How to Synthesize a Large-Scale and Trainable Micro-Expression Dataset?

Yuchi Liu[1], Zhongdao Wang[2], Tom Gedeon[1], and Liang Zheng[1]

[1] Australian National University, Canberra, Australia,
{firstname.lastname}@anu.edu.au
[2] Tsinghua University, Beijing, China wcd17@mails.tsinghua.edu.cn
https://github.com/liuyvchi/MiE-X

Abstract. This paper does not contain technical novelty but introduces our key discoveries in a data generation protocol, a database and insights. We aim to address the lack of large-scale datasets in micro-expression (MiE) recognition due to the prohibitive cost of data collection, which renders large-scale training less feasible. To this end, we develop a protocol to automatically synthesize large scale MiE training data that allow us to train improved recognition models for real-world test data. Specifically, we discover three types of Action Units (AUs) that can constitute trainable MiEs. These AUs come from real-world MiEs, early frames of macro-expression videos, and the relationship between AUs and expression categories defined by human expert knowledge. With these AUs, our protocol then employs large numbers of face images of various identities and an off-the-shelf face generator for MiE synthesis, yielding the MiE-X dataset. MiE recognition models are trained or pre-trained on MiE-X and evaluated on real-world test sets, where very competitive accuracy is obtained. Experimental results not only validate the effectiveness of the discovered AUs and MiE-X dataset but also reveal some interesting properties of MiEs: they generalize across faces, are close to early-stage macro-expressions, and can be manually defined [3].

Keywords: Micro-expression, action units, facial expression generation

1 Introduction

Micro-Expressions (MiEs) are transient facial expressions that typically last for 0.04 to 0.2 seconds [23,9]. Unlike conventional facial expressions (or Macro-Expressions, MaEs) that last for longer than 0.2 seconds, MiEs are involuntary. They are difficult to be pretended, and thus more capable of revealing people's genuine emotions. MiE recognition underpins various valuable applications such as lie detection, criminal justice and psychological consultation.

The difficulty in collecting and labeling MiEs poses huge challenges in building MiE recognition datasets [3]. First, collecting *involuntary* MiEs is strenuous,

[3] This work was supported by the ARC Discovery Early Career Researcher Award (DE200101283) and the ARC Discovery Project (DP210102801).

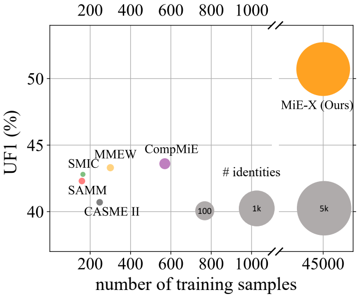


Fig. 1. We present a large-scale synthetic MiE training dataset, MiE-X, created by the proposed protocol. It is two magnitudes larger than existing real MiE recognition datasets in terms of number of MiE samples and number of identities. Compared with existing real-world MiE datasets, MiE-X allows the MiE classifier [21] to achieve consistently higher accuracy evaluated on the real-world MiE dataset CompMiE [33].

even in a controlled environment [3]. Unlike MaEs, which participants can easily "perform", MiEs are too vague and subtle to precisely interpret. Second, correctly labeling MiEs is difficult. It usually requires domain knowledge from psychology experts, and oftentimes even experts cannot guarantee a high accuracy of annotations. As a consequence, scales of existing MiE recognition datasets are severely limited: they typically consist of a few hundreds of samples from dozens of identities (refer Fig. 1 for an illustrative summary). Shortage of training data would compromise the development of MiE recognition algorithms.

In this work, we aim to address the data shortage issue by proposing a useful protocol for *synthesizing* MiEs. This protocol has three steps. First, we conveniently obtain a large number of faces from existing face datasets. Second, we compute sensible AUs. Third, we employ a conditional generative model to "add" MiEs onto these faces. Conditional facial expression generation is a well-studied problem, and we adopt an off-the-shelf algorithm, GANimation [29], which employs coefficients of Action Units (AUs) as the generative conditions.

At the core of this synthesis protocol, we contribute in finding three types of AUs helpful in the second step. The **first** type, intuitively, are AUs extracted from real-world, annotated MiE datasets. Specifically, we extract AU coefficients of annotated MiE samples and use these AU coefficients as conditions to transfer corresponding MiEs to faces of other identities. The **second** type are AUs extracted from early-stage MaEs. The formation of macro-expressions consists of a process of facial muscle movements, and we find early stages of these movements usually share similar values of AUs to those of MiEs. The **third** type are AU combinations given by expert knowledge. For example, human observations suggest that AU12 (`Lip Corner Puller`) is often activated when the subject is "happy", so we set AU12 to be slightly greater than 0 when synthesizing a "happy" MiE. In this regard, this work is an early attempt to explore the underlying *computational* mechanism of micro-expressions, and it would be of value for the community facilitating the understanding of micro-expressions and the design of learning algorithms.

Using the proposed three types of AUs, our protocol allows us to create a large-scale synthetic dataset, **MiE-X**, to improve the accuracy of data-driven MiE recognition algorithms. As shown in Fig. 1, MiE-X is two orders of magnitude larger than existing real-world datasets. Notably, despite being synthetic, MiE-X can be effectively used to train MiE recognition models. When the target application has the same label space as MiE-X, we can directly use MiE-X to train a recognition model, achieving competitive results to those trained on real-world data. Otherwise, MiE-X can be used for pre-training, and its pre-training quality outperforms ImageNet [6]. Our experiment shows that MiE-X consistently improves the accuracy of frame-based MiE recognition methods and a state-of-the-art video-based method.

- We introduce a large-scale MiE training dataset created by a useful protocol, for training MiE recognition models. The database will be released.
- We identify three types of AUs that allow for synthesizing trainable MiEs in the protocol. They are: AUs extracted from real MiEs, mined from early-stage of MaEs and provided by human experts of facial expressions.
- Our experiments reveal interesting properties of MiEs: they generalize across identities, are close to early-stage MaEs, and can be manually defined.

2 Related Work

Facial micro-expression recognition. Many MiE recognition systems use handcrafted features, such as 3DHOG [27], FDM [38] and LBP-TOP [42] descriptors. They describe facial texture patterns. Variants and extensions of LBP-TOP have also been proposed [37,13,14]. Afterwards, deep learning based solutions were proposed [24,17,11,25,16,20]. Petal *et al.* [24] use the VGG model pretrained on ImageNet [6] and perform fine-tuning for MiE recognition. In ELRCN [16], the network input is enriched by the concatenation of the RGB image, optical flow and derivatives of optical flow [34]. To reduce computation cost and prevent overfitting, it is common to use representative frames as model input. For example, Peng *et al.* [26] and Li *et al.* [19] select the onset frame, apex frame and offset frame in each micro-expression video. Branches [21] uses the onset and apex frame as model input. Following this practice, we focus on synthesising representative frames for MiEs.

Deep learning from synthetic data. Deep learning using synthetic data has drawn recent attention. Many works use graphic engines to generate virtual data and corresponding ground truths. Richter *et al.* [30] use a 3D game engine to simulate training images with pixel-level label maps for semantic segmentation. In [32], prior human knowledge is used to constrain the distribution of synthetic target data. Tremblay *et al.* [35] randomize the parameters of the simulator to force the model to handle large variations in object detection. Learning-based approaches [15,31,40] try to find the best parameter ranges in simulators so that the domain gap between generated content and the real-world data is minimized. Another line of works uses generative adversarial networks (GANs) to generate images for learning. For example, the label smoothing regularization technique is adopted for generated images [43]. Camstyle [44] trains camera-to-camera person appearance translation to generate new training data. CYCADA [12] reconstructs images and introduces semantic segmentation loss on these generated images to maintain consistent semantics.

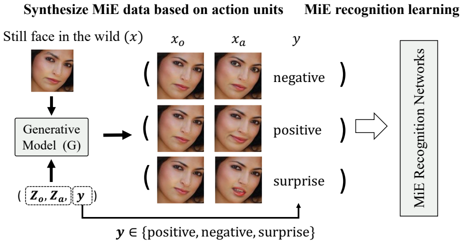


Fig. 2. Overview of the proposed protocol for synthesizing our MiE recognition dataset. We generate MiE samples (a triplet containing onset frame $\mathbf{x_o}$, apex frame $\mathbf{x_a}$ and the emotion label y) with a pretrained GANimation [29] model G, faces in the wild and AU vectors ($\mathbf{z}_o$, $\mathbf{z}_a$) introduced in Section 4.2.

Action Units (AUs) in facial analysis. Action Units are defined according to the Facial Action Coding System (FACS) [8], which categorizes the fundamental facial muscles movements by their appearance on the face. Correlations between Action Units and emotions are widely discussed in literature [7,9,28]. This work uses such correlations where we look for and validate effective AUs as generative conditions to synthesis realistic and trainable MiEs.

3 Preliminaries

MiE recognition aims to classify emotion categories of a given face video clip. In practice, the video clips should be first processed by a *spotting* algorithm to determine the onset (starting time), apex (time of the highest expression intensity) and offset (ending time) frames. In this work, we assume all data have been processed by spotting algorithms [3,33] and focus on the recognition task.

Emotion labels in existing datasets are usually different, ranging from 3 to 8 categories. In this work, we use a unified and balanced label space to synthesize MiE-X. Specifically, during synthesis, we choose the most basic categories (`positive`, `negative`, `surprise`, as defined in MEGC) and merge other emotion labels into these three categories. If the label space in the target dataset is different from MiE-X, we need to fine-tune the model further.

In the following sections, when mentioning action units (AUs), we by default refer to the AU coefficient vector $\mathbf{z} \in [0,1]^d$. Each dimension in vector $\mathbf{z}$ indicates the intensity of a specific action unit. There are usually $d = 17$ dimensions [29,1].

4 Synthesizing Micro-Expressions

4.1 The Proposed Protocol

Given a face image, an emotion label $y \in \{\texttt{positive}, \texttt{negative}, \texttt{surprise}\}$, and an onset-apex AU pair $(\mathbf{z_o}, \mathbf{z_a})$, our protocol uses GANimation [29] to generate an MiE sample consisting of two representative frames (refer Fig. 2).

First, we randomly select an "in-the-wild" face image $\mathbf{x}$ from a large pool of identities (we use the EmotionNet [10] dataset) as the template face upon which we add MiEs. Then, we find an onset AU $\mathbf{z_o}$, an apex AU $\mathbf{z_a}$, and the corresponding emotion label y. A triplet of $(\mathbf{z_o}, \mathbf{z_a}, y)$ could be computed from three different sources, which are elaborated in Section 4.2. Finally, a conditional generative model G is employed to transfer the onset and apex AUs to the template face $\mathbf{x}$, producing an onset frame $\mathbf{x_o} = G(\mathbf{x}, \mathbf{z_o})$ and an apex frame $\mathbf{x_a} = G(\mathbf{x}, \mathbf{z_a})$, whose emotion label is y (same as the label of $\mathbf{x}$). Here, we adopt GANimation [29] as G, which identity-preserving and only changes facial muscle movements. Training details of GANimation are provided in supp. materials.

Please note that the protocol uses existing techniques and that we do not claim it as our main finding. Also note that we do not synthesize entire video sequences of MiEs, but only the onset (the beginning) and apex (most intensive) frames. The motivation is three-fold. First, a full MiE clip may contain up to 50 frames, so a dataset of full MiEs can be 25 times as large as a dataset of representative frames (2 frames per MiE). Second, recent literature on MiE recognition (*e.g.*, [26,19,21]) indicate that using representative frames suffice to obtain very competitive accuracy. Last, synthesizing video sequences in a realistic way is much more challenging than static frames, requiring smooth motions and consistency over time. We leave video-level MiE generation to future work.

4.2 Major Finding: Action Units That Constitute Trainable MiEs

In the protocol, we make the major contribution in finding three sources of AUs that are most helpful to define the onset and apex AUs, to be described below.

AUs extracted from real MiEs. An intuitive source of MiE AUs are, of course, real-world MiE data. Assume we have a real-world MiE dataset with M MiE videos, where each video is annotated with the onset and apex frames. For each video, we extract the onset and apex AUs and record the emotion label, forming a set of AUs $\mathcal{Z}^{\mathrm{MiE}} = \{(\mathbf{z_o}^{(m)}, \mathbf{z_a}^{(m)}\}_{m=1}^{M}$ and labels $\mathcal{Y}^{\mathrm{MiE}} = \{y^{(m)}\}_{m=1}^{M}$. Here, AU coefficients are extracted with the OpenFace toolkit [1]. When synthesizing MiEs with a certain emotion category based on $\mathbf{z}^{\mathrm{MiE}}$, we randomly draw a pair of AUs from $\mathcal{Z}^{\mathrm{MiE}}$ that have the desired emotion label.

Discussion. Despite being a valuable source of MiEs AUs, existing real-world MiE data are severely limited in size, so $\mathcal{Z}^{\mathrm{MiE}}$ is far from being sufficient. If we had more MiE data, it would be interesting to further study whether our method can synthesize a better dataset. At this point, to include more MiE samples in our synthetic training set, we find another two AU sources below.

AUs extracted from early-stage of real MaEs. Abundant MaE videos exist in the community, which have a similar set of emotion labels with MiE datasets. These MaE videos usually start from a neutral expression, leak subtle muscle movements in early frames, and present obvious expressions later. In our preliminary experiments, we observe that AUs extracted from early frames of MaE videos have similar values as those of MiE clips. This suggests that MiEs and *early-stage* of MaEs have similar intensities in muscle movements, rendering the latter a potential source to simulate MiEs.

In leveraging MaE videos as an AU source, we regard the first frame of MaE clips, which usually has a neutral expression, as our onset frame. The selection of the apex frame is more challenging. However, we empirically observe that existing MaE clips usually present MiE-liked AU intensities in the first half

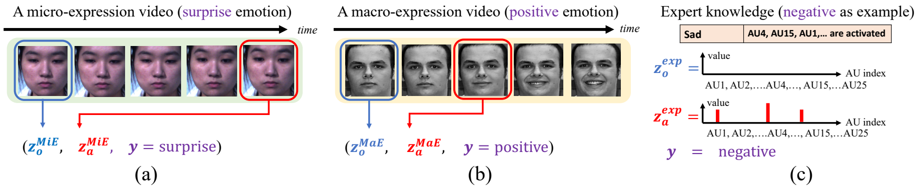

A micro-expression video (surprise emotion)
time
$(z_o^{MiE}$, z_a^{MiE}, y = surprise)
(a)
A macro-expression video (positive emotion)
time
$(z_o^{MaE}$, z_a^{MaE}, y = positive)
(b)
Expert knowledge (negative as example)
Sad
AU4, AU15, AU1,... are activated
z_o^{exp} =
value
AU index
AU1, AU2,....AU4,..., AU15,...AU25
z_a^{exp} =
value
AU index
AU1, AU2,....AU4,..., AU15,...AU25
y = negative
(c)

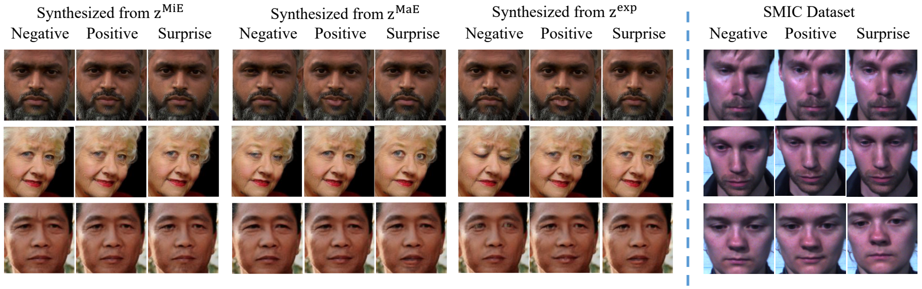
Fig. 3. Examples of how to compute $\mathbf{z}^{\mathrm{MiE}}$, $\mathbf{z}^{\mathrm{MaE}}$ and $\mathbf{z}^{\mathrm{exp}}$. **(a)** We compute $\mathbf{z}^{\mathrm{MiE}}$ from representative frames (*i.e.*, the onset frame and the apex frame) of real-world MiE videos. **(b)** Early frames in real-world macro-expression videos are used to obtain $\mathbf{z}^{\mathrm{MaE}}$. The hyperparameters of choosing the frame indices are selected in Section 5.3. **(c)** We specify an emotion type (*e.g.*, sad) and then the AU distribution from the Expert Mapping table [7], which determine the activated AU entries. Then we assign activated AU entries with intensity values (red bars) and others with 0. The hyperparameters of constraining intensity values are experimented in Section 5.3.

of the video. Therefore, we use two hyperparameters to find the apex frame approximately. Suppose an MaE clip has n frames. An apex frame is randomly drawn from frame index $\lfloor \alpha \times n \rfloor$, where $\lfloor \cdot \rfloor$ rounds a number down to the nearest integer. The selections of α and β are briefly discussed in Section 5.3.

Discussion. Different MaE datasets may be different in the frame index of the onset and apex frames, so in practice we need to do a rapid scanning to roughly know them. But this process is usually quick, and importantly reliable, because 1) a certain dataset usually follows a stable pattern in terms of the onset and apex positions and 2) onset and apex states usually last for a while. As such, while this procedure requires a bit manual work, it is still very valuable considering the gain it brings (large-scale MiE data).

AUs defined by expert knowledge. Studies reveal strong relationships between AUs and emotions [7,9,28]. Some explicitly summarize the posterior probability of each AU entry being activated for each emotion label: $P(z_i > 0|y)$, where z_i indicates the i-th entry of AU vector $\mathbf{z}$. The posterior probabilities, for simplicity, are usually modeled with a Bernoulli distribution [7], *i.e.*, $P(z_i > 0|y) = p$ and $P(z_i = 0|y) = 1 - p$. We find the AU distribution summarized by experts another effective source of AUs for synthesizing trainable MiEs.

We use the expert knowledge mainly to find the apex AUs, where we resort to a mapping table [7] that describes the aforementioned posterior probabilities. Given an emotion label, when generating the apex AUs $\mathbf{z}_a^{\mathrm{exp}}$, we first decide which entries in $\mathbf{z}_a^{\mathrm{exp}}$ should be activated (> 0) by drawing samples from the Bernoulli distribution. We then determine the intensities of the activated entries by randomly sampling from a uniform distribution with a fixed interval $[\mu, \nu]$. The selection of hyperparameters μ, ν is briefly discussed in Section 5.3. On the other hand, for the onset AUs $\mathbf{z}_\mathbf{o}^{\mathrm{exp}}$, we set them to zero vectors, which means


Synthesized from z^{MiE}
Negative Positive Surprise
Synthesized from z^{MaE}
Negative Positive Surprise
Synthesized from z^{exp}
Negative Positive Surprise
SMIC Dataset
Negative Positive Surprise

(a) Synthetic data (b) Real data

Fig. 4. Examples of MiE apex frames from **(a)** synthetic (MiE-X) and **(b)** real-world (the SMIC dataset [18]) micro-expression data. In (a), we show three columns of synthesized MiE apex frames corresponding to three types of Action Units (AUs), *i.e.*, $\mathbf{z}^{\mathrm{MiE}}$,$\mathbf{z}^{\mathrm{MaE}}$, $\mathbf{z}^{\mathrm{exp}}$ described in Section 4.2. Both real-world data and synthetic data the shown under classes labels `positive`, `negative`, and `surprise`.

that no action unit is activated, thus representing a neutral face. Examples of how to compute the above three types of AUs are provided in Fig. 3.

Discussion. We use three basic expression categories (`positive`, `negative`, `surprise`) when synthesizing MiE-X, because these three classes form the largest common intersection between the label sets from the three sources. If we could have more fine-grained label space, it would be interesting to further explore how the label space affects the training quality of MiE-X.

4.3 The MiE-X dataset

With the above three types of AUs and a large pool of in-the-wild faces, we eventually are able to synthesize a large-scale MiE recognition dataset, coined MiE-X. MiE-X contains 5,000 identities, each with 9 MiE samples[4], resulting in 45,000 samples in total. To our knowledge, MiE-X is the first large-scale MiE dataset and is more than two orders of magnitude larger than existing real-world MiE datasets. Visualization of the generated apex frames in MiE-X is provided in Fig. 4; comparisons with existing MiE datasets are illustrated in Fig. 1.

The strength of MiE-X as training data comes from its diversity in identity and MiE patterns[5]. For instance, it contains 5,000 human identities, encouraging models to learn identity-invariant expression features. At the same time, the three sources of AUs are complementary, provide a wide range of AU values, and sometimes have random AU perturbations. MiE-X alleviates overfitting risks and allows algorithms to consistently improve their accuracy.

[4] For each ID and each of the three classes `positive`, `negative`, and `surprise`, we generate three MiE samples corresponding to three types of AUs. Each sample has an onset and an apex frames, totaling 9 MiE samples and 18 frames per ID.

[5] We also acknowledge GANimation that provides us with realistic facial images.

5 Experiment

5.1 Experimental setups

Baseline classifiers. Two image-based MiE recognition methods are mainly evaluated in this paper: the **Branches** [21] and **ApexME** [19]. Both are trained for 80 epochs. More details are provided in supplementary materials.

Real-world datasets. We report experimental results on commonly-used real-world datasets: **CompMiE [33]**, **MMEW** and **SAMM**. CompMiE is proposed by the MiE recognition challenge MEGC2019 [33] which merges three existing real MiE datasets into one. The three component datasets are CASME II [39], SAMM [5,4], and SMIC [18], respectively. CompMiE has the same label space (Section 4.2) as MiE-X and consists of 442 samples from 68 subjects in total. MMEW and SAMM have 234 and 72 samples, respectively, and their label spaces are different with MiE-X[6]. The MaE dataset **CK+** [22] is a commonly used real-world MaE dataset containing 327 videos. Its label space is also merged into the same one as CompMiE. When generating MiE-X (see Section 4), we extract $\mathbf{z}^{\mathrm{MiE}}$ and $\mathbf{z}^{\mathrm{MaE}}$ from CompMiE and CK+, respectively.

Evaluation protocols. We use subject-wise k-fold cross-validation, commonly performed in the community [3,19,16]. Specifically, when real-world data are used in testing, we split them into k subsets. Each time, we use $k-1$ subsets for training and the rest 1 subset for testing. The average accuracy of the k tests is reported. For CompMiE, $k = 3$; for MMEW and SAMM, $k = 5$. To evaluate the effectiveness of MiE-X, we replace real training sets (*i.e.*, $k-1$ subsets) with MiE-X when MiE-X is used for direct deployment. Note that, for each fold, MiE-X samples whose AUs (*i.e.*, $\mathbf{z}^{\mathrm{MiE}}$) are computed from real MiE samples in the test subset will not be used in training. If MiE-X is used for pre-training, where a fine-tuning stage is required, the $k-1$ subsets will be used for fine-tuning. Other real-world datasets (*e.g.*, MMEW, SMIC) are also used for pre-training to form comparisons with MiE-X[7] Experiment is categorized as follows.

- Pre-training with MiE-X (or other competing datasets) and fine-tuning on target training set. We adopt this setting especially when the source domain has a different label space from the target domain.
- Training (or fine-tuning) with MiE-X (or other competing datasets) followed by direct model deployment. If the target domain and training dataset share the same label space, models obtained from the training set can be directly used for inference on the target test set.

Metrics. We mainly use unweighted F1-score (UF1) and unweighted average recall (UAR) [33]. UF1 and UAR indicate the average F1-score and recall, respectively, over all classes. We also report the conventional recognition rate on the MMEW [3] and SAMM [4] datasets to compare with the state of the art.

[6] Label space of MMEW: `happiness`, `surprise`, `anger`, `disgust`, `fear`, `sadness`; Label space of SAMM: `happiness`, `surprise`, `anger`, `disgust`, `fear`.

[7] We discard those samples in real-world datasets that overlap with the test subset.

Table 1. Effectiveness of MiE-X in model (pre-)training. Models are pre-trained using MiE-X or other real-world datasets and then fine-tuned on real-world training data *i.e.*, CompMiE, or the combination of CompMiE and CK+ [22]. UF1 (%) and UAR (%) are reported on the CompMiE dataset after three-fold cross-validation. ApexME [19] and Branches [21] are used as baselines. We observe consistent accuracy improvement when models are pre-trained with MiE-X. In addition, when directly deploying the MiE-X pretrained model, the accuracy is also competitive.

Pre-training	Fine-tuning		ApexME [19]		Branches [21]	
MiE data	CompMiE	CK+	UF1	UAR	UF1	UAR
-	✓		41.8 ± 0.7	41.9 ± 0.7	43.6 ± 0.5	44.6 ± 0.6
-	✓	✓	45.0 ± 0.5	45.5 ± 1.0	45.2 ± 0.5	47.0 ± 0.6
SMIC [18]	✓		45.0 ± 1.7	44.8 ± 1.9	42.8 ± 0.8	41.4 ± 0.9
CASME [39]	✓		44.0 ± 1.2	45.1 ± 0.5	40.7 ± 0.9	41.4 ± 0.9
SAMM [5]	✓		43.7 ± 0.7	42.8 ± 0.5	42.3 ± 1.4	42.9 ± 1.7
MMEW [3]	✓		43.3 ± 0.8	44.4 ± 1.2	43.3 ± 1.3	44.1 ± 1.5
MiE-X			45.2 ± 0.5	46.3 ± 0.5	47.7 ± 0.5	48.9 ± 0.8
MiE-X	✓		46.9 ± 0.9	**48.3** ± 0.9	50.7 ± 0.9	52.1 ± 1.4
MiE-X	✓	✓	**47.0** ± 0.8	48.2 ± 0.4	**52.3** ± 0.7	**52.3** ± 0.4

By default, we run each experiment (k-fold cross-validation) 3 times and report the mean and standard variance of the results in the last epoch. Moreover, we provide the best accuracy among all epochs for reference (Table 2).

5.2 Effectiveness of the Synthetic Database

Effectiveness of MiE-X in training models for direct deployment. MiE-X has the same label space with CompMiE. So models trained with MiE-X can be directly evaluated on the CompMiE. In Table 1, ApexME and Branches trained with MiE-X alone produce an UF1 of 45.2% and 47.7%, respectively, which outperforms the training set composed of CompMiE and CK+.

Effectiveness of MiE-X in model pre-training. First, when using MiE-X for model pre-training, we observe consistent improvement over not using it (Table 1). For example, when we perform fine-tuning on CompMiE using the ApexME method, pre-training with MiE-X brings 5.1% and 7.1% improvement in UF1 and UAR, respectively, over not using MIE-X. Second, we compare MiE-X with existing datasets (*i.e.*, SMIC, CASME, SAMM, and MMEW) of their effectiveness as a pre-training set, on which we train the baseline MiE classifiers (*i.e.*, ApexME, Branches). We do three-fold cross-validation on CompMiE. For each fold, we use the dataset (e.g., SMIC) we would like to evaluate as the pre-training data. Samples are removed from the training set if they also appear in the test subset of CompMiE in the current fold. Then we fine-tune the model on the training subset of CompMiE. Results are shown in both Table 1 and Fig. 1. We observe that the model pre-trained on MiE-X significantly outperforms those pre-trained on other datasets. For instance, when we pre-train Branches on MiE-X, the final fine-tuning results on CompMiE in UF1 and UAR are 7.4% and 8.0%

Table 2. Comparison with the state-of-the-art MiE recognition methods on MMEW and SAMM datasets. We re-implement ApexME, Branches and DTSCNN, which are pretrained with either ImageNet or MiE-X (grey). We report the mean recognition accuracy (%) and standard variance. † donates vide-based methods. "Last" means test result in the last epoch, and "Best" refers to the best accuracy among all epochs.

Methods	MMEW		SAMM	
	Last	Best	Last	Best
FDM [38]	34.6	–	34.1	–
LBP-TOP [42]	38.9	–	37.0	–
DCP-TOP [2]	42.5	–	36.8	–
ApexME [19]	48.5 ± 0.6	58.3 ± 0.9	41.3 ± 0.6	54.9 ± 0.7
ApexME + **MiE-X**	55.9 ± 2.0	61.4 ± 0.8	46.4 ± 0.7	60.3 ± 1.1
Branches [21]	50.1 ± 0.6	58.3 ± 0.6	44.5 ± 0.7	53.3 ± 0.5
Branches + **MiE-X**	56.8 ± 1.1	61.5 ± 1.0	48.7 ± 1.0	56.3 ± 0.8
TLCNN† [36]	–	69.4	–	73.5
DTSCNN† [25]	60.9 ± 1.3	71.1 ± 1.1	51.6 ± 1.8	60.6 ± 1.1
DTSCNN† + **MiE-X**	63.1 ± 1.0	74.3 ± 0.5	55.5 ± 1.4	73.9 ± 0.9

higher than using MMEW as the pre-training data. This phenomenon validates the effectiveness of our dataset and the proposed synthesis procedure.

Positioning within the state of the art. We follow a recent survey [3] and compare with the state of the art on two datasets, MMEW [3] and SAMM [4], all under 5-fold cross validation. Results are summarized in Table 2. We re-implemented three baselines (ApexME, Branches and DTSCNN), pretrained on either ImageNet or MiE-X. To pretrain the video-base method DTSCNN, we use a simple variant of MiE-X where each sample has multiple frames. Specifically, when computing $\mathbf{z}^{\mathrm{MiE}}$ and $\mathbf{z}^{\mathrm{MaE}}$, we extract AUs for all the frames between the onset and apex frames. All these extracted AUs are used for frame generation. For $\mathbf{z}^{\mathrm{exp}}$, we linearly interpolate 8 AU vectors between the onset and apex AU vectors, thus generating 10 frames per sample.

Table 2 clearly informs us that MiE-X pre-training improves the accuracy of all the three methods. Importantly, when MiE-X is used for pre-training, MiE recognition accuracy is very competitive: DTSCNN achieves accuracy (best epoch) of 74.3 ± 0.5 % and 73.9 ± 0.9 % on MMEW and SAMM, respectively.

5.3 Further Analysis

All experiments in this section are performed on the Branches baseline [21].

Comparisons of various AU combinations. Fig. 5 evaluates various AU combinations on CompMiE. We have the following observations. **First**, none of the three types of AUs are dispensable. We observe that the best recognition accuracy is obtained when all three types of AUs are used, which outperforms training with CompMiE+CK+ by 1.7% and 2.0% in UF1 and UAR, respectively. Importantly, if we remove any single type of AUs, the UF1 and UAR scores de-

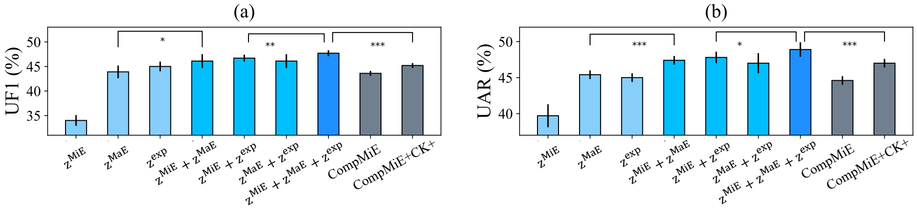


Fig. 5. Comparing training effectiveness of real-world data and various synthetic datasets sourced from different combinations of AUs. We compare UF1 **(a)** and UAR **(b)** on CompMiE. "n.s." means the difference is not statistically significant (*i.e.*, p-value > 0.05). $*$ denotes statistically significant (*i.e.*, $0.01 < p$-value < 0.05). $**$ and $***$ mean statistically very significant (*i.e.*, $0.001 < p$-value < 0.01) and statistically extremely significant (*i.e.*, p-value < 0.001), respectively. We observe decreased accuracy if we remove any of the three types of AUs. When all the three types are used for database creation, both UF1 and UAR exceed results obtained by training on real-world data, with very high statistical confidence.

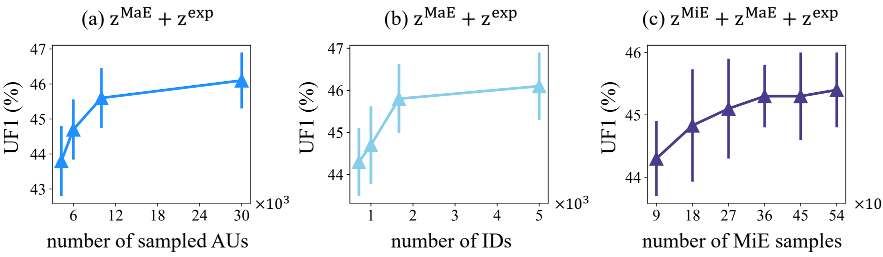
crease. For example, when removing $\mathbf{z}^{\mathrm{MiE}}$, $\mathbf{z}^{\mathrm{MaE}}$, $\mathbf{z}^{\exp}$ one at a time, the decrease in UF1 score is 1.6%, 1.0% and 1.6%, respectively.

Second, using two types of AUs outperforms using only a single type with statistical significance. For example, when using $\mathbf{z}^{\mathrm{MiE}}$ and $\mathbf{z}^{\mathrm{MaE}}$, UF1 is higher than using $\mathbf{z}^{\mathrm{MaE}}$ alone by 2.15%. In fact, the three AU types come from distinct and trustful sources, allowing them to be complementary and effective. This also explains why all three AU types are better than any combination of two.

Third, when using a single type of AUs, we find $\mathbf{z}^{\mathrm{MaE}}$ or $\mathbf{z}^{\exp}$ produces much higher UF1 and UAR than $\mathbf{z}^{\mathrm{MiE}}$. Their superiority could be explained by their diversity. Compared with $\mathbf{z}^{\mathrm{MiE}}$, MiEs generated from $\mathbf{z}^{\mathrm{MaE}}$ and $\mathbf{z}^{\exp}$ are much more diverse. Specifically, when constructing $\mathbf{z}^{\mathrm{MaE}}$, the index of apex frame is randomly drawn from a range $\lfloor \alpha \times n \rfloor$ and $\lfloor \beta \times n \rfloor$. Similarly, the randomness of AU intensities is also introduce by hyperparameter μ and ν when generating $\mathbf{z}^{\exp}$. In contrast, the index of the apex frame is fixed when constructing $\mathbf{z}^{\mathrm{MiE}}$.

Lastly, we compare results that employ two real-world training datasets. The first is CompMiE, described as in Section 5.1, and the second is a combination of CompMiE and CK+. It is shown that CompMiE + CK+ outperforms CompMiE by an obvious margin, suggesting that *early-stage of MaEs highly correlate with MiEs.* These results motivated us to mine effective AUs ($\mathbf{z}^{\mathrm{MaE}}$) from MaEs.

Impact of the number of AUs, IDs and MiE samples in MiE-X. For MiE-X, the IDs, AUs and MiE samples are all important, and we now investigate how their quantities influence MiE recognition accuracy by creating MiE-X variants with different numbers of IDs, AU triplets and samples. Here, please note that the diversity is highly relevant to the number of distinct IDs/AUs/samples, so sometimes we use number and diversity interchangeably. When studying AU and ID diversity, we set the AU combination to be $\mathbf{z}^{\mathrm{MaE}} + \mathbf{z}^{\exp}$ because their


(a) $\mathrm{z}^{\mathrm{MaE}} + \mathrm{z}^{\mathrm{exp}}$

(b) $\mathrm{z}^{\mathrm{MaE}} + \mathrm{z}^{\mathrm{exp}}$

(c) $\mathrm{z}^{\mathrm{MiE}} + \mathrm{z}^{\mathrm{MaE}} + \mathrm{z}^{\mathrm{exp}}$

Fig. 6. (a)-(b): Impact of the number of AU triplets **(a)**, IDs **(b)** and MiE samples **(c)**. In **(a)-(b)**, we use $\mathbf{z}^{\mathrm{MaE}}$ and $\mathbf{z}^{\mathrm{exp}}$ for database synthesis, while in **(c)** all three types AUs are used. We employ the Branches method [21]. When we gradually increase the numbers, the three-fold cross validation accuracy (UF1, %) on CompMiE first improves and then remains stable in all the three subfigures.

diversity can be easily changed by specifying the number of sampling times from the uniform distributions (refer Section 4.2). When investigating the number of MiE samples, we use all three types of AUs.

To evaluate the influence of **AU** diversity, we set the number of MiE samples and IDs to 30,000 and 5,000 (6 samples per ID), respectively in all the dataset variations. The AU diversity can be customized by allowing multiple identities to share the same AU triple. Specifically, the number of AU triplets is set to 4,000, 6,000, 10,000, 30,000 and From the experimental results in Fig. 6 (a), we observe the effectiveness of synthetic data generally increases when AU diversity is improved. For example, the UF1 score increases by 1.8%, when the number of distinct AU triplets increases from 4,000 to 10,000. When the number of AUs is greater than 10,000, the curve reaches saturation.

To study the diversity of **IDs**, we fix the number of MiE samples and AU triplets in MiE-X to be 30,000. We set the ID number as 700, 1,000, 1,700, and 5,000, achieved by randomly selecting face images from the EmotionNet [10] dataset[8]. In this experiment, an ID generates more than 6 MiE samples using AU triplets randomly drawn from the pool of 30,000. Results in Fig. 6 (b) show that more IDs leads to a higher recognition accuracy. For example, UF1 of synthetic dataset increases from 44.4% to 45.8% when the number of IDs increases from 700 to 1,700. When the number of IDs exceeds 1,700, the curve becomes stable.

To study the impact of the number of **MiE samples**, we fix the number of AU triplets to 9,000 and the number of IDs to 1,000. We then gradually increase the generated samples from 9,000 to 54,000 by reusing more AU triplets on each ID. Experimental results are shown in Fig. 6 (c). We find the effectiveness of the synthetic training set generally increases when more samples are included and that curve becomes flat when the number of samples are greater than 36k. For

[8] Note that each image in EmotionNet usually denotes a different identity.

example, the UF1 is improved by 1.0%, when the number of samples increases from 9k to 36k. When the number of samples increases from 36k to 54k, there is a slight UF1 improvement of 0.2%. This observation is expected because when the number of IDs and AUs are fixed, the total information contained in the dataset is constrained. From the above experiments, we conclude that MiE-X benefits from more AUs, IDs and samples within a certain range.

Table 3. Performance comparison between training with and without side faces. Evaluation is on the CompMiE dataset.

	w/ side	*w/o* side
UF1 (%)	47.7 ± 0.5	47.4 ± 0.8

Impact of face poses. We use 5,000 IDs with frontal faces to synthesize a training set variant which is compared with MiE-X composed of faces of various poses. To find the frontal faces, we manually select 10 frontal faces in the EmotionNet dataset as queries and for each search for 500 faces with similar facial landmarks detected by a pretrained MTCNN landmark detector [41]. Table 3 summarizes the results on CompMiE, where we do not observe obvious difference between the two training sets. This can possibly be explained by the fact that real-world MiE datasets mostly contain frontal faces collected in laboratory environments. Therefore, pose variance in MiE-X may not significantly influence performance on existingtests. Nevertheless, we speculate using various poses to generate MiE-X would benefit MiE recognition in uncontrolled environments.

Analysis of other hyperparameters. Due to the lack of validation data in real-world MiE datasets, we mostly used prior knowledge and intuition to choose the hyperparameters. Specifically, we chose $\alpha = 0.3$, $\beta = 0.5$ and $\mu = 0.1$, $\nu = 0.3$ in experiments. Here, we briefly analyze these two sets of hyperparameters involved in the AU computation on CompMiE using cross-validation. $[\alpha, \beta]$ is the interval from which the apex frames for computing $\mathbf{z}^{\mathrm{MaE}}$ are randomly selected. Specifically, we analyze three options: ($\alpha = 0.1$, $\beta = 0.3$), ($\alpha = 0.3$, $\beta = 0.5$) and ($\alpha = 0.5$, $\beta = 0.7$). The number of identities is 5,000. Recognition accuracy of the three options is given by Fig. 7 (a), where $\alpha = 0.3$, $\beta = 0.5$ produces the highest UF1 score. This result is in accordance with our intuition: the first 30% to 50% frames of an MaE would be more similar to an MiE.

$[\mu,\ \nu]$ is the interval from which the intensities of expert-defined AUs are uniformly sampled. Similarly, we analyze three options, *i.e.*, ($\mu = 0.1$, $\nu = 0.3$), ($\mu = 0.3$, $\nu = 0.5$) and ($\mu = 0.5$, $\nu = 0.7$). This is inspired by observing AU coefficients of real MiEs: the intensity of each action unit is not large, *i.e.*, < 0.7 in most cases, because micro-expressions have subtle facial muscle movements. Results are shown in Fig. 7 (b): the intensity range $[0.1, 0.3]$ is superior. Because the highest value of an MaE AU is 1.0, the value of $[\mu,\nu]$ delivers another intuitive message: facial AU intensities of MiEs are around 10% to 30% those of MaEs.

5.4 Understanding of MiEs: A Discussion

MiEs generalize across faces. AUs extracted from real MiEs provide closest resemblance to true MiEs and are thus indispensable. These AUs $\mathbf{z}^{\mathrm{MiE}}$ are generalizable because they can be transplanted to faces of different identities. The fact that a higher number of face identities generally leads to a higher accuracy indicates the benefit of adding AUs $\mathbf{z}^{\mathrm{MiE}}$ to sufficiently many faces to improve MiE recognition towards identity invariance.

Early-stage MaEs resemble real MiEs. To our knowledge, we make very early attempt to leverage MaEs for MiE generation. Although the two types of facial expressions differ significantly in their magnitude of facial movement, we find AUs in initial stages of MaEs are effective approximations to those in MiEs.

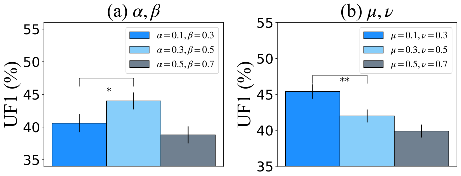


Fig. 7. Impact of hyperparameters in computing $\mathbf{z}^{\mathrm{MaE}}$ and $\mathbf{z}^{\mathrm{exp}}$. UF1 (%) on the CompMiE dataset is reported in each subfigure. **(a):** MiE-X is composed by $\mathbf{z}^{\mathrm{MaE}}$ only. Three groups of α and β values are tested. **(b):** MiE-X is made from $\mathbf{z}^{\mathrm{exp}}$ only. Three groups of μ and ν are investigated. $*$ and $**$ have the same meaning as Fig. 5.

Expert knowledge is transferable to MiEs. While AUs annotated by experts are used to describe MaEs, we find expert AUs with reduced magnitude are effective in synthesizing MiEs. We therefore infer from a computer vision viewpoint that MiEs are related to normal expressions but with lower intensity. Moreover, by examining the complementary nature of the three types of AUs, we infer that expert knowledge adds some useful computational cues, which do not appear in MaEs and real MiEs but can be humanly defined. Nevertheless, our work is limited in that the psychological aspects of MiEs are not considered, which will be studied in future with cross-disciplinary collaborations.

6 Conclusion

This paper addresses the data lacking problem in MiE recognition. An important contribution is the introduction of a large-scale synthetic dataset, MiE-X, with standard emotion labels to improve MiE model training. In the synthesis protocol, we feed faces in the wild, desired emotion labels and AU triplets (our focus) to a generation model. Specifically, sourced from real MiEs, early-stage MaEs, and expert knowledge, three types of AUs are identified as useful and complementary to endorse an effective protocol. This understanding of the role of AUs in effective MiE synthesis is another contribution of this work. Experiment on real-world MiE datasets indicates MiE-X is a very useful training set: models (pre-)trained with MiE-X consistently outperform those (pre-)trained on real-world MiE data. In addition, this paper reveals some interesting computational properties of MiEs, which would be of value for further investigation.

References

1. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P.: Openface 2.0: Facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 59–66. IEEE (2018)
2. Ben, X., Jia, X., Yan, R., Zhang, X., Meng, W.: Learning effective binary descriptors for micro-expression recognition transferred by macro-information. Pattern Recognition Letters **107**, 50–58 (2018)
3. 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)
4. Davison, A., Merghani, W., Yap, M.: Objective classes for micro-facial expression recognition. Journal of Imaging **4**(10), 119 (2018)
5. Davison, A.K., Lansley, C., Costen, N., Tan, K., Yap, M.H.: Samm: A spontaneous micro-facial movement dataset. IEEE Transactions on Affective Computing **9**(1), 116–129 (2016)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. Du, S., Tao, Y., Martinez, A.M.: Compound facial expressions of emotion. Proceedings of the National Academy of Sciences **111**(15), E1454–E1462 (2014)
8. Eckman, P., Friesen, W.: Facial action coding system (facs): A technique for the measurement of facial action. A8@ 5 **3**, 56–75 (1978)
9. Ekman, P., Rosenberg, E.L.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
10. 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)
11. Hao, X.l., Tian, M.: Deep belief network based on double weber local descriptor in micro-expression recognition. In: Advanced Multimedia and Ubiquitous Engineering, pp. 419–425. Springer (2017)
12. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A.A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. arXiv preprint arXiv:1711.03213 (2017)
13. Huang, X., Wang, S.J., Zhao, G., Piteikainen, M.: Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 1–9 (2015)
14. Huang, X., Zhao, G., Hong, X., Zheng, W., Pietikäinen, M.: Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. Neurocomputing **175**, 564–578 (2016)
15. Kar, A., Prakash, A., Liu, M.Y., Cameracci, E., Yuan, J., Rusiniak, M., Acuna, D., Torralba, A., Fidler, S.: Meta-sim: Learning to generate synthetic datasets. arXiv preprint arXiv:1904.11621 (2019)
16. Khor, H.Q., See, J., Phan, R.C.W., Lin, W.: Enriched long-term recurrent convolutional network for facial micro-expression recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 667–674. IEEE (2018)

17. Kim, D.H., Baddar, W.J., Ro, Y.M.: Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In: Proceedings of the 24th ACM international conference on Multimedia. pp. 382–386. ACM (2016)
18. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikäinen, M.: A spontaneous micro-expression database: Inducement, collection and baseline. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). pp. 1–6. IEEE (2013)
19. Li, Y., Huang, X., Zhao, G.: Can micro-expression be recognized based on single apex frame? In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 3094–3098. IEEE (2018)
20. Liong, S.T., Gan, Y., Yau, W.C., Huang, Y.C., Ken, T.L.: Off-apexnet on micro-expression recognition system. arXiv preprint arXiv:1805.08699 (2018)
21. 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)
22. 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)
23. Matsumoto, D., Yoo, S.H., Nakagawa, S.: Culture, emotion regulation, and adjustment. Journal of personality and social psychology **94**(6), 925 (2008)
24. Patel, D., Hong, X., Zhao, G.: Selective deep features for micro-expression recognition. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 2258–2263. IEEE (2016)
25. Peng, M., Wang, C., Chen, T., Liu, G., Fu, X.: Dual temporal scale convolutional neural network for micro-expression recognition. Frontiers in psychology **8**, 1745 (2017)
26. 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)
27. Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor (2009)
28. Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expression detection in hi-speed video based on facial action coding system (facs). IEICE transactions on information and systems **96**(1), 81–92 (2013)
29. 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)
30. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: European Conference on Computer Vision. pp. 102–118. Springer (2016)
31. Ruiz, N., Schulter, S., Chandraker, M.: Learning to simulate. arXiv preprint arXiv:1810.02513 (2018)
32. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision **126**(9), 973–992 (2018)
33. 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)

34. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: Face and Gesture 2011. pp. 51–56. IEEE (2011)
35. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S.: Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 969–977 (2018)
36. Wang, S.J., Li, B.J., Liu, Y.J., Yan, W.J., Ou, X., Huang, X., Xu, F., Fu, X.: Micro-expression recognition with small sample size by transferring long-term convolutional neural network. Neurocomputing **312**, 251–262 (2018)
37. Wang, Y., See, J., Phan, R.C.W., Oh, Y.H.: Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In: Asian conference on computer vision. pp. 525–537. Springer (2014)
38. Xu, F., Zhang, J., Wang, J.Z.: Microexpression identification and categorization using a facial dynamics map. IEEE Transactions on Affective Computing **8**(2), 254–267 (2017)
39. Yan, W.J., Li, X., Wang, S.J., Zhao, G., Liu, Y.J., Chen, Y.H., Fu, X.: Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. PloS one **9**(1), e86041 (2014)
40. Yao, Y., Zheng, L., Yang, X., Naphade, M., Gedeon, T.: Simulating content consistent vehicle datasets with attribute descent. In: ECCV (2020)
41. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (Oct 2016). https://doi.org/10.1109/LSP.2016.2603342
42. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE transactions on pattern analysis and machine intelligence **29**(6), 915–928 (2007)
43. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3754–3762 (2017)
44. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5157–5166 (2018)