

SF-Net

(1) Actionness module $A \in \mathbb{R}^{N \times T}$

每一帧有无动作的概率

$A^l \in \mathbb{R}^K$ of labeled frames

$$L_{\text{actionness}} = -\frac{1}{K} \sum \log \sigma(A^l) - \frac{1}{\eta K} \sum \log(1 - \sigma(A^b))$$

(2) $A^b \in \mathbb{R}^{n \times K}$ of ~~pseudo~~ pseudo background frames

$$L_{\text{video}} = -\frac{1}{TV} \sum_{i=1}^V \sum_{j=1}^{N_c} q_{ij} \log \frac{\exp(r_{ij})}{\sum_{N_c+1} \exp(r_{ij})}$$

(2) Classification module $C \in \mathbb{R}^{V \times T \times (N_c+1)}$

* 检测每一帧所属的动作类别

Pseudo Label Mining and Training Objective

① Action classification at Labeled Frames.

$$L_{\text{frame}}^l = -\frac{1}{K} \sum_i y_i \log p_i^l \quad p_i^l \in \mathbb{R}^{K \times N_c+1}$$

p_i^l denote the prediction for i th labeled action frame

Pseudo labeling of Frames 标记帧的

① Action frame mining: 何例扩展 若得分大于 δy_i

则记为 Action frame

② Background frame mining

将得分低为前 ηK 的记为伪帧

$$L_{\text{frame}}^b = -\frac{1}{\eta K} \sum \log p_i^b$$

$$L_{\text{frame}} = L_{\text{frame}}^l + \frac{1}{N_c} L_{\text{frame}}^b$$

