

# SPATIALLY-AWARE PHOTO-REALISTIC FACE RELIGHTING USING JOINT EMBEDDING OF LIGHT PROPERTIES

**Anonymous authors**

Paper under double-blind review

## 1 ABLATION STUDY

We performed an ablation study to show the benefits of our network design choices. We replaced specific modules or inputs and retained everything else. This enabled us to evaluate the performance drop-off without specific modules. We evaluated the following design choices

1. Using SH coordinates for light feature embedding<sup>1</sup> instead of our novel light feature embedding
2. Not using any MDHA modules
3. Not using MDHA modules with cross-attention layers
4. Using only global losses ( $\lambda_3, \lambda_4 = 0$ )
5. Using only local losses ( $\lambda_1, \lambda_2, \lambda_6 = 0$ )

We evaluated the performance of the model with various ablations on the real human test dataset (RH). The quantitative results are shown in Table 1. We observed that using the MDHA modules (transformer blocks) are very critical to model performance. Cross-attention layers are also important to improve the visual quality of the final results. Using SH coordinates as light feature embedding worsens the model performance and global loss improves model performance more than local loss. The best performance was observed by the full model using all our design choices.

We observed that the performance of the models in Table 1 was better than the prior works in Table 1 of the main paper. This could be due to the training dataset strategy<sup>2</sup> which enables training a model with good performance despite removing important modules in the network architecture.

Novel light embedding	MDHA modules	Cross-attention layers	Global loss	Local loss	MSE ↓	DSSIM ↓	LPIPS ↓
✗	✓	✓	✓	✓	0.0055	0.0367	0.0732
✓	✗	✗	✓	✓	0.0111	0.0542	0.0868
✓	✓	✗	✓	✓	0.0049	0.0361	0.0729
✓	✓	✓	✗	✓	0.0046	0.0338	0.0723
✓	✓	✓	✓	✗	0.0061	0.0369	0.0733
✗	✗	✗	✓	✓	0.0133	0.0742	0.1243
✓	✓	✗	✗	✓	0.0122	0.0681	0.1045
✓	✗	✗	✓	✗	0.0141	0.0784	0.1468
✓	✓	✓	✓	✓	<b>0.0043</b>	<b>0.0307</b>	<b>0.0701</b>

Table 1: Ablation study to evaluate the benefit of various design choices for the image relighting network.

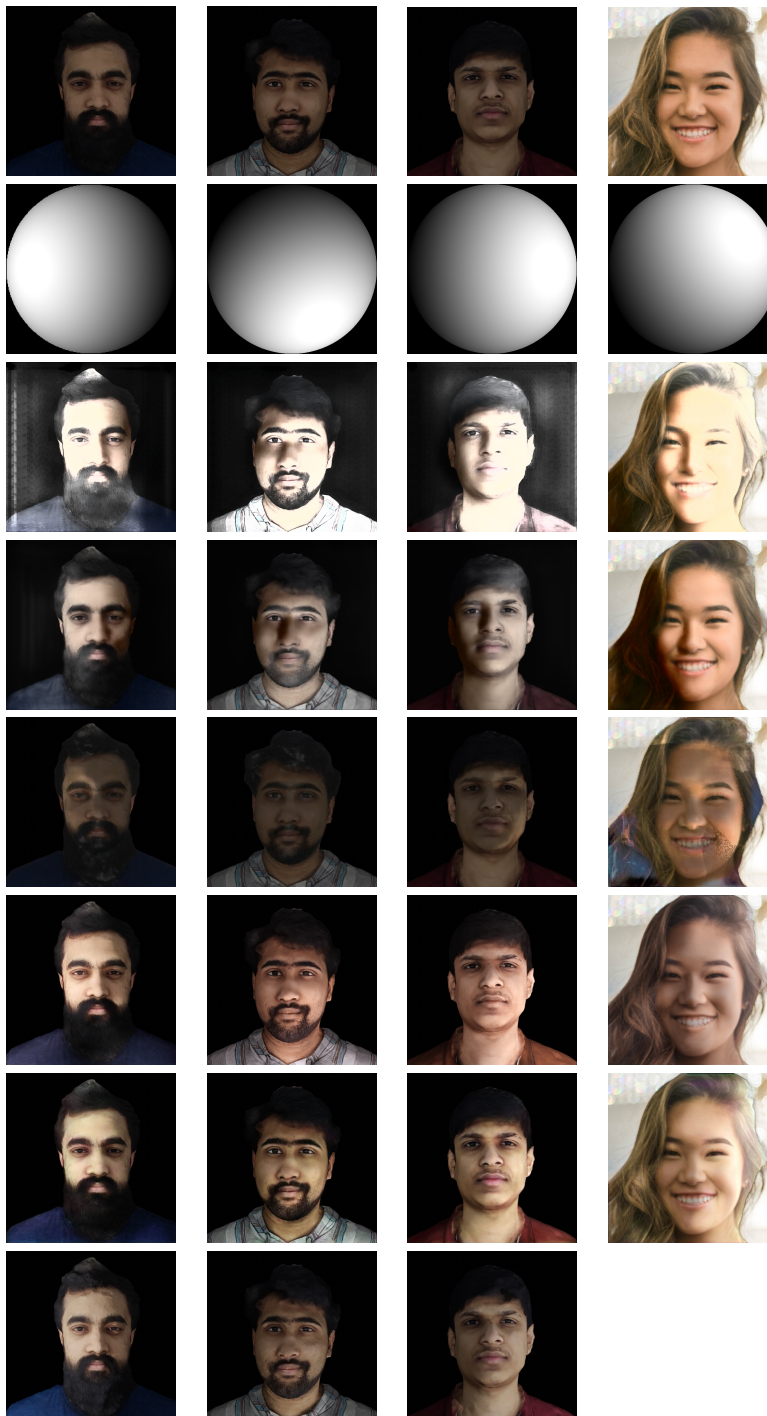
## 2 ADDITIONAL COMPARISON AGAINST PRIOR WORKS

Fig 1 shows some additional comparison results against prior works. We observe that the results from our model are significantly more accurate and photo-realistic than prior works. Our method is

<sup>1</sup>The lighting network was trained to estimate the 9-dimensional SH coordinates.

<sup>2</sup>We mostly used a similar strategy to that proposed by Pidaparthi et al. (2024), but made a few important modifications as discussed in the main paper.

054 able to easily handle variations in input image such as changes to ambient lighting, facial structure  
 055 and ethnicities, directional shadows in the input image, etc.  
 056



102  
103 Figure 1: Qualitative comparison of our model against other methods on the real human test dataset  
 104 (RH) (columns 1, 2), Celeb-FFHQ dataset (columns 3-5) and Multi-pie dataset (MP) (last column).  
 105 Rows 1 and 2 are the input image and light source position; rows 3-6 are the results from Zhou et al.  
 106 (2019), Hou et al. (2021), Hou et al. (2022) and Pidaparthy et al. (2024); rows 7 and 8 are our results  
 107 and ground truth relit images. We do not have ground truth relit images on the Celeb-FFHQ dataset.  
 Images are best viewed in colour.

108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161

## REFERENCES

Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiyong Tong, and Xiaoming Liu. Towards high fidelity face relighting with realistic shadows. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, pp. 14719–14728, 2021.

Andrew Hou, Michel Sarkis, Ning Bi, Yiyong Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, pp. 4217–4226, 2022.

Hemanth Pidaparthi, Abhay Chauhan, and Pavan Sudheendra. Multi-level attention aggregation for aesthetic face relighting. In *Proceedings of Winter Conference on Applications of Computer Vision (WACV)*, pp. 4057–4066, 2024.

Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of International Conference on Computer Vision (ICCV)*, pp. 7194–7202, 2019.