Supplementary Material: Supervised Contrastive Learning with Identity-Label Embeddings for Facial Action Unit Recognition

Tangzheng Lian
lian.tangzheng@kcl.ac.uk
David Adama
David.adama@ntu.ac.uk
Pedro Machado
pedro.machado@ntu.ac.uk
Doratha Vinkemeier
doratha.vinkemeier@ntu.ac.uk

Department of Computer Science
Computing and Informatics Research Centre
Nottingham Trent University
Nottingham, UK

1 Experiments on λ, λ_1, λ_2, and λ_3

1.1 Evaluation of λ

We evaluated the hyperparameter λ present in Eq. (3) specifically on the BP4D dataset. For this investigation, while retaining all other parameters constant, we systematically varied the value of λ from 0.2 to 0.8 in increments of 0.2. As the results in Table 5 delineate, a λ value of 0.4 produces the most optimal outcome in terms of the average F1 score across all 12 AUs in BP4D.

λ	0.2	0.4	0.6	0.8
F1 (Avg.)	63.6	**64.4**	64.2	63.9

Table 5: Evaluation results of hyperparameter λ on the BP4D dataset. The best performance is highlighted in bold.

1.2 Evaluation on λ_1, λ_2, and λ_3

Similarly, we conducted evaluations on λ_1, λ_2, and λ_3 in Eq. (6) on the BP4D dataset. Ensuring the consistency of our analysis, all other parameters were kept invariant throughout the evaluation. The baseline for this set of evaluations was chosen with λ_1, λ_2, and λ_3 values set at 1, 0.5, and 0.5 respectively. Despite time constraints restricting us to 11 distinct experimental setups, our results, as outlined in Table 6, indicate that the most effective combination for these hyperparameters stands at 1, 0.5, and 0.4.

© 2023. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

Experiments	λ_1	λ_2	λ_3	F1 (Avg.)
E_0	1	0.5	0.5	64.1
E_1	0.9	-	-	63.6
E_2	0.8	-	-	63.2
E_3	-	0.7	-	63.6
E_4	-	0.6	-	63.4
E_5	-	0.4	-	63.8
E_6	-	0.3	-	63.1
E_7	-	-	0.7	63.7
E_8	-	-	0.6	64.1
E_9	-	-	0.4	**64.6**
E_{10}	-	-	0.3	64.3

Table 6: Evaluation results of hyperparameters λ_1, λ_2, and λ_3 on the BP4D dataset. The "-" symbol signifies the retention of the original value from the baseline experiment. The best performance is highlighted in bold.

2 Calculations of the correlation matrix

Given a matrix F with dimensions $n \times m$, where n represents the number of frames and m denotes the total number of AUs, each entry $F_{i,j}$ indicates the presence (1) or absence (0) of the j-th AU in the i-th frame. Recognizing the scarcity of positive samples for certain AUs, it is crucial to assess dependencies encompassing both positive and negative samples, thus mitigating potential imbalances. To elaborate on these dependencies, we introduce the conditional probabilities as represented by Eq. (7) and (8), which epitomize co-occurrence and co-absence respectively:

$$P_{occ}(j=1|i=1) = \frac{\sum_{k=1}^{n} I(F_{k,i}=1 \wedge F_{k,j}=1)}{\sum_{k=1}^{n} F_{k,i}} \tag{7}$$

$$P_{abs}(j=0|i=0) = \frac{\sum_{k=1}^{n} I(F_{k,i}=0 \wedge F_{k,j}=0)}{n-\sum_{k=1}^{n} F_{k,i}} \tag{8}$$

Interpreting from Eq. (7), a conditional probability $P_{occ}(j=1|i=1)=0.5$ conveys that when the i-th AU is activated, the probability of the j-th AU's occurrence is on par with its non-occurrence. Such a scenario alludes to the inference that the activation of the i-th AU doesn't significantly inform the activation tendencies of the j-th AU. The same applies to when $P_{occ}(j=0|i=0)=0.5$. Therefore, to further delineate this interrelationship, we introduce the metric $P_{\text{correlation}}(i,j)$ as defined in Eq. (9). Here, variations from the neutral point of 0.5 reflect the intensity of the inter-AU correlation. By emphasizing absolute values, we ensure that the metric is non-negative, allowing for representation in heatmap visualizations, with values ranging from 0 (indicating a neutral or non-correlational relationship) to 1 (signifying a potent relationship, whether through co-occurrence or co-absence).

$$P_{\text{correlation}}(i,j) = |P_{occ}(j=1|i=1)-0.5| + |P_{abs}(j=0|i=0)-0.5| \tag{9}$$