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 largescale 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-ofthe-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 | Нарру | 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.

| | Input | Anger | Fear | Happy | Neutral | Sad | Surprise | Input | Anger | Fear | Happy | Neutral | Sad | Surprise |
|---|-------|---|---|---|--|--|---|---|---|-------|-------|---|-----|----------|
| StarGAN | 6 : 4 | 010 | 00 | 00 | 6 24 | | 6.1 | StarGAN | | | | | | |
| Our Method | 6 : 4 | 10-10 | 10 | 010 | 01 | | 6.0 | Our Method | | | | | | - |
| StarGAN | 20 | Z | 00 | Z | 20 | E) | 00 | StarGAN | OC al | - | 6 | 6 | a a | 6 |
| Our Method | E.C. | 25 | 00 | 25) | 30 | 2P | 000 | Our Method | (B) (B) | 60 | 6 | 6 | | 6 |
| StarGAN | 0.0 | | 0 | 26 | 6.6 | | | StarGAN | 36.38 | | 00 | 00 | | |
| Our Method | 030 | 36 | 0 | 20 | 8 | T | 0 | Our Method | 000 | | | | | |
| StarGAN | 100 | 003 | 10 | 10 10 | 4 | The second | 100 | StarGAN | 0 | 00 | | 0 | | |
| Our Method | 100 | C at | 0 0 | te e | 10 | | (0) E | Our Method | | 0 | | 0 | | |
| StarGAN | 25 | None None | 16 1 | 25 | 0 | 25 | (e . 1) | StarGAN | 620 | THE P | 6 | 6 | | (i) (i) |
| Dur Method | 25 | 20 | 10-1 | 25 | | ZE | 2.0 | Our Method | 30 | | 6 | 6 | 0 | (B) (B) |
| 0 | | | | | | A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER | A DECEMBER OF A | | | | | | | |
| Ű | Input | Anger | Fear | Нарру | Neutral | Sad | Surprise | Input | Anger | Fear | Нарру | Neutral | Sad | Surprise |
| StarGAN | Input | Anger | Fear | Нарру | Neutral | Sad | Surprise | StarCAN | Anger | Fear | Нарру | Neutral | Sad | Surprise |
| Our Method StarGAN 0 | Input | Anger | Fear | Happy | Neutral | Sad | Surprise | Our Method StarCAN | Anger | Fear | Нарру | Neutral | Sad | Surprise |
| StarGAN Our Method StarGAN 0 | Input | Anger | Fear | нарру | Neutral | Sad | Surprise | starcAN Our Method StarcAN Our Method StarcAN | Anger | Fear | Happy | Neutral | Sad | Surprise |
| Our Method StarGAN Our Method StarGAN 0 | Input | Anger | Fear | Happy | Neutral | Sad | Surprise | Our method StarGAN Our method StarGAN Our method StarGAN Our method StarGAN | Anger | Fear | Happy | Neutral | Sad | Surprise |
| StarGAN Our Method StarGAN Our Method StarGAN 0 | Input | Anger | Fear | Happy H | Neutral | Sad | Surprise | StarcAN arcan | Anger | Fear | Happy | Neutral | Sad | Surprise |
| Our Method StarGAN Our Method StarGAN Our Method StarGAN C | Input | Anger | Fear Image: I | Happy H | Neutral | Sad | Surprise | Our method StartSAN Our method StartSAN | Anger | Fear | Happy | Neutral | Sad | Surprise |
| StarGAN Our Method StarGAN Our Method StarGAN Our Method StarGAN (| Input | Anger Image: I | Fear Image: I | Happy ()) ()) ()) ()) ()) ()) ()) () | Neutral Neutral <td< th=""><th>Sad</th><th>Surprise</th><th>Input Input Input StarGAN Our Method StarGAN Imput StarGAN Imput StarGAN Imput Imput Imput Imput</th><th>Anger</th><th>Fear</th><th>Happy</th><th>Neutral</th><th>Sad</th><th>Surprise</th></td<> | Sad | Surprise | Input Input Input StarGAN Our Method StarGAN Imput StarGAN Imput StarGAN Imput Imput | Anger | Fear | Happy | Neutral | Sad | Surprise |
| Our Method StarGAN Our Method StarGAN Our Method StarGAN Our Method StarGAN O | Input | Anger Image: I | Fear Image: Im | Happy ()) ()) ()) ()) ()) ()) ()) () | Neutral Neutral <td< th=""><th>Sad</th><th>Surprise</th><th>Our wentod starGAN Our wentod starGAN Our metod starGAN</th><th>Anger</th><th>Fear</th><th>Happy</th><th>Neutral Image: Image:</th><th>Sad</th><th>Surprise</th></td<> | Sad | Surprise | Our wentod starGAN Our wentod starGAN Our metod starGAN | Anger | Fear | Happy | Neutral Image: | Sad | Surprise |
| StarGAN Our Method StarGAN Our Method StarGAN Our Method StarGAN Our Method StarGAN C | Input | Anger Image: I | Fear Image: Im | Happy | Neutral Neutral <td< th=""><th>Sad ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) (()))) (())) (())) (()))) (())) (()))) (()))) (()))) (()))) (())))(())) (())))(())))(()))(()))(())))(())))(()</th><th>Surprise</th><th>Input Input Input StarGAN Our Method StarGAN <t< th=""><th>Anger Image: I</th><th>Fear</th><th>Happy</th><th>Neutral Image: Image:</th><th>Sad</th><th>Surprise</th></t<></th></td<> | Sad ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) ((())) (()))) (())) (())) (()))) (())) (()))) (()))) (()))) (()))) (())))(())) (())))(())))(()))(()))(())))(())))(() | Surprise | Input Input Input StarGAN Our Method StarGAN <t< th=""><th>Anger Image: I</th><th>Fear</th><th>Happy</th><th>Neutral Image: Image:</th><th>Sad</th><th>Surprise</th></t<> | Anger Image: I | Fear | Happy | Neutral Image: | Sad | Surprise |

Figure 4. Additional Qualitative Results on AffectNet.