GlassesGAN: Supplementary Material

CVPR 2023, Paper ID: 3276

In the main part of the paper, we presented a wide range of results to highlight the merits of the proposed GlassesGAN framework for personalization of glasses in virtual try-on settings. In this *Supplementary material*, we now show additional results to further highlight the capabilities of GlassesGAN. Specifically, we: (i) demonstrate the robustness of the overall processing pipeline on another dataset (MetFaces) with out-of-domain images, (ii) show additional continuous multi-style editing results, (iii) provide additional details on the user-study conducted, (iv) describe ablation experiments with respect to the Targeted Subspace Modelling (TSM), and (v) provide information on the reproducibility of our results.

A. Out of Domain Edits

To demonstrate the robustness of the processing pipeline of GlassesGAN, we generate a few example edits on the MetFaces dataset. MetFaces [22] is a dataset of human faces extracted from works of art. Because the images in this dataset do not correspond to real faces, they come with vastly different characteristics than the facial images used to train our framework. As can be seen from Figure 12, GlassesGAN is able to apply clear and tinted glasses to the artwork without any changes to the overall framework. We observe that even with these challenging images, the edits appear highly realistic and visually convincing.

B. Additional Visual Results

In Figures 13 and 14 we show additional high-resolution results of continuous multi-style edits with a couple of sample images and clear glasses. Similarly, in Figures 15 and 16 we demonstrate continuous multi-style edits for tinted glasses. In order to better demonstrate the continuous edit-capability of GlassesGAN, we also provide a video demonstration of our editing as part of the supplementary material. We observe that GlassesGAN is able to smoothly transition between different eyeglass styles in the latent space while preserving the realism of the edit.

C. Additional User Survey Details

Evaluators were shown randomly selected probe images and randomly ordered glasses edit images from each comparison method and asked to choose the edit image that best meets the criteria of the prompt question. Prompt questions were $Q1$: Which of the edited images preserves the identity of the original image the most?, $Q2$: Which of the edited images is of the highest quality overall?, $Q3$: Which of

Figure 12. **Out of domain editing with GlassesGAN.** Additions of clear and tinted glasses to images from the MetFaces dataset are presented. Observe the realism and detail of the generated edits.

the edited images has the most realistic looking and visually convincing glasses? (consider realism, shadows, frame shape, fit on ears, etc.), and $Q4$: Which of the edited images is closest to what you would consider a good try on result?

D. TSM Ablation Experiment

Targeted Subspace Modeling (TSM), introduced in this paper, is a key component of GlassesGAN that allows us to capture the simulated variations in the appearance of eyeglasses in a small number of principal subspace axes. In this section, we investigate the importance of Targeted Subspace Modeling (TSM) in the methodology of GlassesGAN. To do this, we explore an alternate procedure where the desired personalization to eyeglasses is performed to the augmentation masks rather than in the learned glasses subspace of the generator's latent space, thereby bypassing TSM. Instead of learning the glasses subspace, this alternative method creates the desired glasses shape in the augmentation mask, sequentially encodes and then decodes the probe image with the mask applied, and blends the result into the original image using the face parser S. We found that this alternate methodology is able to provide a virtual-try-on experience, but suffers when it comes to flexibility of edits, off-angle robustness, and inference-time computational complexity. We elaborate on these characteristics in the following sections.

D.1. Edit Flexibility

As described in Section 3.2, the Synthetic Appearance Discovery (SAD) mechanism begins with a set of hand-drawn eyeglass templates and then augments them using morphological operations, such as dilation and erosion. With only these augmented templates available, the alternative methodology without TSM is substantially limited in the type of continuous edits available. In Figure 17, we show our ability to thin and thicken frames (17A) and shrink and enlarge frames (17B) without TSM. With use of TSM, on the other hand, GlassesGAN is able to learn six distinct edit types including the rounding and squaring of glasses

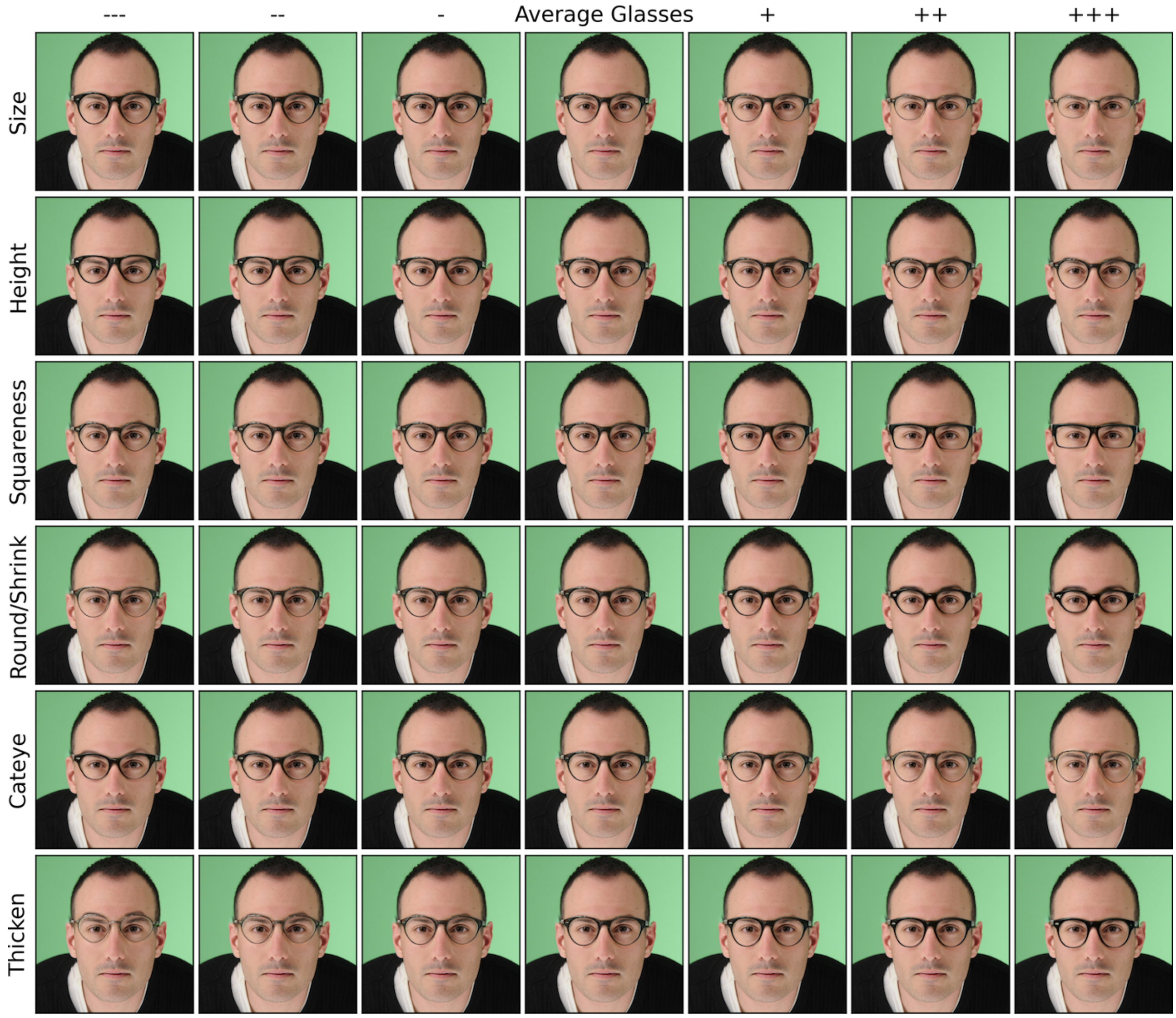

Figure 13. Extended visualization of continuous multi-style edits with clear glasses on sample from SiblingsDB-HQf.

and cat-eye appearances as shown in Figures 13, 14, 15, 16.

D.2. Off-angle Robustness

The removal of TSM from the methodology also comes at the expense of the robustness of glasses edits to off-angle scenarios, as shown in Figure 18. As the eyeglasses positioning and size are based on the coordinates of the temples, there is no mechanism to warp the mask to fit the face correctly in severely off-angle poses. This causes eyeglasses to be poorly embedded (18A), misoriented (18B), or correctly oriented when the augmentation positioning is close enough (18C).

D.3. Inference Computation

During inference-time sequential edits without TSM are more time-intensive. This is because every edit without TSM requires a run from the Encoder model E. Alternatively, the use of TSM moves the editing step into the latent space, thereby only requiring encoding to occur during the first run. Since GlassesGAN is designed for continuous edits to eyeglass style, this extra latency for sequential edits can quickly become very apparent.

E. Reproducibility

All of our experiments are fully reproducible. The models used for the implementation of GlassesGAN are all publicly available from the official repositories, i.e.:

Figure 14. Extended visualization of continuous multi-style edits with clear glasses on sample from CelebA-HQ.

- StyleGAN2: https://github.com/NVlabs/stylegan2
- E4E encoder: https://github.com/omertov/encoder4editing
- DatasetGAN: https://github.com/nv-tlabs/datasetGAN_release
- DeepLabV2: https://github.com/tensorflow/models/tree/master/research/deeplab
- Dlib: http://dlib.net/
- ArcFace: https://github.com/deepinsight/insightface

Additionally, we also plan to publicly release the GlassesGAN source code, including all training and testing scripts, once the review procedure is completed.

Figure 15. Extended visualization of continuous multi-style edits with tinted glasses on sample from SiblingsDB-HQf.

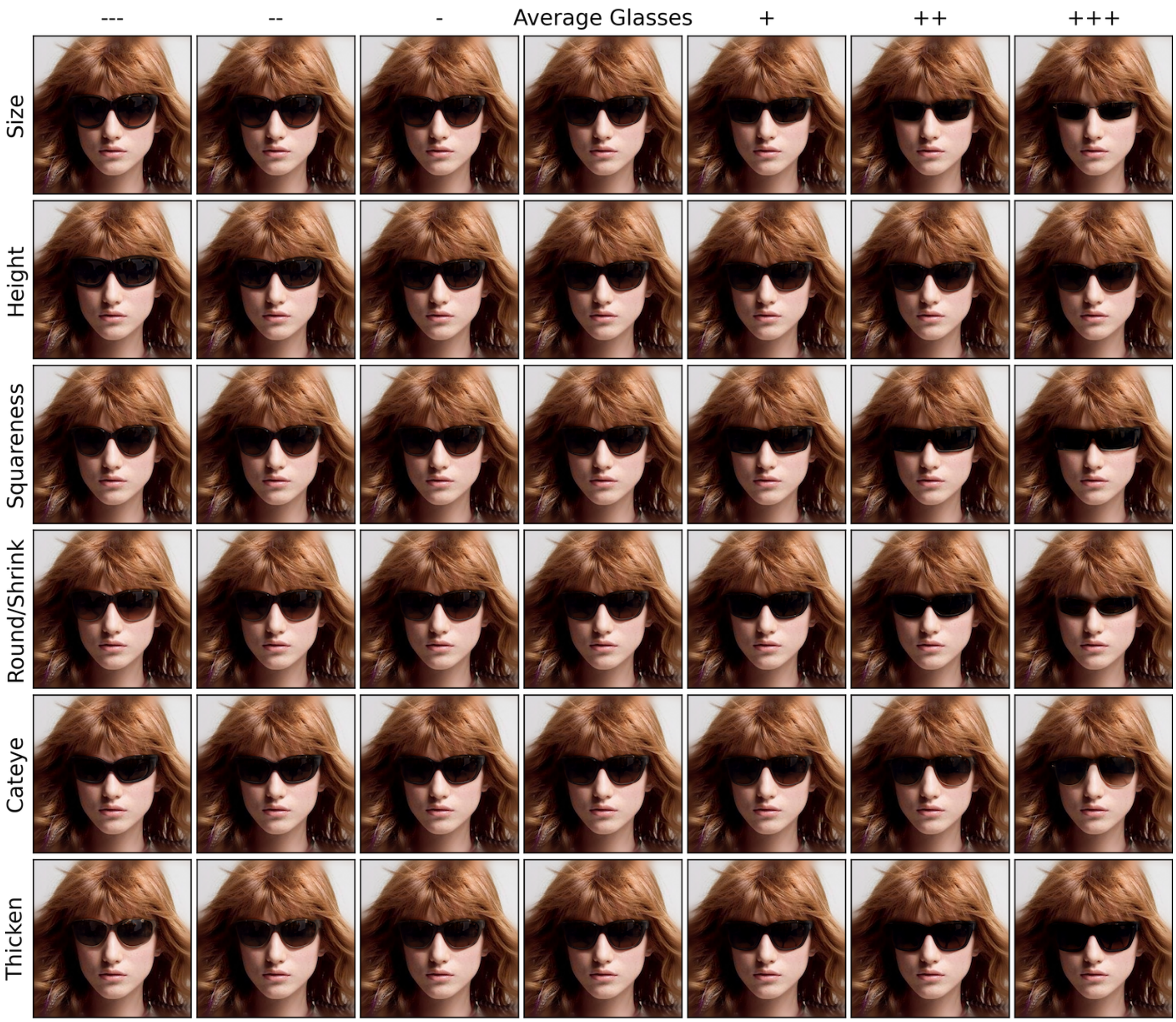

Figure 16. Extended visualization of continuous multi-style edits with tinted glasses on sample from CelebA-HQ.

Figure 17. **Targeted Subspace Modeling (TSM) Ablation Results Demonstrating Edit Flexibility Without TSM.** Without TSM, glasses personalization via continuous edits is substantially limited. In this figure, we show that our template augmentation procedure is only able to (A) thin and thicken or (B) shrink and enlarge glasses frames, as compared to the six distinct edit styles available when using TSM. The upper row in each group shows the augmented image while the lower row shows the result after encoding and blending.

Figure 18. **Targeted Subspace Modeling (TSM) Ablation Results for Off-angle Samples.** The first column shows the original probe images while the second shows the intermediate augmentation image from the eyeglasses masking. The third and forth columns show the final edited image without and with TSM respectively. Group A and B show samples where removing TSM caused glasses to be poorly embedded and misoriented respectively. The samples in group C show the result when the augmentation alignment is correct.