real-world images.

6 Additional Qualitative Results

We provide additional qualitative results to further demonstrate the effectiveness of our method. Fig. 4 shows results on the Synthetic CG dataset. Figs. 5 and 6 present results on CosmicMan-HQ 1.0. Fig. 7 reports results on in-the-wild Internet images, and Fig. 8 shows results on captured images. These examples further demonstrate that our method recovers detailed normals, produces diffuse albedo with improved illumination disentanglement, and synthesizes realistic relighting results. Additional outputs of our method are provided in Fig. 9; dynamic relighting results are included in the supplementary video.

Table 5: Prior sensitivity for normal and diffuse albedo on OpenHumanBRDF.

Method	Normal			Diffuse Albedo		
	Angle↓	RMSE↓	PSNR↑	SSIM↑	PSNR↑	LPIPS↓
Noised Prior	11.25	0.0740	22.72	0.876	27.74	0.0192
w/o IAFI	10.12	0.0648	23.91	0.878	27.89	0.0189
Ours (Stage I)	**7.68**	**0.0521**	**25.76**	**0.898**	**28.52**	**0.0149**

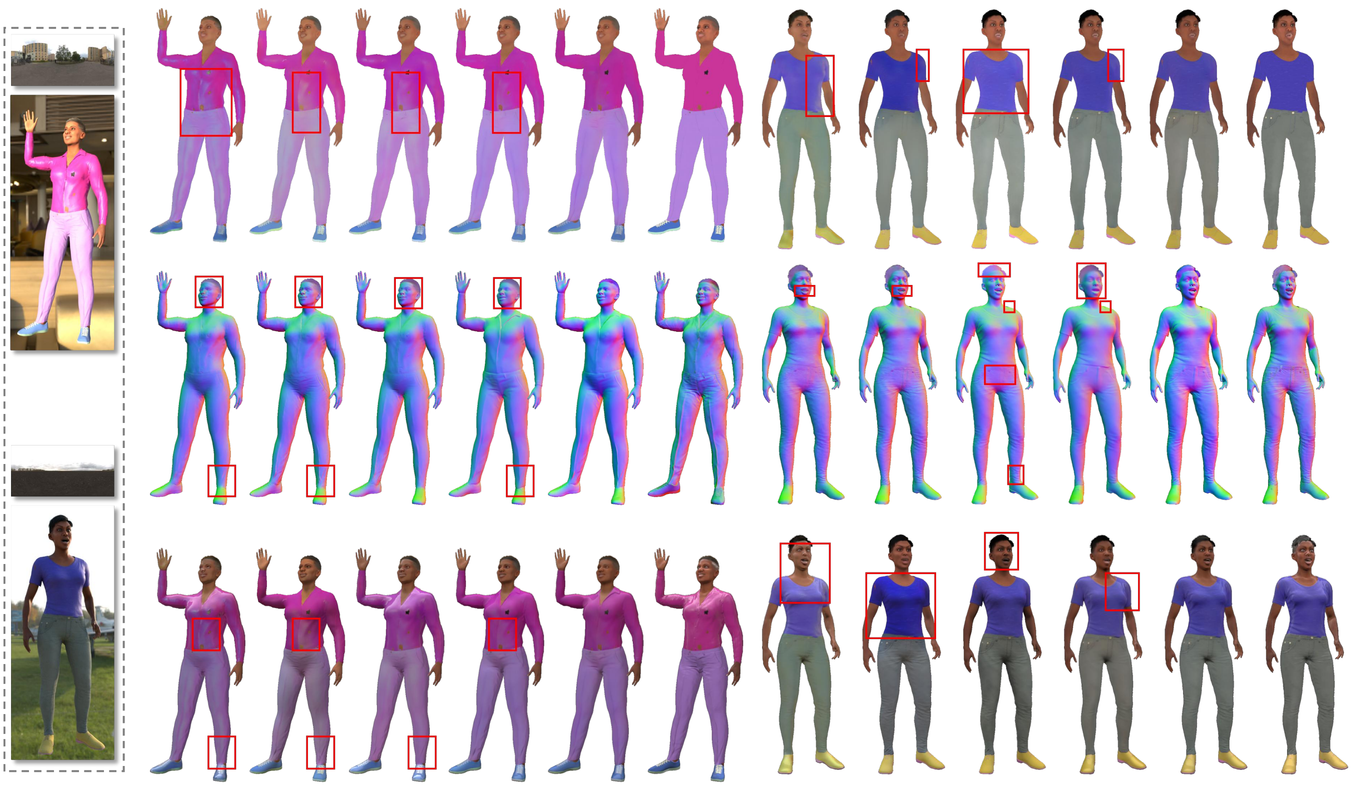

Figure 4: Material estimation and relighting comparison on Synthetic CG dataset. From top to bottom are the "Diffuse albedo", "Normal", and "Relighting" results, respectively.

Table 6: Prior sensitivity for subsurface scattering on OpenHumanBRDF.

Method	SSIM↑	PSNR↑	LPIPS↓
Noised Prior	0.974	44.82	0.0216
w/o IAFI	0.988	48.51	0.0034
Ours (Stage I)	**0.993**	**48.99**	**0.0025**

Table 7: Prior sensitivity for relighting on OpenHumanBRDF.

Method	Real Relighting			Synthetic Relighting		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Noised Prior	0.816	22.72	0.0359	0.837	24.05	0.0288
w/o IAFI	0.811	22.79	0.0355	0.814	23.99	0.0317
Ours (Stage I)	**0.821**	**23.14**	**0.0335**	**0.844**	**24.17**	**0.0243**

Figure 5: Material estimation and relighting comparison on CosmicMan-HQ 1.0 dataset.

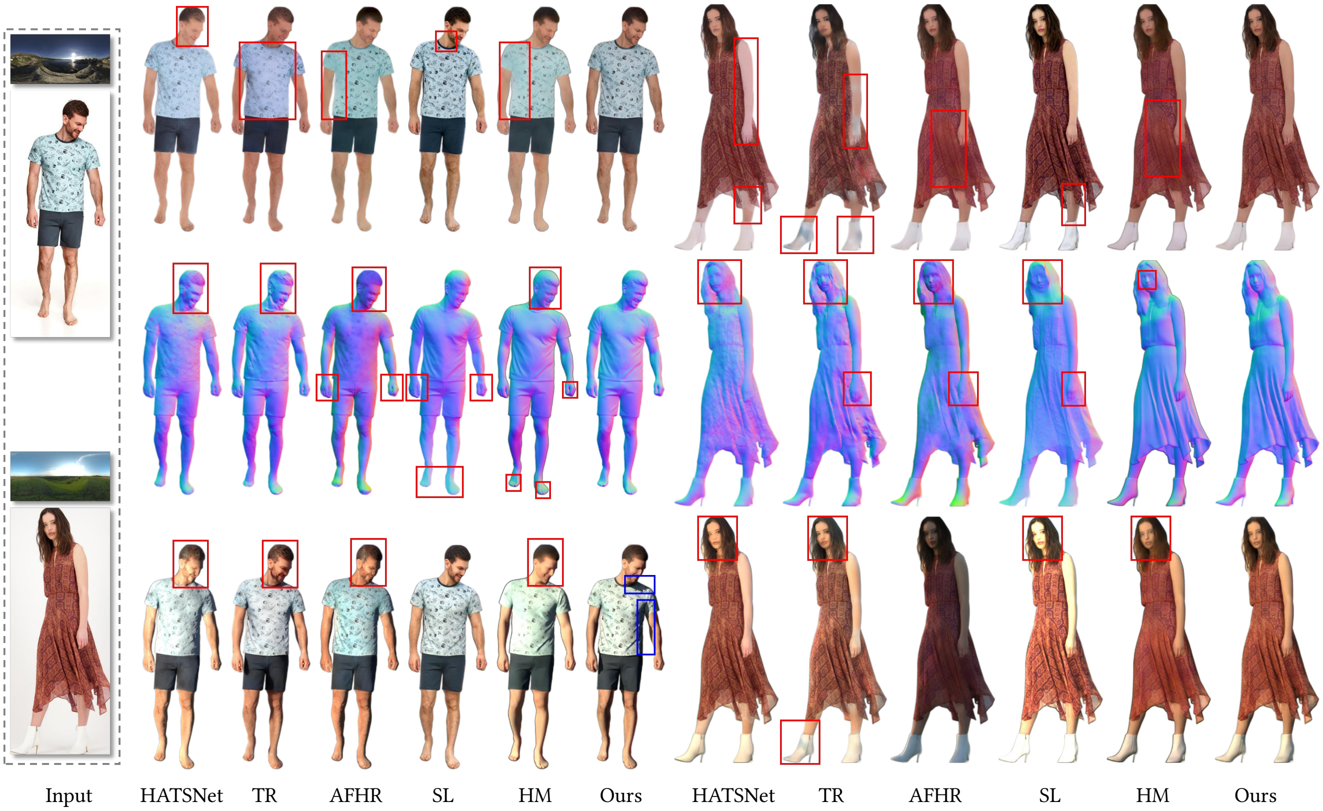

Figure 6: Material estimation and relighting comparison on CosmicMan-HQ 1.0 dataset.

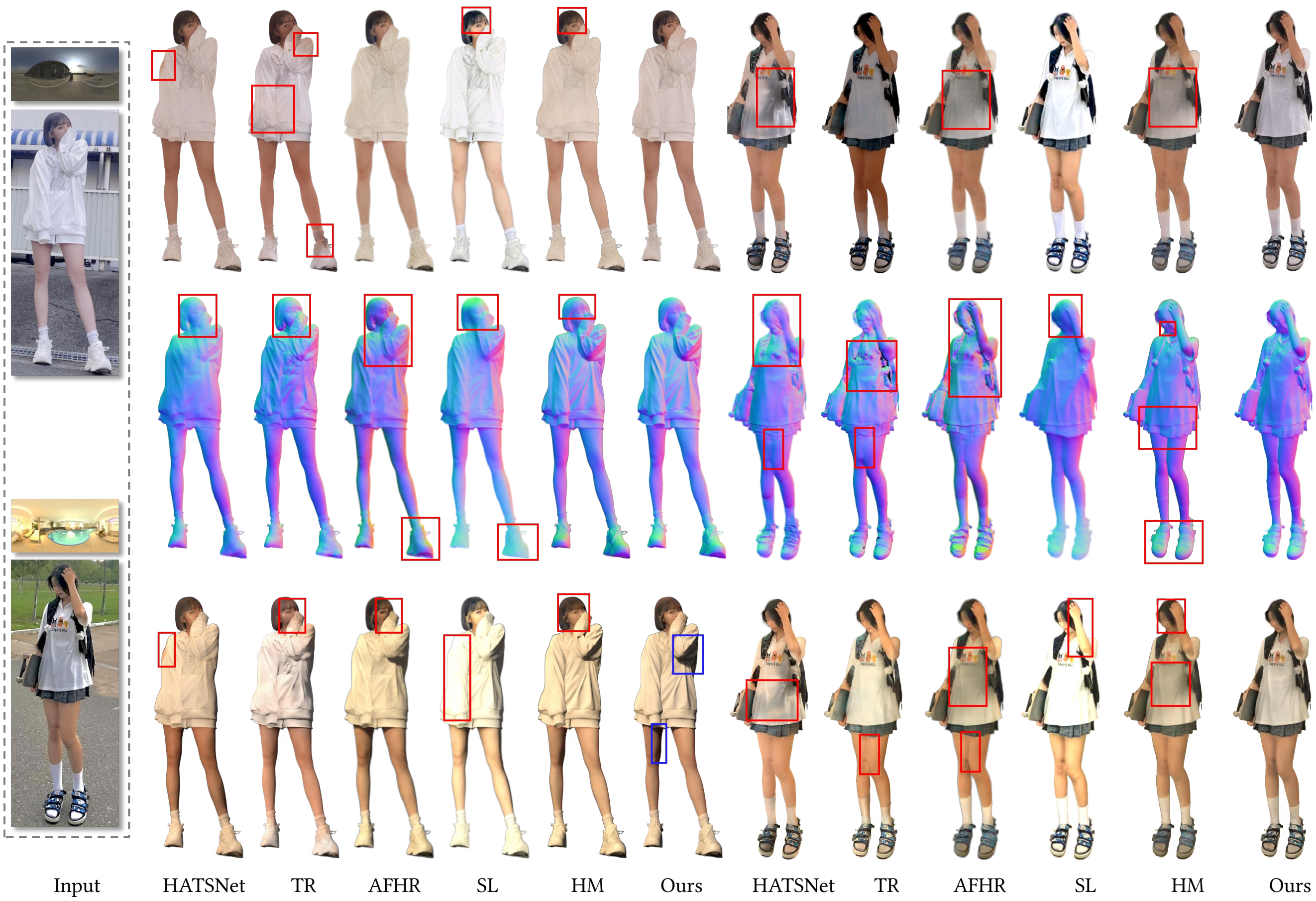

Figure 7: Material estimation and relighting comparison on in-the-wild images collected from the Internet.

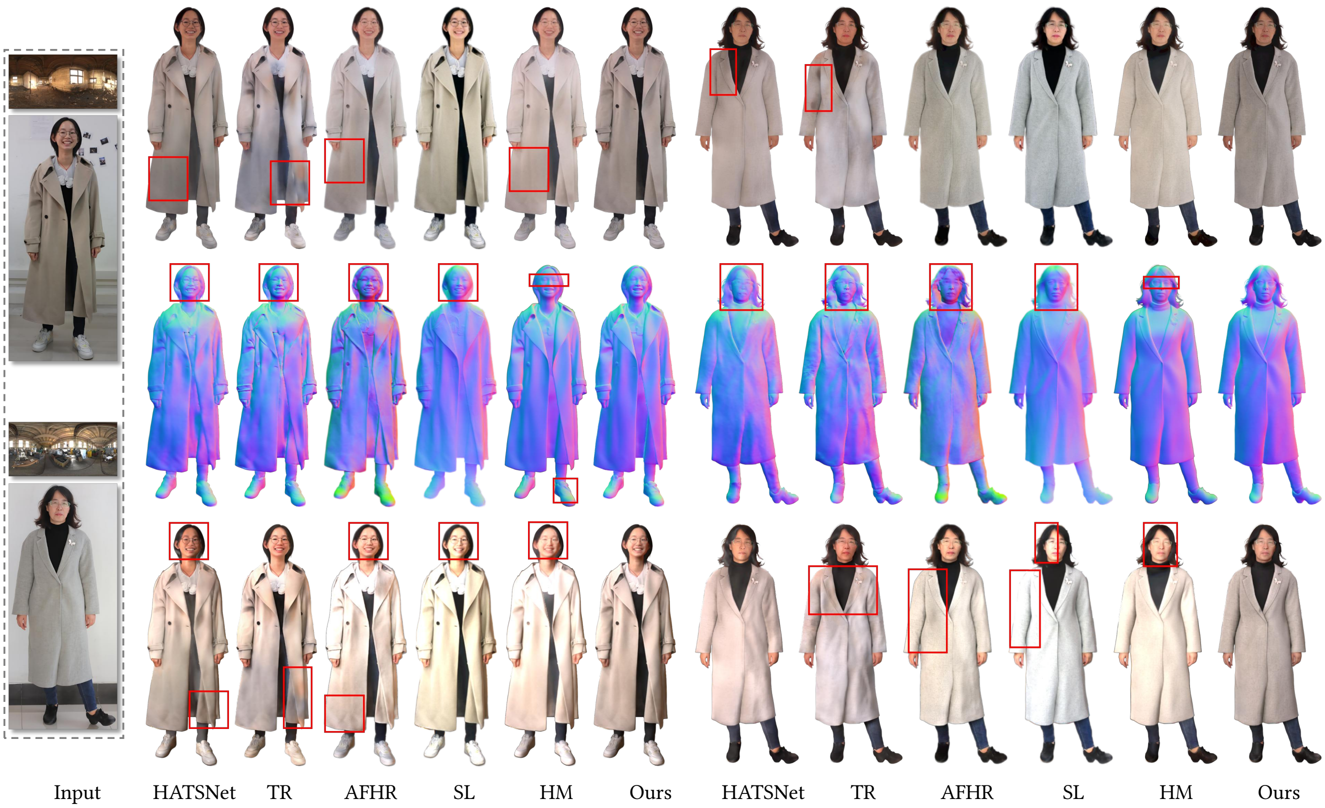

Figure 8: Material estimation and relighting comparison on our real captured images.

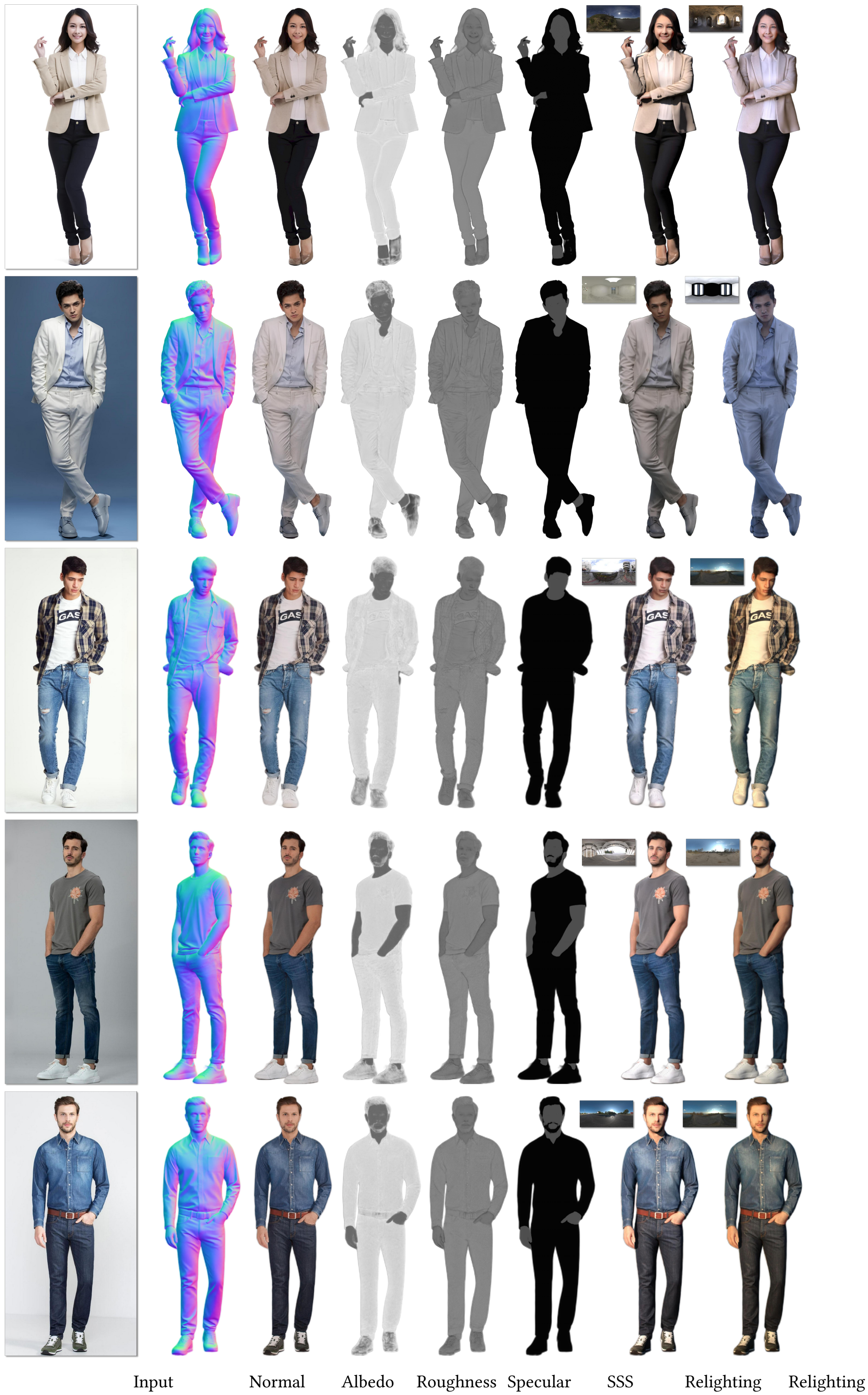

Figure 9: More material-estimation and relighting results on real images. Dynamic relighting results are provided in the supplementary video.

7 Detailed Ablation Settings

We evaluate several variants of our model: *simple injection*, which replaces the proposed illumination-aware injection with direct feature concatenation; *w/o light correct*, which removes lighting correction and directly uses the illumination estimated by the frozen lighting estimator; *w/o physical*, which keeps pixel-level fusion but removes physics-inspired cues; *w/o prior fusion*, which removes pixel-level prior fusion entirely; *w/o finetune* (*Ours (Stage I)*), which is trained only on synthetic data; *full finetune*, which updates all model parameters during fine-tuning; *w/o reg*, which removes teacher–student regularization during fine-tuning; and *Ours*, the full model. Variants (1)–(5) are evaluated on OpenHumanBRDF, while variants (5)–(8) are evaluated on CosmicMan-HQ 1.0.

8 User Study

We conduct a pairwise user study to evaluate the estimated PBR materials and relighting realism. The study focuses on the accuracy of estimated surface normals and diffuse albedo, and on the perceptual realism of rendered images under novel illumination.

For each method, we randomly sample 100 test images from CosmicMan-HQ 1.0 to cover diverse materials and lighting conditions. All methods are compared in a round-robin manner with an equal number of trials. In each trial, two results are displayed side by side in randomized order, and participants select the result that better preserves the underlying PBR material properties or appears more realistic after relighting. In total, 30 participants complete 50 trials each, resulting in 1,500 pairwise comparisons. We compute the vote percentage of each method as

$$\mathrm{VoteRatio}_i = \frac{v_i}{\sum_k v_k}, \tag{2}$$

where v_i is the number of votes received by method i. Table 8 shows that our method obtains the highest vote ratio in all three categories.

Table 8: User-study preference ratios.

Method	Normal	Diffuse Albedo	Relighting
HATSNet [4]	1.4%	1.0%	0.9%
TR [11]	1.6%	1.3%	1.2%
AFHR [2]	10.5%	10.6%	10.5%
SL [7]	11.8%	11.2%	11.3%
HM [5]	13.7%	11.0%	11.1%
Ours	**61.0%**	**64.9%**	**65.0%**

9 Limitations and Future Work

A promising direction is to enrich the training data with more diverse and complex clothing textures and to introduce stronger regularization for normal estimation to reduce texture leakage.

References

[1] Brent Burley and Walt Disney Animation Studios. 2012. Physically-based shading at disney. In *Acm Siggraph*, Vol. 2012. vol. 2012, 1–7.

[2] Yuki Endo Daichi Tajima, Yoshihiro Kanamori. 2025. All-frequency Full-body Human Image Relighting. *Computer Graphics Forum (Proc. of Eurographics 2025)* 44, 2 (2025), e70007.

[3] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-Image SVBRDF Capture with a Rendering-Aware Deep Network. *ACM Transactions on Graphics (TOG)* 37, 4, Article 128 (jul 2018), 15 pages. doi:10.1145/3197517.3201378

[4] Jie Guo, Shuichang Lai, Chengzhi Tao, Yuelong Cai, Lei Wang, Yanwen Guo, and Ling-Qi Yan. 2021. Highlight-Aware Two-Stream Network for Single-Image SVBRDF Acquisition. *ACM Transactions on Graphics (TOG)* 40, 4, Article 123 (jul 2021), 14 pages. doi:10.1145/3450626.3459854

[5] Yu Jiang, Jiahao Xia, Yusen Wang, Tuo Cao, and Chunxia Xiao. 2025. HumanMaterial: Human Material Estimation from a Single Image via Progressive Training. *arXiv preprint arXiv:2507.18385* (2025).

[6] Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. 2024. Sapiens: Foundation for Human Vision Models. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part IV.* Springer-Verlag, Berlin, Heidelberg, 206–228. doi:10.1007/978-3-031-73235-5_12

[7] Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, Sanghyun Woo, and Beeble Ai. 2024. SwitchLight: Co-design of Physics-driven Architecture and Pre-training Framework for Human Portrait Relighting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Jan 2024), 25096–25106.

[8] Manuel Lagunas, Xin Sun, Jimei Yang, Ruben Villegas, Jianming Zhang, Zhixin Shu, Belen Masia, and Diego Gutierrez. 2021. Single-image Full-body Human Relighting. In *Eurographics Symposium on Rendering (EGSR)*. The Eurographics Association. doi:10.2312/sr.20211301

[9] Shikai Li, Jianglin Fu, Kaiyuan Liu, Wentao Wang, Kwan-Yee Lin, and Wayne Wu. 2024. Cosmicman: A text-to-image foundation model for humans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 6955–6965.

[10] Ilya Loshchilov and Frank Hutter. 2017. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations (ICLR)*. https://api.semanticscholar.org/CorpusID:53592270

[11] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total Relighting: Learning to Relight Portraits for Background Replacement. *ACM Transactions on Graphics (TOG)* 40, 4. doi:10.1145/3450626.3459872