HiGlassRM: Learning to Remove High-prescription Glasses via Synthetic Dataset Generation

Supplementary Material

A. Details of the Domain Adaptation Network

To mitigate the domain gap between synthetic (I) and real (R) images, our Domain Adaptation (DA) network maps both to a shared 64-channel feature space (I_{fm}, R_{fm}) to learn domain-invariant representations. The network architecture consists of an initial frozen layer from a pre-trained VGG encoder, followed by six trainable ResNet blocks. Following the approach of Lyu et al. [23], we train the network adversarially on the CelebA dataset [21] using the LS-GAN objective [35]:

$$\begin{aligned} \mathcal{L}_D &= D(I_{fm})^2 + (D(R_{fm}) - 1)^2 \\ \mathcal{L}_G &= (D(I_{fm}) - 1)^2. \end{aligned} \tag{S1}$$

This process encourages the feature maps of real and synthetic images to be indistinguishable, enhancing the model's overall generalization performance.

B. Training and Evaluation Details

Implementation Details. Our framework is implemented in PyTorch. All training and inference tasks were conducted on a single NVIDIA RTX 2080 GPU. The full training of HiGlassRM for 90 epochs takes approximately 80 hours. At inference, the model processes a 256x256 image in an average of 60.5 ms.

Resource comparisons. The HiGlassRM model has 35.7M parameters and requires 165.62 GFLOPs for a single forward pass. This indicates that HiGlassRM is more efficient than the previous state-of-the-art method, take-off-eyeglasses [23], which requires 195.26 GFLOPs.

Training Protocol We use the Adam optimizer [34] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, an initial learning rate of 1×10^{-4}, and a batch size of 32. The training proceeds in two stages to ensure stability: first, the Glass Mask Network is trained independently until convergence. Afterwards, its weights are frozen, and the remaining networks (Flowmap, De-Shadow, De-Glass) are trained jointly.

C. Additional Qualitative Comparisons

The figures below provide additional qualitative examples that supplement the results presented in the main paper, further illustrating the common failure modes of baseline methods and the robustness of our approach.

Results on Synthetic Data Figure S1 visualizes the intermediate predictions and final outputs of our HiGlassRM model on the HiGlass test set, comparing them against the ground truth. The second column of each block shows that the predicted eyeglass-frame mask ($\widehat{M}$) accurately localizes the frame region without bleeding into the face. The third column displays the predicted displacement flowmap ($\widehat{F}$), which successfully learns the complex, non-uniform patterns of the lens-induced geometric warp. By leveraging these accurate intermediate predictions, the final de-glassed output from HiGlassRM (fourth column) is nearly identical to the ground-truth target (O), demonstrating a high-fidelity restoration of the subject's true facial structure.

Results on Real Data Figure S2 presents more results on real-world images [11], demonstrating the strong generalization of our method. Image-to-image models show clear limitations: CycleGAN [33] often leaves faint remnants of the glasses, while pix2pix [13] fails to generalize from our synthetic data, resulting in severe artifacts. While other baselines, particularly take-off-eyeglasses [23], also struggle to handle real-world optical distortions, HiGlassRM robustly removes the eyeglasses while restoring the natural scale of the eyes and correcting the warped facial periphery, yielding the most visually coherent results.

D. Limitations and Future Work

Our method has limitations that suggest avenues for future research, as shown in Figure S3. The current model can fail to remove strong specular reflections on lens surfaces, as this effect is not explicitly modeled in our synthesis pipeline. Additionally, on some challenging real images, the predicted flowmap can over-correct distortion, leading to unnatural eye enlargement. This may stem from a domain gap with real-world factors like complex lens thickness. Future work could address these issues by incorporating reflection-aware modeling and stronger geometric priors for flow estimation.

The scope of our work is currently focused on transparent lenses. Promising future directions include extending our dataset and model to handle more complex cases like tinted lenses and sunglasses, which introduce challenges of color filtering and severe occlusion. Furthermore, developing a lightweight variant for efficient deployment on edge devices and enhancing the dataset with more diverse real-world convex distortion examples remain important next steps for the field.

Figure S1. Qualitative results and intermediate predictions of HiGlassRM on the HiGlass test set. Columns show the input (I), predicted frame mask ($\widehat{M}$), predicted flowmap ($\widehat{F}$), our final output ($\widehat{O}$), and the ground-truth target (O).

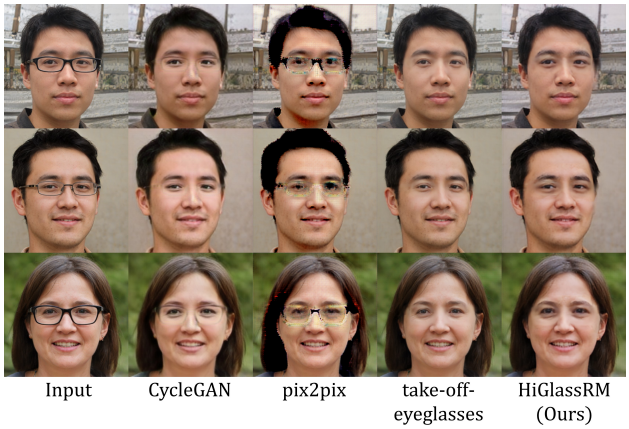

Figure S2. Qualitative comparison on real-world images from the Glasses or No Glasses dataset [11]

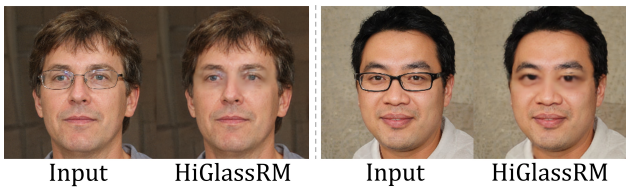

Figure S3. Examples of failure cases. Left: Incomplete removal of strong specular reflections. Right: Over-correction of distortion leading to unnatural eye enlargement.

References

[34] Kingma DP Ba J Adam et al. A method for stochastic optimization. *arXiv:1412.6980*, 1412(6), 2014. 1

[35] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2794–2802, 2017. 1