applied sciences

MDPI

Article

AU-Guided Unsupervised Domain-Adaptive Facial Expression Recognition

Xiaojiang Peng [1,*], Yuxin Gu [2] and Panpan Zhang [3]

[1] College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China
[2] Alibaba Group, Hangzhou 310011, China; jixin.gyx@alibaba-inc.com
[3] National University of Singapore, Singapore 119077, Singapore; panpan1821768@gmail.com
* Correspondence: pengxiaojiang@sztu.edu.cn

Abstract: Domain diversities, including inconsistent annotation and varied image collection conditions, inevitably exist among different facial expression recognition (FER) datasets, posing an evident challenge for adapting FER models trained on one dataset to another one. Recent works mainly focus on domain-invariant deep feature learning with adversarial learning mechanisms, ignoring the sibling facial action unit (AU) detection task, which has obtained great progress. Considering that AUs objectively determine facial expressions, this paper proposes an AU-guided unsupervised domain-adaptive FER (AdaFER) framework to relieve the annotation bias between different FER datasets. In AdaFER, we first leverage an advanced model for AU detection on both a source and a target domain. Then, we compare the AU results to perform AU-guided annotating, i.e., target faces that own the same AUs as source faces would inherit the labels from the source domain. Meanwhile, to achieve domain-invariant compact features, we utilize an AU-guided triplet training, which randomly collects anchor–positive–negative triplets on both domains with AUs. We conduct extensive experiments on several popular benchmarks and show that AdaFER achieves state-of-the-art results on all these benchmarks.

Keywords: facial expression recognition (FER); action units; unsupervised cross-domain FER

Citation: Peng, X.; Gu, Y.; Zhang, P. AU-Guided Unsupervised-Domain Adaptive Facial Expression Recognition. *Appl. Sci.* **2022**, *12*, 4366. https://doi.org/10.3390/app12094366

Academic Editors: Tao Lei, Xianye Ben, Peng Zhang and Lei Chen

Received: 16 March 2022
Accepted: 21 April 2022
Published: 26 April 2022

1. Introduction

Facial expression is one of the most important modalities in human emotional communication. Accurately recognizing facial expressions helps individuals understand various human emotions and intents, which is applied in a wide range of applications, such as human–computer interaction [1], service robots [2], and medicinal treatments [3]. Both in industrial and academic areas, in past decades, many well-labeled datasets [4–12] and high-performance algorithms [13–15] have been proposed to automatically recognize facial expressions.

In general, deep learning-based facial expression recognition (FER) methods [9–11,13–15] achieve high performance only when the training domain is identical or similar to the testing domain. However, due to the diversities of inconsistent annotation and different image collection conditions, there inevitably exists annotation biases (domain gaps) among different datasests, which poses an evident challenge for adapting FER models trained on one dataset to another one. As shown in Figure 1, a naive cross-domain method that deploys a trained model on a different target domain often fails. To this end, many methods have been presented methods for mitigating the domain shift in FER [16–18], though almost all the methods follow general domain adaption algorithms, ignoring the sibling facial action unit (AU) detection task. Action units (AUs) represent the movement of facial muscles, which have lower bias than subjective facial expression annotations. Intuitively, AUs can be regarded as auxiliary cues to alleviate the annotation and data biases (domain gaps) among different FER datasets.

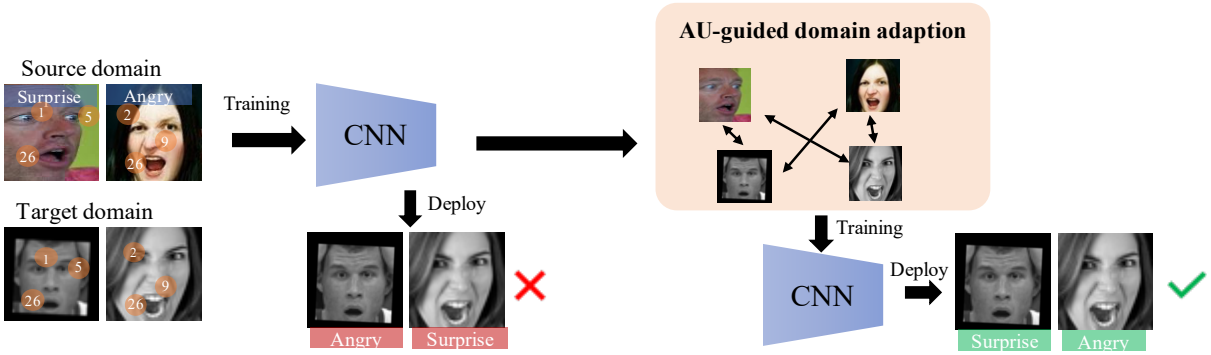


Figure 1. Illustration of unsupervised domain-adaptive facial expression recognition. A model trained on a source domain usually fails on a target domain due to subjective inconsistent annotations and different imaging conditions. AdaFER leverages the objective facial action units for auxiliary training, effectively relieving the annotation bias between source and target domains.

In this paper, considering the remarkable progress and stable performance of AU detection [19,20], we propose an AU-guided unsupervised domain-adaptive FER (AdaFER). The AdaFER consists of two crucial modules: AU-guided annotating (AGA) and AU-guided triplet training (AGT). Given two groups of images from a source domain and a target domain, we first utilize an advanced pretrained AU detection model to extract the AUs' coding from both domains. Then, we perform AU-guided annotating as follows: for each image on the target domain, we use its AU's coding to query the source domain. All the images on a source domain with the same AU coding as the query image will be used for annotating. By default, the query image is assigned with a soft label, which is the statistic label distribution of the retrieval images. Further, to achieve structure-reliable and compact facial expression features, we perform the AU-guided triplet training. Each triplet is generated by comparing the AU coding of source and target images. For example, when a source image is used as an anchor, we randomly sample a positive (negative) sample that has same (different) AUs as the anchor from all target images. With AdaFER, we are able to fine-tune a source-pretrained model on a target domain with both pseudo-soft labels and triplet loss, which effectively prevents FER performance degradation on the target domain.

Overall, our contributions can be summarized as follows:

- We heuristically utilize the relationship between action units and facial expressions for cross-domain facial expression recognition, and propose an AU-guided unsupervised domain-adaptive FER (AdaFER) framework;
- We elaborately design an AU-guided annotation module to assign soft labels for a target domain and an AU-guided triplet training module to learn structure-reliable and compact facial expression features;
- We conduct extensive experiments on several popular benchmarks and significantly outperform the state-of-the-art methods.

2. Related Work

In general, facial expression recognition (FER) can be conducted for static images or dynamic videos. Facial expressions may refer to micro expressions [12] and basic macro expressions. In this paper, we focus on the static image-based FER for basic macro expressions.

2.1. Facial Expression Recognition

Recently, facial expression recognition has achieved significant progress due to well-designed feature extraction methods and high-performance algorithms. They first detect and align faces using several popular face detectors, such as MTCNN [21] and Dlib [22]. For feature extraction, a large number of methods focus on modeling the facial geometry and appearance features to help facial expression recognition. From the feature-type view, these features can be generally divided into hand-crafted features and deep-learning

features. A hand-crafted feature usually contains texture-based features and geometry-based features [23–25]. Sometimes, they are used in a combination called hybrid features. For deep-learning features, Tang [26] and Kahou et al. [27] utilized deep CNNs for feature extraction, and won the FER2013 and Emotiw2013 challenges, respectively. Zhou et al. [28] achieved a remarkable result in the Emotiw2019 multi-modal emotion recognition challenge by using an audio-video deep fusion method. To address the pose variant and occlusion in FER, Wang et al. [13] and Li et al. [14] designed a region-based attention network. Wang et al. [15] proposed a self-cure network to suppress uncertainty samples in FER datasets. Liu et al. [29] introduced a facial action units (FAUs)-based network for expression recognition. Daniel et al. [30] presented a cross-platform real-time multi-face expression recognition toolkit using FAUs, which shows the robustness of FAUs in FER systems.

2.2. Action Units Detection

A facial action coding system (FACS) [31] uses action units (AUs) to represent facial muscle movements. Facial action units detection has been widely used in the face perception and emotion analysis research areas, including deception detection [32], diagnosing mental health [33], and improving e-learning experiences [34]. The impressive progress of AUs detection are due mainly to large-scale datasets and well-designed methods. Large-scale datasets include two categories: one is collected from a controlled or a laboratory environment, the other is collected from the Internet or is collected "in the wild". The BP4D [35], DISFA [36], and GFT [37] datasets are collected from controlled conditions. BP4D contains 41 subjects from 18 to 29 years old. They require subjects to join eight different tasks (one-on-one interviews (inducing pleasure), watching movie clips (inducing sadness), suddenly hearing a voice (inducing surprise), improvising a funny song (inducing embarrassment), feeling threatened (inducing fear), putting one's hands into ice water, experiencing an insult from the experimenter (inducing anger), and smelling some peculiar smell (inducing disgust)) and record the changes of the subjects. The BP4D dataset simultaneously records 2D and 3D facial expression videos, in which 2D videos contain more than 160,000 face images. The DISFA [36] dataset collects 27 videos when subjects watch movies. Each video consists of 4845 frames. All videos are annotated with AUs encoding, as well as with the five levels of intensity. GFT [37], as the first dataset to include the conversations of multiple people, records 240 videos of 3 people communicating. They annotate 20 AUs with a 0–1 code and divide the GFT into training and validation sets. EmotioNet [10] downloads 1,000,000 images from the Internet and manually annotates 50,000 images with 11 AUs for training and testing. In addition, the 900,000 images of EmotioNet are automatically annotated by the pretrained model using manual annotation for 50,000 images.

The methods of AUs detection can be summarized as follows. Supervised methods: Walecki et al. [38] propose a Copula CNN deep learning approach by combining conditional random field (CRF)-encoded AU dependencies with deep learning. Ji [19] propose a dynamic threshold for each AU to improve the performance of AU detection. Semi-supervised methods: Peng [39] utilize the relationships between AUs encoding and expression categories to automatically annotate the AU labels for facial images with only expression labels. Wu et al. [40] use the restricted Boltzmann machine to model the AU distribution, which is further used to train the AU classifiers with partially labeled data. Niu et al. [20] propose two networks to generate multi-view features for both labeled and unlabeled face images, and utilize multi-label co-regularization loss to minimize the the distance between the predicted AU probability distributions of the two views.

2.3. Cross-Domain FER

It is inevitable that distribution divergences among different facial expression datasets will exist, due to variant collecting conditions and annotating subjectiveness. In past decades, cross-domain FER (CD-FER) has recieved more attention [16,18,41–48]. Generally, the CD-FER can be divided into semi-supervised-based, unsupervised-based, and

generation-based methods. Semi-supervised-based methods [45,49] apply a convolutional neural network (CNN) model to train a classification model using limited labeled samples from a target domain. For the unsupervised-based methods, Valstar et al. [50] use a Gabor feature-based landmark detector to localize facial points and track them in facial sequences to model temporal facial activation for facial expression recognition. They trained the recognition model using the CK database and performed the test in the MMI database for cross-validation. Zheng et al. [48] propose a transductive transfer subspace learning method using labeled source domain images and unlabelled auxiliary target domain images to jointly learn a discriminative subspace. For generation-based methods, Zong et al. [18] propose a domain regeneration framework (DR) that aims at learning a domain regenerator to regenerate samples from source and target databases, respectively. Wang et al. [51] introduce an unsupervised domain-adaptation method using a generative adversarial network (GAN) on the target dataset, and dynamically assign the unlabelled GAN-generated samples distributed pseudo-labels according to the current prediction probabilities. In order to understand the conditional probability distributions' differences between the source and target datasets, Li et al. [42] develop a deep emotion-conditional adaption network that simultaneously considers the conditional distribution bias and expression class imbalance problem in CD-FER. Chen et al. [17] propose an adversarial graph representation adaptation (AGRA) framework that unifies graph representation propagation with adversarial learning for cross-domain holistic–local feature co-adaptation. Different from the above works, our work utilizes AU information as auxiliary cues to bridge the gap between different FER datasets, and is expected to learn a generic feature space for a source and a target dataset.

3. Methodology

3.1. Overview of AdaFER

The goal of our method is to mitigate the domain gap, including inconsistent annotations and different imaging conditions. Considering the relationship between subjective facial expressions and objective action units, we propose a simple yet efficient AU-guided unsupervised domain-adaptive facial expression recognition (AdaFER) method. Figure 2 illustrates the pipeline of our AdaFER, which mainly consists of two crucial modules: (i) an AU-guided annotation (AGA) module and (ii) an AU-guided triplet training (AGT) module. Given images from a source domain and a target domain, we first utilize a pretrained AU detection model to extract the AUs coding for images from both domains. Then, the AGA module assigns a soft/hard pseudo-label for each image in the target domain by comparing the AUs between the source and target domains. For example, one of the AGA strategies is to assign a target image with a hard label that is the same as a source image if both images have equal AUs. Meanwhile, we mine triplets among the source and target domains according to the AUs. For example, given an anchor image in the target domain, a positive image in the source domain is the one with equal AUs, and a negative image is the one with different AUs. We jointly train the FER model on the source domain with ground truths and on the target domain with pseudo-labels and triplets.

3.2. The AU Distributions of FER Datasets

To check whether AUs have low bias among FER datasets (RAF-DB, FERPlus, ExpW, CK+), we visualize the AU distributions of each facial expression category. Specifically, we first utilize a pretrained AU detection model to extract AUs for each image. Then, we make statistics of AU occurrence numbers over each category. After normalizing, we show the AU statistics in Figure 3. We observe that (i) the AU distributions of the same categories are very similar among different datasets, and (ii) different facial expressions own very different AU distributions, which indicates that AUs offer discriminate cues.

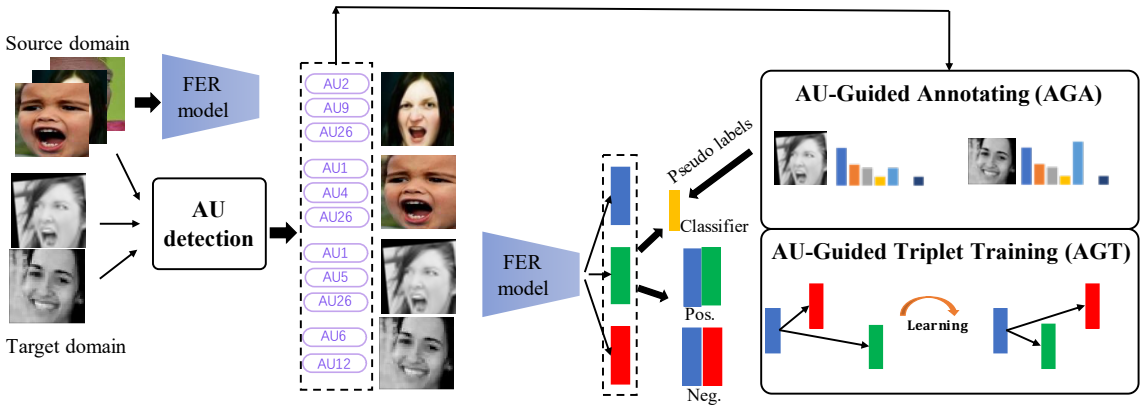
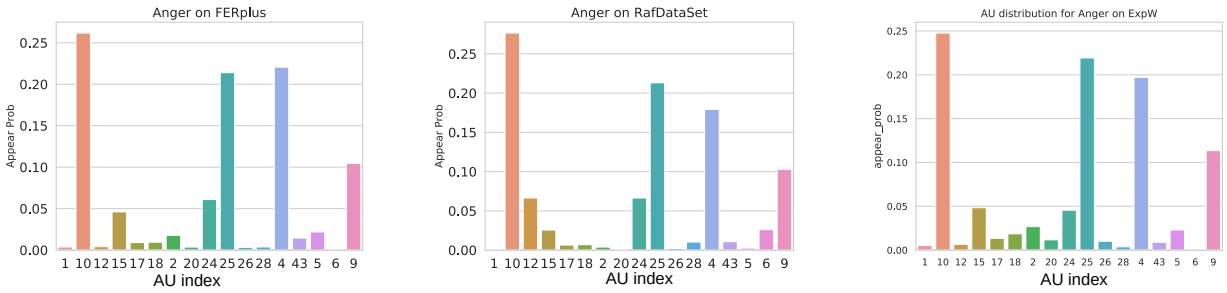
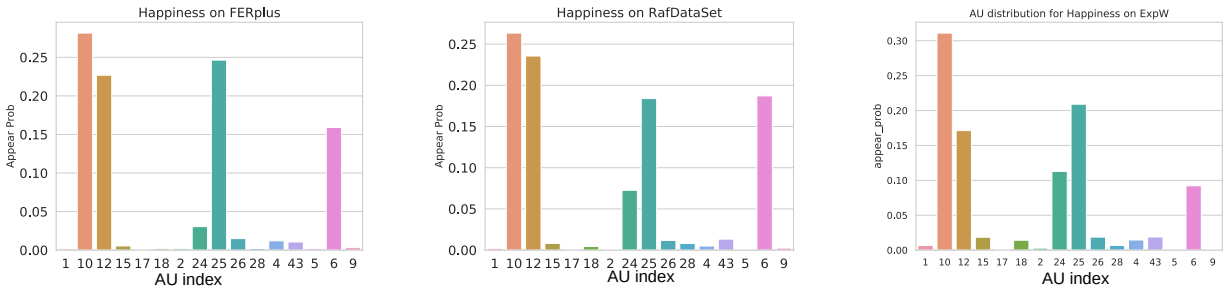


Figure 2. The pipeline of our AdaFER. First, a pretrained AU detection model is used to extract facial action units, and then an AU-guided annotation (AGA) module assigns pseudo-labels for joint training on the source domain and the target domain. AdaFER makes full use of the objective AUs to bridge the gap between FER datasets caused by subjective and inconsistent annotations.

The AUs Distributions of Anger Images on FER datasets



The AUs Distributions of Happy Images on FER datasets



The AUs Distributions of Sad Images on FER datasets

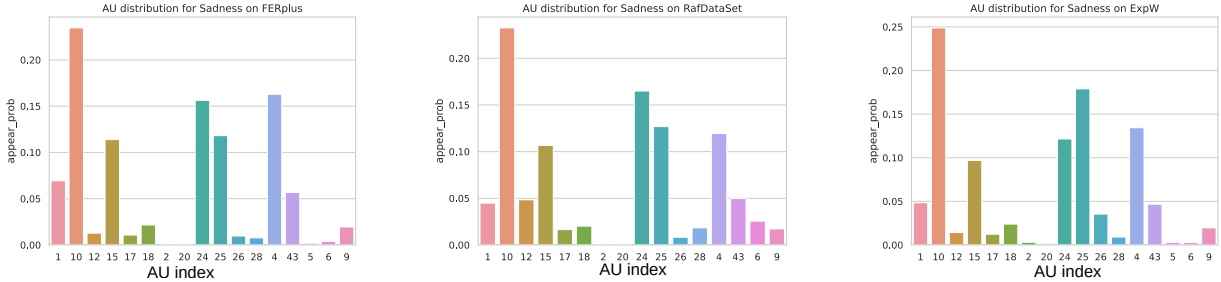


Figure 3. The distributions of AUs in different FER datasets. It can be seen that the AU distributions of a certain class in different datasets are almost consistent.

3.3. AU-Guided Annotating

We introduce the AU-Guided Annotation (AGA) module to assign pseudo-labels for target domain images according to AU detection results. Specifically, we elaborately design several assignment schemes as follows.

Source-based hard label assignment (S-hard). Given an image X_s^i and its detected AUs E_s^i from the source domain, the S-hard scheme utilizes E_s^i as a query to search for target domain images that have same AUs with E_s^i. For a face image X_t^j in the target domain, its label can be defined as follows:

$$y_{X_t^j} = y_{X_s^i} \text{ if } E_t^j \equiv E_s^i \tag{1}$$

where $y_{X_t^j}$ denotes the facial expression label of X_t^j. S-hard assigns all the retrieval images in the target domain with the label of X_s^i.

Target-based hard label assignment (T-hard). Different from the S-hard scheme, the T-hard scheme uses target domain images as query images. Given an image X_t^i and its detected AUs E_t^i from the target domain, it first utilizes the E_t^i to retrieve all the images in source domain, denoting them as $[X_s^1, \ldots, X_s^k]$. Then, with the ground truths of the source domain images, T-hard assigns the most-frequent label to the target domain image X_t^i.

Target-based soft label assignment (T-soft). For a query image from the target domain, unlike the hard assignment scheme, the T-soft scheme directly uses the label distribution of retrieval samples to assign each target domain image a soft label vector. It is worth noting that there does not exist a source-based soft label assignment since we do not have the labels of the target domain images.

Learning with AGA. After assigning pseudo-labels for the target domain, we can train the FER models in traditional ways. Suppose $Y_s \in \mathcal{R}^{1\times M}$ represents the labels of M source domain images, $Y_{\mathrm{S-hard}} \in \mathcal{R}^{1\times N}$, $Y_{\mathrm{T-hard}} \in \mathcal{R}^{1\times N}$, and $Y_{\mathrm{T-soft}} \in \mathcal{R}^{C\times N}$ denote the labels of N target domain images in S-hard, T-hard, and T-Soft schemes; we use the following loss function by default to train with AGA module:

$$L_c = CE(P_s, Y_s) + \beta((CE(P_t, Y_{\mathrm{S-hard}}) + \mathrm{KL}(\mathrm{P_t}, \mathrm{Y_{T-soft}})), \tag{2}$$

where CE denotes the cross-entropy loss function, KL is the KL divergence loss function, and P_t and P_s represent the predictions of the target and source domain images. β is the trade-off ratio between the two loss values, calculated by pseudo-labels and ground truths.

3.4. AU-Guided Triplet Training

To achieve structure-reliable and compact facial features, we perform AU-guided triplet training (AGT) to further narrow the gap among different domains. The key step is to sample triplets from source and target domains.

Triplet selection. Intuitively, we can select triplets from the union of the source domain and target domain. However, considering that CD-FER is a classification task, we ignore the triplets in the source domain since the ground truths are available. Specifically, we keep only those cross-domain triplets. Given an anchor in the source domain (X_s^a) or the target domain (X_t^a), we first use it to retrieve the images of the target domain or the source domain that own the same AUs, and then we randomly select a positive sample from the target domain (X_t^p) or source domain (X_s^p), according to the retrieval samples, and a negative sample X_t^n or X_s^n, according to the rest of the samples from the target or source domain. Thus, we mainly select two kinds of cross-domain triplets: (X_s^a,X_t^p,X_t^n) and (X_t^a,X_s^p,X_s^n). We conduct triplet selection in an offline manner. In addition, we also perform hard-negative mining by sorting the similarities between the AU scores of the anchor and all negative samples. We randomly select a sample from these with AU similarities that is larger than a threshold τ_n (0.5 by default) as a negative sample.

AU-guided triplet loss. After the selection of triplets, we use triplet loss to learn the discriminative and compact features as follows:

$$L_{tri} = max\{0, \gamma - (\|F_a - F_n\| - \|F_a - F_p\|)\}, \quad (3)$$

where F_a, F_p, and F_n represent the L2-normalized features of the anchor, positive, and negative images, respectively. γ is a margin which can be a fixed hyper parameter or a learnable parameter. We evaluate it in the Experiments section. Training with these cross-domain triplets, we can obtain a cross-domain common feature space which makes similar facial images close and dissimilar ones far away. Considering both pseudo-annotations and triplets, the total loss function is $L_{all} = L_c + \epsilon L_{tri}$, where ϵ is a trade-off ratio.

3.5. Implementation Details

AU detection and FER backbone. Face images are detected and aligned by Retinaface [52] and further resized to 224 × 224 pixels. We utilize an advanced AU detector that was pretrained on the EmotiNet dataset using the MLCR [20] algorithm to extract AUs for each image. We then evaluate the effectiveness of AdaFER using ResNet-18 [53]. ResNet-18 is pre-trained on the MS-Celeb-1M face recognition dataset and the facial features for triplet training are extracted from the last pooling layer.

Training. We use the PyTorch toolbox (https://pytorch.org/, accessed on 22 October 2021) to implement our method on a Linux server with 1 Nvidia Tesla V100 GPU. For training, we set the batch size to 128, i.e., 128 triplets with ground truth labels or pseudo-labels. In each iteration, all the images are optimized by cross-entropy loss, KLDiv loss, and AU-guided triplet Loss. The ratio β is defaulted as 1 and evaluated in the ensuing Experiments section. The triplet loss margin γ is set to 0.5 by default. The ratio of L_c and L_{tri} is empirically set to 1:1, and its influence will be studied in the ensuing ablation study in the Experiments section. The leaning rate is initialized as 0.001 with an Adam optimizer using an exponential (gamma = 0.9) scheme to reduce the learning rate. We stop training at the 40th epoch.

4. Experiments

In this section, we first describe the employed datasets. We then demonstrate the robustness of our AdaFER in cross-domain facial expression recognition tasks. Further, we conduct ablation studies to show the effectiveness of each module and the settings of hyper-parameters in AdaFER. After that, We compare AdaFER to related state-of-the-art methods. Finally, to obtain a better understanding of AdaFER, we visualize the statistical distributions of CK+ and FERPlus datasets.

4.1. Datasets

The **RAF-DB** [11] dataset consists of 30,000 facial images annotated with 7 basic and 14 compound facial expressions from 40 trained students. In our experiments, we, by default, used RAF-DB as the source dataset, and only images with 7 basic expressions (neutral, happiness, surprise, sadness, anger, disgust, fear, neutral) were used, which led to 12,271 images for training.

The **FER2013 and FERPlus** [8] datasets contain 28,709 training images, 3589 validation images, and 3589 test images. The image size of these datasets is 48 × 48. FER2013 is collected by Google Image Search API and annotated by seven facial expressions. However, the annotations of FER2013 are not accurate because there are only two annotators. Therefore, the FERPlus dataset is extended from FER2013, as in the ICML 2013 challenges, and it is also re-annotated by 10 annotators, and a contempt category is added.

The **CK+** [9] dataset contains 593 video sequences from 123 subjects. Among these videos, 327 sequences from 118 subjects are labeled with 7 expressions (except neutral), i.e., anger, contempt, disgust, fear, happiness, sadness, and surprise. All the subjects start from neutral and increase their expression intensity to seven expressions. Therefore, we select the last 3 frames with peak formation from each sequence and the first frame (neutral

face) of each sequence, resulting in 1236 images. We follow previous work [9] to choose 1108 images for training and 128 images for testing.

The **ExpW** [54] dataset contains 91,793 images that are annotated by 1 of the 7 expressions. Since the official ExpW dataset does not provide training/testing splits, we follow [17] to select 28848 for training, 28,853 for validating and 28,848 for testing.

The **JAFFE** [55] dataset collects 213 images from 10 Japanese females in lab-controlled conditions. Here, we chose 170 images for training and 43 images for testing.

4.2. AdaFER for Unsupervised CD-FER

To evaluate the effectiveness of AdaFER, we compare several baseline methods with our proposed AdaFER, using RAF-DB as the source dataset and testing on the CK+, JAFFE, ExpW, FER2013, and FERPlus datasets We implement five baseline methods in total, as follows:

- #1: We train the ResNet-18 (also pre-trained on MS-Celeb-1M) model on source data and directly test on target data;
- #2: We first extract AUs for both the source and the target data, and then use the AUs of each image in the target data to query the source data. Finally, we assign the most-frequent category of retrieval images to the target image;
- #3: We first use the trained model on the source data to predict hard pseudo-labels of the target data, then fine-tune the model on target data;
- #4: This method is identical to method #3, except that the predicted pseudo-labels are kept as vectors (i.e., soft labels);
- #5: We use both the images and the detected AUs as inputs to train a classification on the source set, and then fine-tune the model on pseudo-soft labelled target data.

The results are shown in Table 1. Several observations can be concluded, as follows: first, our AdaFER almost outperforms all other baseline methods by a large margin, especially when testing on the lab-controlled CK+ and JAFFE datasets. Second, using pseudo-labels (#3 and #4) to fine-tune model on target data also achieves large improvements over #1. Third, the naive AU-based method (#2) performs better than method #1 on the FERPlus and ExpW datasets, which indicates that AUs are useful among similar data. Last, but not least, the naive AU-based method degrades in the lab-controlled FER datasets, which suggests that there exists an AU domain gap between in-the-wild datasets and the lab-controlled datasets. Nevertheless, our AdaFER mitigates the domain gap and achieves large improvements on both the in-the-wild and the lab-controlled datasets.

Table 1. Performance (%) comparison between the proposed AdaFER and baseline methods. RAF-DB is used as source dataset.

Method	CK+	JAFFE	ExpW	FER2013	FERPlus
#1	70.54	46.51	61.53	52.91	62.40
#2	65.11	27.50	66.52	46.31	68.35
#3	74.42	55.81	68.13	55.89	63.81
#4	73.64	58.14	69.41	54.81	70.02
#5	71.32	53.49	**73.58**	57.15	77.99
AdaFER	**81.40**	**61.37**	70.86	**57.29**	**78.22**

Visualization of AU-guided annotating. To further investigate AdaFER, we visualize the pseudo-labels of the target images annotated by the AGA module and baseline method #3. We use our T-soft assignment and list the top three categories. As shown in Figure 4, AGA achieves better results than baseline method #3. In the first six images, AGA assigns the highest weights on the ground-truth category. The last two bad cases are extremely ambiguous, even for humans, and our AGA seems to assign reasonable categories for them.

Figure 4. Visualizations of pseudo-labels from baseline method #3 and our AGA module. BSA represents baseline method #3.

4.3. Ablation Studies

We conduct ablation studies for the modules of our AdaFER and other hyper parameters.

The three types of label assignments. In AGA, we introduce three kinds of assignment, namely, S-hard, T-hard, and T-soft. We explore individual types and their combination on the CK+, ExpW, FERPlus, and FER2013 datasets. The results are shown in Figure 5. For individual assignments, we observe that the T-soft strategy performs best on average, followed by the S-hard strategy. Combining the T-soft and S-hard strategies further boosts performance in most cases. We use this combination strategy by default in the following experiments.

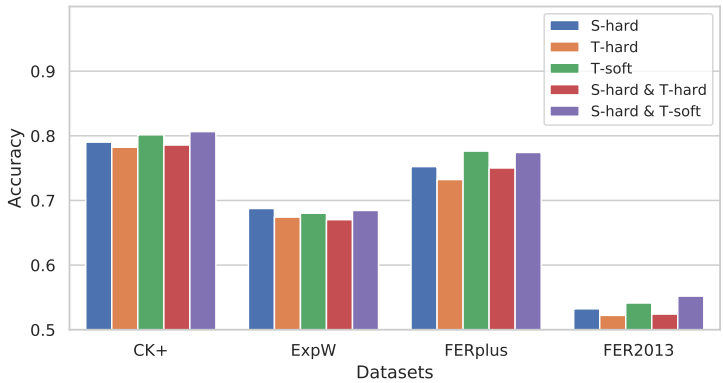


Figure 5. Evaluation of pseudo-assignment strategies.

Anchor images in AGT module. In AU-guided triplet training, both the source and target data can be used as anchor images. We evaluate the effect of the anchor images in Table 2. As can be observed, using source data as anchor images largely outperforms the target anchor scheme on the CK+ dataset, which may be explained by the target CK+ dataset being too small to collect enough triplet samples. Both individual anchor schemes

perform similarly on large-scale datasets, and combining both schemes consistently boosts performance on all datasets.

Table 2. Evaluation of the influences of anchor images on the AGA module.

Source	Target	CK+	ExpW	FER2013	FERPlus
√	×	80.62	69.90	54.36	77.61
×	√	71.23	69.73	55.45	77.93
√	√	**81.40**	**70.86**	**57.29**	**78.22**

AU-guided annotating (AGA) and AU-guided triplet training (AGT) are two crucial modules in AdaFER, which, respectively, leverages pseudo-labels for target domain images and constraints the distance structure of each triplet tuple. To explore the effectiveness of each module, we conduct the evaluation on the CK+, ExpW, FER2013, and FERPlus datasets. As shown in Table 3, both the AGA and AGT modules can individually improve the baseline by a large margin, and they perform similarly on most of the datasets. This may be explained by that both the AGA and AGT modules essentially resort to AU information, and the only difference is that AGA performs supervision on the classifier, while AGT does so on the feature mapping. Nevertheless, from the results of AGA+AGT, it is clear that they compliment each other on all the datasets.

Table 3. Evaluation of AU-guided annotating (AGA) and AU-guided triplet training (AGT) on the CK+, ExpW, FER2013, and FERPlus datasets.

AGA	AGT	CK+	ExpW	FER2013	FERPlus
×	×	70.54	61.53	52.91	62.40
√	×	80.62	68.45	55.20	77.42
×	√	80.28	69.80	56.45	74.50
√	√	**81.40**	**70.86**	**57.29**	**78.22**

The margin γ of the AGT module. γ is a margin parameter to control the distance between the anchor-positive pair and the anchor-negative pair. Theoretically, it can be a learnable parameter in the end-to-end framework. We evaluate it with both a fixed mode and a learnable mode. The results are illustrated in Figure 6. For the fixed mode, we evaluate a margin from 0.25 to 1.5, with 0.25 as the interval. On all of the four FER datasets, our default margin $\gamma = 0.5$ achieves the highest performance. Larger margins make training harder, which degrades performance on the CK+ and ExpW datasets. For the learnable mode, γ, respectively, converges to 0.62 ($\pm$0.034), 0.37 ($\pm$0.067), 0.71($\pm$0.026), and 0.42($\pm$0.033) on the CK+, ExpW, FERPlus, and FER2013 datasets. Meanwhile, the learnable γ also obtains competitive results.

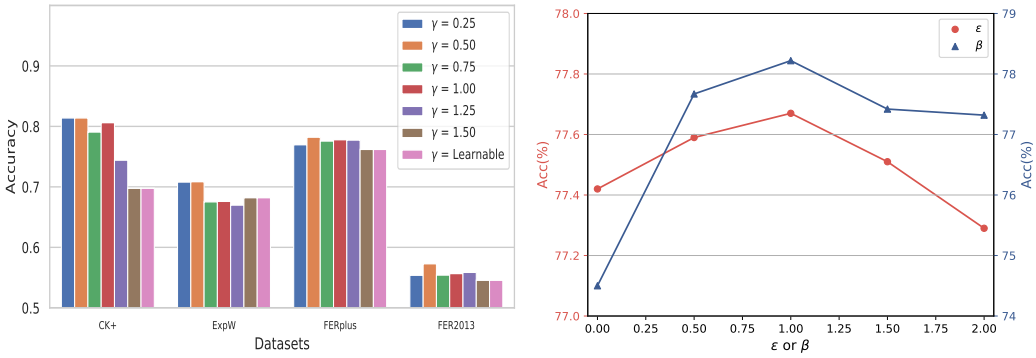


Figure 6. Evaluation of the margin parameter (γ), and the trade-off ratios ϵ and β.

The trade-off ratios β and ϵ. β is the trade-off ratio between two loss values calculated by the pseudo-labels and ground truths in Equation (2). ϵ is the trade-off ratio between L_c and L_{tri}. We evaluate them from 0.0 to 2.0, with 0.25 as the interval on the FERPlus dataset, and present the results in Figure 6. For both β and ϵ, the final performance increases gradually and reaches the peak in value 1.0. Larger values degrade performance dramatically, which illustrates that all the loss items are almost equally important.

The threshold for hard-negative mining of triplet samples. For triplet selection, we sort the AU scores and set a threshold for negative mining. We evaluate the threshold from 0 to 0.75 in Table 4. On all datasets, increasing the threshold from 0 to 0.5 boosts performance largely. A too-small threshold could result in zero triplet loss since the triplet samples are too easy. A too-large threshold may introduce hard-positive samples as negative ones, which is harmful for training.

Table 4. Evaluation of the threshold for hard-negative mining.

Threshold	CK+	ExpW	FER2013	FERPlus
0	68.99	65.24	52.21	66.68
0.25	72.03	68.23	56.23	74.23
0.5	**81.40**	**70.86**	**57.29**	**78.22**
0.75	80.96	69.12	55.80	77.72

4.4. Comparison with State-of-the-Art Methods

Table 5 compares our AdaFER with 13 state-of-the-art (SOTA) methods in the CD-FER task. The methods from the bottom table use RAF-DB as the source dataset, and those from the upper table use other datasets. As shown in Table 5, our AdaFER achieves competitive results compared to the SOTA methods. Zavarez et al. [56] achieve SOTA quality on CK+ using six datasets. ECAN [42] obtain SOTA quality on the JAFFE and FER2013 datasets using a model pretrained on VGGFace2 and then fine-tuned on RAF-DB2.0, which is not publicly available. For fair comparison, the methods in the bottom part of Table 5 use the same source dataset and backbone. The mean accuracy over all target datasets is also computed for easy comparison. Our AdaFER obtains accuracy values of **81.40%**, **61.37%**, **57.29%**, and **70.86%** on the CK+, JAFFE, FER2013, and ExpW datasets, respectively, which are the new state-of-the-art CD-FER results for these datasets. Moreover, our AdaFER does not increase any computing cost in the inference phase. AGRN [17] needs to extract holistic and local features to initialize the nodes of the target domain in the inference stage, which is time-cosuming. LPL [11] relies on the assumption of both the source and target domains.

To evaluate the robustness of our method on different source datasets, we also show the performance of CD-FER using FERPlus as the source dataset. Note that the FER2013 dataset has the same images as FERPlus; therefore, we ignore its performance on FER2013. Our AdaFER also improves the baseline method (#3) by a large margin in terms of mean accuracy.

Understanding AU-guided learning. To better understand the differences between the baseline method (#3) and our AdaFER, we utilize the T-SNE method to illustrate the feature distributions of the test sets of the CK+ and FERPlus datasets. The results are shown in Figure 7. We can find that our AdaFER can learn more compact features than the baseline in unsupervised CD-FER tasks. For CK+, "Neutral" samples are the most scattered ones for the baseline method; AdaFER can cluster them dramatically. For FERPlus, the samples of same categories are clustered compactly, while those of different categories are largely separated. These illustrates the reason why AdaFER improves the baseline dramatically.

Table 5. Comparison of the state-of-the-art methods on the CK+, JAFFE, FER2013, and ExpW datasets. The results of upper part are taken from the corresponding papers, and the results of the bottom part are taken from the implementation of a cross-domain facial expression recognition benchmark [57]. The last column shows the mean accuracy of the performances on all the datasets.

Methods	Source Dataset	Backbones	CK+	JAFFE	FER2013	ExpW	Mean
Da et al. [58]	BOSPHORUS	HOG and Gabor Filters	57.60	36.2	-	-	-
Hasani et al. [59]	MMI and FERA and DISFA	Inception-ResNet	67.52	-	-	-	-
Hasani et al. [60]	MMI and FERA	Inception-ResNet	73.91	-	-	-	-
Zavarez et al. [56]	Six Datasets	VGG-Net	88.58	44.32	-	-	-
Mollahosseini et al. [61]	Six Datasets	Inception	64.20	-	34.00	-	-
DETN [62]	RAF-DB	Manually Designed Net	78.83	57.75	52.37	-	-
ECAN [42]	RAF-DB 2.0	VGG-Net	86.49	61.94	58.21	-	-
CADA [63]	RAF-DB	ResNet-18	73.64	55.40	54.71	63.74	61.87
SAFN [64]	RAF-DB	ResNet-18	68.99	49.30	53.31	68.32	59.98
SWD [65]	RAF-DB	ResNet-18	72.09	53.52	53.70	65.85	61.29
LPL [11]	RAF-DB	ResNet-18	72.87	53.99	53.61	68.35	62.20
DETN [62]	RAF-DB	ResNet-18	64.19	52.11	42.01	43.92	50.55
ECAN [42]	RAF-DB	ResNet-18	66.51	52.11	50.76	48.73	54.52
AGRA [17]	RAF-DB	ResNet-18	77.52	61.03	54.94	69.70	65.79
AdaFER	RAF-DB	ResNet-18	**81.40**	**61.37**	**57.29**	**70.86**	**67.73**
Baseline (#3)	FERPlus	ResNet-18	64.34	41.86	-	66.64	57.87
AdaFER	FERPlus	ResNet-18	65.12	46.51	-	73.58	61.47

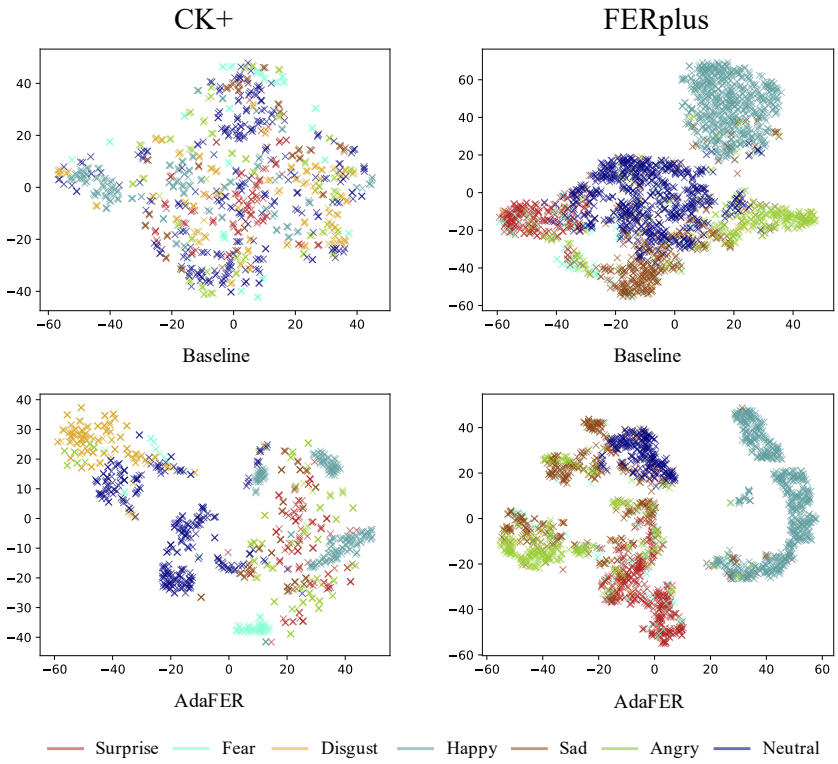


Figure 7. Visualization of the feature distributions on the CK+ and FERPlus datasets. The top and bottom parts show the feature distributions using the baseline method and AdaFER, respectively.

5. Conslusions

In this paper, we address the unsupervised domain-adaptive facial expression recognition task with auxiliary facial action units. Our method is very different from existing cross-domain FER methods, which typically follow generic cross-domain methods. Specifically, we proposed an AU-guided unsupervised domain-adaptive FER (AdaFER) framework, which includes an AU-guided annotation module and an AU-guided triplet training module. We evaluated several AU-guided annotation strategies and triplet selection methods. Extensive experiments on several popular benchmarks have shown the effectiveness of our AdaFER.

Limitations and future work. Since our method resorts to AU for consistent labeling, the accuracy of AU detection may be limited. In addition, our method actually only focuses on the label bias of different datasets; thus, the data bias problem in both FER and AU detection can be our future work.

Author Contributions: Conceptualization, X.P.; methodology, X.P. and Y.G.; software, Y.G. and P.Z.; data curation,Y.G. and P.Z.; writing—original draft preparation, Y.G..; writing—review and editing, X.P.;supervision, X.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the National Natural Science Foundation of China (62176165) and the Stable Support Projects for Shenzhen Higher Education Institutions (SZWD2021011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J.G. Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* **2001**, *18*, 32–80. [CrossRef]
2. Giorgana, G.; Ploeger, P.G. Facial expression recognition for domestic service robots. In *Robot Soccer World Cup*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 353–364.
3. Jiang, W.; Yin, Z.; Pang, Y.; Wu, F.; Kong, L.; Xu, K. Brain functional changes in facial expression recognition in patients with major depressive disorder before and after antidepressant treatment: A functional magnetic resonance imaging study. *Neural Regen. Res.* **2012**, *7*, 1151. [PubMed]
4. 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 Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
5. Valstar, M.; Pantic, M. Induced disgust, happiness and surprise: An addition to the mmi facial expression database. In Proceedings of the 3rd International Workshop on EMOTION: Corpora for Research on Emotion and Affect, Paris, France, 17–23 May 2010; p. 65.
6. Zhao, G.; Huang, X.; Taini, M.; Li, S.Z.; PietikäInen, M. Facial expression recognition from near-infrared videos. *Image Vis. Comput.* **2011**, *29*, 607–619. [CrossRef]
7. Dhall, A.; Goecke, R.; Lucey, S.; Gedeon, T. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Proceedings of the IEEE ICCV Workshops, Barcelona, Spain, 6–13 November 2011; pp. 2106–2112.
8. Barsoum, E.; Zhang, C.; Canton Ferrer, C.; Zhang, Z. Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution. In Proceedings of the ACM ICMI, Tokyo, Japan, 12–16 November 2016.
9. Mollahosseini, A.; Hasani, B.; Mahoor, M.H.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *TAC* **2017**, *10*, 18–31. [CrossRef]
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 CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 5562–5570.
11. Li, S.; Deng, W.; Du, J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In Proceedings of the CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 2852–2861.
12. Ben, X.; Ren, Y.; Zhang, J.; Wang, S.J.; Liu, Y.J. Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**. [CrossRef] [PubMed]
13. Wang, K.; Peng, X.; Yang, J.; Meng, D.; Qiao, Y. Region attention networks for pose and occlusion robust facial expression recognition. *IEEE TIP* **2020**, *29*, 4057–4069. [CrossRef] [PubMed]
14. Li, Y.; Zeng, J.; Shan, S.; Chen, X. Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE TIP* **2018**, *28*, 2439–2450. [CrossRef] [PubMed]
15. Wang, K.; Peng, X.; Yang, J.; Lu, S.; Qiao, Y. Suppressing uncertainties for large-scale facial expression recognition. In Proceedings of the IEEE CVPR, Seattle, WA, USA, 14–19 June 2020; pp. 6897–6906.
16. Yan, K.; Zheng, W.; Zhang, T.; Zong, Y.; Tang, C.; Lu, C.; Cui, Z. Cross-domain facial expression recognition based on transductive deep transfer learning. *IEEE Access* **2019**, *7*, 108906–108915. [CrossRef]
17. Xie, Y.; Chen, T.; Pu, T.; Wu, H.; Lin, L. Adversarial Graph Representation Adaptation for Cross-Domain Facial Expression Recognition. In Proceedings of the 28th ACMMM, Seattle, WA, USA, 12–16 October 2020; pp. 1255–1264.
18. Zong, Y.; Zheng, W.; Huang, X.; Shi, J.; Cui, Z.; Zhao, G. Domain regeneration for cross-database micro-expression recognition. *IEEE Trans. Image Process.* **2018**, *27*, 2484–2498. [CrossRef]

19. Ji, S.; Wang, K.; Peng, X.; Yang, J.; Zeng, Z.; Qiao, Y. Multiple Transfer Learning and Multi-Label Balanced Training Strategies for Facial AU Detection in the Wild. In Proceedings of the IEEE CVPR Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 414–415.
20. Niu, X.; Han, H.; Shan, S.; Chen, X. Multi-label co-regularization for semi-supervised facial action unit recognition. *arXiv* **2019**, arXiv:1910.11012.
21. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
22. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. Openface: A general-purpose face recognition library with mobile applications. *CMU Sch. Comput. Sci.* **2016**, *6*, 20.
23. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef] [PubMed]
24. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* **2009**, *27*, 803–816. [CrossRef]
25. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE CVPR, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
26. Tang, Y. Deep learning using linear support vector machines. *arXiv* **2013**, arXiv:1306.0239.
27. Kahou, S.E.; Pal, C.; Bouthillier, X.; Froumenty, P.; Gülçehre, Ç.; Memisevic, R.; Vincent, P.; Courville, A.; Bengio, Y.; Ferrari, R.C.; et al. Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction, Sydney, Australia, 9–13 December 2013; pp. 543–550.
28. Zhou, H.; Meng, D.; Zhang, Y.; Peng, X.; Du, J.; Wang, K.; Qiao, Y. Exploring emotion features and fusion strategies for audio-video emotion recognition. In Proceedings of the 2019 ICMI, Suzhou, China, 14–18 October 2019; pp. 562–566.
29. Liu, M.; Li, S.; Shan, S.; Chen, X. Au-inspired deep networks for facial expression feature learning. *Neurocomputing* **2015**, *159*, 126–136. [CrossRef]
30. McDuff, D.; Mahmoud, A.; Mavadati, M.; Amr, M.; Turcot, J.; Kaliouby, R.E. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In Proceedings of the Association for Computing Machinery, CHI EA '16, New York, NY, USA, 7–12 May 2016; pp. 3723–3726.
31. 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: Oxford, MI, USA, 1997.
32. Feldman, R.S.; Jenkins, L.; Popoola, O. Detection of deception in adults and children via facial expressions. *Child Dev.* **1979**, *50*, 350–355. [CrossRef]
33. Rubinow, D.R.; Post, R.M. Impaired recognition of affect in facial expression in depressed patients. *Biol. Psychiatry* **1992**, *31*, 947–953. [CrossRef]
34. Niu, X.; Han, H.; Zeng, J.; Sun, X.; Shan, S.; Huang, Y.; Yang, S.; Chen, X. Automatic engagement prediction with GAP feature. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder, CO, USA, 16–20 October 2018; pp. 599–603.
35. Zhang, X.; Yin, L.; Cohn, J.F.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P.; Girard, J.M. Bp4d-spontaneous: A high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.* **2014**, *32*, 692–706. [CrossRef]
36. Mavadati, S.M.; Mahoor, M.H.; Bartlett, K.; Trinh, P.; Cohn, J.F. Disfa: A spontaneous facial action intensity database. *IEEE Trans. Affect. Comput.* **2013**, *4*, 151–160. [CrossRef]
37. Girard, J.M.; Chu, W.S.; Jeni, L.A.; Cohn, J.F. Sayette group formation task (gft) spontaneous facial expression database. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 581–588.
38. Walecki, R.; Pavlovic, V.; Schuller, B.; Pantic, M. Deep structured learning for facial action unit intensity estimation. In Proceedings of the IEEE CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 3405–3414.
39. Peng, G.; Wang, S. Dual semi-supervised learning for facial action unit recognition. In Proceedings of the AAAI, Montreal, QC, Canada, 2–5 July 2019; Volume 33, pp. 8827–8834.
40. Wu, S.; Wang, S.; Pan, B.; Ji, Q. Deep facial action unit recognition from partially labeled data. In Proceedings of the IEEE ICCV, Venice, Italy, 22–29 October 2017; pp. 3951–3959.
41. Chu, W.S.; De la Torre, F.; Cohn, J.F. Selective transfer machine for personalized facial expression analysis. *IEEE TPAMI* **2016**, *39*, 529–545. [CrossRef]
42. Li, S.; Deng, W. A deeper look at facial expression dataset bias. *IEEE Trans. Affect. Comput.* **2020**. 2020.2973158. [CrossRef]
43. Miao, Y.Q.; Araujo, R.; Kamel, M.S. Cross-domain facial expression recognition using supervised kernel mean matching. In Proceedings of the 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 12–15 December 2012; Volume 2, pp. 326–332.
44. Sangineto, E.; Zen, G.; Ricci, E.; Sebe, N. We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer. In Proceedings of the 22nd ACMMM, Orlando, FL, USA, 3–7 November 2014; pp. 357–366.
45. Yan, H. Transfer subspace learning for cross-dataset facial expression recognition. *Neurocomputing* **2016**, *208*, 165–173. [CrossRef]
46. Yan, K.; Zheng, W.; Cui, Z.; Zong, Y. Cross-database facial expression recognition via unsupervised domain adaptive dictionary learning. In *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 427–434.

47. Zhu, R.; Sang, G.; Zhao, Q. Discriminative feature adaptation for cross-domain facial expression recognition. In Proceedings of the 2016 International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–7.
48. Zheng, W.; Zong, Y.; Zhou, X.; Xin, M. Cross-Domain Color Facial Expression Recognition Using Transductive Transfer Subspace Learning. *IEEE Trans. Affect. Comput.* **2016**, *9*, 21–37. [CrossRef]
49. Levi, G.; Hassner, T. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In Proceedings of the 2015 ACM ICMI, Seattle, WA, USA, 9–13 November 2015; pp. 503–510.
50. Valstar, M.F.; Pantic, M. Fully Automatic Recognition of the Temporal Phases of Facial Actions. *IEEE Trans. Syst. Man, Cybern. Part B (Cybernetics)* **2012**, *42*, 28–43. [CrossRef] [PubMed]
51. Wang, X.; Wang, X.; Ni, Y. Unsupervised domain adaptation for facial expression recognition using generative adversarial networks. *Comput. Intell. Neurosci.* **2018**, *2018*, 7208794. [CrossRef]
52. Deng, J.; Guo, J.; Zhou, Y.; Yu, J.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-stage dense face localisation in the wild. *arXiv* **2019**, arXiv:1905.00641.
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
54. Zhang, Z.; Luo, P.; Loy, C.C.; Tang, X. From facial expression recognition to interpersonal relation prediction. *IJCV* **2018**, *126*, 550–569. [CrossRef]
55. Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with gabor wavelets. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 200–205.
56. Zavarez, M.V.; Berriel, R.F.; Oliveira-Santos, T. Cross-database facial expression recognition based on fine-tuned deep convolutional network. In Proceedings of the 30th SIBGRAPI, Niterói, Brazil, 17–20 October 2017; pp. 405–412.
57. Chen, T.; Xie, Y.; Pu, T.; Wu, H.; Lin, L. Cross-Domain Facial Expression Recognition:A Unified Evaluation Benchmark and Adversarial Graph Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**. [CrossRef]
58. Da Silva, F.A.M.; Pedrini, H. Effects of cultural characteristics on building an emotion classifier through facial expression analysis. *J. Electron. Imaging* **2015**, *24*, 023015. [CrossRef]
59. Hasani, B.; Mahoor, M.H. Facial expression recognition using enhanced deep 3D convolutional neural networks. In Proceedings of the IEEE CVPR Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 30–40.
60. Hasani, B.; Mahoor, M.H. Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields. In Proceedings of the 12th IEEE FG 2017, Washington, DC, USA, 30 May–3 June 2017; pp. 790–795.
61. Mollahosseini, A.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the IEEE WACV, Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.
62. Li, S.; Deng, W. Deep facial expression recognition: A survey. *IEEE Trans. Affect. Comput.* **2020**. 2020.2981446. [CrossRef]
63. Long, M.; Cao, Z.; Wang, J.; Jordan, M.I. Conditional adversarial domain adaptation. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1640–1650.
64. Xu, R.; Li, G.; Yang, J.; Lin, L. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In Proceedings of the IEEE ICCV, Seoul, Korea, 27 October–2 November 2019; pp. 1426–1435.
65. Lee, C.Y.; Batra, T.; Baig, M.H.; Ulbricht, D. Sliced wasserstein discrepancy for unsupervised domain adaptation. In Proceedings of the IEEE CVPR, Long Beach, CA, USA, 15–20 June 2019; pp. 10285–10295.