Uncertain Label Correction via Auxiliary Action Unit Graphs for Facial Expression Recognition

Yang Liu[1,2], Xingming Zhang[1], Janne Kauttonen[3] and Guoying Zhao[2*]
[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, 510006
[2] Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland, FI-90014
[3] Haaga-Helia University of Applied Sciences, Helsinki, Finland, FI-00520

***Abstract*—High-quality annotated images are significant to deep facial expression recognition (FER) methods. However, uncertain labels, mostly existing in large-scale public datasets, often mislead the training process. In this paper, we achieve uncertain label correction of facial expressions using auxiliary action unit (AU) graphs, called ULC-AG. Specifically, a weighted regularization module is introduced to highlight valid samples and suppress category imbalance in every batch. Based on the latent dependency between emotions and AUs, an auxiliary branch using graph convolutional layers is added to extract the semantic information from graph topologies. Finally, a re-labeling strategy corrects the ambiguous annotations by comparing their feature similarities with semantic templates. Experiments show that our ULC-AG achieves 89.31% and 61.57% accuracy on RAF-DB and AffectNet datasets, respectively, outperform the baseline and state-of-the-art methods.**

I. Introduction

Facial expression recognition (FER) plays an essential role in realizing human-computer interaction. It can be used in many practical applications, including health monitors, virtual reality, and social robots. Recently, deep neural networks (DNNs) have become the dominant FER methods and achieved excellent performance based on sufficient annotated images and high-speed computing resources [1], [2], [3].

Since the training of deep models requires massive data, public FER datasets have been collected in both constrained (e.g., CK+ [4] and Oulu-CASIA [5]) and in-the-wild (e.g., RAF-DB [6] and AffectNet [7]) conditions. However, for those samples in real-world datasets, their annotations are difficult to maintain consistency in a large-scale manner. As a result, many labels are ambiguous or even incorrect. These may be due to the subjectivity of the annotators and the natural confusion of certain facial expressions. In Fig. 1, we show several examples in RAF-DB and AffectNet to illustrate that uncertainty is common in images collected on the Internet. For samples on the left column, annotators can easily make consistent labeling. While for the right column, it is obvious that multiple annotators might have various perspectives on the same sample. In other words, this phenomenon may result in two negative impacts on the model learning: 1) the overfitting problem will arise due to the considerable proportion of ambiguous samples in the training set; 2) the incorrect labels will mislead the model learning features of specific facial expressions and decrease recognition performance.

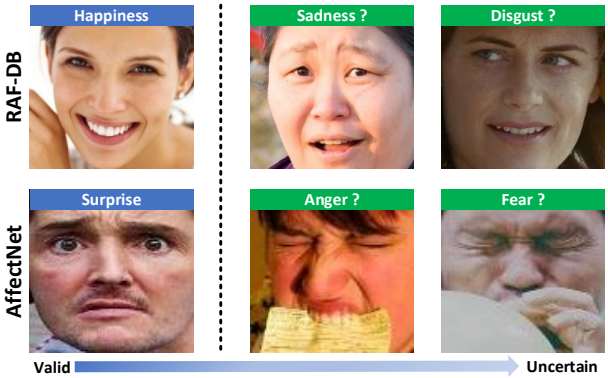

Fig. 1. Examples of valid and uncertain images in RAF-DB and AffectNet datasets.

To this end, methods have been studied to mine the knowledge in label space and alleviate the uncertainties. Wang *et al.* [8] proposed a self-cure network to learn the importance weight of each facial image and suppress uncertain samples by identifying and modifying untruthful labels. Song *et al.* [9] imposed probabilistic masks to capture and weight uncertain samples through an uncertain graph neural network. She *et al.* [10] exploited auxiliary multi-branch distribution learning, and pairwise uncertainty estimation to solve the ambiguity in both the label space and the instance space. Zhang *et al.* [11] formulated a noise modeling network based on a weakly-supervised strategy that learned the mapping from feature space to the residuals between clean and noisy labels.

Recently, the idea of using the relationship among multiple labels has been explored. Chen and Joo [12] incorporated the *triplet* loss into the objective function to embed the dependency between AUs and expression categories. Zhang *et al.* [13] designed a unified adversarial learning framework to link the emotion prediction and the joint distribution of dimensional labels. Alternatively, Chen *et al.* [14] introduced auxiliary label space graphs that cluster samples in neighbor tasks such as landmark detection and AU detection, and

* Corresponding author
This work was supported by the China Scholarship Council under Grant 202006150091, and the Academy of Finland for Academy Professor project EmotionAI (grants 336116, 345122), and the Ministry of Education and Culture of Finland for AI forum project. The authors wish to acknowledge CSC-IT Center for Science, Finland, for generous computational resources.

leverage the distributions to handle the label inconsistency. Using graphs to solve uncertainty problems was also applied in [15]. Similarly, Cui *et al.* [16] extracted the dependency between object-level labels and property-level labels, which could be used to revise and generate labels for new datasets. However, these previous methods are still plagued by uncertain samples for the following two reasons: 1) although knowledge including AUs is applied, semantic information is considered from the label level rather than the feature level; 2) the relabeling strategy without constraints is usually rough, which could decrease the reliability of the generated labels.

In this paper, we perform FER on data with uncertain samples, called **U**ncertain **L**abel **C**orrection via **A**uxiliary AU **G**raphs (ULC-AG). ULC-AG consists of two parts: the target branch and the auxiliary branch. For the former, facial features are first extracted through a backbone DNN for every batch of the training data. A weighted regularization module estimates each sample by learning confidence and encourages the model to focus on images with valid labels while considering category imbalance. For the latter, we utilize the same backbone network but change the task to AU detection. It can be regarded as a form of multi-task learning with shared parameters. Then, a graph convolution block is added with all the AU features as nodes to output the semantic feature of each sample. For those images identified as having uncertain labels, we compare their feature similarity with semantic templates and re-label them under the constraint of semantic preserving. Overall, the main contributions can be summarized as follows:

- The proposed ULC-AG method mitigates the effects of ambiguous samples and category bias for better facial expression feature learning.
- ULC-AG explores semantic information of facial expressions from the auxiliary AU detection task and conducts uncertain label correction under a feature-level constraint.
- Our ULC-AG is an end-to-end framework and can achieve superior performance on large-scale FER benchmarks.

II. Proposed Method

As mentioned above, for public FER datasets, especially with large-scale web images, it is hard to keep all labels high-quality and consistent. To this end, one possible solution is to correct the mislabeled sample with the help of knowledge spaces other than the labels themselves. Inspired by [13], [14], we introduce the idea of the auxiliary task but focus on the similarity in feature level. The main assumption of this work is that labels of similar samples should have an underlying dependency, which will also be reflected in their feature representation. In this section, we first present an overview of ULC-AG and then elaborate on its crucial modules.

A. Overview of ULC-AG

An overview of ULC-AG is illustrated in Fig. 2. The ULC-AG contains: 1) a target branch that takes facial features extracted by a pre-trained DNN and computes the annotation confidence using a self-attention layer. These confidence weights will affect the importance of the sample when calculating the classification loss. In addition, the class-oriented weight is computed to deal with category imbalance in the current batch; 2) an auxiliary branch that follows the idea of multi-task learning exploits the same backbone to get AU features of each facial image and feed them into a two-layer graph convolutional network (GCN) [17] for semantic feature extraction. The re-labeling strategy corrects suspicious labels according to the semantic similarity between the low confidence sample and the templates. The whole ULC-AG is an end-to-end framework and the auxiliary branch will not participate in the testing process.

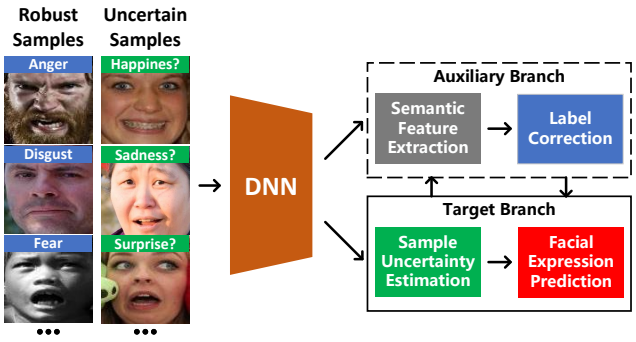

Fig. 2. The framework of ULC-AG. It consists of a target branch and an auxiliary branch. The auxiliary branch will not participate in the testing process. Only samples with low confidence labels will be re-annotated.

B. Weighted Regularization

Fig. 3 illustrates the pipeline of the ULC-AG target branch. To identify the ambiguous samples and estimate their uncertainties, inspired by [8], [18], a self-attention module is employed that consists of a fully connected (FC) layer and the sigmoid function. For a batch of N images, $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_N] \in \mathbb{R}^{D\times N}$ indicates the facial features extracted by the pre-trained DNN, D is the dimension for each facial feature. The confidence weight of the i-th sample can be calculated as:

$$\alpha_i = Sigmoid(\boldsymbol{W}_a^\top \boldsymbol{f}_i), \tag{1}$$

where $\boldsymbol{W}_a^\top$ denotes the parameters of the self-attention layer. In addition, to prevent the uncertainty caused by category imbalance, we introduce class-oriented weights that are computed as:

$$\gamma_j = 1 - \frac{N_j}{N}, j \in \{1, 2, ..., C\}, \tag{2}$$

where N_j is the number of images belonging to class j, and C is the number of classes.

During the model training, it is expected that samples with lower confidence weights should impose less impact, while categories with fewer samples should receive more attention in the current batch. Therefore, we improve the weighted Cross-Entropy (CE) loss proposed in [10]. Specifically, the loss function for facial expression classifier is formulated as:

$$L_{wce} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{\alpha_i \gamma_{y_i} \boldsymbol{W}_{y_i}^\top \boldsymbol{f}_i}}{\sum_{j=1}^{C} e^{\alpha_i \gamma_{y_i} \boldsymbol{W}_j^\top \boldsymbol{f}_i}}, \tag{3}$$

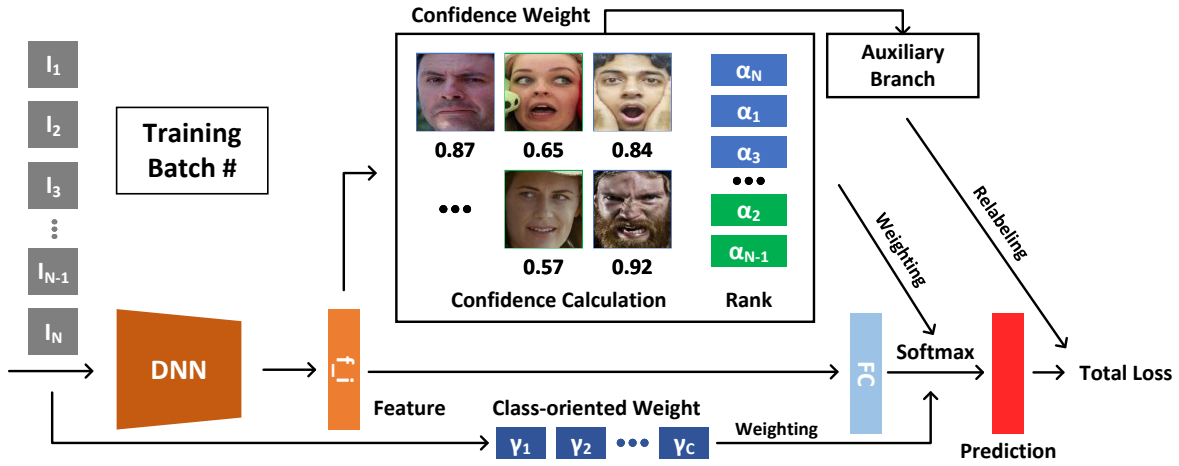

Fig. 3. The pipeline of the ULC-AG Target Branch.

where $\boldsymbol{W}_j^\top$ denotes the parameters of the j-th classifier, f_i is the facial feature, α_i is the confidence weight, y_i and γ_{y_i} are the original label and its corresponding class-oriented weight, respectively. According to [19], L_{wce} and α are positively correlated.

After we obtain the confidence weights, the high and low confidence samples in the current batch are divided with a ratio φ by using a simple rank regularization approach [8], which is formulated as:

$$L_{rr} = max(0, \theta - (Avg_h - Avg_l)), \tag{4}$$

where θ is a margin threshold that can be a fixed hyperparameter or updated with training, Avg_h and Avg_l are mean values of confidence weights in high and low groups, respectively.

C. Auxiliary AU Graph Branch

Recent studies reveal that introducing knowledge of the multi-label space can alleviate the effect of ambiguous facial expressions [13], [20]. In this work, we choose AU detection as our auxiliary task and construct semantically representative AU graphs because the Facial Action Coding System is an affect description model that has latent mappings with expression categories [21], [22], [23]. The AU graph takes individual AU features as graph nodes and the co-occurring AU dependency as graph edges. Fig. 4 illustrates the pipeline of the ULC-AG auxiliary branch.

Following the multi-task learning idea, we can get a set of AU features of each image with the same backbone network, $\boldsymbol{X}^i = [\boldsymbol{x}_1^i, \boldsymbol{x}_2^i, ..., \boldsymbol{x}_M^i] \in \mathbb{R}^{B\times M}$, B and M denote feature dimension and AU number, respectively. Considering the consistency of predefined mappings between expression categories and AUs in large-scale datasets is difficult to guarantee[24], [25], we exploit a data-driven approach based on the conditional probability to obtain co-occurring AU dependencies from the training set as graph edges, which can be calculated as:

$$\boldsymbol{A}_{p,q} = P(AU_p|AU_q) = \frac{OCC_{p\cap q}}{OCC_q}, \tag{5}$$

where $OCC_{p\cap q}$ denotes the number of co-occurrences of AU_p and AU_q, and OCC_q is the total number of occurrences of AU_q. Since the AU co-occurring relationship is actually asymmetry, so $P(AU_p|AU_q) \neq P(AU_q|AU_p)$.

Then, the two are input in a two-layer GCN with each AU as a graph node to extract the semantic feature. Specifically, each graph convolution layer is formulated as:

$$\boldsymbol{X}' = g(\boldsymbol{X}, \boldsymbol{A}) = LeakyRELU(\bar{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_g), \tag{6}$$

where $\bar{\boldsymbol{A}}$ denotes the normalized $\boldsymbol{A}$ with all rows sum to one, $\boldsymbol{W}_g$ is the weight matrix to be learned in the current layer.

All the node features outputted by the GCN are fed into a FC layer with sigmoid functions to predict multiple AUs. The binary CE loss is used to train every AU classifier, and the total balanced group loss is defined for the two-layer GCN as:

$$L_{au} = -\sum_{m=1}^{M} \alpha(z_m \log p_m + (1 - z_m) \log (1 - p_m)), \tag{7}$$

where α is the confidence weight, z_m and p_m are the pseudo label and the prediction of m-th AU, respectively. The feature $\boldsymbol{s}_i \in \mathbb{R}^{1\times M}$ before AU classifiers are treated as the semantic feature of the sample.

D. Relabeling with Semantic Preserving

To determine which labels need to be corrected and which new classes should be assigned, we design a semantic preserving strategy (see Fig. 4). For the divided two sample sets, a weighted semantic template set $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_C] \in \mathbb{R}^{M\times C}$ for every facial expression category is first generated with the semantic features and the confidence weights of valid samples, which can be formulated as:

$$\boldsymbol{T}_j = \frac{1}{K_j} \sum_{k_j=1}^{K_j} \alpha_{k_j} \boldsymbol{s}_{k_j}, \tag{8}$$

where K_j is the number of the samples with the j-th label in the high confidence set. The semantic templates will be dynamically updated throughout the whole training process.

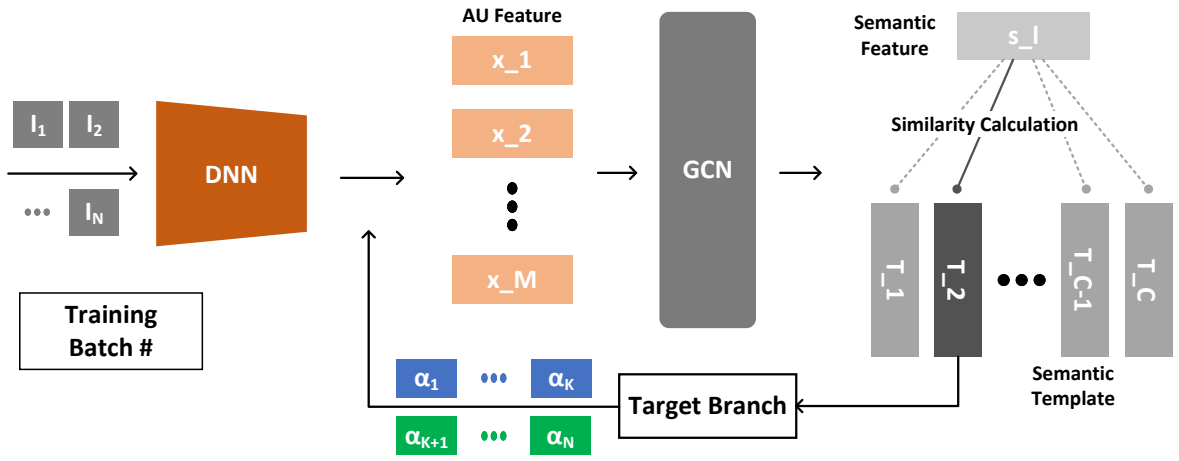

Fig. 4. The pipeline of the ULC-AG Auxiliary Branch.

For the ambiguous samples in the low confidence set, we calculate the cosine distance between $\boldsymbol{s}_l$ ($l \in \{1, 2, ..., N-K\}$) and each of $\boldsymbol{t}_j$ in the semantic template $\boldsymbol{T}$, which can be formulated as:

$$SP_{s_l,j} = 1 - \frac{\boldsymbol{t}_j \times \boldsymbol{s}_l}{\|\boldsymbol{t}_j\| \|\boldsymbol{s}_l\|}, \tag{9}$$

where $\times$ denotes the dot product operation, K is the total number of high confidence samples in each batch.

Next, for every ambiguous sample, we compare its semantic feature with each semantic template in T. The template class with the highest semantic similarity will be assigned to this sample as a new label. Formally, the re-labeling strategy can be defined as:

$$y_i' = \begin{cases} j, & \text{if } SP_{s_l,org} - min(SP_{s_l,j}) > 0 \\ y_i, & \text{otherwise} \end{cases} \tag{10}$$

where y_i' indicates the corrected label, and $j \neq org$, org is the original labeled class.

E. Model Training

Finally, the total loss function of the whole network can be written as:

$$L_{total} = \frac{\lambda_1}{2}(L_{wce} + L_{rr}) + \lambda_2 L_{au}, \tag{11}$$

where λ_1 and λ_2 are the weighted ramp functions that will change with epoch rounds [26], which can be computed as follows:

$$\lambda_1 = \begin{cases} \exp(-(1 - \frac{epoch}{\beta})^2), & epoch \leq \beta \\ 1, & epoch > \beta \end{cases}, \tag{12}$$

$$\lambda_2 = \begin{cases} 1, & epoch \leq \beta \\ \exp(-(1 - \frac{\beta}{epoch})^2), & epoch > \beta \end{cases}. \tag{13}$$

The weighted ramp functions allow ULC-AG to pay more attention to auxiliary branches in the initial training stage. Since the number of samples accumulated at the beginning is insufficient, it is unable to generate robust semantic features. After a certain number of training rounds, the model will focus more on the target branch to extract discriminative features for final predictions.

III. Experiments

A. Datasets

To evaluate the performance of ULC-AG in tackling label uncertainties, we conduct experiments on two popular FER benchmarks, RAF-DB [6] and AffectNet [7]. Both datasets have unconstrained conditions and large-scale samples.

RAF-DB has 15339 face images with annotations of six basic emotions and neutral. In our experiments, 12271 and 3368 samples are used for training and test, respectively.

AffectNet contains close to one million expression images. To ensure a fair comparison, we select samples manually labeled as six basic emotions and neutral for evaluation. The number of images in the training set and the test set is $283,901$ and 3500, respectively. In addition, automatically labeled samples in AffectNet are used as a set of real noisy data, denoted as **AffectNet_Auto**, to verify the ability of ULC-AG in handling uncertain expressions.

Since the AU annotation requires specially trained experts and is time-consuming, it is natural that no AU labels are provided in RAF-DB and AffectNet. To account for this issue, we applied Openface 2.0 [27] to automatically generate pseudo AU labels, similar to [12], [14]. Note that our ULC-AG utilizes feature-level semantic similarity preserving, which can reduce the negative impact of incorrect pseudo AU labels.

B. Implementation Details

The ULC-AG is implemented with the Pytorch platform and trained using two Nvidia Volta V100 GPUs. Face images are obtained using MTCNN [28] and further resized to 224×224 pixels as inputs. For the target branch, we choose the ResNet-18 as the backbone DNN which is pre-trained on the MS-Celeb-1M [29] dataset as previous methods [8], [10], [30]. In every iteration, φ and θ are set as 0.8 and 0.15, respectively. The initial learning rate is 0.01, which is updated to 10^{-3}

TABLE I
EVALUATION OF THE COMPONENTS IN ULC-AG. '*Target Branch*' APPLIES THE WEIGHT REGULARIZATION LOSS FUNCTION. '*Auxiliary Branch*' EXPLOITS AU GRAPHS AND THE SEMANTIC PRESERVING RELABELING.

Target Branch	*Auxiliary Branch*	RAF-DB	AffectNet
×	×	85.82	57.94
✓	×	86.54	58.66
×	✓	87.73	59.34
✓	✓	**89.31**	**61.57**

and 10^{-4} at the 10-th and 20-th epoch, respectively. For the auxiliary branch, each GCN layer has 64 channels, and the decayed learning rate is set as 0.005. The auxiliary branch will not participate in the network optimization until 10 epochs to obtain initial templates. Thus, when the relabeling starts afterward, there is no missing template in the current batch. We choose a batch size 512 to ensure that every template can be effectively updated during the whole training process.

C. Ablation Study

We conduct the ablation experiments to demonstrate the contributions of the proposed modules in this paper.

1) Components evaluation: ULC-AG aims to solve the influence of uncertain samples during feature learning. The target branch is to suppress low-quality inputs through confidence estimation and weighted regularization, and the auxiliary branch corrects uncertain labels by AU graph construction and semantic similarity constraint, both of which can be flexibly combined with various network architectures. In this experiment, we design four different settings for effectiveness verification. Note that the confidence weight will be calculated but not applied for regularization when only the auxiliary branch works. When the two branches are not active, ULC-AG is equivalent to a standard ResNet-18.

As shown in Table I, the independent use of the target branch or the auxiliary branch on the two datasets can significantly enhance the FER performance. In particular, the auxiliary branch makes the greater improvement because it introduces additional semantic information and explicitly manipulates uncertain samples. The best performance is achieved when using the two in collaboration. In other words, ULC-AG can effectively handle the uncertain samples in large-scale data.

2) Evaluation of data-driven edges: The data-driven edges introduce important semantic information about AU co-occurring dependencies into the constructed AU graphs. In this experiment, we randomly initialize A to shield edge attributes. Results in Table II show that the data-driven AU co-occurrence matrix can provide AU relationships that approximate the actual distribution, thereby helping the GCN to better extract affective semantic features from the AU graph.

D. Evaluation of Handling Uncertain Labels

To test our re-labeling strategy, we set up comparative experiments under synthetic uncertainty and real uncertainty, respectively. The ResNet-18 baseline and the SCN [8] also using label repair are selected for comparison.

TABLE II
EVALUATION OF DATA-DRIVEN EDGES IN ULC-AG.

Edges	RAF-DB	AffectNet
Random	85.06	55.79
Data-driven	**89.31**	**61.57**

1) Synthetic uncertain samples: We randomized 10%, 20%, and 30% of the original labels of the training set for RAF-DB and AffectNet, respectively. From Table III, our ULC-AG outperforms another two methods in this experiment. This illustrates the universality of uncertain samples in large-scale facial expression datasets. In addition, as the proportion of uncertain labels increases, the performance degradation of ULC-AG compared to another two methods are also smaller, which further proves the effectiveness of our feature-level semantic similarity constraint.

TABLE III
PERFORMANCE OF ULC-AG ON DATASETS WITH SYNTHETIC UNCERTAIN SAMPLES.

Method	Uncertainty	RAF-DB	AffectNet
Baseline	10%	80.63	57.25
SCN [8]	10%	82.18	58.58
ULC-AG	10%	**83.21**	**59.45**
Baseline	20%	78.06	56.23
SCN [8]	20%	80.10	57.25
ULC-AG	20%	**81.16**	**58.51**
Baseline	30%	75.13	52.60
SCN [8]	30%	77.46	55.05
ULC-AG	30%	**79.01**	**56.45**

2) Real uncertain samples: Apart from manually annotated samples, we also select AffectNet_Auto as a training set that has nature uncertain samples for cross-dataset validation, which is rarely considered by previous studies. The automatic labeling algorithm published in the official document has an accuracy of 65% [7]. As shown in Table IV, ULC-AG performs the best when facing real uncertain samples, and the performance growth exceeds that in the synthetic uncertainty experiment. One possible explanation is that the uncertainty in real data is more caused by the insignificant inter-class difference. Our weighted regularization mitigates ambiguities from imbalanced categories, and the semantic information introduced by the auxiliary AU graph further conducts effective label correction.

E. Visualization

1) Target branch: Fig. 5 depicts the visualization of the confidence estimation in the target branch on two examples in RAF-DB and AffectNet datasets. The proposed ULC-AG

TABLE IV
PERFORMANCE OF ULC-AG ON DATASETS WITH REAL UNCERTAIN SAMPLES.

Method	Uncertainty	AffectNet_Auto
Baseline	Real	53.23
SCN [8]	Real	55.43
ULC-AG	Real	**57.37**

can successfully perform the label correction on synthetic uncertain labels and adaptively update sample confidence. In particular, in the second line of Fig. 5, ULC-AG not only accurately identified the synthetic label but also corrected the originally uncertain sample.

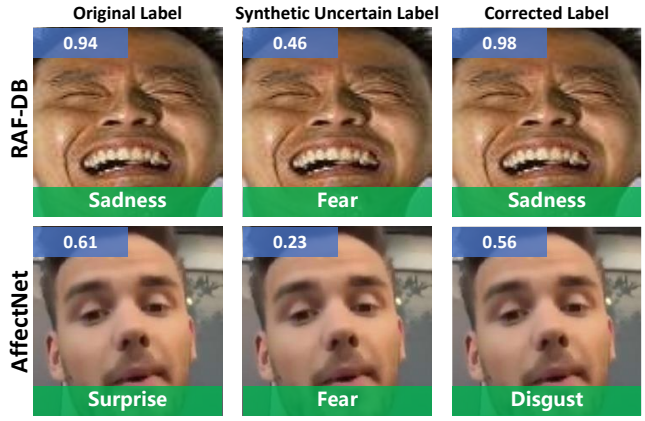

Fig. 5. Visualization of target branch. The blue block in the top left corner denotes the confidence value α, and the green block at the bottom presents the label of the current sample. Zoom in for better view.

2) Auxiliary Branch: To further analyze the actual effect of the semantic preserving relabeling in the auxiliary branch, we visualize the key intermediate values in the auxiliary branch on two examples in RAF-DB and AffectNet datasets. In addition, subjective annotations from twelve volunteers are presented to evaluate the label correction strategy. As shown in Fig. 6, the semantic feature extracted by ULC-AG can increase the inter-class distance, and the predicted emotion categories are similar in distribution to manual annotations. It reveals that the auxiliary branch can effectively handle uncertain samples to improve the final FER performance.

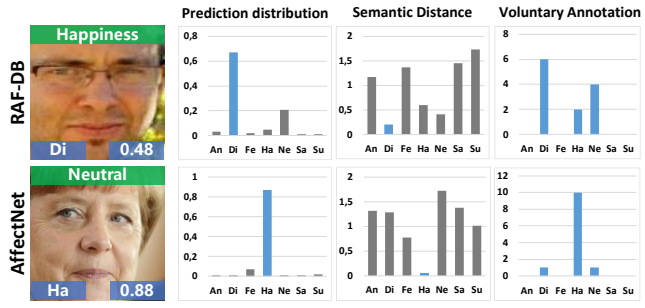

Fig. 6. Visualization of auxiliary branch. The green block at the top indicates the synthetic uncertain label, the blue block at the bottom left presents the label after correction, and the blue block at the bottom right is the confidence value α after correction. Zoom in for better view.

F. Comparison with the State-of-the-art

Table V shows the performance comparisons with the state-of-the-art approaches and Fig. 7 presents the confusion matrices of ULC-AG. To summarize, our method obtains competitive results on both RAF-DB and AffectNet datasets. Although LDL-ALSG [31] and SEIIL [20] introduce different auxiliary tasks to train the network, they only consider the label-level distribution and cannot repair the uncertain samples. In addition, IPA2LT [31], SCN [8], and WSND [11] explicitly deal with ambiguous images, but the label uncertainties can still mislead feature learning without extra knowledge in side-space and cause performance limitations. Benefiting from the confidence estimation, the data-driven AU graph, and the feature-level constrained label correction, the proposed ULC-AG outperforms all the comparison methods, including NMA [13] that is similar but lacks effective facial representation.

TABLE V
COMPARISONS WITH THE STATE-OF-THE-ART METHODS. $*$ MEANS RAF-DB AND AFFECTNET ARE USED FOR TRAINING TOGETHER. † INDICATES THE HANDLING OF AMBIGUOUS LABELS IS INTRODUCED. ‡ DENOTES EXTRA KNOWLEDGE OF AUXILIARY TASKS IS CONSIDERED.

Method	Year	RAF-DB	AffectNet
IPA2LT†[31]	2018	86.77	55.11*
SCN† [8]	2020	88.14*	60.23
RAN [30]	2020	86.90	59.50
LDL-ALSG‡ [14]	2020	85.53	59.35
SPWFA-SE [32]	2020	86.31	59.23
SEIIL‡ [20]	2021	88.23	/
WSND† [11]	2021	88.89	60.04
IDFL [33]	2021	86.96	59.20
NMA†‡ [13]	2021	76.10	46.08
ULC-AG†‡	Ours	**89.31**	**61.57**

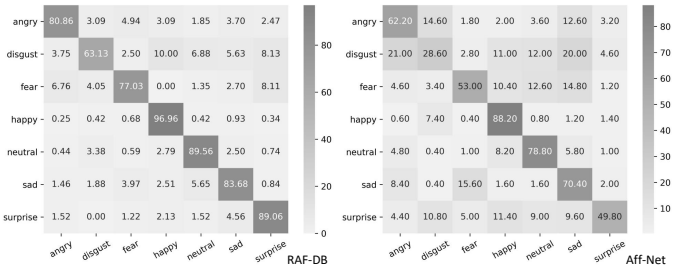

Fig. 7. Confusion matrices of ULC-AG. Zoom in for better view.

IV. CONCLUSION

In this paper, we proposed the ULC-AG framework to alleviate the uncertainties in facial expression images. The weighted regularization helped the model identify ambiguous images and balance categories. The relabeling strategy with semantic preserving corrected the suspicious labels through the auxiliary AU graph. Experiments on two large-scale datasets showed that ULC-AG achieved superior results and was robust to uncertain labels. In the future, other auxiliary tasks such as landmark detection and intensity estimation can be considered, and ULC-AG can be extended to generate annotations for unlabeled data and make multi-task predictions.

REFERENCES

[1] X. Liu, L. Jin, X. Han, J. Lu, J. You, and L. Kong, "Identity-aware facial expression recognition in compressed video," in *25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 7508–7514.

[2] L. Pham, T. H. Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in *25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4513–4519.

[3] J. Zhang, Z. Su, and L. Liu, "Median pixel difference convolutional network for robust face recognition," in *32en British Machine Vision Conference (BMVC)*, 2021. [Online]. Available: https://www.bmvc2021-virtualconference.com/conference/papers/paper_0145.html

[4] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Proceedings of the 2010 IEEE Computer Society Sonference on Computer Vision and Pattern Recognition-workshops (CVPR)*. IEEE, 2010, pp. 94–101.

[5] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. PietikäInen, "Facial expression recognition from near-infrared videos," *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, 2011.

[6] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[7] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.

[8] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 6897–6906.

[9] T. Song, L. Chen, W. Zheng, and Q. Ji, "Uncertain graph neural networks for facial action unit detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 1, 2021.

[10] J. She, Y. Hu, H. Shi, J. Wang, Q. Shen, and T. Mei, "Dive into ambiguity: Latent distribution mining and pairwise uncertainty estimation for facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6248–6257.

[11] F. Zhang, M. Xu, and C. Xu, "Weakly-supervised facial expression recognition in the wild with noisy data," *IEEE Transactions on Multimedia*, 2021. [Online]. Available: https://doi.org/10.1109/TMM.2021.3072786

[12] Y. Chen and J. Joo, "Understanding and mitigating annotation bias in facial expression recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 980–14 991.

[13] S. Zhang, Z. Huang, D. P. Paudel, and L. Van Gool, "Facial emotion recognition with noisy multi-task annotations," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 21–31.

[14] S. Chen, J. Wang, Y. Chen, Z. Shi, X. Geng, and Y. Rui, "Label distribution learning on auxiliary label space graphs for facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 13 984–13 993.

[15] J.-X. Zhong, N. Li, W. Kong, S. Liu, T. H. Li, and G. Li, "Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1237–1246.

[16] Z. Cui, Y. Zhang, and Q. Ji, "Label error correction and generation through label relationships," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 34, no. 04, 2020, pp. 3693–3700.

[17] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[18] W. Hu, Y. Huang, F. Zhang, and R. Li, "Noise-tolerant paradigm for training face recognition cnns," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 887–11 896.

[19] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 212–220.

[20] Y. Li, Y. Gao, B. Chen, Z. Zhang, G. Lu, and D. Zhang, "Self-supervised exclusive-inclusive interactive learning for multi-label facial expression recognition in the wild," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3103782

[21] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, p. 5, 1978.

[22] Y. Liu, X. Zhang, Y. Lin, and H. Wang, "Facial expression recognition via deep action units graph network based on psychological mechanism," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 2, pp. 311–322, 2019.

[23] Y. Liu, X. Zhang, J. Zhou, X. Li, Y. Li, G. Zhao, and Y. Li, "Graph-based facial affect analysis: A review of methods, applications and challenges," *arXiv preprint arXiv:2103.15599*, 2021.

[24] Y. Liu, X. Zhang, J. Zhou, and L. Fu, "Sg-dsn: A semantic graph-based dual-stream network for facial expression recognition," *Neurocomputing*, vol. 462, pp. 320–330, 2021.

[25] L. Lei, T. Chen, S. Li, and J. Li, "Micro-expression recognition based on facial graph representation learning and facial action unit fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1571–1580.

[26] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[27] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2018, pp. 59–66.

[28] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[29] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *Proceedings of the European conference on computer vision (ECCV)*. Springer, 2016, pp. 87–102.

[30] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4057–4069, 2020.

[31] J. Zeng, S. Shan, and X. Chen, "Facial expression recognition with inconsistently annotated datasets," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 222–237.

[32] Y. Li, G. Lu, J. Li, Z. Zhang, and D. Zhang, "Facial expression recognition in the wild using multi-level features and attention mechanisms," *IEEE Transactions on Affective Computing*, 2020. [Online]. Available: https://doi.org/10.1109/TAFFC.2020.3031602

[33] Y. Li, Y. Lu, B. Chen, Z. Zhang, J. Li, G. Lu, and D. Zhang, "Learning informative and discriminative features for facial expression recognition in the wild," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2021.3103760