Open Set Action Recognition via Multi-Label Evidential Learning Supplementary Materials

Chen Zhao, Dawei Du, Anthony Hoogs, Christopher Funk
Kitware Inc.
{chen.zhao, dawei.du, anthony.hoogs, christopher.funk}@kitware.com

1. Notations

Vectors are denoted by lower case bold face letters, *e.g.,* positive and negative evidence $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^k$. Vectors with subscripts of indices, such as α_i, β_i indicate the i-th entry in $\boldsymbol{\alpha}, \boldsymbol{\beta}$. The Euclidean ℓ_2-norm is denoted as $\|\cdot\|$. Scalars are denoted by lowercase italic letters, *e.g.,* $\eta > 0$. Matrices are denoted by capital italic letters, *e.g.,* $X \in \mathbb{R}^{C\times H\times W}$. Proposition $[u]_+$ denotes the projection of $[u]$ on the non-negative orthant.

Notations	Descriptions
N	Number of detected persons
M	Number of detected objects
$\boldsymbol{\theta}$	Parameters of the backbone to calculate ACO-R
$\mathbf{x}$	Input video
X	Context feature
C, H, W	Channel, height and width
A, O	Actor (person) and object feature
F	Actor-context-object relation feature
H	Actor relational feature
ω	Subjective opinion
b, d	Belief and disbelief masses
u	Uncertainty mass
W	Non-informative prior weight
a	Base rate
K	Total number of classes
α, β	Positive and negative evidence
p	Class probability
γ	The independence criterion relaxation of M-EDC
m	primal-dual updating step

Table 1. Important notations and corresponding descriptions.

2. Proof Sketch of Proposition 1

Proposition 1. *(Convergence of Averaged Primal Sequence) Under Assumption 1, when the convex set Θ is compact, let the approximate primal sequence $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ be the running averages of the primal iterates given in Equation (6). Then $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ can converge to its limit $\tilde{\boldsymbol{\theta}}^*$.*

To prove the convergence in Proposition 1, we first prove the below Lemma 1 that

Lemma 1. *The approximate primal sequence $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ given in Equation (6) is a Cauchy sequence. That is $\forall \epsilon > 0$, there is a $Q \in \mathbb{N}$ such that $||\tilde{\boldsymbol{\theta}}^{(m')} - \tilde{\boldsymbol{\theta}}^{(m)}|| \leq \epsilon$, $\forall m', m \geq Q$.*

Proof. Given Equation (6), we derive

$$\begin{aligned}
\tilde{\boldsymbol{\theta}}^{(m+1)} &= \frac{1}{m+1}\sum_{i=1}^{m}\boldsymbol{\theta}^{(i)} \\
&= \frac{1}{m+1}\big(\boldsymbol{\theta}^{(m)} + \sum_{i=1}^{m-1}\boldsymbol{\theta}^{(i)}\big) \\
&= \frac{1}{m+1}\boldsymbol{\theta}^{(m)} + \frac{m}{m+1}\cdot\frac{1}{m}\sum_{i=1}^{m-1}\boldsymbol{\theta}^{(i)} \\
&= \frac{1}{m+1}\boldsymbol{\theta}^{(m)} + \frac{m}{m+1}\tilde{\boldsymbol{\theta}}^{(m)} \\
&= \frac{1}{m+1}\boldsymbol{\theta}^{(m)} + \tilde{\boldsymbol{\theta}}^{(m)} - \frac{1}{m+1}\tilde{\boldsymbol{\theta}}^{(m)}.
\end{aligned} \tag{9}$$

Rearrange the above equation, we have

$$\tilde{\boldsymbol{\theta}}^{(m+1)} - \tilde{\boldsymbol{\theta}}^{(m)} = \frac{1}{m+1}(\boldsymbol{\theta}^{(m)} - \tilde{\boldsymbol{\theta}}^{(m)}). \tag{10}$$

Under Assumption 1, Θ is a compact convex set and $\tilde{\boldsymbol{\theta}}^{(m)}, \boldsymbol{\theta}^{(m)} \in \Theta$. Let $m' > m$ and $||\tilde{\boldsymbol{\theta}}^{(m)}||, ||\boldsymbol{\theta}^{(m)}|| \leq G$, we have

$$\begin{aligned}
||\tilde{\boldsymbol{\theta}}^{(m')} - \tilde{\boldsymbol{\theta}}^{(m)}|| &= ||\tilde{\boldsymbol{\theta}}^{(m')} - \tilde{\boldsymbol{\theta}}^{(m'-1)} + \cdots + \tilde{\boldsymbol{\theta}}^{(m+1)} - \tilde{\boldsymbol{\theta}}^{(m)}|| \\
&= ||\frac{\boldsymbol{\theta}^{(m'-1)} - \tilde{\boldsymbol{\theta}}^{(m'-1)}}{m'} + \cdots + \frac{\boldsymbol{\theta}^{(m)} - \tilde{\boldsymbol{\theta}}^{(m)}}{m+1}|| \\
&\leq \frac{||\boldsymbol{\theta}^{(m'-1)}|| + ||\tilde{\boldsymbol{\theta}}^{(m'-1)}||}{m'} + \cdots + \frac{||\boldsymbol{\theta}^{(m)}|| + ||\tilde{\boldsymbol{\theta}}^{(m)}||}{m+1} \\
&\leq \frac{2G(m'-m)}{m+1}.
\end{aligned} \tag{11}$$

Therefore, for any arbitrary $\epsilon > 0$, let $\frac{2G(m'-m)}{m+1} < \epsilon$, and we have $||\tilde{\boldsymbol{\theta}}^{(m')} - \tilde{\boldsymbol{\theta}}^{(m)}|| \leq \epsilon$. Therefore we conclude that $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ is a Cauchy sequence. $\square$

Next we prove the proposed Proposition 1.

Proof. As stated in Lemma 1 that $\{\tilde{\boldsymbol{\theta}}^{(m)}\}$ in Equation (6) is a Cauchy sequence, it hence is bounded and there exists a subsequence b_n converging to its limit L. For any $\epsilon > 0$, there exists $n, q \geq Q$ satisfying $||\tilde{\boldsymbol{\theta}}^{(n)} - \tilde{\boldsymbol{\theta}}^{(q)}|| < \frac{\epsilon}{2}$. Thus, there is a $b_m = \tilde{\boldsymbol{\theta}}^{(q_m)}$, such that $q_m \geq Q$ and $||b_{q_m} - L|| < \frac{\epsilon}{2}$.

$$\begin{aligned} ||\tilde{\boldsymbol{\theta}}^{(n)} - L|| &= ||\tilde{\boldsymbol{\theta}}^{(n)} - b_m + b_m - L|| \\ &\leq ||\tilde{\boldsymbol{\theta}}^{(n)} - b_m|| + ||b_m - L|| \\ &< ||\tilde{\boldsymbol{\theta}}^{(n)} - \tilde{\boldsymbol{\theta}}^{(q)}|| + \frac{\epsilon}{2} \\ &< \epsilon. \end{aligned} \tag{12}$$

Since ϵ is arbitrarily small, we prove that the sequence $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ converges to its limit $L = \tilde{\boldsymbol{\theta}}^{*}$ asymptotically. □

3. Proof Sketch of Proposition 2

Proposition 2. *(Bounds for $\mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)})$ and the violation of $g(\tilde{\boldsymbol{\theta}}^{(m)})$ [5]) Let the dual sequence $\{\lambda^{(m)}\}_{m=1}^{\infty}$ be generated through Equation (8) and $\{\tilde{\boldsymbol{\theta}}^{(m)}\}_{m=1}^{\infty}$ be the averages in Equation (6). Under Assumption 1, we have*

1. *An upper bound on the amount of constraint violation of $\tilde{\boldsymbol{\theta}}^{(m)}$ that $\left\|\left[g(\tilde{\boldsymbol{\theta}}^{(m)})\right]_+\right\| \leq \frac{\lambda^{(m)}}{m\eta_2}$.*
2. *An upper bound on $\mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)})$ that $\mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) \leq f^* + \frac{(\lambda^{(0)})^2}{2m\eta_2} + \frac{\eta_2 L^2}{2}$, where $\left\|g(\tilde{\boldsymbol{\theta}}^{(m)})\right\| < L$ and $L > 0$.*
3. *A lower bound $\mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) \geq f^* - \lambda^* \cdot \left\|\left[g(\tilde{\boldsymbol{\theta}}^{(m)})\right]_+\right\|$.*

where $[u]_+$ denotes the projection of $[u]$ on the nonnegative orthant. f^* is the optimal solution of Equation (5) and λ^* denotes the optimal value of the dual variable.

Proof. 1. According to Equation (8), we have

$$\lambda^{(m)} \geq \lambda^{(m-1)} + \eta_2 \cdot \left(g(\tilde{\boldsymbol{\theta}}^{(m)}) - \gamma - \delta\lambda^{(m-1)}\right).$$

Under Assumption 1, $g(\boldsymbol{\theta})$ is convex, we have

$$\begin{aligned} g(\tilde{\boldsymbol{\theta}}^{(m)}) &\leq \frac{1}{m}\sum_{i=1}^{m-1} g(\boldsymbol{\theta}^{(i)}) \\ &= \frac{1}{m\eta_2}\sum_{i=1}^{m-1} \eta_2 g(\boldsymbol{\theta}^{(i)}) \\ &\leq \frac{1}{m\eta_2}(\lambda^{(m)} - \lambda^{(0)}) \\ &\leq \frac{\lambda^{(m)}}{m\eta_2}, \quad \forall m \geq 1. \end{aligned} \tag{13}$$

Since $\lambda^{(m)} \geq 0$, we derive $||[g(\tilde{\boldsymbol{\theta}}^{(m)})]_+|| \leq \frac{\lambda^{(m)}}{m\eta_2}$.

2. Under Assumption 1 and Equations (7) and (8), we have $q^* = f^*$. Together with the condition that $f(\boldsymbol{\theta})$ is convex and $\boldsymbol{\theta} \in \Theta$, we have

$$\begin{aligned} \mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) &\leq \frac{1}{m}\sum_{i=0}^{m-1} \mathcal{L}(\boldsymbol{\theta}^{(i)}) \\ &= \frac{1}{m}\sum_{i=0}^{m-1}\left(\mathcal{L}(\boldsymbol{\theta}^{(i)}) + \lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)}) - \lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)})\right) \\ &= \frac{1}{m}\sum_{i=0}^{m-1}\left(\mathcal{L}(\boldsymbol{\theta}^{(i)}) + \lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)})\right) - \frac{1}{m}\sum_{i=0}^{(m-1)} \lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)}) \\ &\leq q^* - \frac{1}{m}\sum_{i=0}^{m-1} \lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)}). \end{aligned} \tag{14}$$

From Equation (8), we have

$$\begin{aligned} (\lambda^{(i+1)})^2 &= \left(\left[\lambda^{(i)} + \eta_2\left(g(\tilde{\boldsymbol{\theta}}^{(i+1)}) - \gamma - \delta\lambda^{(i)}\right)\right]_+\right)^2 \\ &\leq \left(\lambda^{(i)} + \eta_2 g(\tilde{\boldsymbol{\theta}}^{(i+1)})\right)^2 \\ &\leq (\lambda^{(i)})^2 + 2\eta_2\lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)}) + \left(\eta_2||g(\tilde{\boldsymbol{\theta}}^{(i+1)})||\right)^2 \end{aligned} \tag{15}$$

Rearrange the above equation, we have

$$-\lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)}) \leq \frac{(\lambda^{(i)})^2 - (\lambda^{(i+1)})^2 + \left(\eta_2||g(\tilde{\boldsymbol{\theta}}^{(i+1)})||\right)^2}{2\eta_2}. \tag{16}$$

Taking $-\lambda^{(i)} g(\tilde{\boldsymbol{\theta}}^{(i+1)})$ back to $\mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)})$, we have

$$\begin{aligned} \mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) &\leq q^* + \frac{1}{m}\sum_{i=0}^{m-1} \frac{(\lambda^{(i)})^2 - (\lambda^{(i+1)})^2 + \left(\eta_2||g(\tilde{\boldsymbol{\theta}}^{(i+1)})||\right)^2}{2\eta_2} \\ &= q^* + \frac{1}{m}\sum_{i=0}^{m-1} \frac{(\lambda^{(i)})^2 - (\lambda^{(i+1)})^2}{2\eta_2} + \frac{1}{m}\sum_{i=0}^{m-1} \frac{\left(\eta_2||g(\tilde{\boldsymbol{\theta}}^{(i+1)})||\right)^2}{2\eta_2} \\ &= q^* + \frac{(\lambda^{(0)})^2 - (\lambda^{(m)})^2}{2m\eta_2} + \frac{\eta_2}{2m}\sum_{i=0}^{m-1} ||g(\tilde{\boldsymbol{\theta}}^{(i+1)})||^2 \\ &\leq f^* + \frac{(\lambda^{(0)})^2}{2m\eta_2} + \frac{\eta_2 L^2}{2}. \end{aligned} \tag{17}$$

3. By definition, $\forall \boldsymbol{\theta} \in \Theta$, we have

$$\mathcal{L}(\boldsymbol{\theta}) + \lambda^* \cdot g(\boldsymbol{\theta}) \geq \mathcal{L}(\boldsymbol{\theta}^*) + \lambda^* \cdot g(\boldsymbol{\theta}^*) = q(\lambda^*). \tag{18}$$

Since $\tilde{\boldsymbol{\theta}} \in \Theta$, $\forall m \geq 1$, we have

$$\begin{aligned} \mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) &= \mathcal{L}(\tilde{\boldsymbol{\theta}}^{(m)}) + \lambda^* \cdot g(\tilde{\boldsymbol{\theta}}^{(m)}) - \lambda^* \cdot g(\tilde{\boldsymbol{\theta}}^{(m)}) \\ &\geq q(\lambda^*) - \lambda^* \cdot g(\tilde{\boldsymbol{\theta}}^{(m)}) \\ &\geq q(\lambda^*) - \lambda^* \cdot \left[g(\tilde{\boldsymbol{\theta}}^{(m)})\right]_+ \\ &\geq f^* - \lambda^* \cdot \left\|\left[g(\tilde{\boldsymbol{\theta}}^{(m)})\right]_+\right\|. \end{aligned} \tag{19}$$

□

m iterations	Method	PE / NE / PNE / Belief Error ↓	AUROC ↑	AUPR ↑	FPR at 95% TPR ↓
$m=2$	MULE	11.22 / 40.14 / 12.18 / 27.91	**86.92** / **63.54** / 85.25 / 83.41	96.18 / 89.90 / 93.46 / 90.90	4.98 / 5.09 / 4.51 / 9.15
	w/o ACO-R	50.12 / 50.04 / 12.45 / 39.84	51.23 / 51.12 / 84.88 / 82.41	86.11 / 86.36 / 88.13 / **92.35**	15.23 / 15.56 / 34.44 / 9.57
	w/o Beta-ENN	35.12 / **35.31** / 35.12 / 34.78	57.10 / 57.59 / 57.88 / 57.01	85.71 / 85.65 / 85.66 / 85.66	12.03 / 13.00 / 14.98 / 15.32
	w/o M-EDC	13.16 / 46.32 / 12.16 / 38.23	86.00 / 50.13 / 85.12 / 83.05	95.12 / 90.67 / 92.45 / 89.19	6.31 / 5.05 / 5.01 / 9.59
$m=3$	MULE	11.01 / 45.05 / 11.17 / 27.86	86.81 / 58.23 / 85.25 / 85.18	98.48 / 95.65 / **99.41** / 90.44	4.58 / **5.00** / 3.98 / 9.09
	w/o ACO-R	48.34 / 50.32 / 17.77 / 47.79	52.78 / 50.15 / 83.34 / 79.26	86.39 / 83.39 / 93.93 / 88.93	13.35 / 15.01 / 25.93 / 10.00
	w/o Beta-ENN	32.64 / 45.12 / 50.14 / 21.71	61.41 / 55.23 / 80.12 / 83.23	83.67 / 83.39 / 84.03 / 88.66	12.21 / 15.23 / 12.24 / 22.22
	w/o M-EDC	**10.12** / 48.12 / 11.86 / 38.43	80.12 / 52.01 / 84.38 / 73.94	93.34 / 93.41 / 98.92 / 89.33	5.15 / 6.03 / 7.41 / 10.70
$m=4$	MULE	11.28 / 47.12 / 11.17 / 27.52	86.56 / 59.01 / 85.13 / 84.72	99.43 / 88.98 / 99.40 / 90.42	4.64 / 5.02 / 3.21 / 9.09
	w/o ACO-R	50.21 / 51.71 / 17.64 / 49.35	52.35 / 50.01 / 84.17 / 80.33	83.36 / 83.49 / 94.97 / 88.21	11.02 / 12.65 / 33.33 / 10.23
	w/o Beta-ENN	42.04 / 50.15 / 48.14 / 29.21	83.01 / 51.36 / **89.33** / 59.63	87.21 / 93.39 / 84.29 / 86.48	10.01 / 12.12 / 14.00 / 9.70
	w/o M-EDC	13.35 / 48.01 / **10.86** / 34.28	82.51 / 50.21 / 85.51 / 78.26	93.34 / 93.93 / 98.94 / 89.17	5.69 / 6.16 / 3.73 / 9.62
$m=5$	MULE	10.22 / 44.32 / 11.15 / 27.49	86.86 / 58.87 / 85.30 / 84.66	**99.52** / **96.35** / **99.41** / 90.41	**4.32** / **5.00** / **3.09** / 9.09
	w/o ACO-R	49.61 / 50.00 / 16.61 / 48.82	51.01 / 51.97 / 84.78 / 81.29	82.16 / 83.33 / 98.05 / 88.09	10.23 / 12.68 / 37.04 / 9.16
	w/o Beta-ENN	41.34 / 49.67 / 49.21 / **10.75**	84.83 / 52.41 / 82.91 / **90.44**	86.13 / 93.40 / 87.27 / 90.29	9.97 / 11.98 / 12.68 / 13.01
	w/o M-EDC	11.09 / 48.01 / **10.86** / 32.14	79.34 / 50.98 / 85.12 / 75.23	93.33 / 94.02 / 98.94 / 91.14	5.39 / 5.79 / 3.70 / **7.11**

Table 2. Exploration of different component in MULE with $m = 2, 3, 4, 5$.

Backbone	Pre-train	PE / NE / PNE / Belief Error ↓	AUROC ↑	AUPR ↑	FPR at 95% TPR ↓	Closed Set mAP ↑
ACAR, R-50 [6]	K400	35.16 / 35.56 / 23.12 / 30.15	52.89 / 58.48 / 60.16 / 57.17	79.15 / 80.17 / 83.48 / 81.48	14.16 / 15.49 / 20.19 / 11.15	28.84
ACAR, R-101 [6]	K700	32.26 / **34.12** / 20.54 / 30.16	55.18 / 58.98 / 63.18 / 60.18	82.15 / 81.48 / 85.48 / 90.01	10.16 / 15.01 / 19.48 / 10.00	**33.30**
AIA, R-101 [9]	K700	35.14 / 34.56 / 25.01 / 32.14	54.17 / 59.48 / 59.78 / 59.69	78.49 / 80.17 / 86.15 / **91.48**	10.15 / 13.98 / 18.48 / 9.68	32.30
Slowfast, R-101 [3]	K600	60.12 / 59.12 / 23.15 / 40.15	50.15 / 52.23 / 69.14 / 68.15	70.15 / 75.15 / 69.68 / 68.21	20.17 / 19.56 / 23.12 / 26.15	29.00
DEAR, R-50 [1]	K400	23.22 / 42.15 / 23.15 / 30.19	82.12 / 60.12 / 80.48 / 83.59	83.15 / 88.14 / 90.15 / 85.49	8.45 / 8.48 / 6.30 / 13.15	18.51
AFAC, R-101 [?]	K600	53.14 / 49.41 / 40.15 / 59.17	79.69 / **65.89** / 80.48 / 79.79	90.79 / 89.18 / 88.18 / 85.15	7.15 / 8.79 / **4.01** / **8.98**	30.20
Ours, R-50	K400	11.22 / 40.14 / 12.18 / 27.91	86.92 / 63.54 / 85.25 / 83.41	96.18 / 89.90 / 93.46 / 90.90	4.98 / **5.09** / 4.51 / 9.15	27.80
Ours, R-101	K700	**10.12** / 40.15 / **10.56** / **25.02**	**88.75** / 65.36 / **89.48** / **84.26**	**98.18** / **89.95** / **94.74** / 90.49	**4.17** / 5.28 / 4.25 / 10.01	29.87

Table 3. Comparison with state-of-the-art on AVA [4]. Ours is highlighted in green. Best value is in **bold**.

4. Proof Sketch of Proposition 3

The Beta loss in Equation (2) is equivalent to the loss function used in DEAR when each actor is associated with one action only. The Evidential Neural Network (ENN), which was initially introduced in [7] and further adopted by DEAR in open-set action recognition, is limited to classifying an actor associated with only one action. The key idea of ENN is to replace the output of a classification network with the parameters $\{\alpha_i\}_{i=1}^{K}$ of K Dirichlet densities. To detect a novel actor with multiple actions, in this work, we modify ENN by estimating parameter pairs $\{\alpha_i, \beta_i\}_{i=1}^{K}$ of K Beta distributions. Proposition 3 shows that the Beta loss of an actor proposed in Equation (3) is equivalent to the loss function in DEAR when $K = 1$.

Proposition 3. *Denote $\mathcal{L}'_j(\boldsymbol{\theta})$ as the loss function introduced in [7],* i.e.,

$$\mathcal{L}'_j(\boldsymbol{\theta}) = \sum_{i=1}^{K'} y_{ij}\Big(\psi(\sum_{i=1}^{K'}\alpha_{ij}) - \phi(\alpha_{ij})\Big), \tag{20}$$

where K' is the total number of classes in a multi-class classification task. Denote $\mathcal{L}_j(\boldsymbol{\theta})$ is the loss function proposed in Equation (3). We have $\mathcal{L}'_j(\boldsymbol{\theta}) = \mathcal{L}_j(\boldsymbol{\theta})$ when $K = 1$ (i.e., *$K' = 2$).*

Proof. When $K = 1$,

$$\mathcal{L}_j(\boldsymbol{\theta}) = \sum\nolimits_{i=1}^{1} \int \Big[\mathbf{BCE}(y_{ij}, p_{ij})\Big]\mathbf{Beta}(p_{ij}; \alpha_{ij}, \beta_{ij})dp_{ij}. \tag{21}$$

To simplify, we omit the subscript i and rewrite

$$\begin{aligned}\mathcal{L}_j(\boldsymbol{\theta}) &= \int \Big[\mathbf{BCE}(y_j, p_j)\Big]\mathbf{Beta}(p_j; \alpha_j, \beta_j)dp_j \\ &= y_j\Big(\psi(\alpha_j + \beta_j) - \psi(\alpha_j)\Big) + (1 - y_j)\Big(\psi(\alpha_j + \beta_j) - \psi(\beta_j)\Big).\end{aligned} \tag{22}$$

As for $\mathcal{L}'_j(\boldsymbol{\theta})$, $K = 1$ indicates binary classification which refers $K' = 2$.

$$\begin{aligned}\mathcal{L}'_j(\boldsymbol{\theta}) &= \int \Big[\mathbf{CE}(y_j, p_j)\Big]\mathbf{Dir}(p_j; \alpha_j, \beta_j)dp_j \\ &= \sum_{i=1}^{2} y_{ij}\Big(\psi(\sum_{i=1}^{2}\alpha_{ij}) - \psi(\alpha_{ij})\Big).\end{aligned}$$

Backbone	Pre-train	PE / NE / PNE / Belief Error ↓	AUROC ↑	AUPR ↑	FPR at 95% TPR ↓	Closed Set mAP ↑
AFAC, R-101 [?]	K600	7.00 / 26.62 / 8.88 / 19.17	82.20 / 69.62 / 82.15 / 87.00	90.15 / 85.12 / 90.15 / 89.15	35.19 / 29.55 / **23.15** / 31.15	48.10
AFAC, CSN-152 [?]	IG-65M	6.18 / 19.78 / 8.49 / **15.59**	80.12 / 61.02 / **90.48** / 84.88	90.79 / 86.36 / 91.15 / 85.67	30.15 / 33.15 / 28.42 / 30.01	**50.30**
CSN, CSN-152 [10]	IG-65M	6.89 / 25.18 / 9.02 / 18.90	80.15 / 70.15 / 90.15 / 87.01	**92.61** / 90.17 / 90.99 / 85.15	35.16 / 30.15 / 35.98 / 26.69	46.40
Slowfast, R-101 [3]	K600	25.15 / 30.15 / 46.12 / 23.00	79.12 / 75.12 / 75.36 / 78.89	79.15 / 80.15 / 80.46 / 80.08	50.15 / 56.15 / 46.12 / 29.16	45.20
DEAR, R-50 [1]	K400	12.15 / 25.59 / 11.11 / 16.98	86.15 / **79.15** / 82.46 / 82.55	92.35 / 90.33 / 89.26 / 90.19	29.96 / 26.16 / 23.99 / 28.98	38.12
X3D-XL [2]	K600	8.45 / 25.85 / 10.15 / 15.51	82.15 / 66.64 / 85.00 / 82.15	86.49 / 80.16 / 90.15 / 88.15	39.98 / 36.14 / 26.15 / 30.55	47.20
Ours, R-50	K400	**6.15** / 23.15 / 7.89 / 17.11	85.49 / 65.17 / 88.79 / **88.49**	90.78 / **92.78** / 94.15 / **95.03**	25.98 / 25.29 / 25.49 / **23.64**	45.33
Ours, R-101	K700	6.23 / **22.15** / **7.02** / 20.15	**88.49** / 65.01 / 89.41 / 82.16	91.15 / 90.78 / **95.42** / 90.15	**25.15** / **21.16** / 26.46 / 28.98	47.21

Table 4. Comparison with state-of-the-art on Charades [8]. Ours is highlighted in green. Best value is in **bold**.

We complete the proof by setting $\alpha_{1j} = \alpha_j$ and $\alpha_{2j} = \beta_j$. □

5. Detailed Ablation Study

In Table 2, we provide a detailed ablation study on the AVA dataset [4] to explore the contributions of different components in our framework. Along with the primal-dual updating step $m = 2, 3, 4, 5$, the performance is slightly improved but nearly saturated when $m = 2$.

6. Full Experimental Results

As presented in Table 3 and Table 4, we report the results of compared methods by using all four novelty score estimation mechanism. It is worth mentioning that other state-of-the-arts perform also well by using the belief based score in terms of some metrics.

Furthermore, we show more visual results of compared methods in Figure 1 and Figure 2. It can be seen that our method performs better than other methods on two datasets for single/multi-actor settings.

References

[1] Wentao Bao, Qi Yu, and Yu Kong. Evidential deep learning for open set action recognition. In *ICCV*, pages 13329–13338, 2021. 3, 4

[2] Christoph Feichtenhofer. X3d: Expanding architectures for efficient video recognition. In *CVPR*, pages 203–213, 2020. 4

[3] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. Slowfast networks for video recognition. In *ICCV*, pages 6201–6210, 2019. 3, 4

[4] Chunhui Gu, Chen Sun, David A. Ross, Carl Vondrick, Caroline Pantofaru, Yeqing Li, Sudheendra Vijayanarasimhan, George Toderici, Susanna Ricco, Rahul Sukthankar, Cordelia Schmid, and Jitendra Malik. AVA: A video dataset of spatio-temporally localized atomic visual actions. In *CVPR*, pages 6047–6056, 2018. 3, 4, 5

[5] Angelia Nedic and Asuman E. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optim.*, 19(4):1757–1780, 2009. 2

[6] Junting Pan, Siyu Chen, Mike Zheng Shou, Yu Liu, Jing Shao, and Hongsheng Li. Actor-context-actor relation network for spatio-temporal action localization. In *CVPR*, pages 464–474, 2021. 3

[7] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *NeurIPS*, volume 31, 2018. 3

[8] Gunnar A. Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, and Abhinav Gupta. Hollywood in homes: Crowdsourcing data collection for activity understanding. In *ECCV*, volume 9905, pages 510–526, 2016. 4, 6

[9] Jiajun Tang, Jin Xia, Xinzhi Mu, Bo Pang, and Cewu Lu. Asynchronous interaction aggregation for action detection. In *ECCV*, pages 71–87, 2020. 3

[10] Du Tran, Heng Wang, Lorenzo Torresani, and Matt Feiszli. Video classification with channel-separated convolutional networks. In *ICCV*, pages 5552–5561, 2019. 4

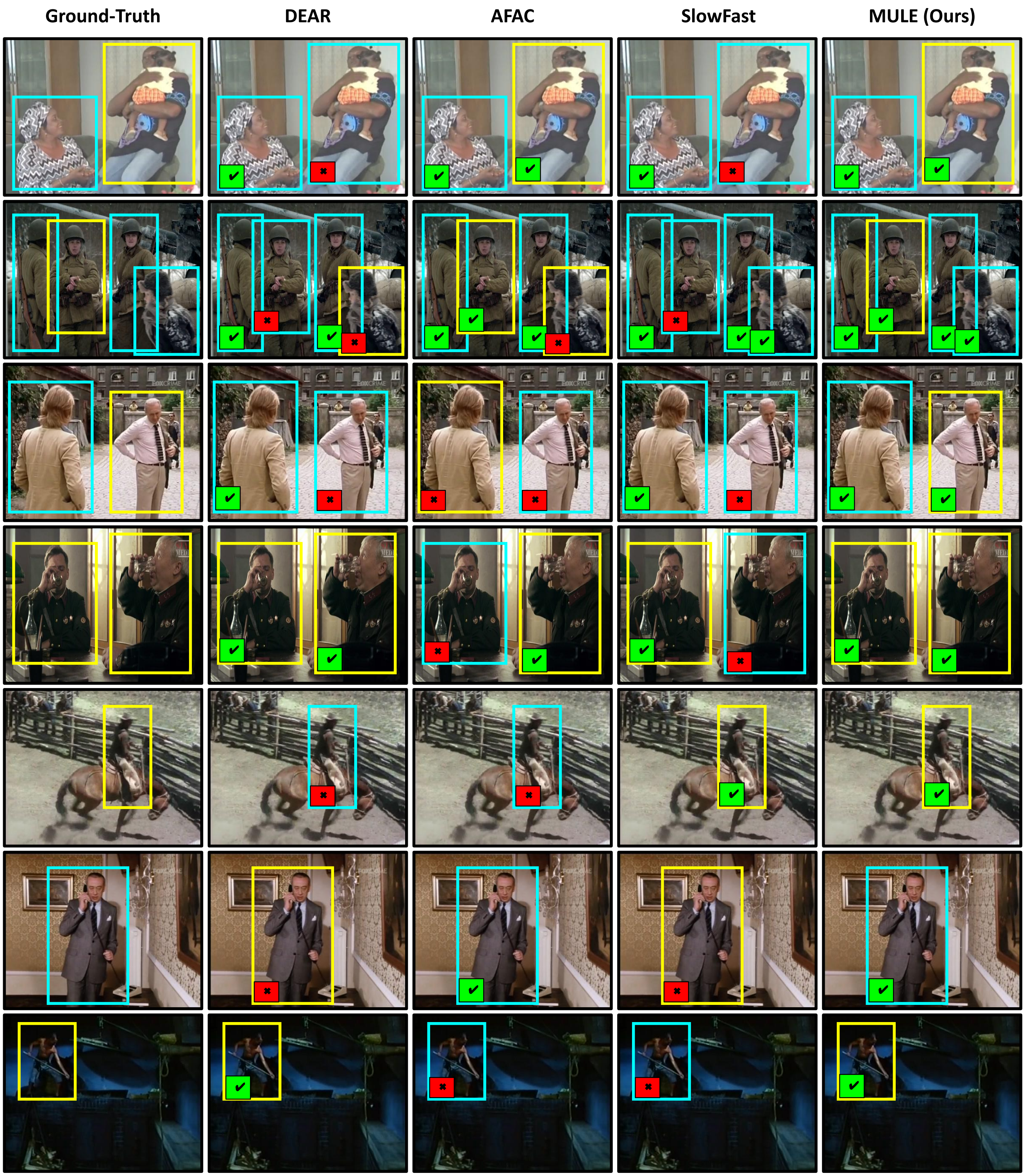

Figure 1. Visual comparison with our method and state-of-the-art on AVA [4]. Cyan and yellow boxes denote the predictions of actors with known and novel actions, respectively. ✓ marks and ✗ marks indicate correct and false predictions, respectively.

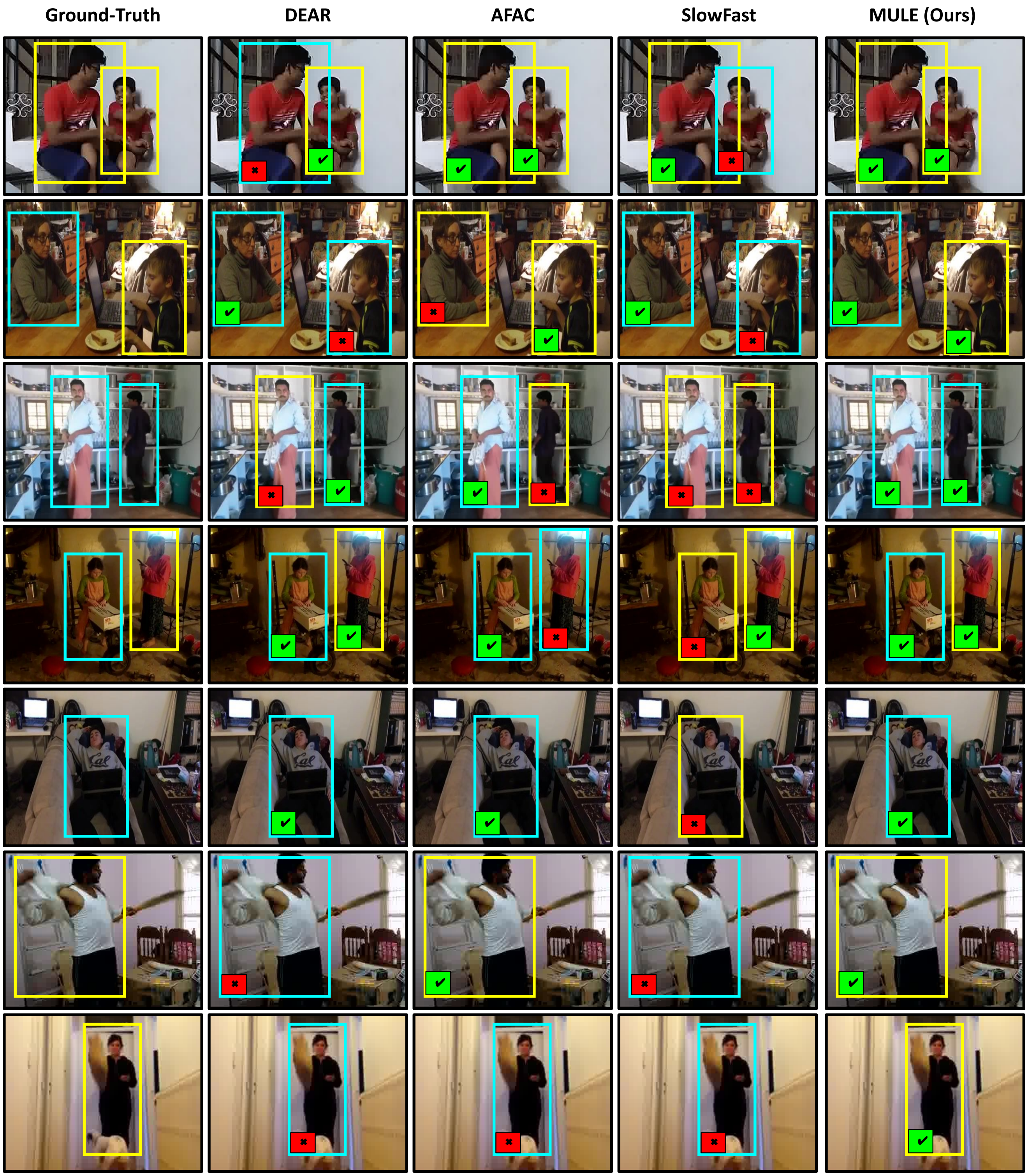

Figure 2. Visual comparison with our method and state-of-the-art on Charades [8]. Cyan and yellow boxes denote the predictions of actors with known and novel actions, respectively. ✓ marks and ✗ marks indicate correct and false predictions, respectively.