7. Supplementary Material

In this section, we provide supplementary material to enhance the understanding of EmoStyle and its applications. Fig 8 displays the cropped mouths of two sample points—one depicting anger and the other excitement—illustrating the precision of facial manipulation in portraying distinct emotions. Additionally, Fig 11, 12 provides extra sample outputs for reference. Fig 13 also includes more real images from our personalized track.

7.1. Diversity in Facial Expression Generation

Our generator produces a broader spectrum of facial expressions, spanning a wider range of emotions in contrast to the original StyleGAN2 generator, which predominantly generates neutral to happy faces. To visually illustrate this distinction, we have presented a heatmap in Figure 10, showcasing the distribution of VA values for 10,000 images generated using both StyleGAN2 and EmoStyle. We can observe that our generator excels in generating a more diverse array of emotional faces.

7.2. W and $W+$ Latent Spaces

Our project involved the utilization of both W and $W+$ latent spaces for inversion and training purposes. While $W+$ generated more visually appealing results, we chose to base our paper on the outcomes obtained from the W latent space, given the existing research and analyses on this particular latent space.

Furthermore, as mentioned in the personalization section, we emphasize the integration of the $W+$ latent space during the training process. Specifically, we optimized the weight parameters of a stack of 18 MLP networks, with each network dedicated to a distinct style. This approach helped us do facial editing within the $W+$ space.

7.3. Limitations

While our generator effectively produces facial expressions with accuracy, it currently falls short in generating emotions with arousal lower than -0.7 and positive valence. Interestingly, we can generate faces in the w+ latent space even within that specific region. For a visual comparison, refer to Fig 9.

7.4. Detailed Experimental Setup

For the quantitative analysis comparing facial expression edits across various methods, including L2L [15], Inter-FaceGAN [31], StyleFlow [1], GANSpace [11] and GAN-mut [7] a deliberate selection of eight distinct valence and arousal values was made. This selection aimed to comprehensively encompass each quadrant within the Valence-Arousal (VA) space, ensuring a comprehensive representation of diverse emotional states. The chosen emotional
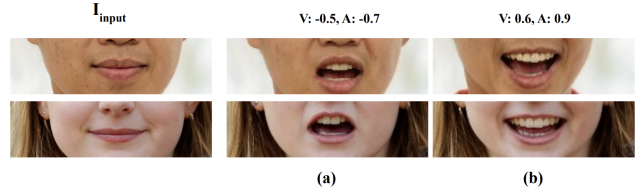

Figure 8. Cropped mouths of two sample points generated by EmoStyle showing (a) anger and (b) excitement.

Figure 9. Comparing the generated faces from (a) W and (b) $W+$ latent space within the challenging VA Region (valence: 0.3, arousal: -0.7)

states featured are anger (-0.6, 0.6), fear (-0.7, 0.3), relaxation (0.3, -0.5), bliss (0.4, -0.2), depression (-0.3, -0.3), sadness (-0.5, -0.2), excitement (0.4, 0.7), and happiness (0.3, 0.5). This strategic choice of values allowed us to thoroughly evaluate the performance of our framework. Further insights and results can be explored further in the experiment section of the paper, as well as being presented concisely in Table 1.

7.5. Ethical Risks

EmoStyle has great potential in various fields like psychology, entertainment, and the VFX industry. However, it is crucial to consider the ethical aspects and potential risks associated with its use. The authors are aware of these ethical implications and are committed to responsible practices. We strongly condemn any misuse of EmoStyle that might compromise individual privacy or spread false information. Moreover, we acknowledge the potential for harm with EmoStyle, as it could be used to create misleading or offensive images, causing distress to the individuals involved. We are dedicated to the ethical and responsible use of EmoStyle and promote its use for research and applications for the benefit of society.
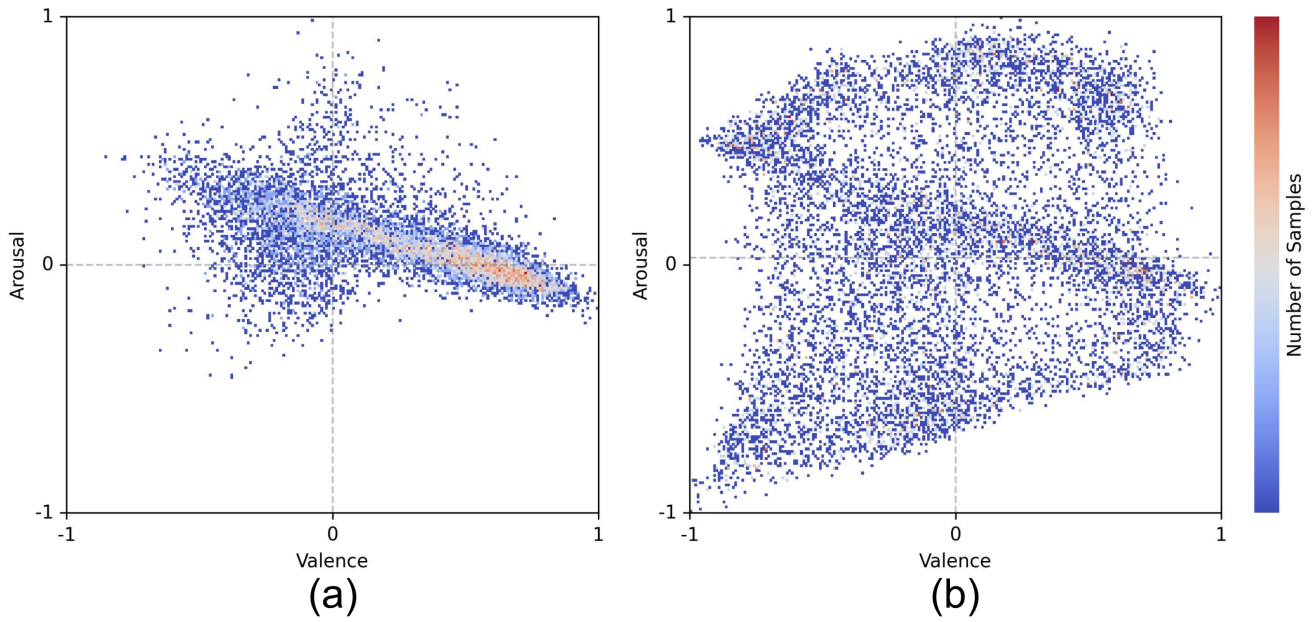
Figure 10. Distribution of VA values for 10k images in a) StyleGAN2 generated images and b) same images edited with EmoStyle given random VA values between -1 and 1
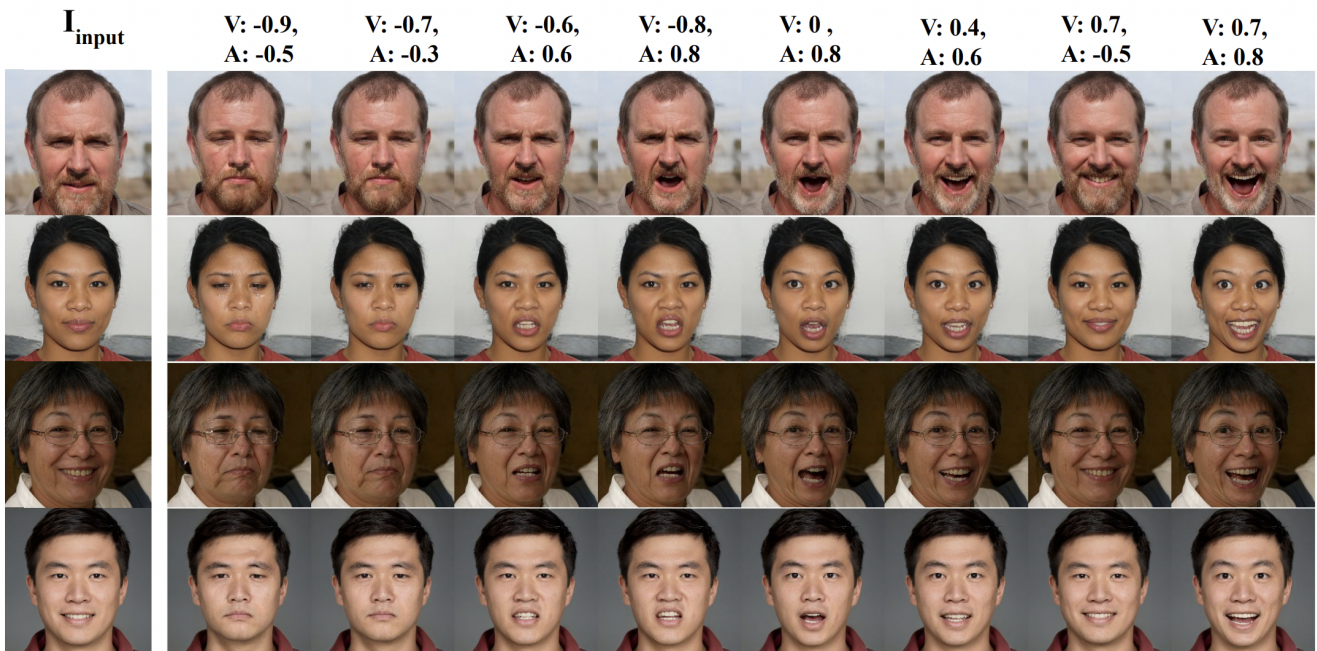

Figure 11. More Sample outputs of EmoStyle, with each output corresponding to a different combination of input valence and arousal.
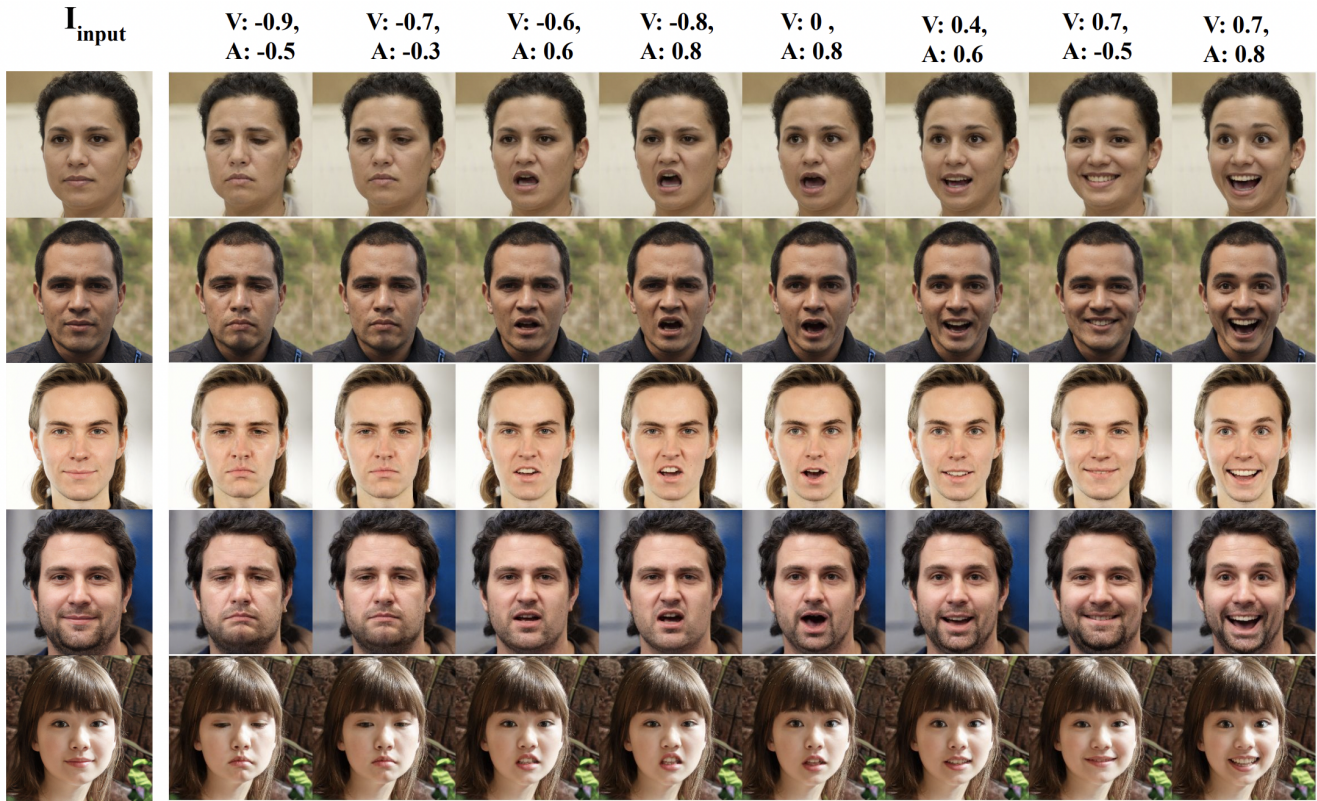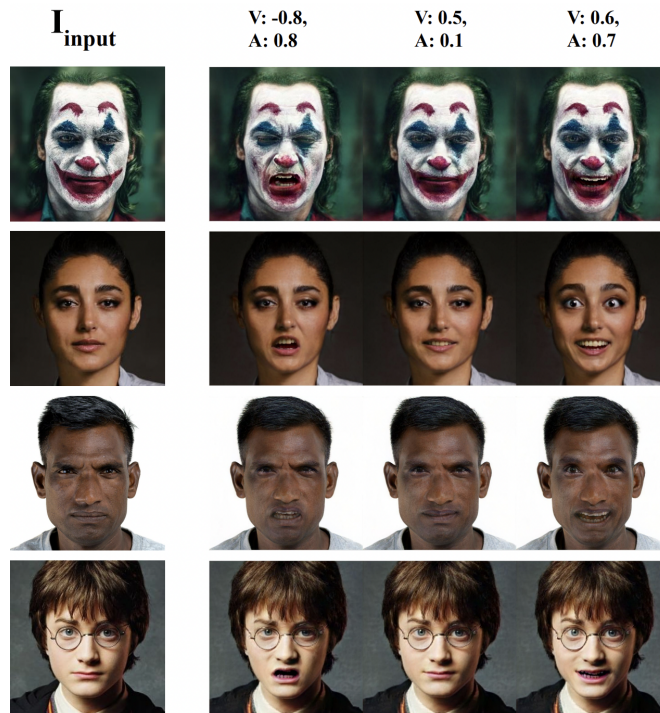
Figure 12. More Sample outputs of EmoStyle.

Figure 13. Real images from the EmoStyle personalized track, edited using different valence and arousal values.