Supplementary Materials for MMAL: Multi-Modal Analytic Learning for Exemplar-Free Audio-Visual Class Incremental Tasks

Anonymous Authors

1 PROOF OF THEOREM 1

At step $k-1$, we have

$$\hat{W}_{k-1}^{av} = ((X_{0:k-1}^{av})^T X_{0:k-1}^{av} + \eta I)^{-1} (X_{0:k-1}^{av})^T Y_{0:k-1}^{train}, \tag{1}$$

Hence, at step k we have

$$\hat{W}_{k}^{av} = ((X_{0:k}^{av})^T X_{0:k}^{av} + \eta I)^{-1} (X_{0:k}^{av})^T Y_{0:k}^{train}, \tag{2}$$

In the paper, we let

$$R_{k-1}^{av} = ((X_{0:k-1}^{av})^T X_{0:k-1}^{av} + \eta I)^{-1}, \tag{3}$$

To facilitate subsequent calculations, here we let,

$$Q_{k-1}^{av} = (X_{0:k-1}^{av})^T Y_{0:k-1}^{train}, \tag{4}$$

Thus we can rewrite (1) as

$$\hat{W}_{k-1}^{av} = R_{k-1}^{av} Q_{k-1}^{av}, \tag{5}$$

Therefore, at step k we have

$$\hat{W}_{k}^{av} = R_{k}^{av} Q_{k}^{av}, \tag{6}$$

From (3), we can recursively calculate R_k^{av} from R_{k-1}^{av}, i.e.,

$$R_{k}^{av} = ((R_{k-1}^{av})^{-1} + (X_k^{av})^T X_k^{av})^{-1}, \tag{7}$$

According to the Woodbury matrix identity, we have

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(VA^{-1}U + C^{-1})VA^{-1}, \tag{8}$$

Let $A = (R_{k-1}^{av})^{-1}$, $U = (X_k^{av})^T$, $C = I$, $V = X_k^{av}$, in (7), we have

$$R_k^{av} = R_{k-1}^{av} - R_{k-1}^{av}(X_k^{av})^T (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I)^{-1} X_k^{av} R_{k-1}^{av}, \tag{9}$$

Hence, R_k^{av} can be recursively updated using its last-step counterpart R_{k-1}^{av} and data from the current step (i.e., X_k^{av}). This proves the recursive calculation of R_k^{av}.

Next, we derive the recursive calculation of $\hat{W}_k^{av}$. To this end, we first recursively calculate Q_k^{av}, i.e.,

$$Q_k^{av} = (X_{0:k}^{av})^T Y_{0:k}^{train} = Q_{k-1}^{av} + (X_k^{av})^T Y_k^{train} \tag{10}$$

Let $K_k = (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I)^{-1}$. Since

$$I = K_k K_k^{-1} = K_k (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I) \tag{11}$$

we have $K_k = I - K_k X_k^{av} R_{k-1}^{av} (X_k^{av})^T$. Therefore,

$$\begin{aligned} & R_{k-1}^{av}(X_k^{av})^T (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I)^{-1} \\ &= R_{k-1}^{av}(X_k^{av})^T K_k \\ &= R_{k-1}^{av}(X_k^{av})^T (I - K_k X_k^{av} R_{k-1}^{av} (X_k^{av})^T) \\ &= (R_{k-1}^{av} - R_{k-1}^{av}(X_k^{av})^T K_k X_k^{av} R_{k-1}^{av})(X_k^{av})^T \\ &= R_k^{av}(X_k^{av})^T, \end{aligned} \tag{12}$$

Hence, $\hat{W}_k^{av}$ can be rewritten as

$$\begin{aligned} \hat{W}_k^{av} &= R_k^{av} Q_k^{av} \\ &= R_k^{av}(Q_{k-1}^{av} + (X_k^{av})^T Y_k^{train}) \\ &= R_k^{av} Q_{k-1}^{av} + R_k^{av}(X_k^{av})^T Y_k^{train}, \end{aligned} \tag{13}$$

By substituting (9) into $R_k^{av} Q_{k-1}^{av}$, we have

$$\begin{aligned} & R_k^{av} Q_{k-1}^{av} = R_{k-1}^{av} Q_{k-1}^{av} \\ & - R_{k-1}^{av}(X_k^{av})^T (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I)^{-1} X_k^{av} R_{k-1}^{av} Q_{k-1}^{av} \\ & = \hat{W}_{k-1}^{av} - R_{k-1}^{av}(X_k^{av})^T (X_k^{av} R_{k-1}^{av} (X_k^{av})^T + I)^{-1} X_k^{av} \hat{W}_{k-1}^{av}, \end{aligned} \tag{14}$$

According to (12), (14) can be rewritten as

$$R_k^{av} Q_{k-1}^{av} = \hat{W}_{k-1}^{av} - R_k^{av}(X_k^{av})^T X_k^{av} \hat{W}_{k-1}^{av}, \tag{15}$$

By inserting (15) into (13), we have

$$\hat{W}_k^{av} = \hat{W}_{k-1}^{av} - R_k^{av}(X_k^{av})^T X_k^{av} \hat{W}_{k-1}^{av} + R_k^{av}(X_k^{av})^T Y_k^{train},$$

which proves the recursive calculation of $\hat{W}_k^{av}$.

2 PARAMETER STUDIES

In this section, we explore the impact of different settings of λ_a and λ_v. Our experimental results, which can be found in Table 1, 2 and 3 for AVE, Kinetics-Sounds, and VGGSound100 datasets respectively, demonstrate the effects of different settings of λ_a and λ_v on the performance of the model.

Table 1: Parameter studies on the AVE dataset.

λ_a	λ_v	Accuracy				Mean Acc.
		step 1	step 2	step 3	step 4	
0.1	1.0	79.81	73.33	73.97	70.56	74.42
0.3	1.0	79.81	74.76	75.56	72.08	75.55
0.5	1.0	79.81	75.71	76.19	72.33	76.01
0.8	1.0	80.77	76.19	76.20	71.57	76.18
1.0	1.0	79.81	76.67	76.68	71.21	76.09
0.8	0.9	80.77	77.14	76.83	72.08	76.71
0.8	0.7	80.77	76.19	74.60	71.07	75.66
0.8	0.5	81.73	75.71	73.33	70.30	75.27
0.8	0.3	81.73	74.76	72.38	69.03	74.48
0.8	0.1	81.73	73.81	71.43	68.53	73.88

Table 2: Parameter studies on the Kinetics-Sounds dataset.

λ_a	λ_v	Accuracy					Mean Acc.
		step 1	step 2	step 3	step 4	step 5	
0.1	1.0	94.89	82.73	75.65	69.71	67.57	78.11
0.3	1.0	95.16	83.38	76.08	70.36	68.33	78.66
0.5	1.0	95.70	82.73	76.17	71.00	69.31	78.98
0.8	1.0	95.43	82.34	75.65	71.07	69.20	78.73
1.0	1.0	95.69	81.82	75.48	70.61	68.54	78.42
0.5	0.9	95.69	82.33	76.08	70.61	69.00	78.74
0.5	0.8	95.43	81.95	75.56	70.61	68.84	78.48
0.5	0.5	94.89	81.43	75.22	69.51	66.60	77.53
0.5	0.3	95.16	80.00	73.48	68.28	65.42	76.47
0.5	0.1	94.89	78.44	71.23	65.95	62.97	74.69

Table 3: Parameter studies on the VGGSound100 dataset.

λ_a	λ_v	Accuracy										Mean Acc.
		step 1	step 2	step 3	step 4	step 5	step 6	step 7	step 8	step 9	step 10	
0.1	1.0	89.80	86.90	84.53	79.05	75.52	70.80	68.28	65.75	63.78	63.14	74.75
0.3	1.0	89.80	86.80	85.60	79.85	76.60	71.53	69.03	66.70	64.64	63.88	75.44
0.5	1.0	90.20	87.10	85.87	80.45	77.36	72.07	69.71	67.20	65.18	64.44	75.96
0.7	1.0	90.40	87.30	85.80	80.55	77.68	72.30	69.80	67.67	65.69	64.66	76.19
1.0	1.0	90.40	87.40	85.80	80.40	77.72	72.20	69.68	67.68	65.51	64.60	76.13
0.5	0.9	90.40	87.30	85.67	80.20	77.60	72.20	69.88	67.35	65.40	64.54	76.05
0.5	0.8	90.20	87.10	85.33	79.85	77.64	72.17	69.83	67.25	65.56	64.60	75.95
0.5	0.5	90.40	87.30	84.60	79.25	77.12	72.00	69.48	67.10	65.29	64.28	75.68
0.5	0.3	90.60	87.00	84.53	79.05	76.48	71.53	69.34	66.90	64.64	63.88	75.39
0.5	0.1	90.60	86.50	84.13	78.65	75.72	70.93	68.57	65.77	63.73	62.96	74.76