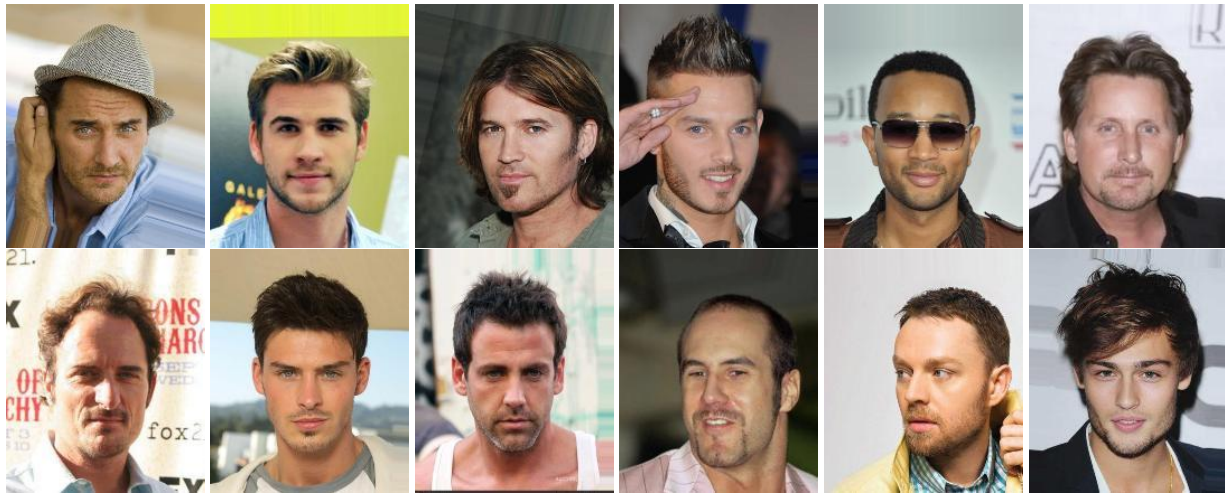


Supplementary Material - Consistency and Accuracy of CelebA Attribute Values

This supplementary material provides the examples of the errors in Section 3.3, which are caused by the short length of facial hair (Figure 1), less information (Figure 2), and inconsistent definition (Figure 3). Also, the examples of the third annotation option - information not visible - are shown in Figure 4. In the CelebA dataset, there are some images that are not usable for face attribute classification, like corrupted image quality, face covered by the heavy face painting, multiple faces, etc. The examples are shown in Figure 5. As mentioned in Section 4.2, the edge images are difficult for people to mark them consistently, they are at the boundary of the definition, shown in Figure 6. Finally, the Score-CAMs in Figure 7 are created to show where the networks look at when they predict the mouth slightly open (MSO). The CAMs show that both models look at the meaningful position when predicting MSO=True. However, when MOON-original predicts MSO=False, instead of looking at mouth, it looks at eye area. Our model looks at the correct position when it predicts MSO=False. Therefore, the cleaned annotations can guide the model to the correct direction.



(a) Error samples of goatee=true in original dataset



(b) Error samples of no Beard=false in original dataset

Figure 1. Error samples that are caused by the short length of facial hair.



(a) Error samples of Wearing_Necklace=true in original dataset



(b) Error samples of Wearing_Necktie=true in original dataset

Figure 2. Error samples that are caused by the less information. The images are shown in pairs. For each image pair, the aligned image is on the left, the original image is on the right.



(a) Error samples of bald=true in original dataset



(b) Error samples of gray_hair=true in original dataset

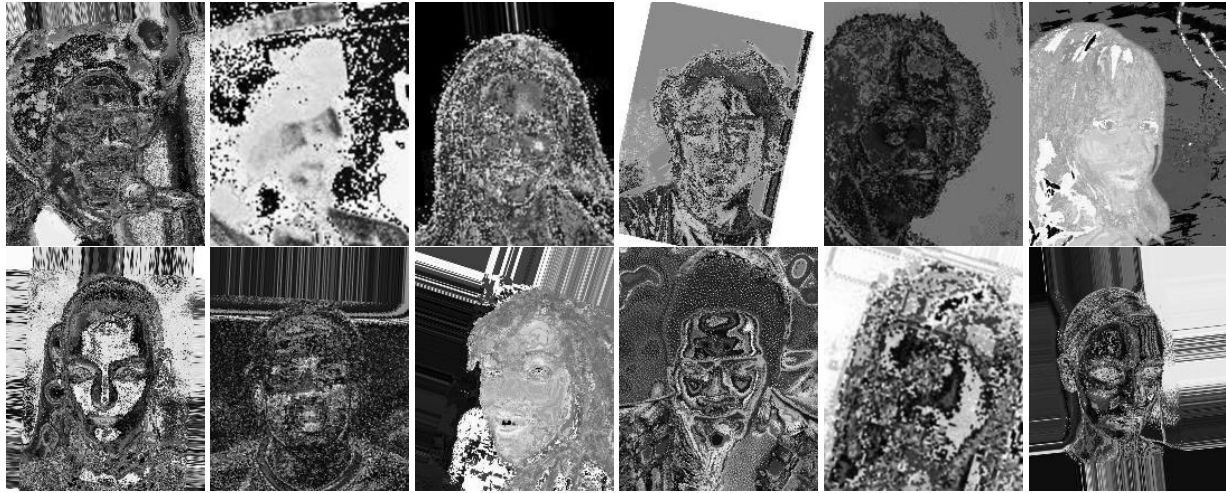


(c) Error samples of MSO=false in original dataset.

Figure 3. Error samples that are caused by the inconsistent definition.



Figure 4. Examples of the third annotation option: information not visible.



(a) Examples of corrupted images in the original dataset.



(b) Examples of heavy-face-painting or face-mask images in the original dataset.

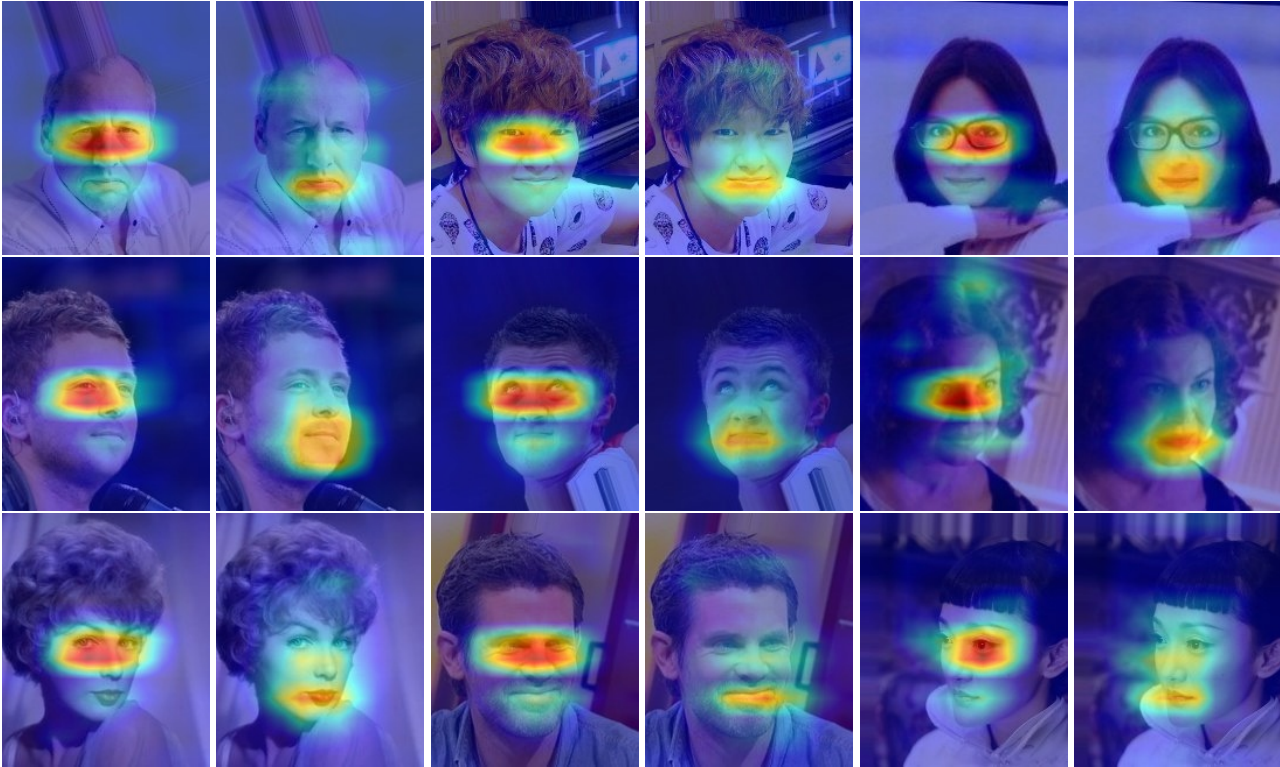


(c) Examples of two-faces images in the original dataset.

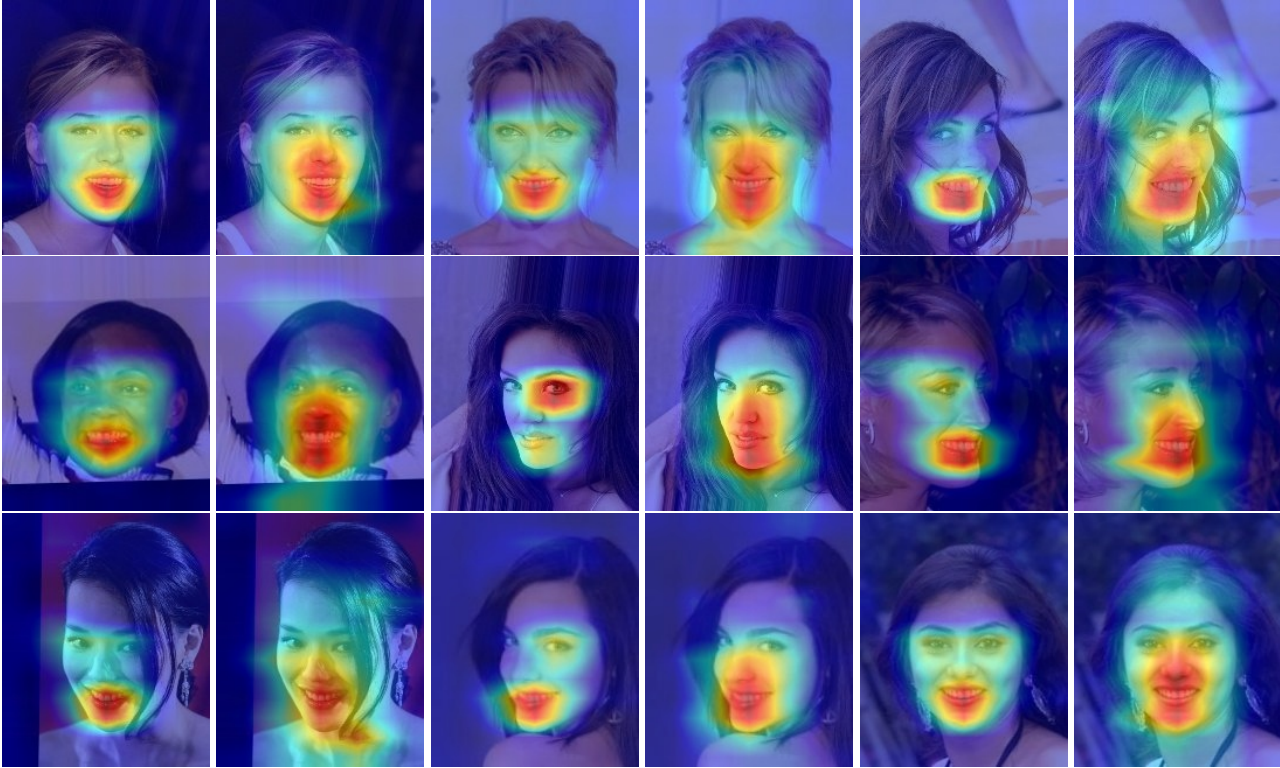
Figure 5. Examples of the part of the unusable images. The main issue of the unusable images is no faithful facial attributes, or more than one faces in the image.



Figure 6. Examples of the edge cases of MSO attribute.



(a) Samples of Score-CAM when MOON-original and MOON-ours predict MSO=False



(b) Samples of Score-CAM when MOON-original and MOON-ours predict MSO=True

Figure 7. Examples of Score-CAMs. The images are shown in pairs. For each image pair, the image on the left is from the model created using original dataset, the image on the right is from the model created using our dataset.