MAGR: Manifold-Aligned Graph Regularization for Continual Action Quality Assessment

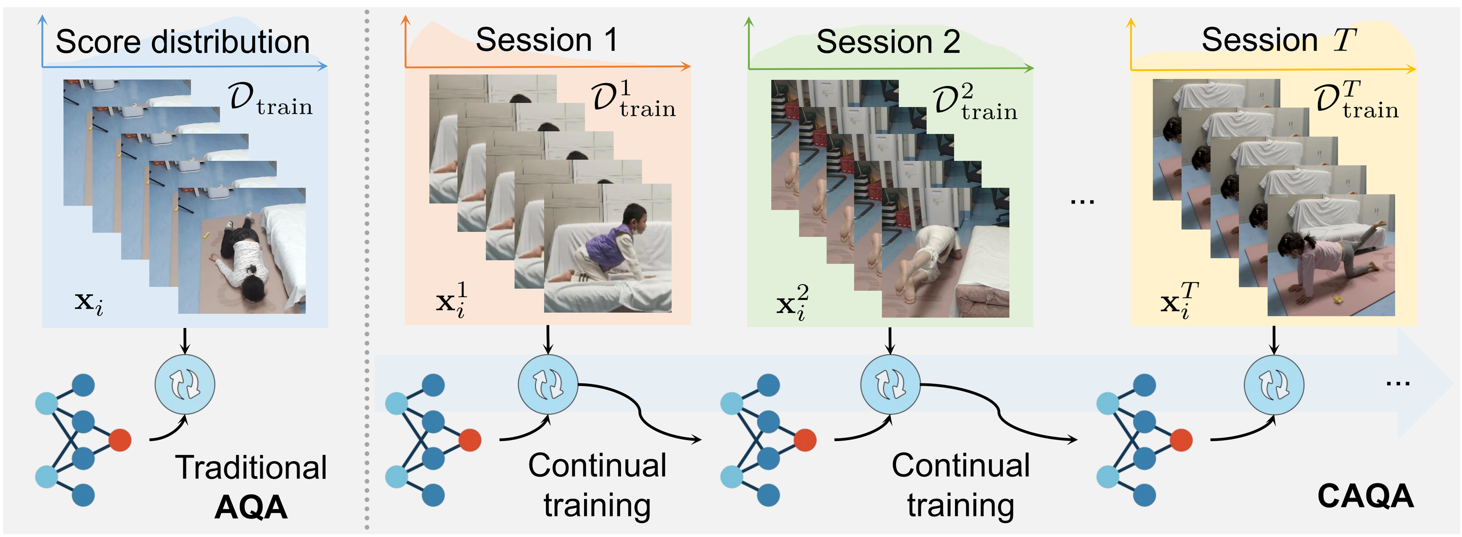
Kanglei Zhou[1,3] Liyuan Wang[2] Xingxing Zhang[2]
Hubert P. H. Shum[3] Frederick W. B. Li[3] Jianguo Li[4] Xiaohui Liang[1]

Code & Data & Model: ZhouKanglei/MAGR_CAQA

New Task, New Challenges, New Solutions, and Core Contributions

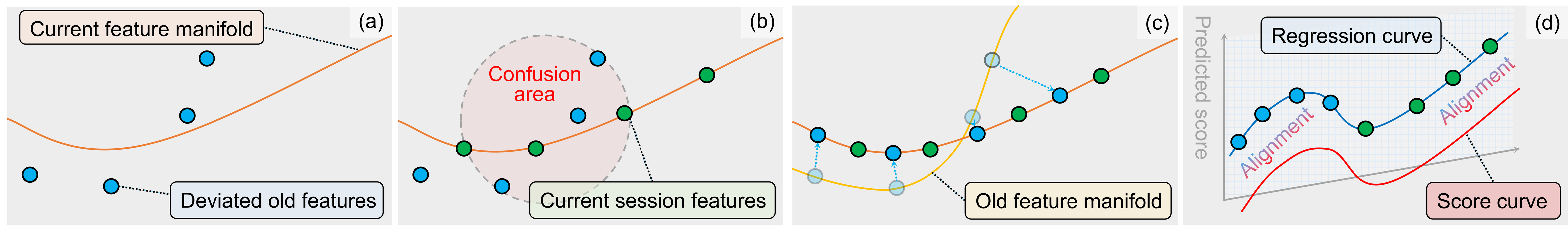
New task to address real-world issues:

We propose the novel Continual Action Quality Assessment (CAQA) to address non-stationary skill variations without exhaustive retraining.



The cause and solution for catastrophic forgetting in CAQA:

Considering privacy concerns in medical care and model adaptability for lifelong adaptation, adopting feature replay while updating backbones is critical. A key challenge in CAQA is the misalignment between static old features and the evolving feature manifold, resulting in catastrophic forgetting (see (a) and (b)). To tackle the misalignment, MAGR adopts a two-step alignment process by dynamically translating deviated features (see (c)) and aligning the feature distribution (see (d)).



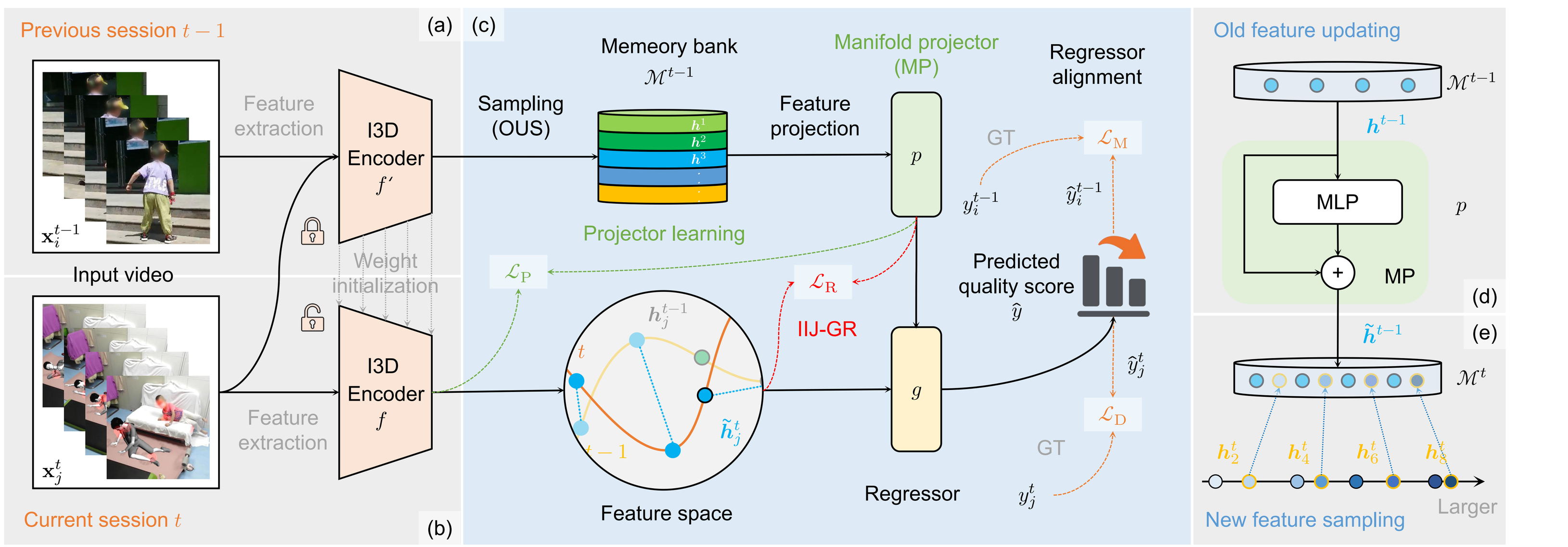
Our core contributions:

- We are the first to introduce CAQA to enable efficient AQA model refinement, addressing the unique challenges versus traditional classification tasks in CL.
- To address the misalignment, we propose MAGR as a novel solution, aligning old features to the current manifold without raw inputs and ensuring alignment between feature and quality score distributions.
- We validate MAGR on multiple AQA split datasets, demonstrating superior performance over recent strong baselines and establishing its effectiveness for continual performance assessment.

Manifold-Aligned Graph Regularization (MAGR)

Framework overview:

We consider two consecutive sessions: (a) At the end of session $t-1$, representative samples are chosen and stored in the memory bank $\mathcal{M}^{t-1}$, and the feature extractor f' is frozen. (b) Throughout session t, MP translates old features to the current manifold, while IIJ-GR regulates the entire feature space to align with the quality space. (c) After that, old features are first updated. (d) Then, new features are selected for the updated memory bank, denoted as $\mathcal{M}^t$, where the superscript indicates the update session.



Problem formulation:

CAQA processes sequentially obtained datasets $\{\mathcal{D}^t_{\text{train}}\}_{t=1}^T$ over T sessions. A key challenge is catastrophic forgetting when learning new sessions. To address this, CAQA employs feature replay utilizing a memory bank $\mathcal{M}^{t-1}$ to store old features. The objective is formulated as:

$$
\begin{aligned}
\min_{\Theta} \quad & \mathcal{L}_{\text{D}} + \mathcal{L}_{\text{M}} + \lambda_{\text{P}}\mathcal{L}_{\text{P}} + \lambda_{\text{R}}\mathcal{L}_{\text{R}}, \\
\text{s.t.} \quad & \hat{y}^s_i = g_{\theta_g}(p_{\theta_p}(\boldsymbol{h}^s_i)),\ (\boldsymbol{h}^s_i, y^s_i) \in \mathcal{M}^t, \\
& \hat{y}^t_j = g_{\theta_g}(f_{\theta_f}(\mathbf{x}^t_j)),\ (\mathbf{x}^t_j, y^t_j) \in \mathcal{D}^t_{\text{train}},
\end{aligned} \tag{1}
$$

where $\boldsymbol{h}^s_i = f_{\theta_f}(\mathbf{x}^s_i)$ denotes old features, $\Theta = \{\theta_f, \theta_p, \theta_g\}$ is the parameter set, and $\mathcal{L}_{\text{P}}$ and $\mathcal{L}_{\text{R}}$ encourage correcting deviated features and regulating the feature space, respectively. λ_{P} and λ_{R} balance the two constraints.

Experiments & Results

Comparison with recent strong baselines:

Experimental results for CAQA models. The primary metric considered is ρ_{avg}. We opt not to incorporate the difficulty label in MTL-AQA and the dive number in FineDiving for consistency to maintain a fair evaluation protocol. This validates MAGR for continual assessment challenges arising from non-stationary variations.

Method	Publisher	Memory	MTL-AQA			FineDiving			UNLV-Dive			JDM-MSA		
			ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)	ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)	ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)	ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)
Joint Training	-	None	0.9360	-	-	0.9075	-	-	0.8460	-	-	0.7556	-	-
Sequential FT	-	None	0.5458	0.1524	0.0538	0.7420	0.1322	0.2135	0.6307	0.2135	**0.3595**	0.5080	0.1029	0.5431
SI [1]	ICML'17	None	0.5526	0.2677	0.0350	0.6863	0.2330	0.1938	0.1519	0.3822	0.0220	0.4804	0.2198	0.5431
EWC [2]	PNAS'17	None	0.2312	0.1553	0.0343	0.5311	0.3177	0.1776	0.4096	0.2576	0.3039	0.3889	0.1690	0.3120
LwF [3]	TPAMI'17	None	0.4581	0.1894	0.0490	0.7648	0.0807	0.2894	0.6081	0.1578	0.3230	0.6441	0.1127	0.2423
MER [4]	ICLR'19	Raw Data	0.8720	0.1303	0.0625	0.8276	0.1446	0.2806	0.7397	0.1321	0.0465	0.6689	**0.0635**	0.3841
DER++ [5]	NeurIPS'20	Raw Data	0.8334	0.1775	0.0433	0.8285	0.1523	0.2851	0.7206	0.1382	-0.1773	0.5364	0.0835	0.5759
TOPIC [6]	CVPR'20	Raw Data	0.7693	0.1427	0.1391	0.8006	0.1344	0.2744	0.4085	0.2647	0.1132	0.6575	0.2184	0.5492
GEM [7]	ICCV'21	Raw Data	0.8583	0.0950	0.1429	0.8309	0.0721	0.2883	0.6538	0.2322	0.0270	0.6084	0.0499	0.3566
Feature MER	-	Feature	0.7283	0.2255	0.0535	0.4914	0.2354	0.2344	0.5675	0.1322	0.1558	0.6295	0.1597	**0.6446**
SLCA [8]	ICCV'23	Feature	0.7223	0.1852	0.1665	0.8130	0.0920	0.2453	0.5551	0.1085	0.3200	0.6173	0.1705	0.4457
NC-FSCIL [9]	ICLR'23	Feature	0.8426	0.1146	0.0718	0.8087	0.0203	**0.3404**	0.6458	**0.0637**	-0.1677	0.6571	0.1295	0.4957
MAGR (Ours)	-	Feature	**0.8979**	**0.0223**	**0.1914**	**0.8580**	**0.0167**	0.2952	**0.7668**	0.0827	0.1227	**0.7166**	0.1069	0.4957

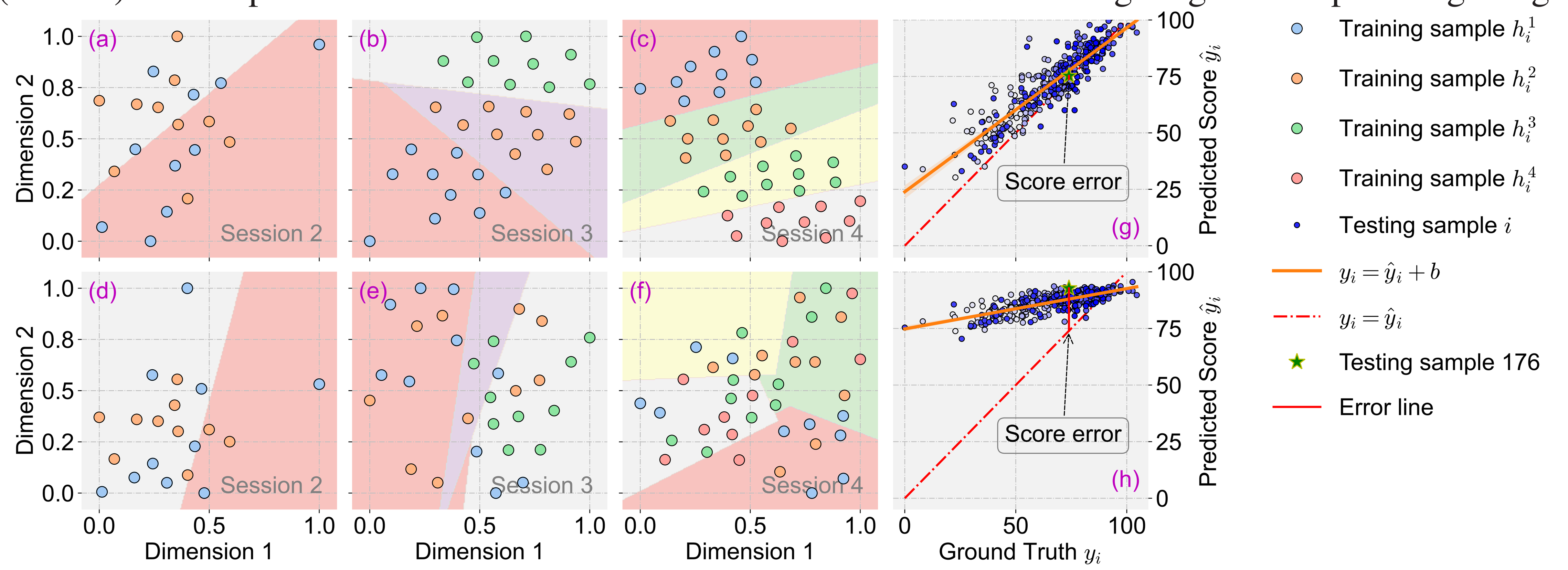
Ablation study on MTL-AQA:

The first row represents the performance of the entire model. Each subsequent row delineates the impact on performance metrics by removing a component. OUS is the proposed sampling strategy.

Setting	ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)
MAGR (Ours)	0.8979	0.0223	0.1914
w/o MP	0.6949 $^{\downarrow 23\%}$	0.1325 $^{\uparrow 494\%}$	0.0814 $^{\downarrow 57\%}$
w/o MP's res.	0.8391 $^{\downarrow 7\%}$	0.0232 $^{\uparrow 4\%}$	0.1743 $^{\downarrow 9\%}$
w/o II-GR	0.8463 $^{\downarrow 6\%}$	0.0970 $^{\uparrow 335\%}$	0.1062 $^{\downarrow 45\%}$
w/o J-GR	0.7839 $^{\downarrow 13\%}$	0.1053 $^{\uparrow 372\%}$	0.1005 $^{\downarrow 48\%}$
w/o IIJ-GR	0.7362 $^{\downarrow 18\%}$	0.1232 $^{\uparrow 452\%}$	0.0883 $^{\downarrow 54\%}$
w/o KL (MSE)	0.8447 $^{\downarrow 6\%}$	0.0265 $^{\uparrow 16\%}$	0.1890 $^{\downarrow 1\%}$
w/o OUS	0.8619 $^{\downarrow 4\%}$	0.0876 $^{\uparrow 293\%}$	0.1027 $^{\downarrow 46\%}$

Visualization of mitigating catastrophic forgetting on MTL-AQA:

Visualizations of feature distribution (a-f) and score correlation (g-h): MAGR (top) and Feature MER (bottom). The explicit division validates the effectiveness of MAGR in mitigating catastrophic forgetting.



Case study on MTL-AQA:

Qualitative comparison of sample #176 (a-d) and outputs of Feature MER (e) and MAGR (f). μ and σ represent the mean and the standard deviation, respectively.