Supplemental

1 INPAINTING MASKS

The segmentation masks of a pair of reference images will not always completely cover the entire target segmentation mask. In this case, there are some uncovered regions, indicated in white in Fig. S-1(a), that need to be in-painted in order to create a complete target mask. In addition, the hair region is complicated in that some portions of the hair belong *behind* the figure, and some portions of hair should occlude the figure. An example of our approach both *without* and *with* inpainting is shown in Fig. S-3 - note that without inpainting we labeled uncovered pixels as *background* which could cause the background to show through the hair where it should not (middle row). It is interesting to us that the alignment process, which uses the StyleGAN W+ space as a prior, does not seem to have the capacity to match the erroneous background regions in the target masks near the subject's forehead.

When dealing with hair transfer, it is useful to relabel the segmentation masks using three labels of *hair*, *background*, and *other*, where the last label includes the skin, eyes, nose, clothing, etc. We create the following masks: $\mathbf{M}_{\text{hair}}^{\text{behind}}$ is a mask labeled as *background* wherever both references were background, and labeled as *hair* wherever the reference image for the hair was labeled as *hair*. The remaining pixels of $\mathbf{M}_{\text{hair}}^{\text{behind}}$ are unknown, and they may be portions of hair that pass behind the subject. Therefore, we inpaint $\mathbf{M}_{\text{hair}}^{\text{behind}}$ using the fast-marching method of [Telea 2004], which is implemented by OpenCV. Next, we create a mask $\mathbf{M}_{\text{other}}^{\text{middle}}$ using the segmentation regions of the *other* reference image, except that its original *hair* region is inpainted using the same exact approach. Finally, we construct a mask $\mathbf{M}$ in three layers: we first initialize the mask with $\mathbf{M}_{\text{hair}}^{\text{behind}}$, and then we transfer the labels *other* than background from $\mathbf{M}_{\text{other}}^{\text{middle}}$, and finally we set any pixel that was hair in the original reference image for hair to also have the label *hair* in $\mathbf{M}$ so that we retain the bangs, or locks of hair which pass in front of the face or shoulders. The heuristic approach we used is not capable of generating segmentation masks for completely occluded features (such as eyes or ears) that were covered by hair, however GAN-based inpainting approaches for the masks themselves are a subject of future work.

2 SENSITIVITY TO POSE AND SPATIAL ALIGNMENT

The proposed approach works for cropped portrait images – these images are always somewhat aligned, with a single dominant face in the center and a frontal or three-quarters vies of a face. This is both due to a preference for this by photographers, but also that the datasets are collected by automatically cropping the images using a facial alignment net such as DLIB or FAN [Bulat and Tzimiropoulos 2017]. The use of face and pose detection networks could allow one to filter out incompatible reference-images and thus mitigate issues with spatial alignment. We did not do this filtering in our user study, so errors caused by misalignment were included in our evaluations. It is therefore useful to understand how sensitive the proposed approach is to changes in the spatial alignments of reference images. In order to demonstrate the qualitative effect of our approach to misalignment of the mask, we translated the hair region in Fig. S-3 when generating the target mask.

3 MANUALLY EDITING MASKS

The main focus of this paper is on completely automated hair transfer, however it is possible to overcome many of the challenges and limitations of an automatic approach if user edits are allowed. For example, by allowing a very limited set of user interactions (dragging, scaling, and flipping) of the hair region we can achieve results shown in Fig. S-4.

4 COMPARISON TO CONCURRENT WORK

Concurrently with our work, StyleMapGan [Kim et al. 2021] is also capable of doing face transfer. We compare against prior work on this task in Fig. 7 We illustrate the differences between our results and StyleMapGan qualitatively in Fig. S-2, which demonstrates the results of eyes and eyebrows (top row) and entire faced (bottom row). We observe that our proposed solution is capable of preserving the details of the composited parts.

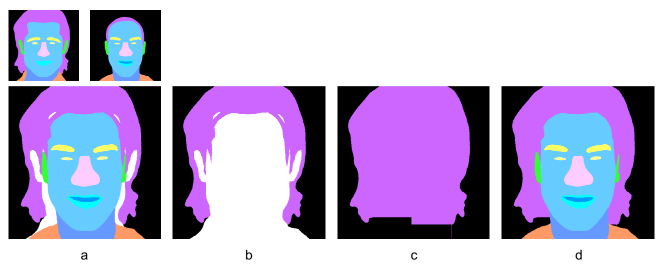

Fig. S-1. Mask inpainting. The semantic segmentations of two reference images for *hair* and *other* are shown on the first row. The second row shows (a) a composite mask without inpainting and disoccluded pixels shown in white; (b) the hair region before inpainting, (c) the result of inpainting the hair mask; (d) the result of filling-in disoccluded regions of (a) using the mask from (c).

REFERENCES

Adrian Bulat and Georgios Tzimiropoulos. 2017. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In *International Conference on Computer Vision.*

Hyunsu Kim, Yunjey Choi, Junho Kim, Sungjoo Yoo, and Youngjung Uh. 2021. StyleMapGAN: Exploiting Spatial Dimensions of Latent in GAN for Real-time Image Editing. *arXiv preprint arXiv:2104.14754* (2021).

Alexandru Telea. 2004. An image inpainting technique based on the fast marching method. *Journal of graphics tools* 9, 1 (2004), 23–34.

Author's address:

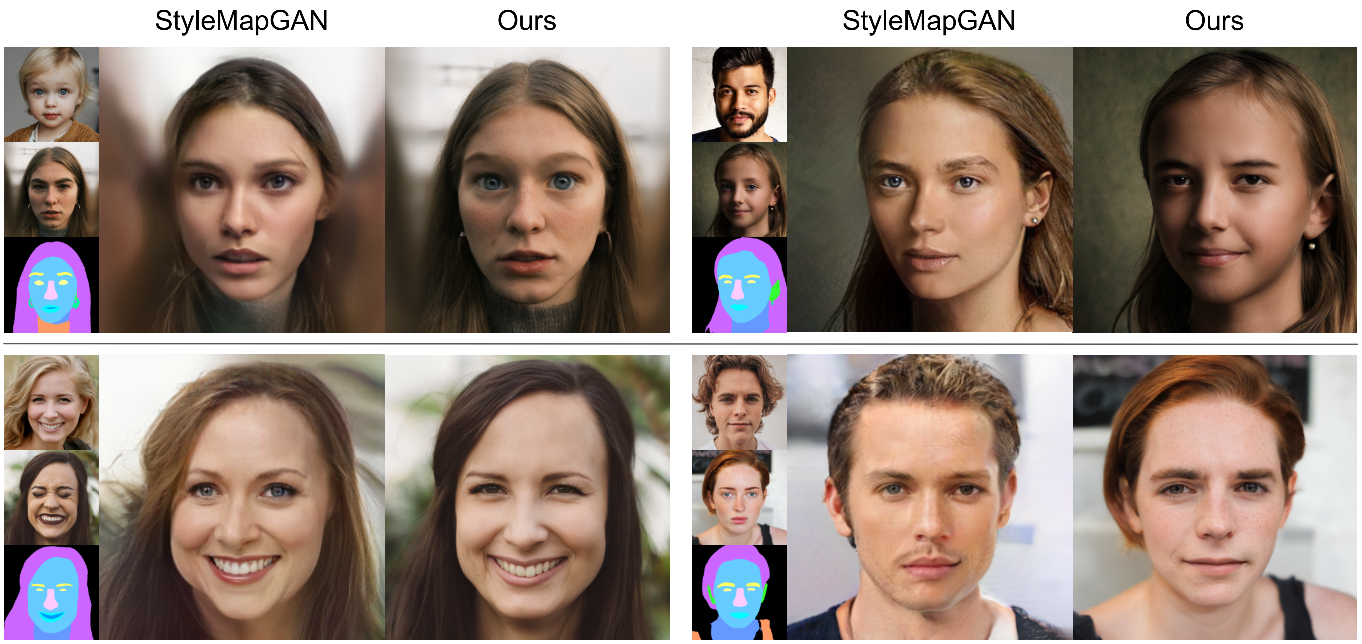

Fig. S-2. Comparison with StyleMapGAN [Kim et al. 2021]. First row: examples of eye and eyebrow transfer; second row: examples of face swapping. Note that ours successfully edits the portrait locally, while StyleMapGAN provides a completely different person.

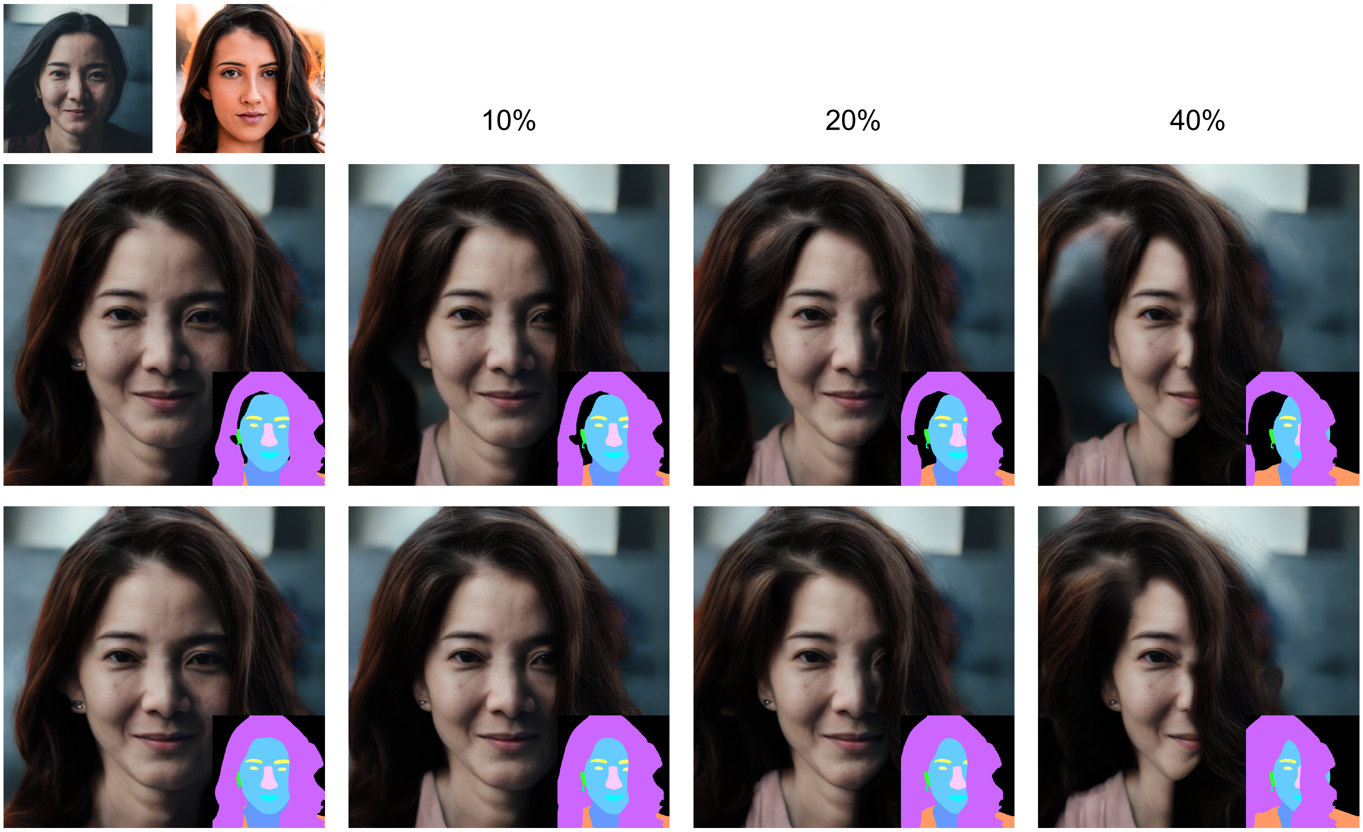

Fig. S-3. Misalgin the target segmentation mask by shifting. First row: translate the target hair region without preprocessing the segmentation mask; Second row: use the segmentation mask preprocessing step. Please note the artifacts between the hair and neck.

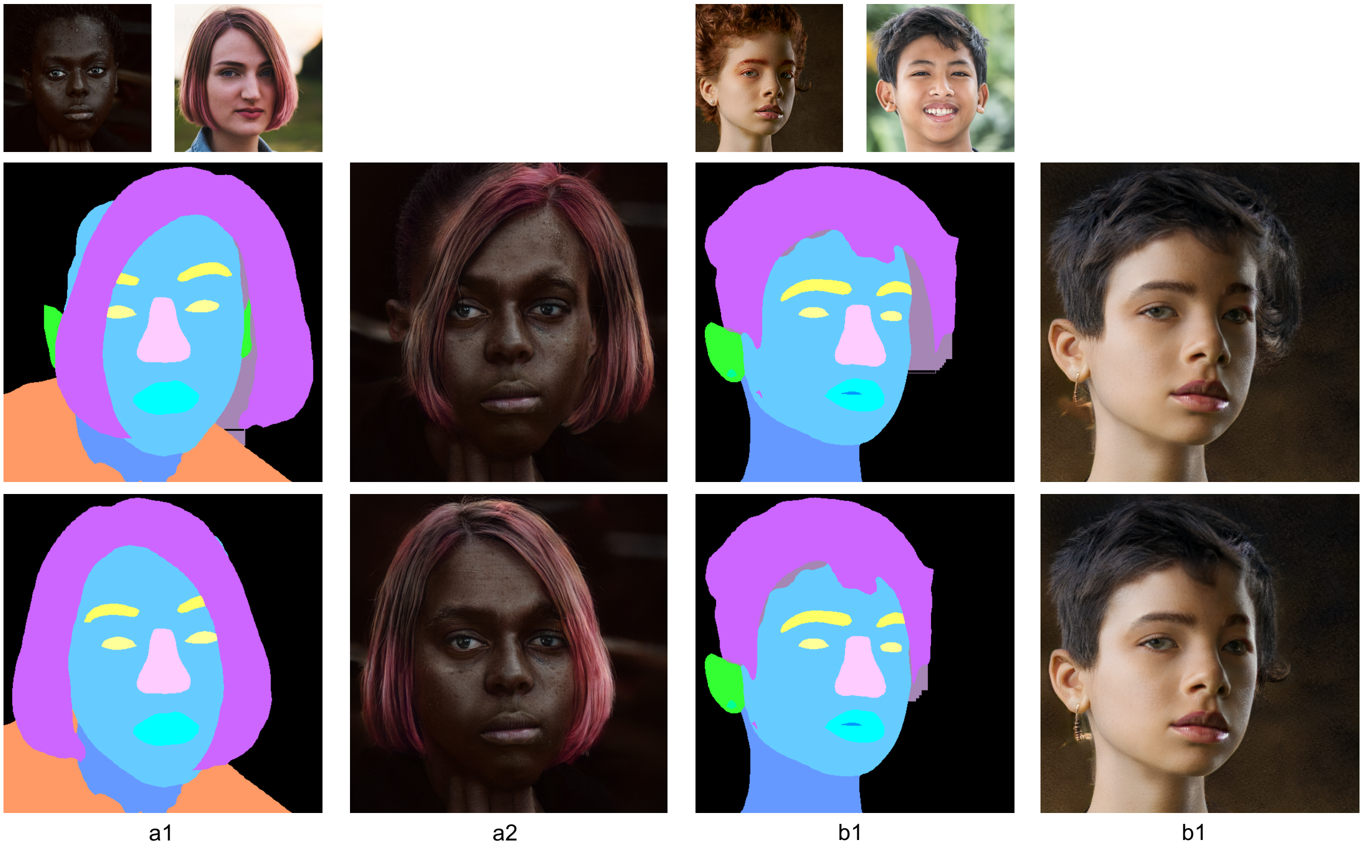

Fig. S-4. The second row shows the results of manually editing the target segmentation mask. The left portion of the figure shows an example where the hair and the face could be aligned by flipping the hair segmentation mask. The right portion shows an example in which the regions could be better aligned by translating them