Supplementary material
Improving Weakly Supervised Temporal Action Localization by Bridging Train-Test Gap in Pseudo Labels

Jingqiu Zhou [1] Linjiang Huang[1,2 *] Liang Wang[4] Si Liu[5] Hongsheng Li[1,2,3]

[1]CUHK-SenseTime Joint Laboratory, The Chinese University of Hong Kong

[2]Centre for Perceptual and Interactive Intelligence, Hong Kong

[3]Xidian Uniersity

[4]Institute of Automation Chinese Academy of Science

[5]Beihang University

1155167063@link.cuhk.edu.hk, ljhuang524@gmail.com, hsli@ee.cuhk.edu.hk

1. Experiments

1.1. Implementation Details

Ablation on Different Distributions. In our ablation study, we discussed the performance of using different distributions in our Gaussian weighted instance fusion module. Here we detailly describe the implementation of different distributions. **For simplicity, we use the notation ▲ to represent the action boundary start point s, action boundary end point e, and action score q.**

Gaussian Distribution. As our defaulting template fusion distribution, is sampled from the Gaussian distribution of the form:

$$N(\blacktriangle) = \frac{1}{\sigma_\blacktriangle \sqrt{2\pi}} \exp(-\frac{(\blacktriangle - \mu_\blacktriangle)^2}{\sigma_\blacktriangle^2}), \tag{1}$$

Next we regard the confidence scores as the unnormalized logits of action instances, then we can compute the probability of sampling the k-th action instance as:

$$g_k = \frac{\exp{(q_k/T)}}{\sum_{i \in \mathcal{I}_*} \exp{(q_i/T)}}, \tag{2}$$

We fit the template distribution to the sampled distribution $\{g_k\}$ by minimizing the cross-entropy loss between them. That is:

$$\begin{aligned} loss &= \sum_{k=1}^{M} -g_k \text{log}(\frac{1}{\sigma_\blacktriangle \sqrt{2\pi}} \exp(-\frac{(\blacktriangle_k - \mu_\blacktriangle)^2}{\sigma_\blacktriangle^2})) \\ &= \frac{1}{\sigma_\blacktriangle^2} \sum_{k=1}^{M} g_k (\blacktriangle_k - \mu_\blacktriangle)^2 + \sum_{k=1}^{M} g_k \log(\frac{1}{\sigma_\blacktriangle \sqrt{2\pi}}) \\ &= \frac{1}{\sigma_\blacktriangle^2} (\sum_{k=1}^{M} g_k (\blacktriangle_k - \mu_\blacktriangle)^2) + \log(\frac{1}{\sigma_\blacktriangle \sqrt{2\pi}}) \end{aligned} \tag{3}$$

Where M is the number of action instances. Now we take directive of $loss$ with respect to $\mu_\blacktriangle$ and let the derivative equal to zero, we find the optimal fusion formulas as:

$$\mu_\blacktriangle = \sum_{k=1} M g_k \blacktriangle_k \tag{4}$$

Equation (4) is our fusion formulas for the Gaussian weighted instance fusion.

Uniform Distribution. In this paragraph, we study the fusion formulas when uniform distribution is used as the template distribution. Assuming we are sampling the action instances from a uniform distribution:

$$U(\blacktriangle) = \frac{1}{b - a} \tag{5}$$

where $\blacktriangle \in [a, b]$ is the definition field of this uniform distribution. For this distribution to be well defined, we must have $a \geq \min \blacktriangle_k$ and $b \leq \max \blacktriangle_k$.

Then we can write the formulas (10) as:

$$\begin{aligned} loss &= \sum_{k=1}^{M} -g_k \text{log}(\frac{1}{b - a}) \\ &= \log(b - a) \\ &\geq \log(\max \blacktriangle_k - \min \blacktriangle_k) \end{aligned} \tag{6}$$

For the equality to hold we have $a = \min \blacktriangle_k$ and $b = \max \blacktriangle_k$, and then the fused average ▲ is:

$$\mu_\blacktriangle = \frac{1}{2}(\min \blacktriangle_k + \max \blacktriangle_k) \tag{7}$$

Exponential Distribution. In this paragraph, we consider using the exponential distribution as the template distortion. However, this distribution is not defined on the whole

*Corresponding author.

real number domain. It requires the random variable to be greater than 0. To address this issue, we assume $\blacktriangle - \min \blacktriangle_k$ satisfy the exponential distribution, which is:

$$E(\blacktriangle) = \lambda_{\blacktriangle} \exp(-\lambda_{\blacktriangle}(\blacktriangle - \min \blacktriangle_k)), \tag{8}$$

Accordingly, the sampling probability g_k is computed as:

$$g_k = \frac{\exp\left((q_k - \min q_j)/T\right)}{\sum_{i \in \mathcal{I}_*} \exp\left((q_k - \min q_j)/T\right)}, \tag{9}$$

According to (8), we can rewrite (10) as:

$$\begin{aligned} loss &= \sum_{k=1}^{M} -g_k \log(\lambda_{\blacktriangle} \exp(-\lambda_{\blacktriangle}(\blacktriangle_k - \min \blacktriangle_j)) \quad (10) \\ &= \sum_{k=1}^{M} g_k(\lambda_{\blacktriangle}(\blacktriangle_k - \min \blacktriangle_j)) - \sum_{k=1}^{M} g_k \log(\lambda_{\blacktriangle}) \\ &= \sum_{k=1}^{M} g_k(\lambda_{\blacktriangle}(\blacktriangle_k - \min \blacktriangle_j)) - \log(\lambda_{\blacktriangle}) \end{aligned}$$

Taking the derivative of $loss$ with respect to $\lambda_{\blacktriangle}$ and let the derivative be 0, we get:

$$\lambda_{\blacktriangle} = \frac{1}{\sum_{k=1}^{M} g_k(\blacktriangle_k - \min \blacktriangle_j)} \tag{11}$$

Then the fusion average of $\blacktriangle$ is:

$$\begin{aligned} \mu_{\blacktriangle} &= \frac{1}{\lambda_{\blacktriangle}} + \min \blacktriangle_j \quad (12) \\ &= \sum_{k=1}^{M} g_k \blacktriangle_k \end{aligned}$$

Note that (12) is different from (4) because the sampling probability g_k is not the same.

T-Distribution. The T-distribution is a very important and common distribution in statistics. It takes the form:

$$F(t) = \frac{1}{\sqrt{v} B(\frac{1}{2}, \frac{v}{2})} (1 + \frac{t^2}{v})^{\frac{-v-1}{2}} \tag{13}$$

where v is the parameter we are to extimate and B is the beta function. For this distribution, the analytic fusion form does not exists, however, we can estimate it's fused average through Newton iteration of solving it's extreme value. This process is quite time consuming because it requires iterations.