Breaking Spurious Correlations: Uncertainty-Driven Causal Transformers for AU Detection

Supplementary Material

1. Overview of Uncertainty-Driven Causal Transformers framework for AU Detection

Algorithm 1 provides a detailed description of the proposed uncertainty-aware AU detection framework. The algorithm outlines the full training workflow, including feature extraction, AU-specific feature generation, uncertainty-aware Transformer reasoning, causal intervention, and uncertainty-weighted optimization. Specifically, the model first extracts spatial facial representations and derives AU-specific feature vectors through the AFG module. These features are then processed by the Uncertainty-Aware Transformer (UAT), which models attention distributions probabilistically to capture AU dependencies under uncertainty. To mitigate spurious correlations, each AU representation is further refined by the causal deconfounding module (CD). Finally, the quantified uncertainty is used to re-weight the loss for robust optimization under imbalanced and noisy annotations. The inference stage follows the same forward pipeline without uncertainty-based reweighting. This algorithmic overview complements the main paper by presenting the complete procedural flow of the proposed method.

Algorithm 1 Uncertainty-Aware Action Unit Detection

Input: Training dataset $\{(X_s, y_s)\}$
Output: Predicted AU probabilities for all AUs: $P_1, P_2, \cdots, P_N$
for each training sample (X_s, y_s) **do**
  Feature Extraction:
    $F \in \mathbb{R}^{C\times H\times W} \leftarrow \text{Backbone}(X_s)$
  AU-Specific Feature Generation:
    $\{V_1, V_2, \ldots, V_N\} \leftarrow \text{AFG}(F)$
  Uncertainty Modeling:
    $\{f_1, f_2, \ldots, f_N\} \leftarrow \text{UAT}(V_1, V_2, \ldots, V_N)$
  Causal Deconfounding:
  for $j = 1$ **to** N **do**
    $\hat{f}_s^j \leftarrow \text{CD}(f_s^j)$ ▷ Apply causal deconfounding to each AU feature
  end for
  Uncertainty-Weighted Optimization:
    Compute uncertainty weight ω using Eq. (13)
    Update model parameters: Θ
end for
Inference:
Generate final AU predictions: $P_1, P_2, \cdots, P_N$

Table 1. Detailed ablation results on DISFA (F1-score, %), where B denotes the backbone-only baseline.

AU	B	B+DT	B+UAT	B+UAT+CD
AU1	44.8	57.05	64.83	71.14
AU2	36.9	50.25	52.05	54.81
AU4	64.5	75.44	73.64	74.69
AU6	46.6	53.57	48.93	53.81
AU9	42.0	49.98	51.09	61.51
AU12	73.9	72.42	73.07	71.76
AU25	91.4	89.95	91.67	92.25
AU26	47.3	55.09	59.97	58.91
Avg	55.9	62.97	64.41	67.36

2. Additional Ablation Studies

In the main paper, we report the ablation results on DISFA in terms of the average F1-score. In this supplementary material, we provide a more detailed breakdown of the ablation study, including per-AU results on DISFA and the full ablation results on BP4D.

Specifically, we compare four model variants: Backbone, Backbone + DT, Backbone + UAT, and Backbone + UAT + CD. Here, DT denotes the standard deterministic Transformer, UAT denotes our uncertainty-aware probabilistic Transformer, and CD denotes the proposed causal deconfounding module. All results are reported in F1-score (%).

2.1. Detailed Ablation on DISFA

Table 1 presents the detailed ablation results on DISFA for each AU. These results complement the average F1-score comparison reported in the main paper by showing how each component affects individual AU recognition performance.

The detailed AU-wise results show that the gains are particularly noticeable on several AUs, such as AU1, AU2, and AU9.

2.2. Ablation on BP4D

Table 2 reports the full ablation results on BP4D, including the F1-score of each AU and the overall average. These results provide additional evidence that the effectiveness of the proposed components is consistent across datasets.

On BP4D, UAT improves the average F1-score from 61.71 to 62.45, and the full model further improves it to 62.59 after adding CD module.

Table 2. Ablation results on BP4D (F1-score, %), where B denotes the backbone-only baseline.

Model	B	B+DT	B+UAT	B+UAT+CD
AU1	47.3	52.04	49.83	49.84
AU2	42.7	44.40	44.48	44.23
AU4	50.0	53.68	55.98	56.34
AU6	75.9	77.50	78.14	78.47
AU7	75.0	74.18	76.77	76.85
AU10	81.7	83.38	83.15	83.11
AU12	87.4	87.15	87.56	87.56
AU14	62.5	64.93	63.76	64.11
AU15	43.9	49.00	49.05	49.34
AU17	62.5	61.55	63.98	64.01
AU23	44.3	45.41	47.65	47.39
AU24	45.9	47.32	48.99	49.82
Avg	59.9	61.71	62.45	62.59

These supplementary ablation results further support the effectiveness of the proposed design. They show that UAT consistently improves AU relation modeling over the deterministic Transformer, while CD module provides additional gains by alleviating spurious correlations.