

A SUPPLEMENTARY MATERIAL: CEFDET: COGNITIVE EFFECTIVENESS NETWORK BASED ON FUZZY INFERENCE FOR ACTION DETECTION

A.1 Evaluation indicator

The evaluation metric in this study is mAP , which is defined by the following formula:

$$AP = \sum_{k=1}^K \max_{\hat{k} \geq k} P(\hat{k}) \Delta r(k), \quad (1)$$

$$mAP = \frac{\sum_{i=1}^j AP_i}{j}. \quad (2)$$

where AP is the area under the precision and recall curves, k is the threshold index, $\max_{\hat{k} \geq k} P(\hat{k})$ represents the value with the highest precision among the k -th point and all subsequent points, $\Delta r(k)$ denotes the change in recall from $k - 1$ to k points, mAP refers to the average of all categories of AP , and j is the number of action categories.

A.2 Qualitative analysis

To further explore the effectiveness of Cefdet in updating the detection results of cognitive abnormalities, the experimental results of Cefdet on the JHMDB and UCF101-24 datasets are visualized. The results are presented in Figures 1 and 2.

To evaluate the effectiveness of FCM in locating detection results with cognitive abnormalities, two sets of data are employed to demonstrate the feasibility of Cefdet. Figure 1 illustrates the capabilities of locating frames with highly similar actions and frames that do not conform to human action norms. Figure 1 (a) shows the ability to locate frames with highly similar actions on the JHMDB dataset. The detector misclassifies “pick” as “run” in complex scenarios. It is difficult to accurately locate the position of frames with cognitive abnormalities using only confidence. However, Cefdet combines human action features with a fuzzy system to simulate a cognition-based detection process and obtain effective detection results. Video frames are categorized into frames with high and low-level cognition based on their effectiveness. Cefdet effectively locates the positions of frames with abnormally high-similarity actions using the effectiveness.

Figure 1 (b) shows the capability of locating frames that do not conform to human action norms on the UCF101-24 dataset. It is unreasonable for a single frame to represent a complete action as a continuous human action. The adjacent frames of “clip” are “wave,” which does not conform to human action norms. The effect of using confidence alone to locate such frames is not significant. However, Cefdet combines the correlation between continuous actions with a fuzzy system to annotate the possibility of continuous actions and effectively locates the positions of frames that do not conform to human action norms.

To assess the feasibility of FCS in repairing detection results with cognitive abnormalities, two sets of data are used to validate the conclusion. Figure 2 shows the results of repairing frames with high-similarity actions and frames that do not conform to human

action norms. Figure 2 (a) shows the results of repairing frames with high-similarity actions on the JHMDB dataset. HIT mistakenly classifies “pick” as “run,” which not only causes errors in the detection of the current frame, but also interferes with the results of subsequent detection. However, Cefdet effectively aggregates the features of frames with high-level cognition to re-detect frames with low-level cognition. This reduces the noise and interference from misjudgments, thereby improving the accuracy of high-similarity actions.

Figure 2 (b) shows the results of repairing frames that do not conform to human action norms on the UCF101-24 dataset. One frame does not occur as a complete action according to human cognition, and HIT does not have constraints on frames that do not conform to human action norms. Therefore, a constraint rule is required to limit noncompliant human cognitive behaviors. Cefdet fully utilizes the correlation of actions in human cognition to constrain detection results and effectively repairs behaviors that do not conform to human action norms.

The experimental results further prove the feasibility of the FCM and FCS modules, while demonstrating the superiority of Cefdet.

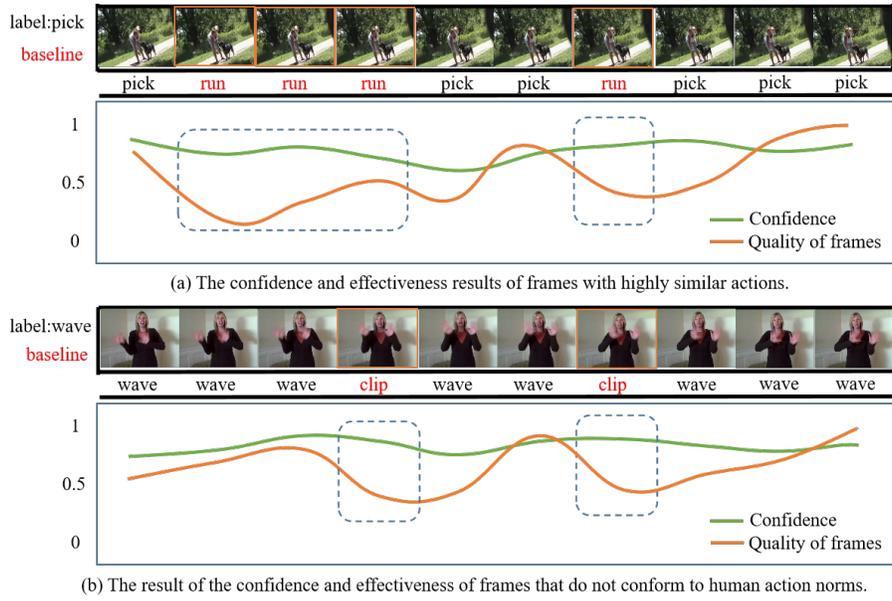


Figure 1: Comparison of the quality and confidence of Cefdet and HIT on public datasets. (a) denotes that Cefdet demonstrates more distinguishable results in frames with highly similar actions. (b) represents that Cefdet has superior differentiation results in frames that do not conform to human action norms.

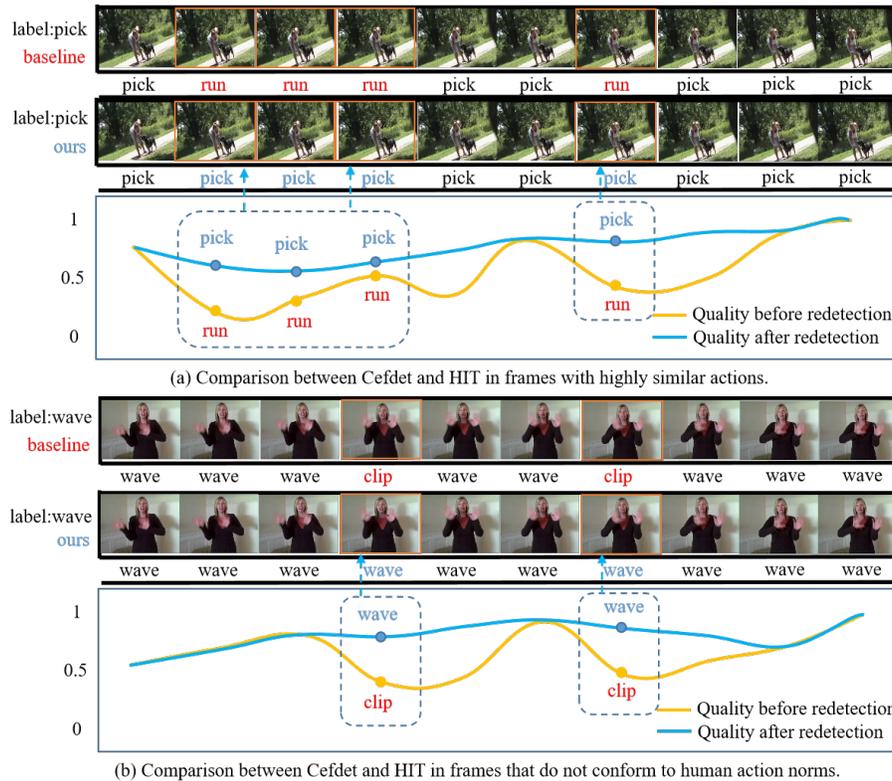


Figure 2: Comparison of the quality of Cefdet and HIT before and after re-detection. (a) indicates that Cefdet can effectively correct the detection results of HIT in frames with highly similar actions. (b) demonstrates the ability of Cefdet to sensitively capture and correct frames that do not conform human action norms.