

Supplementary Materials

Learning with Alignments: Tackling the Inter- and Intra-domain Shifts for Cross-multidomain Facial Expression Recognition

Author(s) Names(s): Anonymous ACM MM Submission

1. Implementation Details of The Global and Local Branches

The global branch consists of a series of convolutional layers and Convolutional Block Attention Modules (CBAMs) [1], while the local branch only consists of a series of convolutional layers. Concretely, the global module expands the dimension of the features from 128 to 1024 after the shared encoder; the local module crops the features from the shared encoder into four regions, and expands the dimensions of these regions from 128 to 256, finally merging them. The final model codes and algorithm codes will be released publicly once the paper is accepted.

2. Impact of Selections of Reliability Threshold ϵ

In our inter-domain alignment, we select only those samples with probabilities exceeding the reliability threshold ϵ which balances the positive knowledge transfer and the number of aligned samples. We conduct the threshold selection experiments on the ‘ \rightarrow CK+’ and ‘ \rightarrow FER-2013’ tasks and the results are tabulated in Tab.S.1. As seen, our LA-CMFER achieves a relatively robust performance under different reliability thresholds and when ϵ is set as 0.4, the best accuracies on the two tasks are obtained. Therefore, we set ϵ as 0.4 in all our experiments.

Table S.1: Experimental results about different selections of reliability threshold ϵ on the ‘ \rightarrow CK+’ and ‘ \rightarrow FER-2013’ tasks.

	Threshold ϵ			
	0.0	0.2	0.4	0.6
Accuracy (%) on ‘ \rightarrow CK+’	89.71	90.19	90.48	90.29
Accuracy (%) on ‘ \rightarrow FER-2013’	56.78	57.26	57.40	56.84

3. Impact of the Number of Source Domains

We also study how the number of source domains impacts the performance. Here, we progressively add the number of source domains on the ‘ \rightarrow CK+’ and ‘ \rightarrow RAF-DB’ tasks, and the classification accuracies are displayed in Tab. S.2. As observed, as the number of source domains increases, the domain shifts become increasingly complex. Even under such challenging circumstances, our proposed LA-CMFER can effectively align the data

Table. S.2: Analytical experiment of impacts on different number of source domains.

Tasks	Accuracy (%)	Tasks	Accuracy (%)
A \rightarrow C	81.71	A \rightarrow R	70.27
A, R \rightarrow C	86.10	A, J \rightarrow R	71.00
A, R, J \rightarrow C	86.48	A, J, F \rightarrow R	74.86
A, R, J, F \rightarrow C	88.00	A, J, F, O \rightarrow R	75.72
A, R, J, F, O \rightarrow C	90.48	A, J, F, O, C \rightarrow R	77.86

distributions among several sources and deeply mine their beneficial expression features, thus finally gaining gradually enhanced accuracies on both ‘ \rightarrow CK+’ and ‘ \rightarrow RAF-DB’ tasks.

4. Quantitative Analysis of Different Cross-view Consistency Constraints

To intuitively evaluate the effectiveness of different cross-view consistency Constraints, we give the quantitative results of different cross-view consistency constraints in Table.S.3. As depicted in Table.S.3, the model’s performance declines to varying degrees when using these alternative strategies instead of our \mathcal{L}_{mcc} . Concretely, while KL and L1 exhibit comparable performance in simple ‘ \rightarrow C’ and ‘ \rightarrow R’ tasks, they show significant performance deterioration in more complex tasks like ‘ \rightarrow A’ and ‘ \rightarrow F’ tasks. This underscores the unreliability of solely considering absolute prediction distances or differences in relative entropy to address intra-domain shifts across FER domains. Additionally, due to substantial intra-domain shifts, the global and local branches may exhibit significant prediction biases for challenging samples near decision boundaries, leading to notable performance fluctuations with the MSE loss which is more sensitive to outliers. In summary, with our multi-view clustering technique, our model can better foster consistency between two branches, thus decreasing the prediction uncertainty and promoting intra-domain alignments.

Table S.3: Quantitative results (%) of different cross-view consistency constraints.

Variants	\rightarrow J	\rightarrow R	\rightarrow C	\rightarrow O	\rightarrow A	\rightarrow F	Avg
(A) KL	69.95	77.30	89.33	62.15	47.76	48.82	65.89
(B) L1	69.01	76.67	86.95	64.96	45.91	45.70	64.87
(C) MSE	51.64	61.04	77.62	55.22	31.50	33.44	51.74
(D) \mathcal{L}_{mcc} (Ours)	70.42	77.86	90.48	66.50	53.26	57.40	69.32

5. Analysis of Hyper-parameter Sensitivity Tests

We conduct further analysis for hyperparameters sensitivity tests of α , β , and γ . Concretely, the three hyperparameters control the importance of the dual-level inter-domain alignment, the cross-view consistency constraint, and the multi-view voting loss, respectively. We select candidate values for α , β , γ from the set {0.1, 0.2, 0.4, 0.5, 0.8, 1} and conduct hyperparameters sensitivity tests on the relatively simple ‘ \rightarrow CK+’ task and the more challenging ‘ \rightarrow FER-2013’ task. Based on the observations from Fig. 6 in the manuscript, we summarize our findings as follows: (1) The model exhibits robustness to all hyperparameters in the ‘ \rightarrow CK+’ task, while performance shows a decreasing trend with larger values of α and γ in the more complex ‘ \rightarrow FER-2013’ task; (2) In

complex ‘ \rightarrow FER-2013’ task which exits more severe inter-domain shifts, excessively large inter-domain alignment hyperparameter settings may lead the model to over-align on erroneous source-target domain features and learn misleading ‘discriminative’ features inevitably that may damage the model’s discriminative ability; (3) For the challenging ‘ \rightarrow FER-2013’ task, excessive reliance on supervision for unlabeled target samples may lead to learning from noisy labels, thus degrading performance; (4) Finally, to optimize performance, we set α as 0.4 for the hard dataset FER-2013 and AffectNet while 0.1 for the rest. Meanwhile, we set β and γ as 0.5 and 0.1, respectively.

REFERENCES

- [1] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. In Computer Vision – ECCV 2018, Lecture Notes in Computer Science. 3–19. DOI:https://doi.org/10.1007/978-3-030-01234-2_1.