Expanding Object Detector's HORIZON: Incremental Learning Framework for Object Detection in Videos (supplementary materials)

Alina Kuznetsova[1,3], Sung Ju Hwang[2], Bodo Rosenhahn[1], and Leonid Sigal[3]

[1]Leibniz University Hannover , [2]UNIST , [3]Disney Research Pittsburgh

1. Probabilistic LME formulation

The probabilistic interpretation of the similarity $d_{\mathbf{W}}(\mathbf{x}, \mathbf{u}_c)$ between a sample $\mathbf{x}$ and a prototype $\mathbf{u}_c$ in embedding space is given by the equation:

$$p(y = c|\mathbf{x}, d = 1) = \frac{e^{d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_c)/2\sigma^2}}{\sum_{i=1}^{C} e^{d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_i)/2\sigma^2}}. \tag{1}$$

and the large margin constraint is

$$d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_{y_i}) + \xi_{ic} \geq d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_c) + 1, \tag{2}$$
$$i = \{1 \dots N\}, c = \{1 \dots C\}, c \neq y_i,$$

The ratio between $p(y = y_i|\mathbf{x_i}, d = 1)$ and $p(y = c|\mathbf{x_i}, d = 1), c \neq y_i$ can be then rewritten as:

$$\frac{p(y = y_i|\mathbf{x_i}, d = 1)}{p(y = c|\mathbf{x_i}, d = 1)} = \frac{e^{d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_{y_i})/2\sigma^2}}{e^{d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_c)/2\sigma^2}} = \tag{3}$$
$$= e^{d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_{y_i})/2\sigma^2 - d_{\mathbf{W}}(\mathbf{x},\mathbf{u}_c)/2\sigma^2} \tag{4}$$

while the large margin constraint is equivalent to

$$d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_{y_i}) + \xi_{ic} \geq d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_c) + 1 \Leftrightarrow \tag{5}$$
$$d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_{y_i}) - d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{u}_c) \geq 1 - \xi_{ic} \Leftrightarrow \tag{6}$$
$$e^{d_{\mathbf{W}}(\mathbf{x}_i,\mathbf{u}_{y_i}) - d_{\mathbf{W}}(\mathbf{x}_i,\mathbf{u}_c)} \geq e^{1-\xi_{ic}} \approx e - \tilde{\xi}_{ic}, \tag{7}$$

in the neighbourhood of 0 for ξ_{ic}

Therefore, large margin constraint for similarity measure is equivalent to the large margin constraint on the probability measure, given in the following form:

$$\frac{p(y = y_i|\mathbf{x_i}, d = 1)}{p(y = c|\mathbf{x_i}, d = 1)} \geq e - \tilde{\xi}_{ic} \tag{8}$$

Equivalently, given the probabilistic interpretation

$$p(d = 1|\mathbf{x}) = \frac{1}{1 + e^{a d^m_{\mathbf{W}}(\mathbf{x}) + b}} \tag{9}$$
$$d^m_{\mathbf{W}}(\mathbf{x}) = \max_c d_{\mathbf{W}}(\mathbf{x}, \mathbf{u}_c), \tag{10}$$

where $a < 0, b$ parameters, that have the following interpretation: $-\frac{1}{a}$ is the standard deviation and $\frac{b}{a}$ is the mean of the logistic distribution.

Detection constraints

$$d_{\mathbf{W}}(\mathbf{x}^0_j, \mathbf{u}_c) \leq 1 + \xi^0_j, c = \{1, \dots, C\} \Leftrightarrow \tag{11}$$
$$d^m_{\mathbf{W}}(\mathbf{x}^0_j) \leq 1 + \xi^0_j, \tag{12}$$

can be reformulated as:

$$a d^m_{\mathbf{W}}(\mathbf{x}) + b \geq a + a\xi j^0 + b \Leftrightarrow \tag{13}$$
$$e^{a d^m_{\mathbf{W}}(\mathbf{x}) + b} \geq e^{a + a\xi j^0 + b} \Leftrightarrow \tag{14}$$
$$p(d = 1|\mathbf{x}) = \frac{1}{1 + e^{a d^m_{\mathbf{W}}(\mathbf{x}) + b}} \leq \frac{1}{1 + e^{a + a\xi^0_j + b}} \approx \tag{15}$$
$$\approx \frac{1}{1 + e^{a+b}} + \tilde{\xi_j}^0 \tag{16}$$

in the neighbourhood of 0 for ξ^0_j

which states that for *non-object* samples its probability of being an object $p(d = 1|\mathbf{x}) \geq \frac{1}{1+e^{a+b}}$ is pushed towards $\frac{1}{1+e^{a+b}}$ as $\xi^0_j \to 0$. In practice, we set $a = -8$ and $b = 12$ in all experiments, so $\frac{1}{1+e^{a+b}} \ll \frac{1}{2}$, which can be seen as a margin in *object* - *non-object* case.

The minimization functional is then formulated as following:

$$\sum_{i,c:c \neq y_i} \max(\tilde{\xi_{ic}}, 0) + \sum_j \max(\tilde{\xi_j}^0, 0) + \tag{17}$$
$$+ \frac{1}{2}\lambda\|\mathbf{W}\|^2_{FRO} + \frac{1}{2}\gamma\|\mathbf{U}\|^2_{FRO}, \tag{18}$$

2. Full multi-prototype LME formulation

The optimization problem for learning multiple prototypes at once can be formulated as:

$$
\begin{aligned}
&\text{minimize:} \\
&\quad \sum_{i,c:c\neq y_i} \max(\xi_{ic},0) + \sum_{j} \max(\xi_j^0,0) + \qquad (19) \\
&\qquad + \frac{1}{2}\lambda\|\mathbf{W}\|_{FRO}^2 + \frac{1}{2}\gamma\|\mathbf{U}\|_{FRO}^2 \\
&\text{subject to:} \\
&\quad S_{\mathbf{W}}^{\alpha}(\mathbf{x}_i,\mathbf{U}_{y_i}) + \xi_{ic} \geq S_{\mathbf{W}}^{\alpha}(\mathbf{x}_i,\mathbf{U}_c) + 1, \\
&\qquad i=\{1,\ldots,N\}, \quad c=\{1,\ldots,C\}, \quad c\neq y_i, \\
&\quad S_{\mathbf{W}}^{\alpha}(\mathbf{x}_j^0,\mathbf{U}_c) \leq 1 + \xi_j^0, \\
&\qquad c=\{1,\ldots,C\}, \quad j=\{1,\ldots,N_j\},
\end{aligned}
$$

where the classification can be done for example:

$$
y^* = \begin{cases} \operatorname{argmax}_c S_{\mathbf{W}}^{\alpha}(\mathbf{x}^*,\mathbf{U}_c), & S_{\mathbf{W}}^{\alpha}(\mathbf{x}^*,\mathbf{U}_c) \geq \tau, \\ c_0, \quad \forall c=1,\ldots,C: & S_{\mathbf{W}}^{\alpha}(\mathbf{x}^*,\mathbf{U}_c) < \tau, \end{cases} \qquad (20)
$$