1. Appendix

In this Supplementary Material, we present our additional results and provide further discussions on our method. First, we give more information about our process of constructing the RGB-depth pair datasets, AffectNet-D and RaFD-D. Then, we present plots that illustrate the learning behaviour of our method. In Section 1.3, we provide further quantitative results. Finally, we visualise additional qualitative results on AffectNet with a comparison of images generated by our method and StarGAN.

1.1. Generating RGB-Depth Pairs

As we mentioned in our main paper, there is no large-scale dataset with RGB-Depth pairs for expression classification. Hence, we propose to augment existing expression annotated datasets, AffectNet and RaFD, with depth information. To this end, we propose to use an existing state-of-the-art method to reconstruct the 3D models of faces. We carefully investigated the quality of the reconstructed 3D models and discarded the ones which are not fitted well. From these 3D models, we computed the corresponding depth maps and surface normal maps. Figure 1 shows the pipeline to extract these depth and normal maps. Please check Table 1 for the statistics of RGB image and 3D mesh pairs for both constructed datasets, AffectNet-D and RaFD-D.

1.2. Learning Behaviour:

Figure 3 shows the plots of the learning curves for the proposed method. From these plots, we can observe even after introducing the depth adversarial and depth classification loss, the learning curve is stable and matches the trends with the existing standard adversarial learning frameworks. Our method has lower reconstruction error than the compared baseline, which is StarGAN. This validates that our method is able to disentangle the expressions in a better form and is also capable of reconstructing the images with a better quality. This further supports that our method is superior to the compared baseline in various image quality metrics such as SSIM, PSNR, FID (please check main paper). Similarly, the classification loss for synthetic data is, in general, lower when compared to that of the baseline. This shows that, data generated by our method is classified as the target class more confidently. This observation is parallel with the results we obtained when applying an independent classifier on synthetic images (please see experiments section of main paper).

1.3. Additional Quantitative Results:

As described in the main paper, we report expression generation rate in our experiments, which is calculated by applying a classifier, that is independent of all models, on the synthetic test sets. Figure 2 shows a comparison of the confusion matrices of StarGAN and the proposed method with different weights for the depth network and with confident penalty.

1.4. Additional Qualitative Results:

Figure 4 shows a comparison of samples generated by StarGAN and our method. We can observe that, in general, our method outperforms StarGAN.

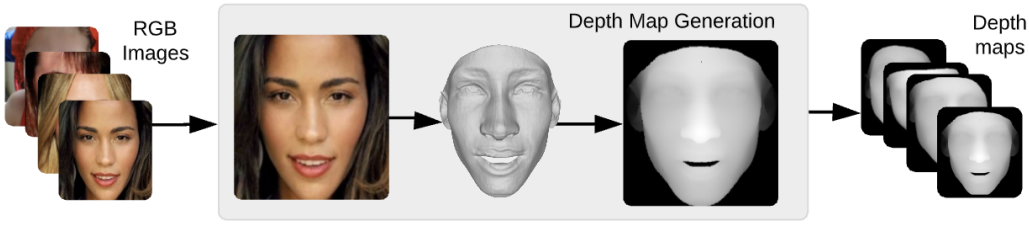

Figure 1. **Pipeline to extract depth maps from RGB images.**

Dataset	Anger	Contempt	Disgust	Fear	Happy	Neutral	Sadness	Surprise	Total
AffectNet-D	15,000	3,703	3,726	6,073	15,000	15,000	15,000	13,604	87,106
RaFD-D	564	580	576	548	557	585	569	515	4,494

Table 1. **RGB-Depth pairs statistics on AffectNet-D and RaFD-D.**

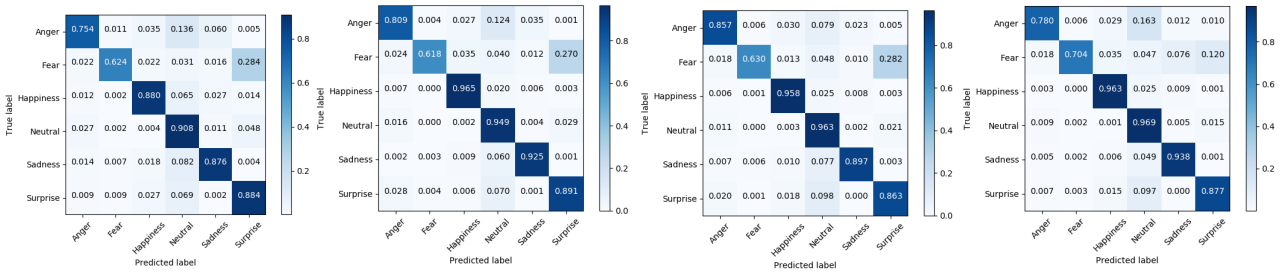

Figure 2. **Confusion Matrices.** The confusion matrices show the performance of StarGAN and our method on AffectNet with different hyper-parameters. The first confusion matrix is for StarGAN, whereas the second and third confusion matrices show the performance of our method with a depth network weight of 0.1 and 0.2, respectively. The last one is obtained by our method with a weight of 0.1 for the depth network with confidence penalty. (Zoom in to view).

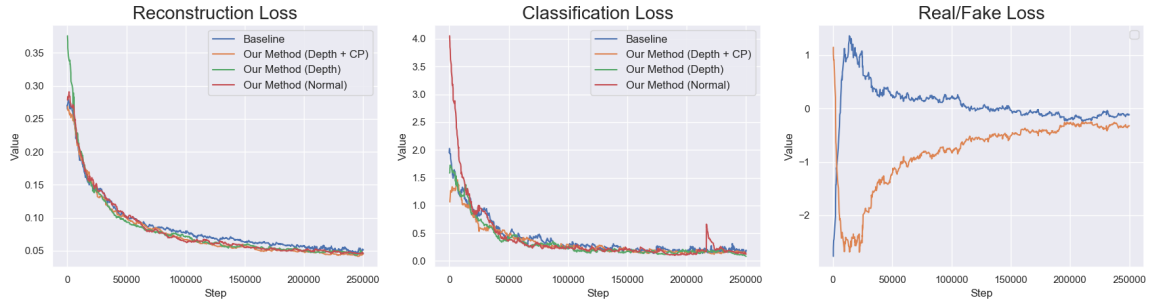

Figure 3. **Learning Curves of Our Method and the Baseline.** The learning curves on the left and in the middle provide a comparison of StarGAN and our method for the reconstruction and expression classification losses of the generator, respectively. The adversarial loss throughout training is shown in the graph on the right-hand side.

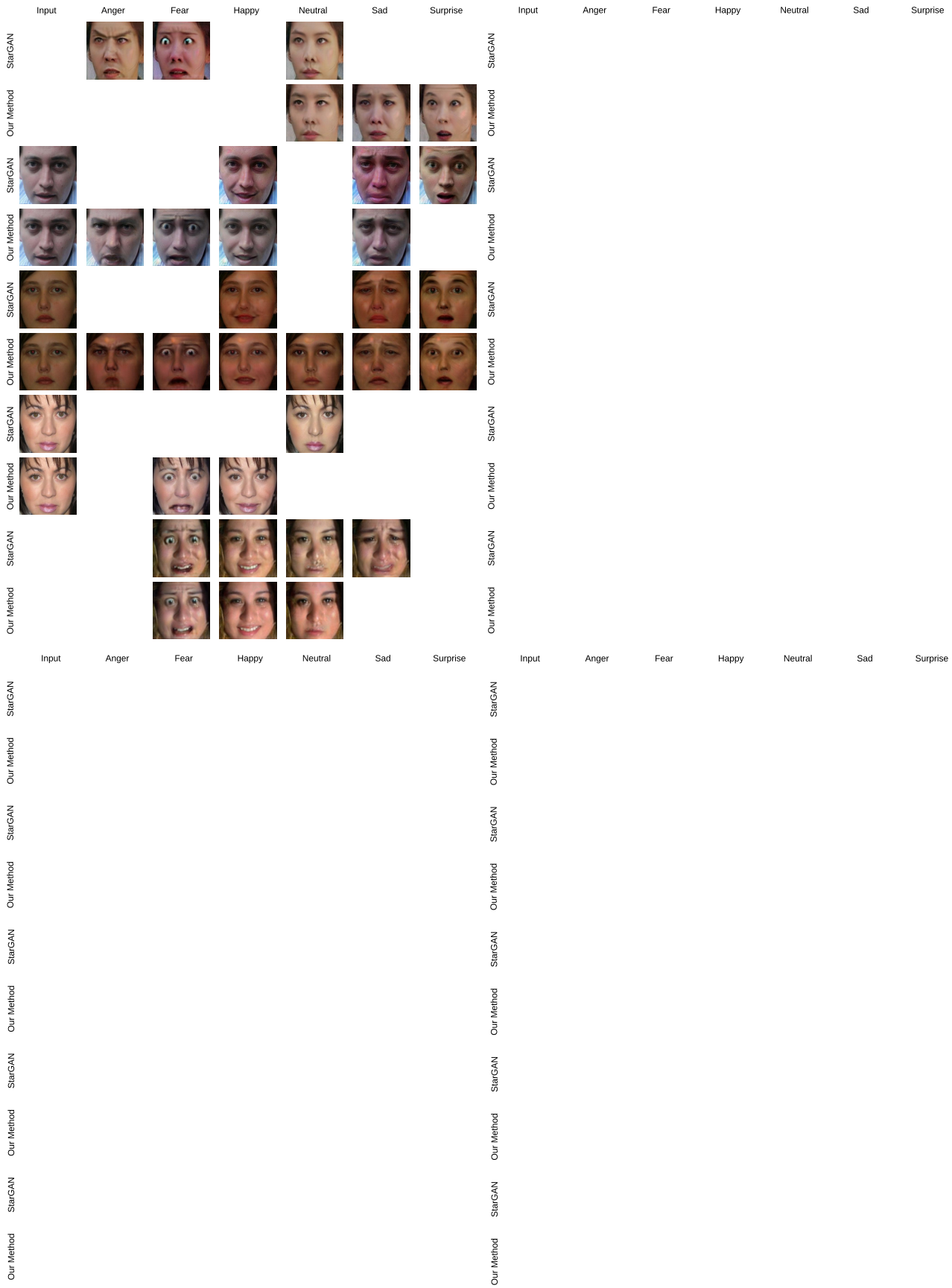

Figure 4. **Additional Qualitative Results on AffectNet.**