

# Supplementary Material for “How to Synthesize a Large-Scale and Trainable Micro-Expression Dataset?”

Yuchi Liu<sup>1</sup>, Zhongdao Wang<sup>2</sup>, Tom Gedeon<sup>1</sup>, and Liang Zheng<sup>1</sup>

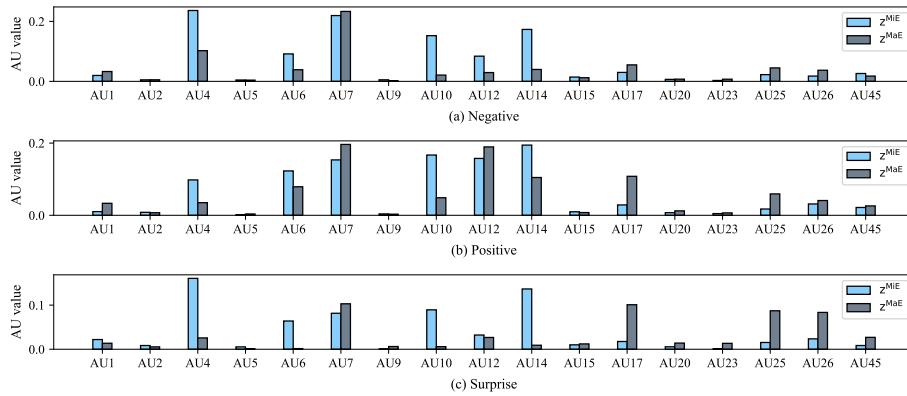
<sup>1</sup> Australian National University, Canberra, Australia,  
{firstname.lastname}@anu.edu.au

<sup>2</sup> Tsinghua University, Beijing, China wcd17@mails.tsinghua.edu.cn  
<https://github.com/liuyvchi/MiE-X>

In the supplementary material, we 1) visualize the values of computed action units of  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$ , 2) show the effectiveness of MiE-X variants using different dataset as ID sources and different datasets as AU sources and 3) the implementation details of models involved in this paper.

## 1 More Analysis on $\mathbf{z}^{\text{MiE}}$ and $\mathbf{z}^{\text{MaE}}$ .

Fig. 1 visualizes the averaged values of Action Unit (AU) vectors of  $\mathbf{z}^{\text{MiE}}$  (extracted from *micro-expressions*) and  $\mathbf{z}^{\text{MaE}}$  (extracted from *early-stage macro-expressions*), respectively, under different micro-expression categories. The setting of AU numbers is same to that used in GANimation [8]<sup>3</sup>.



**Fig. 1.** Averaged values of 17 Action Units calculate on  $\mathbf{z}^{\text{MiE}}$  (extracted from micro-expressions) and  $\mathbf{z}^{\text{MaE}}$  (extracted from early-stage macro-expressions), respectively, under different emotion categories (*i.e.*, **negative**, **positive** and **surprise**). For each category, we have two observations: 1)  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  share similar trends among different AU. For example, both  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  have high values of AU7 for Negative.

<sup>3</sup> 17 representative AU numbers are utilized (*i.e.*, AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, and AU45).

It means they are depicting for the same AU of each kind of expressions, 2)  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  have different values, which means they are complementary to each other.

There are two major findings. First,  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  show similar activation patterns when they have same emotion labels. For example, when the emotion category is positive, both types of AUs have relatively large values in AU7, AU12 and AU14. Second,

$\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  show very different average values for some AU numbers. For instance, their values in AU10, AU14, and AU17 are dissimilar for all the three categories. This suggests  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  complement each other to some extent. This might explain why using synthetic data generated from both  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  for MiE recognition learning achieves higher accuracy than only using one of them (refer Fig. 5 of the main paper).

## 2 Synthesizing MiE-X with A Different ID Source

In the main paper, we use the EmotionNet [2] dataset to sample face IDs. To analyze the generalization ability of the proposed MiE generation protocol on different face datasets, we use CelebA [5] to replace EmotionNet for ID sampling while controlling the numbers of AU triplets and IDs to be unchanged. The generated MiE-X variant whose IDs come from CelebA is evaluated on the CompMiE dataset by three-fold cross validation. Comparative results are shown in Table 1. From the results, we do not observe significant difference in UF1 score after we change the ID source. The results also suggest one of the major findings: MiEs generalize cross faces.

**Table 1.** Comparing two MiE-X variants with different ID sources. We report the three-fold cross-validation results on the CompMiE dataset. EmotionNet and CelebA are compared. The Branches method [4] is used.

ID Source	EmotionNet	CelebA
UF1 (%)	47.7 ± 0.5	47.2 ± 0.7

## 3 Synthesizing MiE-X Using Different Datasets as AU Sources

**Table 2.** Performance comparison between MiE-X variants with different ID sources. We report the three-fold cross-validation results (UF1, %) on the CompMiE dataset. The Branches method [4] is used.

AU Source	$\mathbf{z}^{\text{MiE}}$ source		$\mathbf{z}^{\text{MaE}}$ source		$\mathbf{z}^{\text{MiE}} + \mathbf{z}^{\text{MaE}}$ source	
	CompMiE	MMEW	CK+	Oulu	CompMiE + CK+	MMEW + Oulu
UF1 (%)	34.0 ± 1.1	37.2 ± 1.4	43.9 ± 1.3	42.33 ± 1.4	46.1 ± 1.5	47.5 ± 0.7

In the main paper,  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$  are extracted from the CompMiE [9] dataset and the CK+ [6] dataset, respectively. It is intriguing to investigate the effectiveness of MiE-X when we change its AU sources. Specifically, to compute  $\mathbf{z}^{\text{MiE}}$  and  $\mathbf{z}^{\text{MaE}}$ , we use MMEW [1] and Oulu [10], respectively. The AU computation and data generation protocols remain the same. We compare MiE-X variants with different AU sources in Table 2. We observe that our MiE synthesis protocol can still generate competitive MiE datasets with alternative AU sources. For example, the MiE-X variant with AUs from MMEW and Oulu achieves an UF1 score of 47.5% on CompMiE, while UF1 of the same model trained on CompMiE is 43.6% (refer Table 1 in the main paper).

## 4 Implementation Details of GANimation

This paper employs the GANimation method [8] to synthesize MiEs. It is an image-to-image translation model for face expression manipulation. Given a face image  $\mathbf{x}^{\mathbf{z}}$  with Action Units (AUs)  $\mathbf{z}$  and a target AU set  $\mathbf{z}'$ , GANimation aims to learn a single mapping function  $G : (\mathbf{x}^{\mathbf{z}}, \mathbf{z}') \rightarrow \mathbf{x}^{\mathbf{z}'}$  such that the generated face image not only has the same identity as the original image but also manifests the target AUs. We follow the default settings of GANimation during training. Specifically, we use Adam with a learning rate of 0.0001 and batch size 25. We train for 30 epochs and linearly decay the rate to 0 over the last 10 epochs. The weight coefficients for loss terms are the same as those in the original paper. With a single GTX 2080TI GPU, two days are needed to train this model.

## 5 Implementation Details of The Baseline Classifiers

We adopt the Branches [4] and the Apex [7] as the baseline methods for MiE recognition in the main paper. Branches won the first place in [9] has two branches which do not share weights. An MiE sample has an onset frame and an apex frame. The two frames are fed into the two branches (using ResNet-18 [3] as backbones), respectively, and their embeddings after global average pooling are concatenated. The classifier has two fully connected (FC) layers with dimension 128 and 32, respectively. We use Adam with a learning rate of 0.0001 and batch size 32. We train for 80 epochs on real-world data 30 epochs on MiE-X. Compared with Branches, Apex only has one CNN branch and takes the apex frame as the input. It also uses ResNet-18 as the backbone. The training strategy of Apex is the same as those in training Branches. With a single GTX 2080TI GPU, around 10 hours are needed to train the baseline classifiers.

## References

1. Ben, X., Ren, Y., Zhang, J., Wang, S.J., Kpalma, K., Meng, W., Liu, Y.J.: Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021) 3

2. Fabian Benitez-Quiroz, C., Srinivasan, R., Martinez, A.M.: Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5562–5570 (2016) [2](#)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) [3](#)
4. Liu, Y., Du, H., Liang, Z., Gedeon, T.: A neural micro-expression recognizer. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE (2019) [2](#), [3](#)
5. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (December 2015) [2](#)
6. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. pp. 94–101. IEEE (2010) [3](#)
7. Peng, M., Wu, Z., Zhang, Z., Chen, T.: From macro to micro expression recognition: Deep learning on small datasets using transfer learning. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 657–661. IEEE (2018) [3](#)
8. Pumarola, A., Agudo, A., Martinez, A.M., Sanfeliu, A., Moreno-Noguer, F.: Ganimation: Anatomically-aware facial animation from a single image. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 818–833 (2018) [1](#), [3](#)
9. See, J., Yap, M.H., Li, J., Hong, X., Wang, S.J.: Megc 2019—the second facial micro-expressions grand challenge. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). pp. 1–5. IEEE (2019) [3](#)
10. Zhao, G., Huang, X., Taini, M., Li, S.Z., Pietikäinen, M.: Facial expression recognition from near-infrared videos. *Image and Vision Computing* **29**(9), 607–619 (2011) [3](#)