Inclusive Portrait Lighting Estimation Model Leveraging Graphic-Based Synthetic Data

Supplementary Material

6. Implementation Details

For pertraining, we train the generator and color regressor for 150 epochs with synthetic data only using an SDG optimizer with 0.9 momentum and a linear scheduler, at a starting learning rate of 0.035. We jointly train the generator and the color regressor in each iteration similar to how generators and discriminators are trained in a typical generative adversarial network [10]. We incorporate random horizontal flip, exposure, and white balance augmentations. We scale the generator reconstruction loss by 1, and the generator color loss by 0.1 when calculating the total generator loss.

For fine-tuning, we train the generator for 50 epochs using an SDG optimizer with 0.9 momentum and a linear scheduler, at a starting learning rate of 0.0075. The color regressor is frozen during fine-tuning. The ratio of labeled synthetic data to unlabeled real video frames used during training is two to one. We apply the same set of augmentations to the synthetic and real data as the augmentations used during pertaining. We scale the reconstruction loss by 1, the color loss by 0.1, and the consistency loss by 0.1 when calculating the total generator loss.

7. Ablation Study

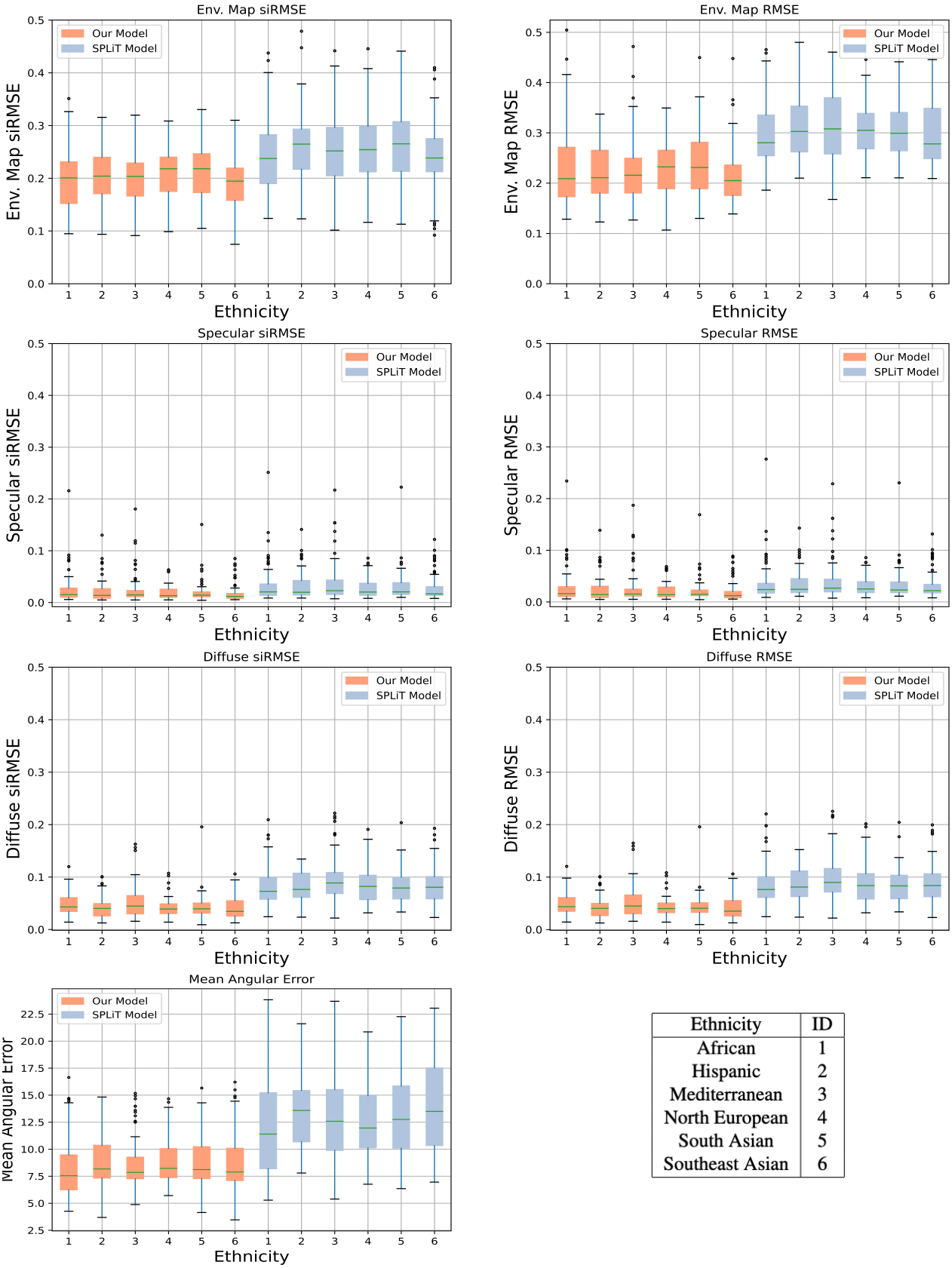
To investigate the impact of having the color regressor and fine-tuning on real videos, we perform an ablation study to compare the following results.

- PT w/o L_{color}: Pretraining the model with reconstruction loss and without color loss
- PT: Pretraining the model with a reconstruction loss and a color regressor loss
- FT: Fine-tuning the model with a consistency loss

The analysis is performed on the real-person images in the LIGHTTEST dataset [5]. The results are shown in Tab. 2.

8. Fairness Analysis

We perform a quantitative and qualitative fairness assessment on our model and on the SPLiT model [5] based on synthetic data. We measure the siRMSE and RMSE of the models directly on environment maps, on rendered specular spheres, and on rendered diffused spheres, as well as, mean angular error (MAE) on the environment map. The results in Fig. 5 and Fig. 6 show that our model's performance fluctuates less on different ethnicities or genders. The box heights of our models are overall shorter and shift less across all categories. The qualitative analysis shown in Fig. 7 also demonstrates that our model is more robust to demographic diversity.



Ethnicity	ID
African	1
Hispanic	2
Mediterranean	3
North European	4
South Asian	5
Southeast Asian	6

Figure 5. Quantitative fairness analysis on ethnicity (#1-6) in comparison with the SPLiT model.

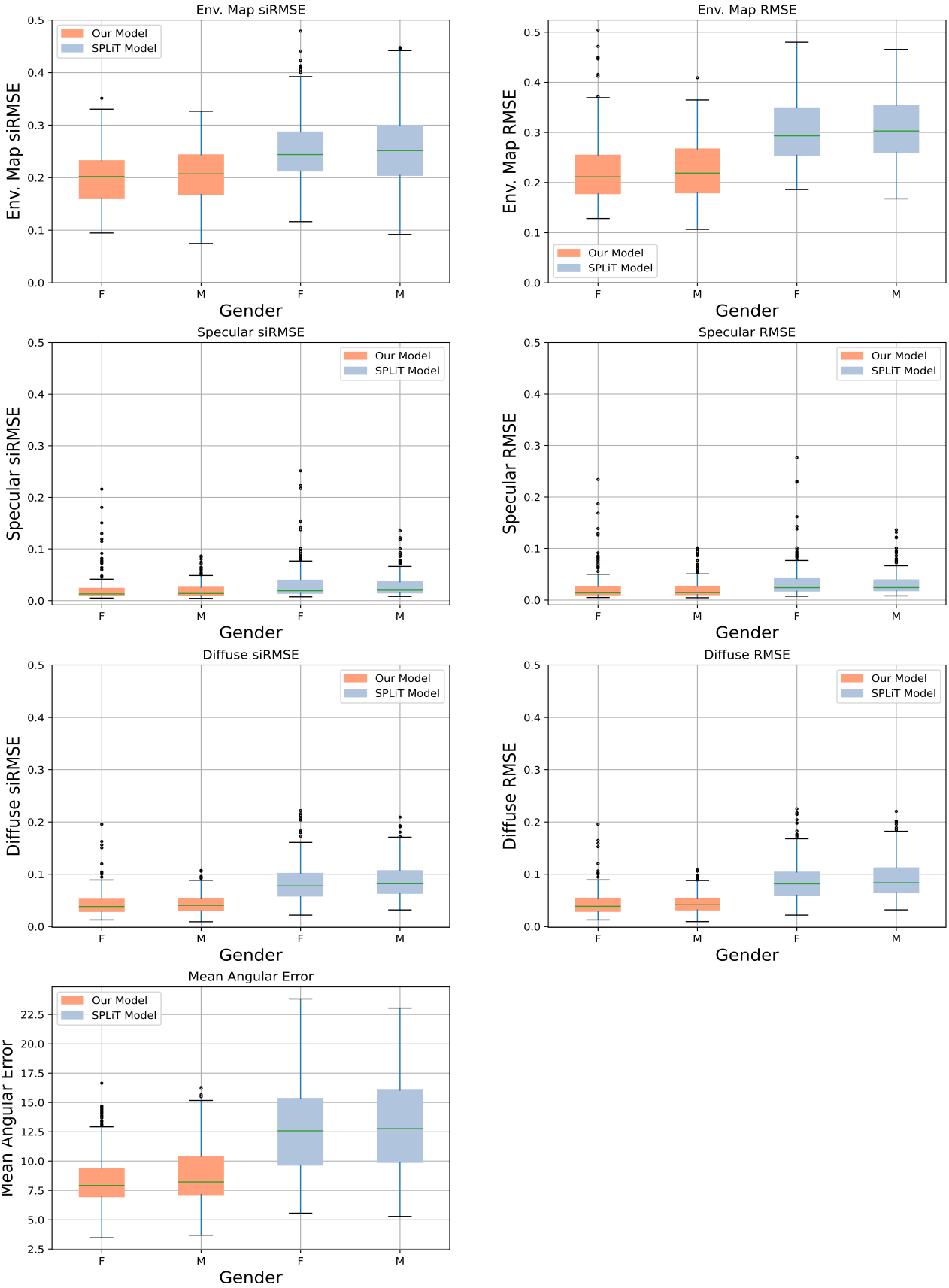


Figure 6. Quantitative fairness analysis on gender in comparison with SPLiT

	Indoor						Outdoor					
	Env. Map ↓		Spec ↓		Diffuse ↓		Env. Map ↓		Spec ↓		Diffuse ↓	
	siRMSE↓	RMSE↓	siRMSE↓	RMSE↓	siRMSE↓	RMSE↓	siRMSE↓	RMSE↓	siRMSE↓	RMSE↓	siRMSE↓	RMSE↓
PT w/o L_{color}	0.235	0.302	0.017	0.030	0.072	0.074	0.171	0.275	**0.015**	0.043	**0.071**	0.076
PT	**0.231**	0.283	0.017	0.024	0.067	0.069	0.171	0.265	**0.015**	0.032	0.072	0.076
FT	**0.231**	**0.276**	**0.016**	**0.023**	**0.065**	**0.067**	**0.170**	**0.253**	**0.015**	**0.028**	**0.071**	**0.074**

Table 2. Ablation study on pre-training only without the color loss, pre-training with the reconstruction and color losses, and fine-tuning.

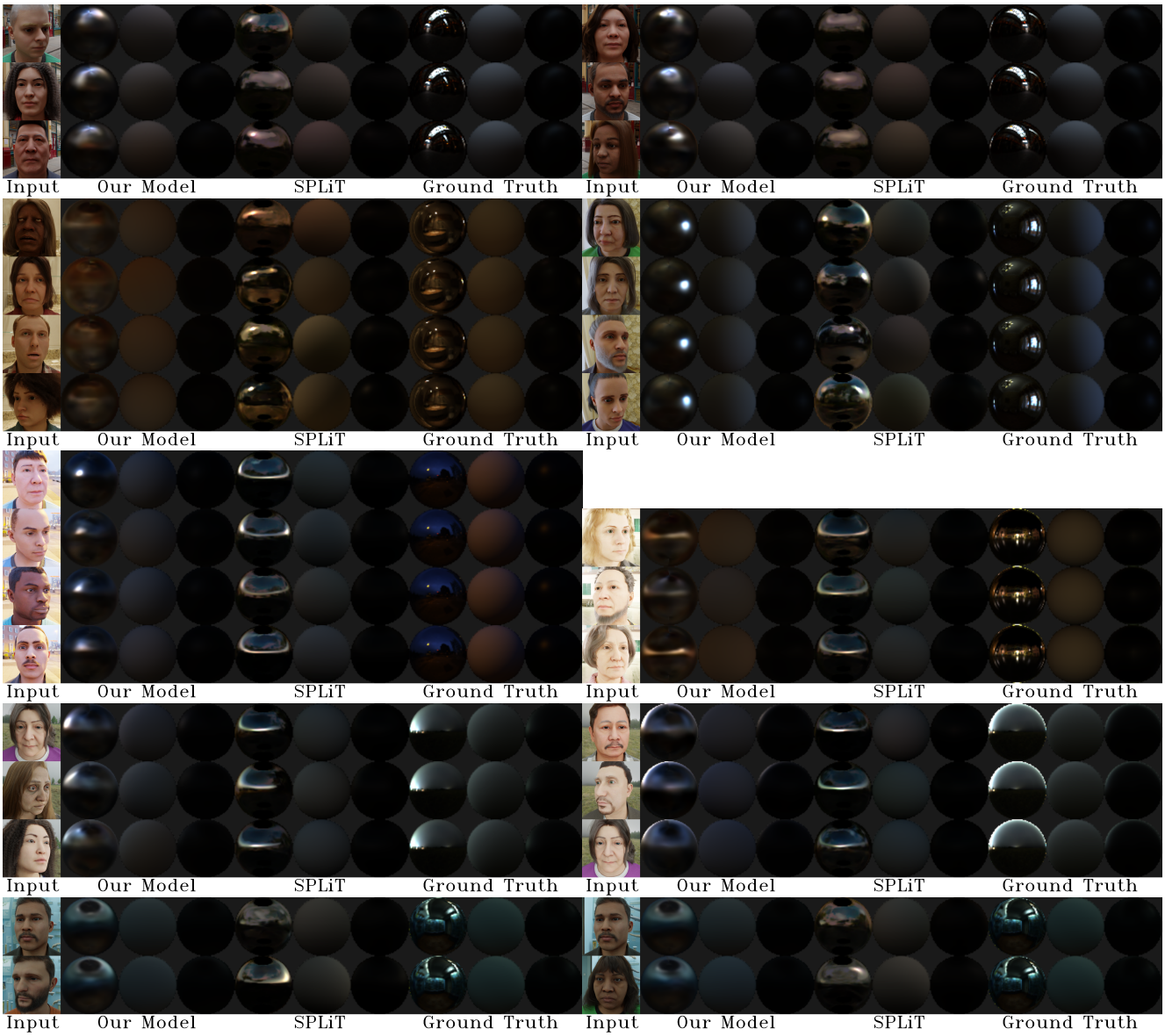


Figure 7. **Fairness qualitative analysis**. To examine if our model is invariant to demographic differences, we utilize synthetic images featuring subjects from diverse ethnicities and genders placed in the same lighting background. We render the predictions on the mirror, specular, and diffused spheres. The objective is to ensure the predictions remain consistent across different faces in the same background. Each image block in the figure contains the predicted light spheres of portrait images with the same lighting background. Within each block, the spheres in the same column should look identical.