7. Details of attribute regression network

The attribute regression network is composed of an untrained, randomly initialized ResNet50 network whose feature output is processed through an additional convolution layer and two fully connected layers (see Figure 5). .

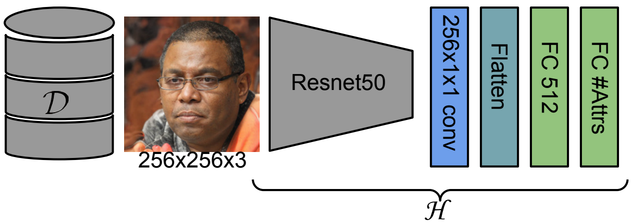

Figure 5: The architecture of the attribute regressor network $\mathcal{H}(I) \rightarrow \widehat{\boldsymbol{a}}$

We trained the attribute regression network on three different datasets as described in the paper. We used a 90/10 split that corresponds to 721718 images for training and 72172 images for the test. All attributes are normalized to the -1 to 1 range. We trained the network both with multi-class loss and mean squared error. Figure 6 shows the L1 error on the 8 attributes used in our quantitative comparisons.

Correcting for unfeasible attribute combinations. As discussed in the paper, our training process relies on randomly changing a single attribute, creating an attribute vector $\boldsymbol{a}'$. A problem with this approach is that $\boldsymbol{a}'$ might correspond to an unfeasible combination of attributes, such as a bearded man with no facial hair.

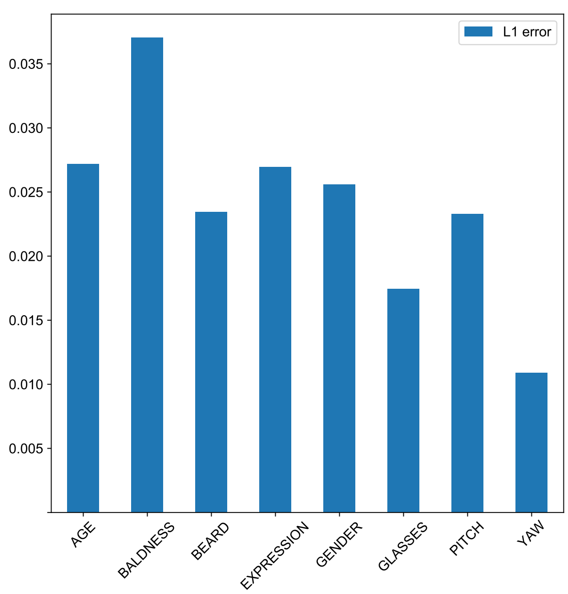

Figure 6: The L1 error of the attribute regression network on the held-out test set for the eight attributes used for quantitative comparisons.

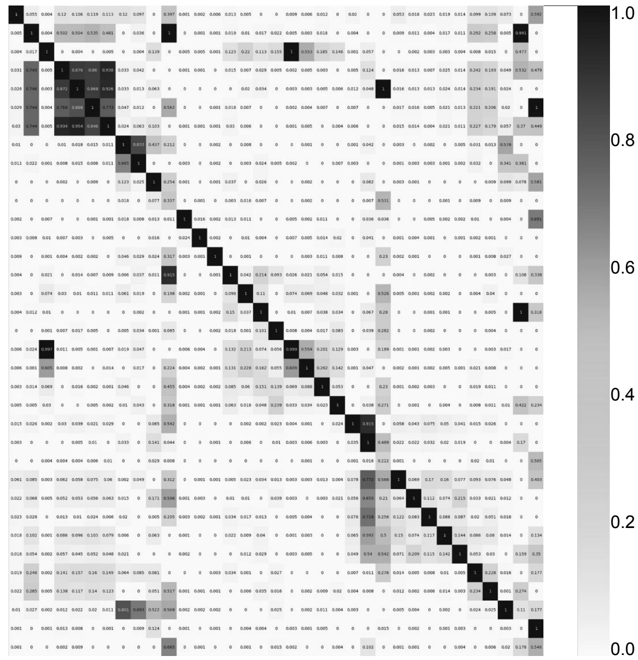

Figure 7: The mutual information matrix between attributes. The clusters of correlated attributes that we observed include: (Beard, Moustache, Facial hair, Sideburns), (No glasses, Glasses, ReadingGlasses), and (Hair invisible, Gray hair, Blond hair, Brown hair, Red hair, Black hair).

To resolve this, we first created a correlation matrix based on all the attributes in the training dataset. Using this data we create a corrected vector $\boldsymbol{a}'_c$ by multiplying the corresponding row elements from the correlation matrix with the $\boldsymbol{a}'$ if the elements are larger than a threshold (which we chose to be 0.8) (see Figure 7).

8. Face identity

In addition to the face identity comparison based on the cosine distance reported in the main paper, Table 2 shows the results using the Euclidean distance. Again, we find that our approach maintains the face identity better than other approaches for most target attributes.

9. Additional qualitative comparisons

In this section, we show several additional qualitative results and compare the range of edits our method and proposed baselines. Figures (8, 9) demonstrate more results of our method.

We compare more examples of changing attribute “Age” with different methods. Figures (10, 11, 12) demonstrate the results of these comparisons. We also show more examples of changing attribute “Beard” (Please see Figures 13 and Figure 14).

Table 2: Comparing the identity loss between our method and the baselines. We report the average Euclidean distance between ground truth and edited image identity features.

ATTRIBUTE	INTERFACE GAN	STYLE FLOW	GAN SPACE	LATENT TO LATENT
AGE	1.10	1.07	**0.91**	1.01
BALDNESS	0.75	1.06	0.92	**0.61**
BEARD	0.96	0.98	0.97	**0.69**
EXPRESSION	0.73	0.59	0.64	**0.50**
GENDER	0.90	0.95	1.02	**0.70**
GLASSES	0.82	0.88	0.67	**0.65**
PITCH	**0.87**	0.93	0.97	0.95
YAW	**0.67**	0.89	0.87	0.90

10. L2L mapper ablation studies

In this section, we perform ablation studies on the mapper network. First, we study the impact of the number of layers of the mapper network on the quality of the edits. We trained 4 models with 1, 2, 3 and 4 hidden layers while keeping all other components of our model the same.

We qualitatively, compared these 4 methods (See Figure 15 top 4 rows). We observe that there is not much benefit from adding more layers to the network. This means we can achieve the same performance with a network with just one hidden layer. Furthermore, we study the impact of different loss terms. We trained a mapper network with 1 hidden layer and without identity loss (l2l-1-id), and another mapper network with 1 hidden layer and without identity and neighborhood loss (l2l-1-id-nb). Figure 15 bottom 2 rows shows the image of a face when changing age with these two networks. We see that trained model without these loss terms performs worse than a model that is trained with these loss terms (l2l-1).

Finally, we performed a user study on 10 faces (see next section for more details) and count the number of undesired changes when changing attributes “age” or “facial hair”. Table 3 shows the results of this user study.

Table 3: The number of annotated attribute changes by user study for different methods: l2l-4-layers / l2l-3-layers / l2l-2-layers / l2l-1-layer / l2l-1-no-same-attrs-no-sparsity / l2l-1-no-same-attrs.

FACES(10)	AGE	FACIAL HAIR
IDENTITY	0 / 0 / 0 / 0 / 2 / 0	3 / 1 / 1 / 0 / 1 / 0
GENDER	1 / 0 / 0 / 0 / 0 / 4	0 / 0 / 1 / 0 / 0 / 0
HAIR COLOR	0 / 0 / 0 / 0 / 1 / 0	1 / 0 / 0 / 4 / 1 / 3
HAIR LENGTH	0 / 0 / 0 / 0 / 0 / 0	0 / 0 / 0 / 0 / 0 / 0
FACIAL HAIR	0 / 0 / 0 / 0 / 1 / 2	1 / 2 / 1 / 2 / 1 / 0
GLASSES	1 / 0 / 0 / 0 / 0 / 0	0 / 0 / 0 / 0 / 0 / 0
EYE OPENNESS	1 / 0 / 0 / 0 / 0 / 1	0 / 0 / 0 / 0 / 0 / 0
UNDESIRED CHANGES	3 / 0 / 0 / 0 / 4 / 7	4 / 1 / 2 / 4 / 2 / 3

11. User study

We selected 35 random face images that are separated from the training images. Following the procedure described in section 4.3, we generate images for all these 35 images for all 8 attributes with 4 different methods (StyleFlow, InterfaceGAN, GANSpace, l2l). We then performed a user study on these generated images by checking what attributes are changed during the edits. For example Figure 16 shows a screenshot of the application we developed for the user study where the “facial hair” attribute has changed. As another example, Figure 17 illustrates how the annotator marked both “Facial Hair” and “Gender” as the attributes that are changed.

After conducting this user study, we count the number of undesired changes and report the detailed results in table 4 for men faces and table 5 for women faces. Note that we do not add facial hair to women, but it is possible to change the gender with our method and then add facial hair.

12. Demo video and codes, amount of compute

Our code is anonymized and submitted with supplemental material. The user-study code will be released as well upon acceptance of the paper. We trained our model on an Amazon P3 instance with 8 GPUs. It takes about 20 hours to train the model. Our web-based demo allows the user to interactively edit faces in real-time. We made a demo video and put it on YouTube anonymously. The video can be found at this link: https://youtu.be/fptbQi_yIDg.

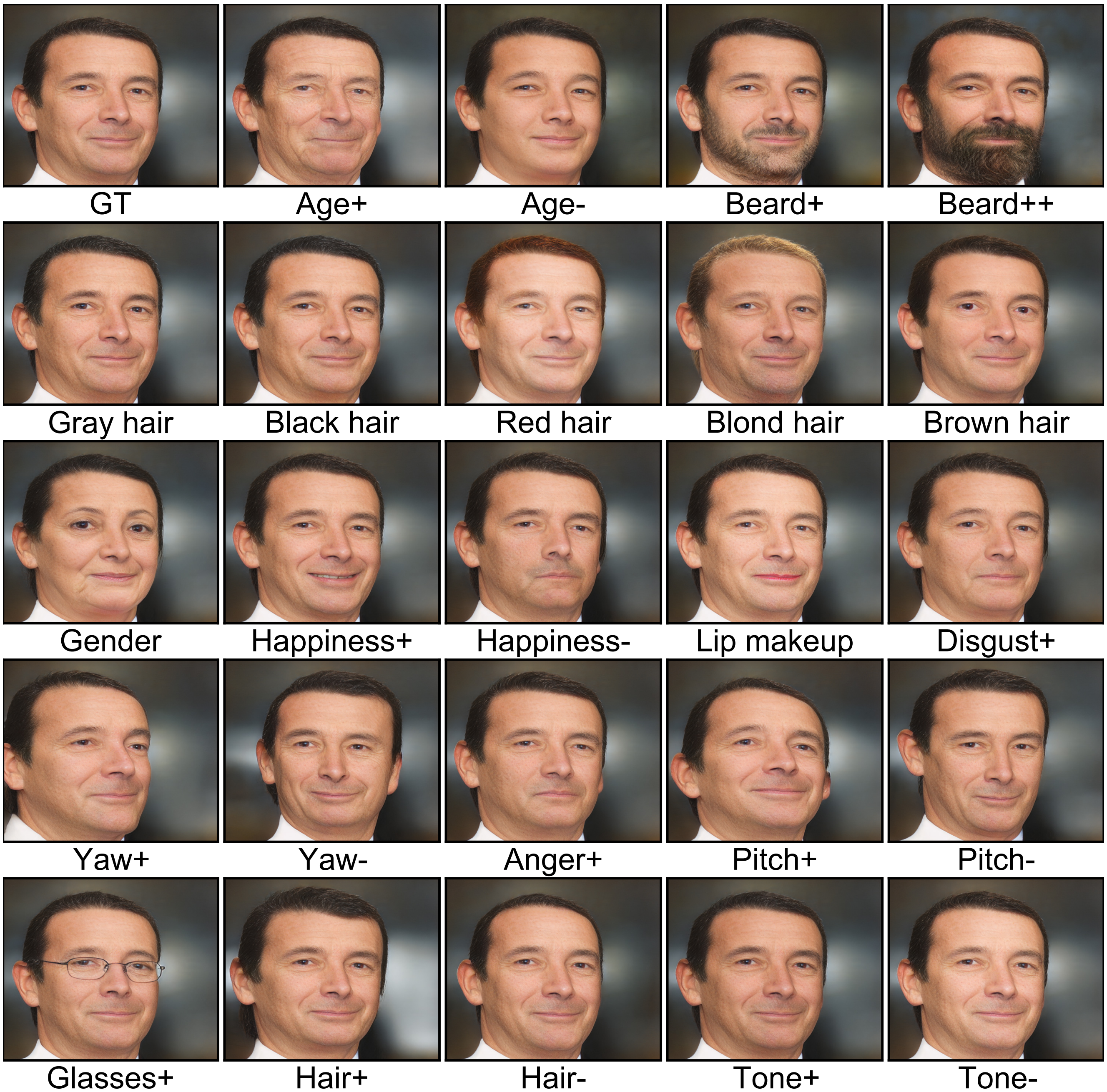

Figure 8: Changing different attributes with latent-to-latent model. GT stands for Ground Truth.

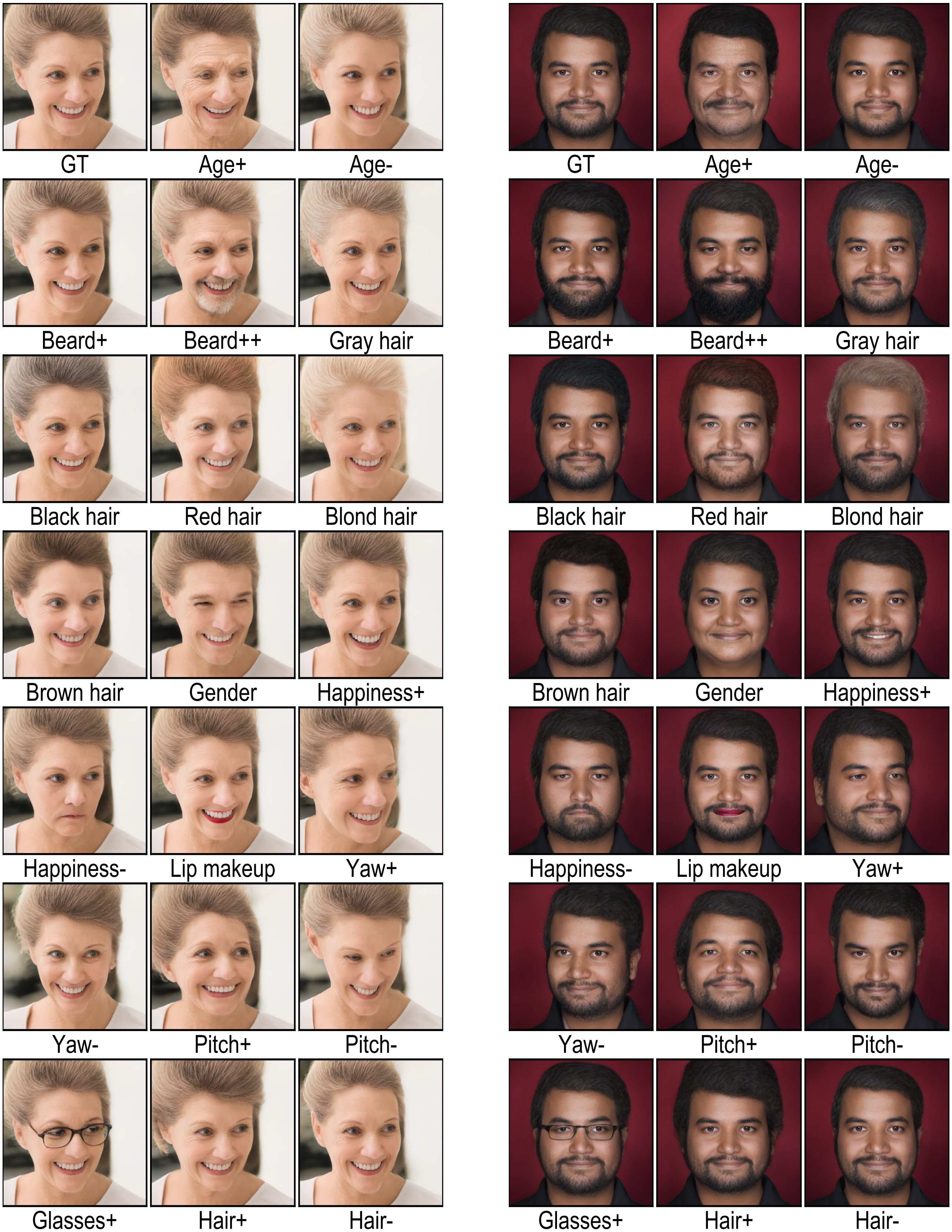

Figure 9: Changing different attributes on different faces. We see that our method performs well on changing various attributes. GT stands for Ground Truth.

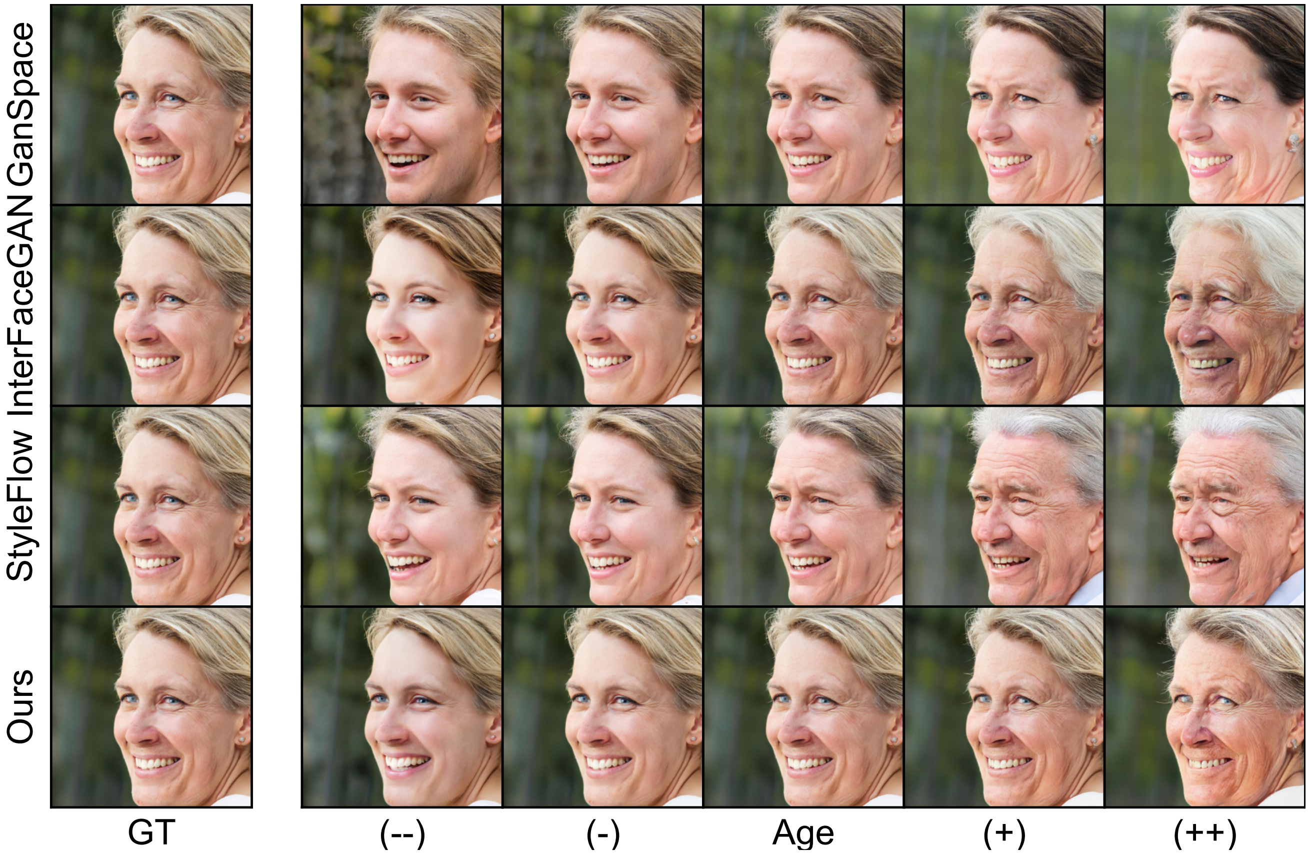

Figure 10: Changing age with different methods.

Table 4: The number of annotated attribute changes by GANSpace / InterfaceGAN / StyleFlow / l2l for males.

MALES(14)	AGE	FACIAL HAIR	GLASSES
IDENTITY	4 / 0 / 0 / 1	2 / 1 / 0 / 0	10 / 1 / 1 / 0
GENDER	6 / 2 / 1 / 3	1 / 10 / 14 / 0	0 / 0 / 0 / 0
HAIR COLOR	0 / 7 / 6 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0
HAIR LENGTH	0 / 1 / 2 / 0	0 / 2 / 0 / 0	1 / 0 / 0 / 0
FACIAL HAIR	4 / 8 / 6 / 5	1 / 11 / 13 / 13	0 / 3 / 0 / 1
GLASSES	1 / 2 / 1 / 0	0 / 0 / 0 / 0	1 / 11 / 14 / 12
UNDESIRED CHANGES	15 / 20 / 16 / 9	3 / 13 / 14 / 0	11 / 4 / 1 / 1

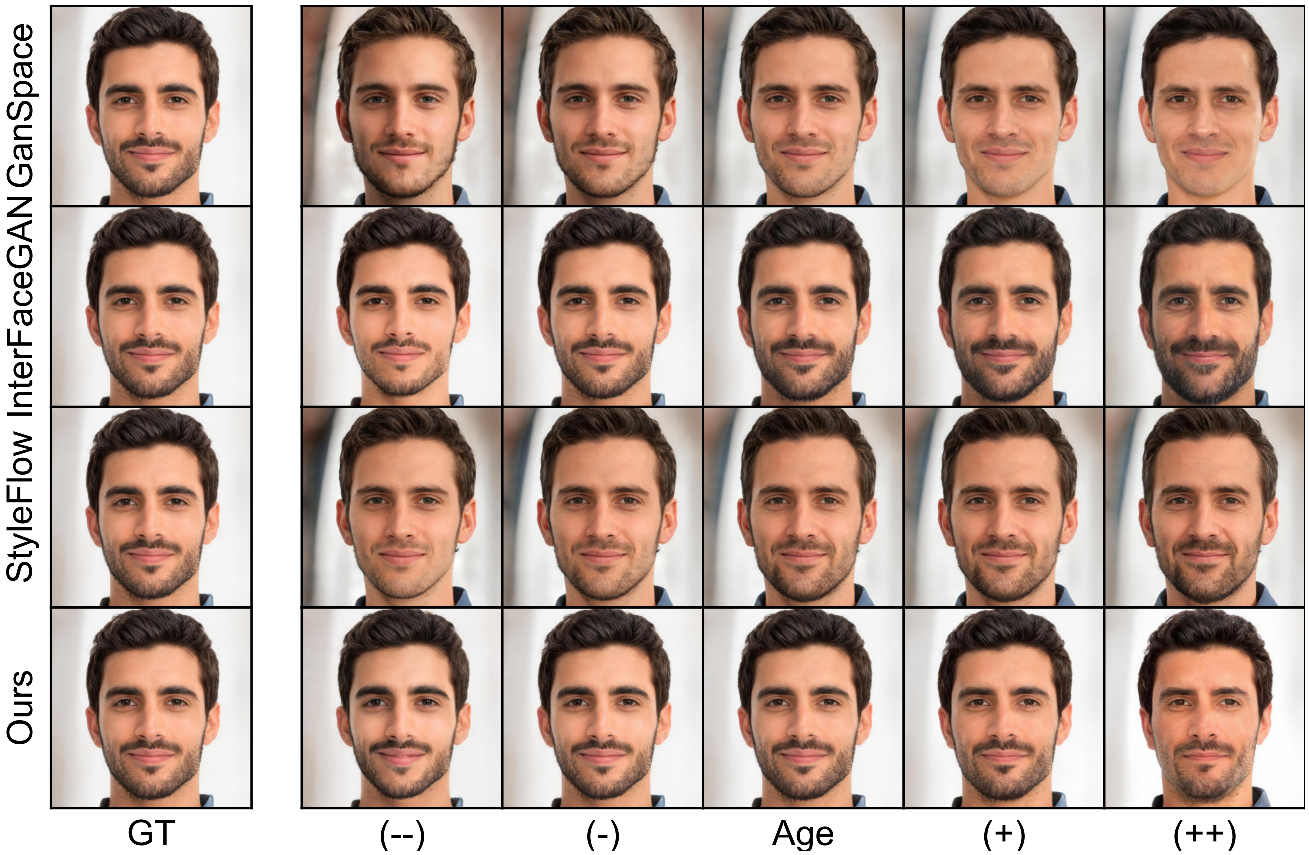

Figure 11: Changing age with different methods. Facial hair does not change with age with our method.

Table 5: The number of annotated attribute changes by GANSpace / InterfaceGAN / StyleFlow / I2I for females.

Females(20)	Age	Facial Hair	Glasses
Identity	11 / 10 / 3 / 6	2 / 0 / 0 / 0	16 / 2 / 2 / 0
Gender	1 / 1 / 4 / 0	5 / 6 / 13 / 0	0 / 2 / 0 / 0
Hair Color	0 / 0 / 10 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0
Hair Length	0 / 1 / 3 / 0	0 / 0 / 0 / 0	1 / 1 / 1 / 0
Facial Hair	0 / 0 / 1 / 0	2 / 9 / 17 / 1	0 / 0 / 0 / 0
Glasses	1 / 3 / 0 / 1	0 / 0 / 0 / 0	1 / 12 / 18 / 12
Undesired changes	13 / 15 / 21 / 7	7 / 6 / 13 / 0	17 / 5 / 3 / 0

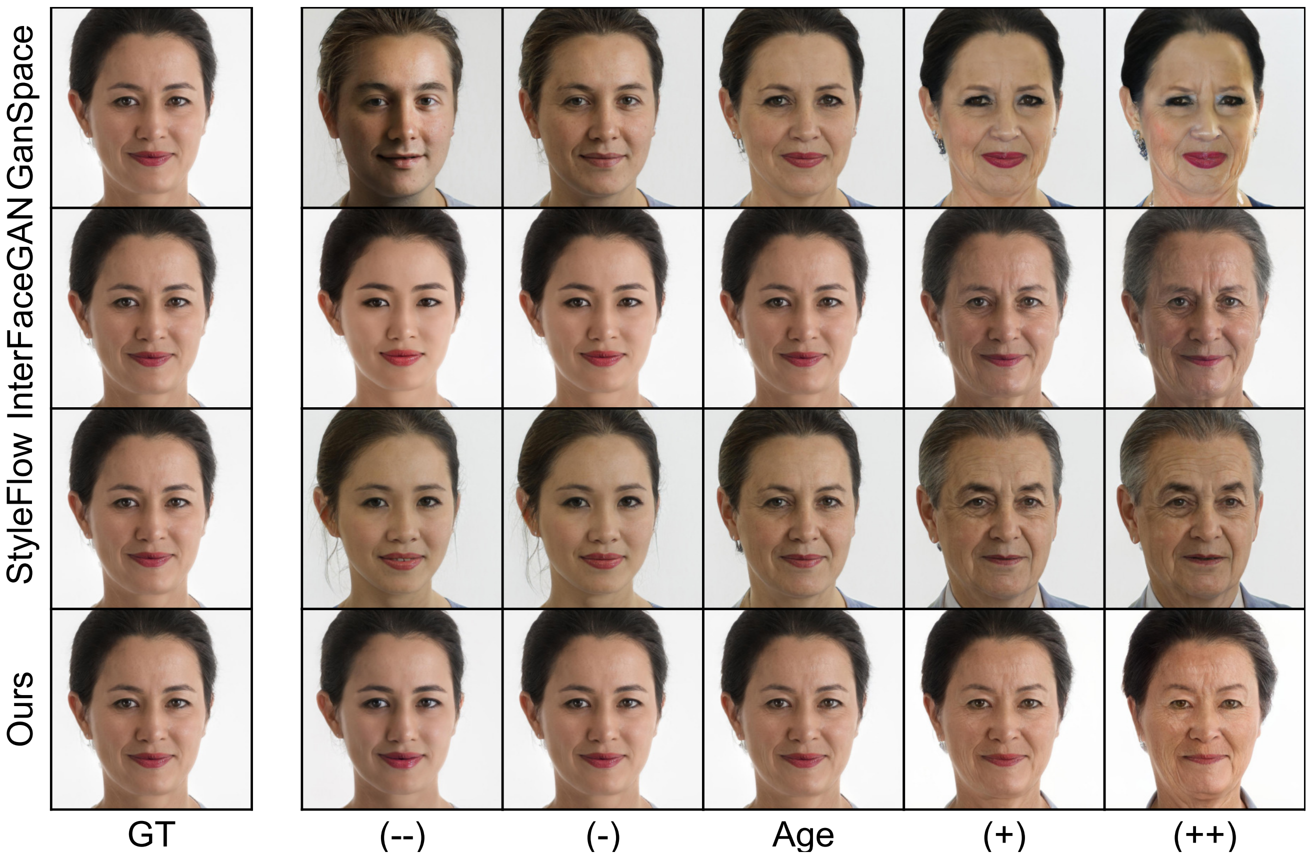

Figure 12: Changing age with different methods.

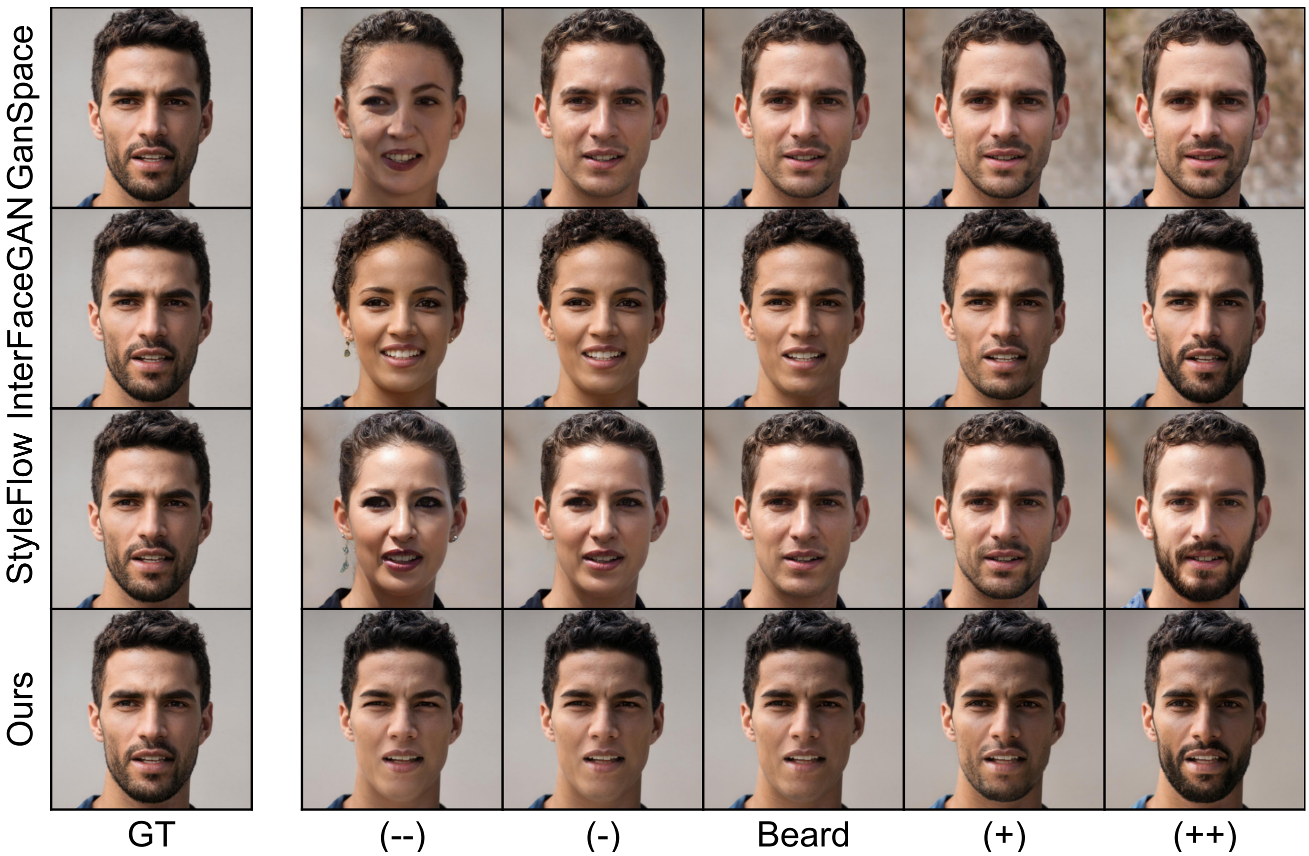

Figure 13: Changing beard with different methods compared with ours.

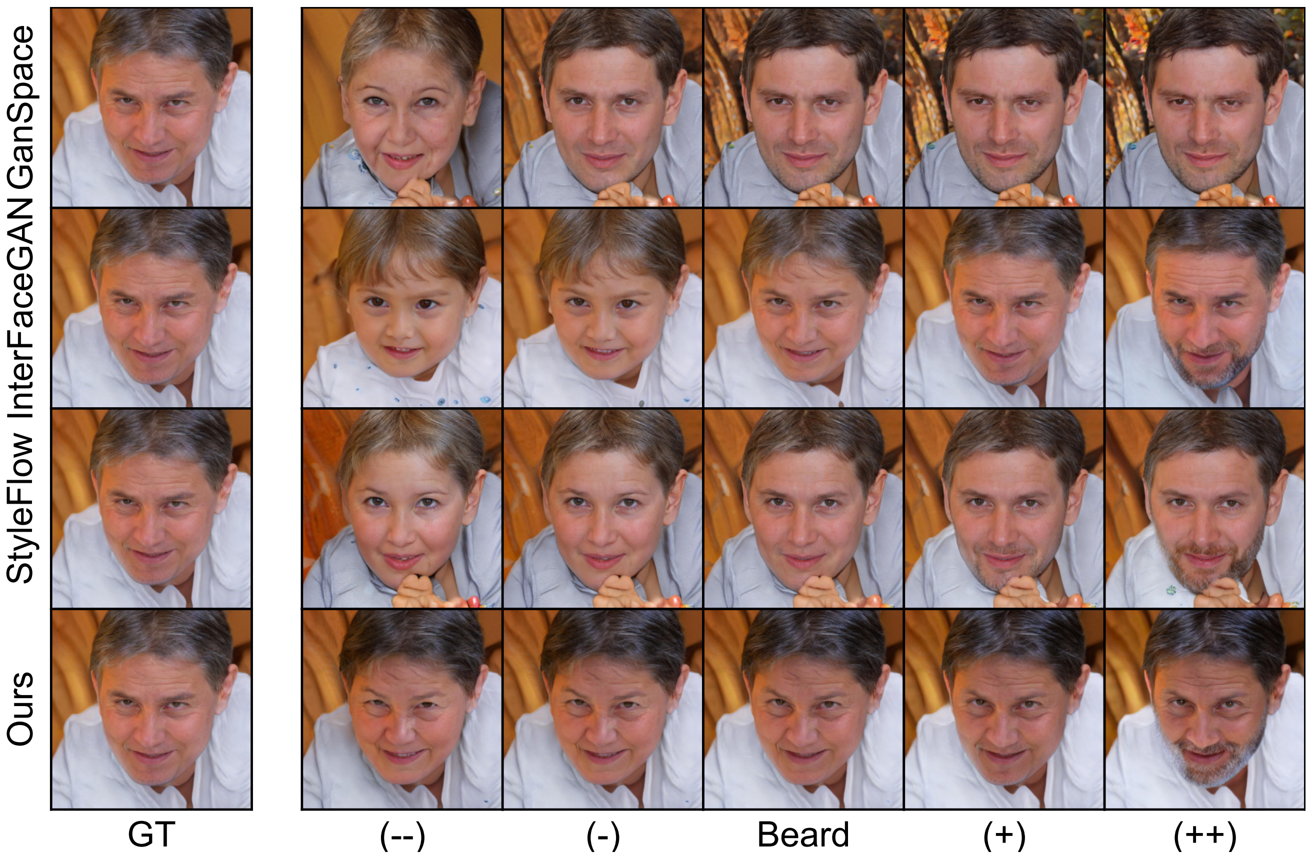

Figure 14: Changing beard with different methods the whole range visualization.

Figure 15: Ablation study example. Changing age with different trained networks.

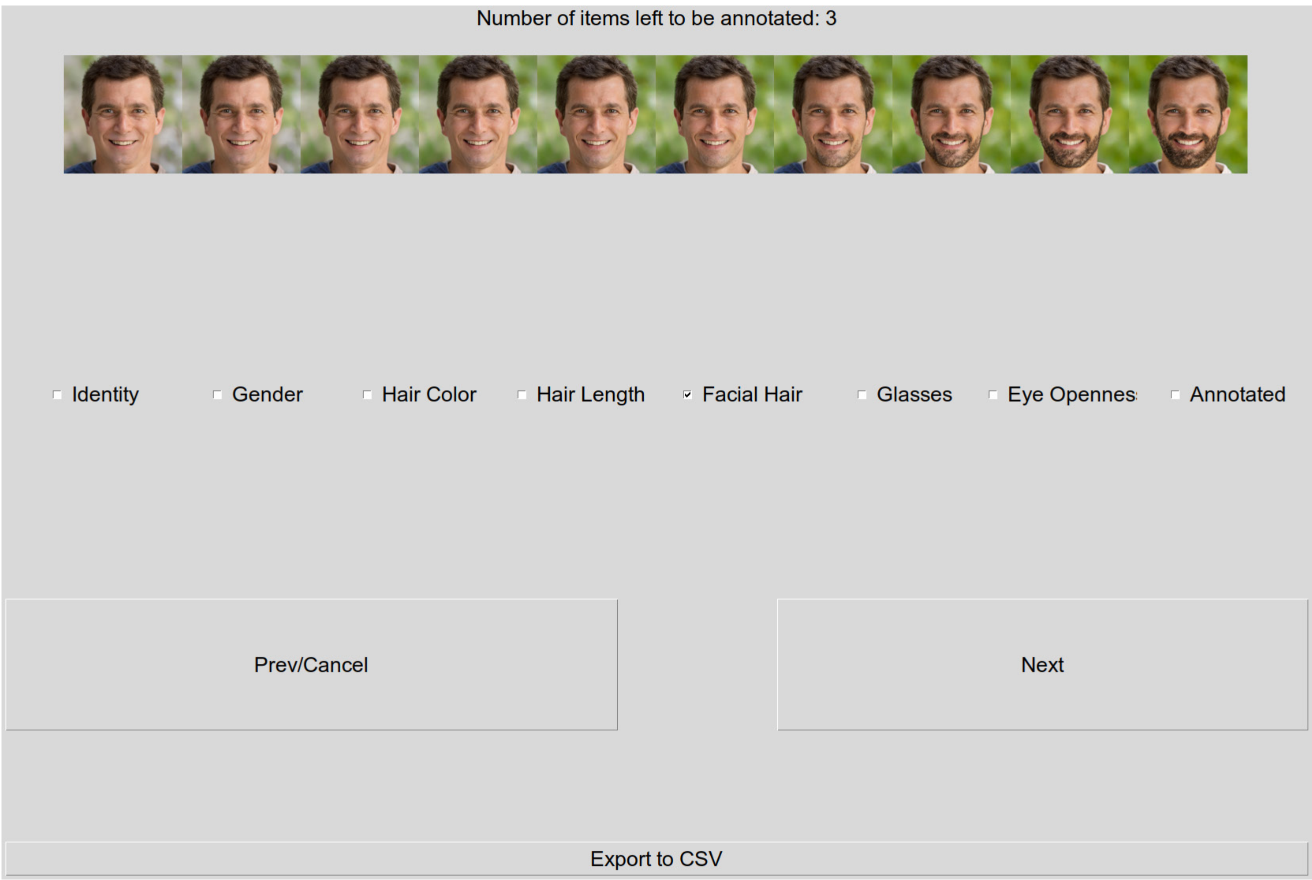

Figure 16: An example of user study application where the “Facial Hair” attribute has changed and marked as changed during the original edits.

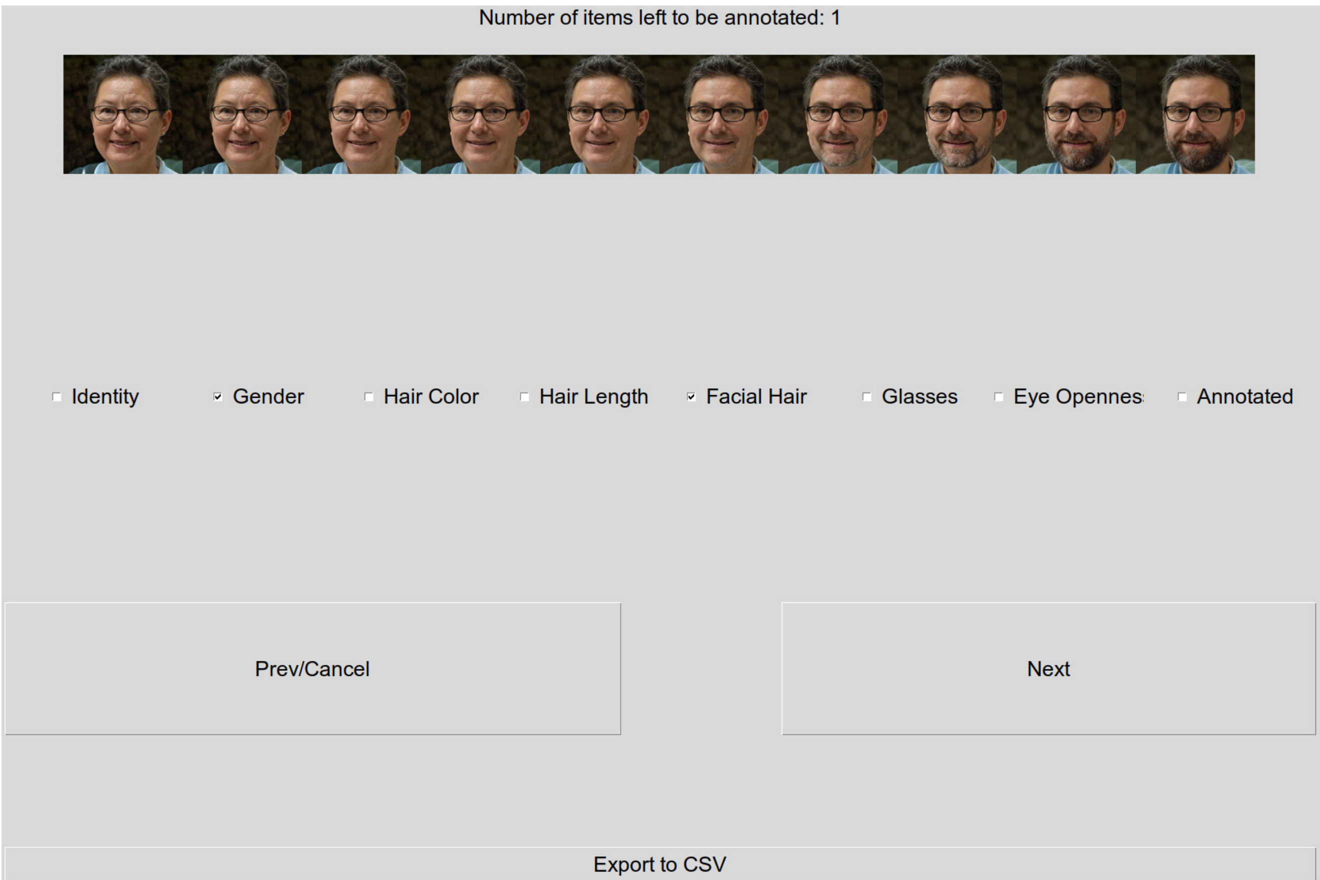

Figure 17: An example of user study application when both attributes “Facial Hair” and “Gender” has changed and marked as changed during the original edits.