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]

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

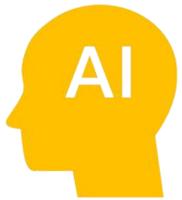
Outline

1. Challenges and Our Core Idea
   - Definition of AQA
   - Issues with traditional AQA
   - New task: CAQA
   - New challenges
   - Core idea of our solutions
2. Our method: MAGR
   - Framework overview
   - Manifold projector
   - IIJ graph regularizer
3. Experiments and Results
   - Comparison with baselines
   - Ablation study
   - Impact of buffer size
   - Visualizations
   - More results
4. Conclusions and Future Work
   - Conclusions
   - Future work

1. Challenges and Our Core Idea

1.1 Action Quality Assessment (AQA): Definition

- AQA aims to evaluate the quantitative performance of performed actions.



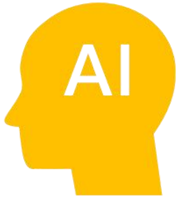
AQA system

1.1 Action Quality Assessment (AQA): Significance

- AQA aims to evaluate the quantitative performance of performed actions.
- Mitigating human judges' biases.

Human judge

VS



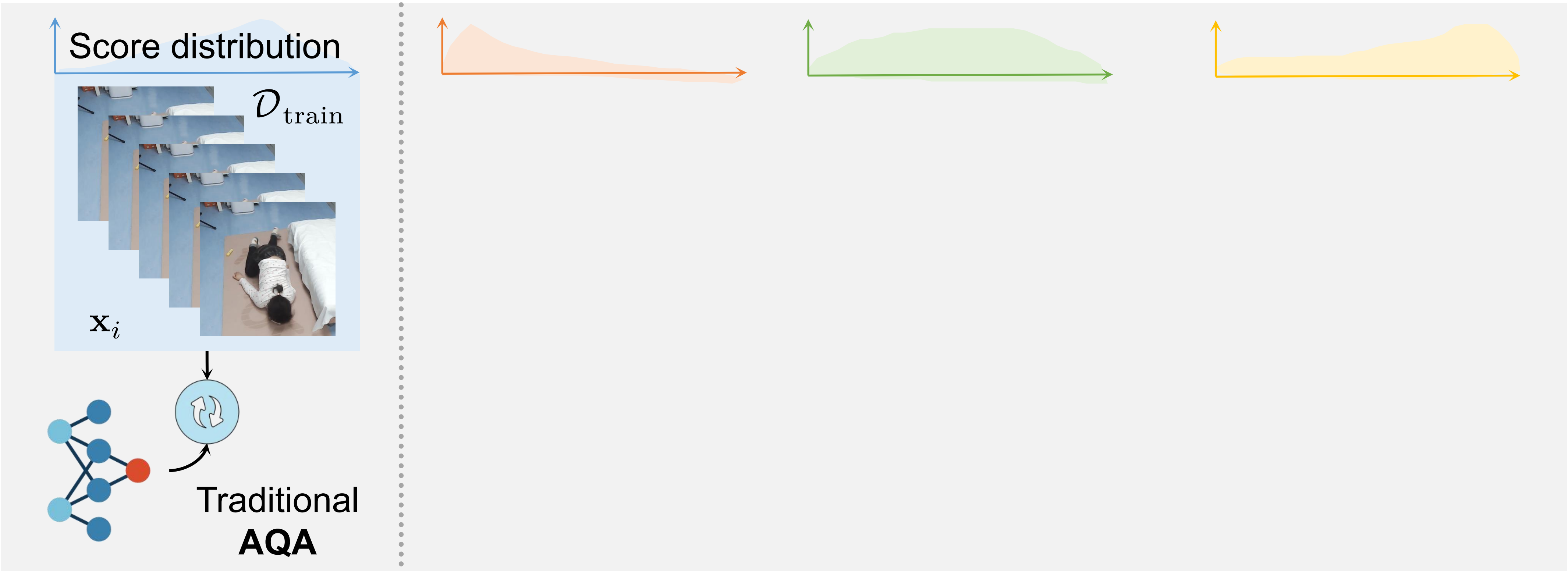
AQA system

1.1 Action Quality Assessment (AQA): Applications

- AQA aims to evaluate the quantitative performance of performed actions.
- Mitigating human judges’ biases.
- Widely used in sports, medical care, etc.

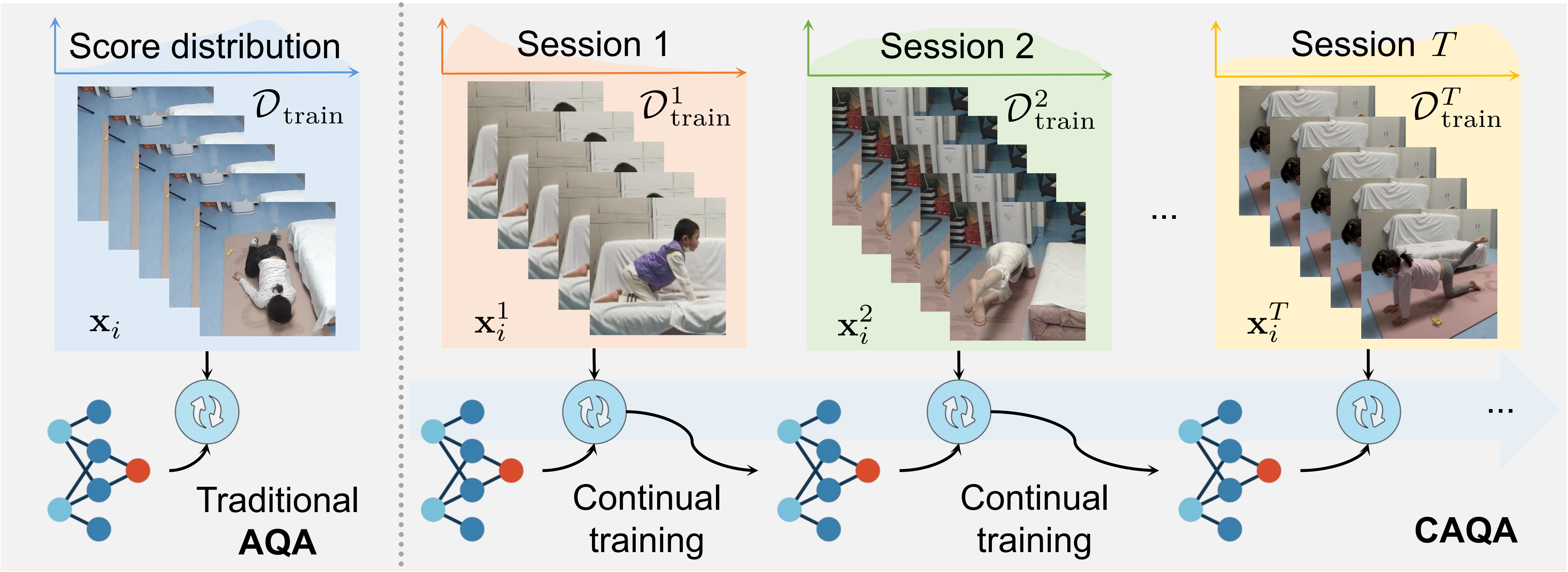
1.2 Issues with Traditional AQA Methods

- Cannot adapt to dynamically evolving changes...

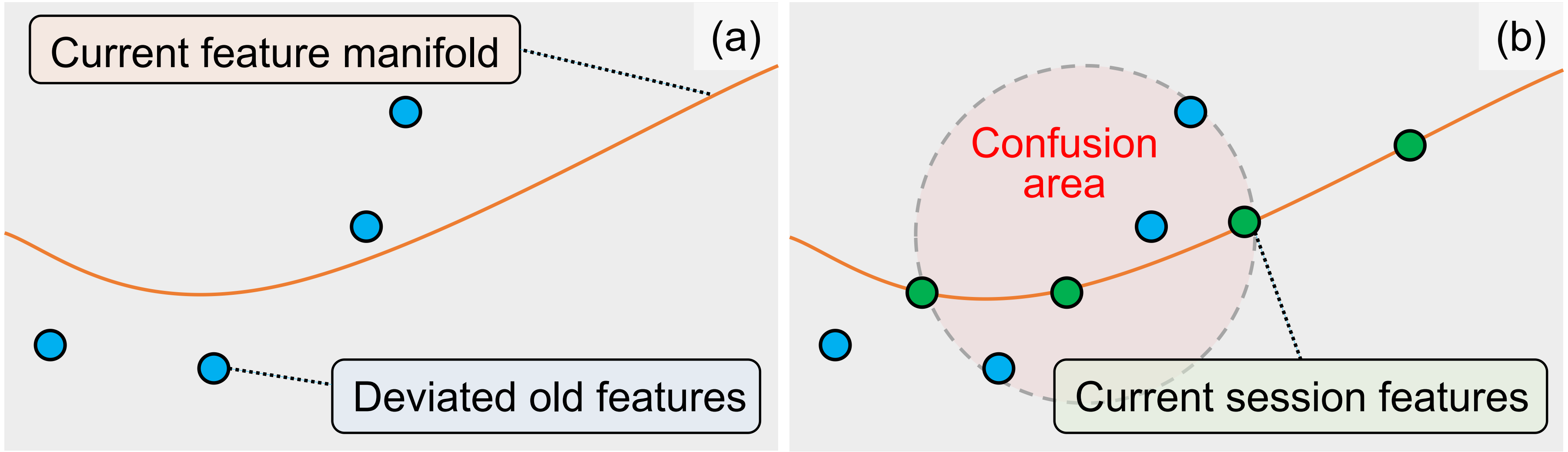


1.3 New Task: Continual AQA

- Integrate continual learning into the AQA framework, protecting user privacy and mitigating severe forgetting.

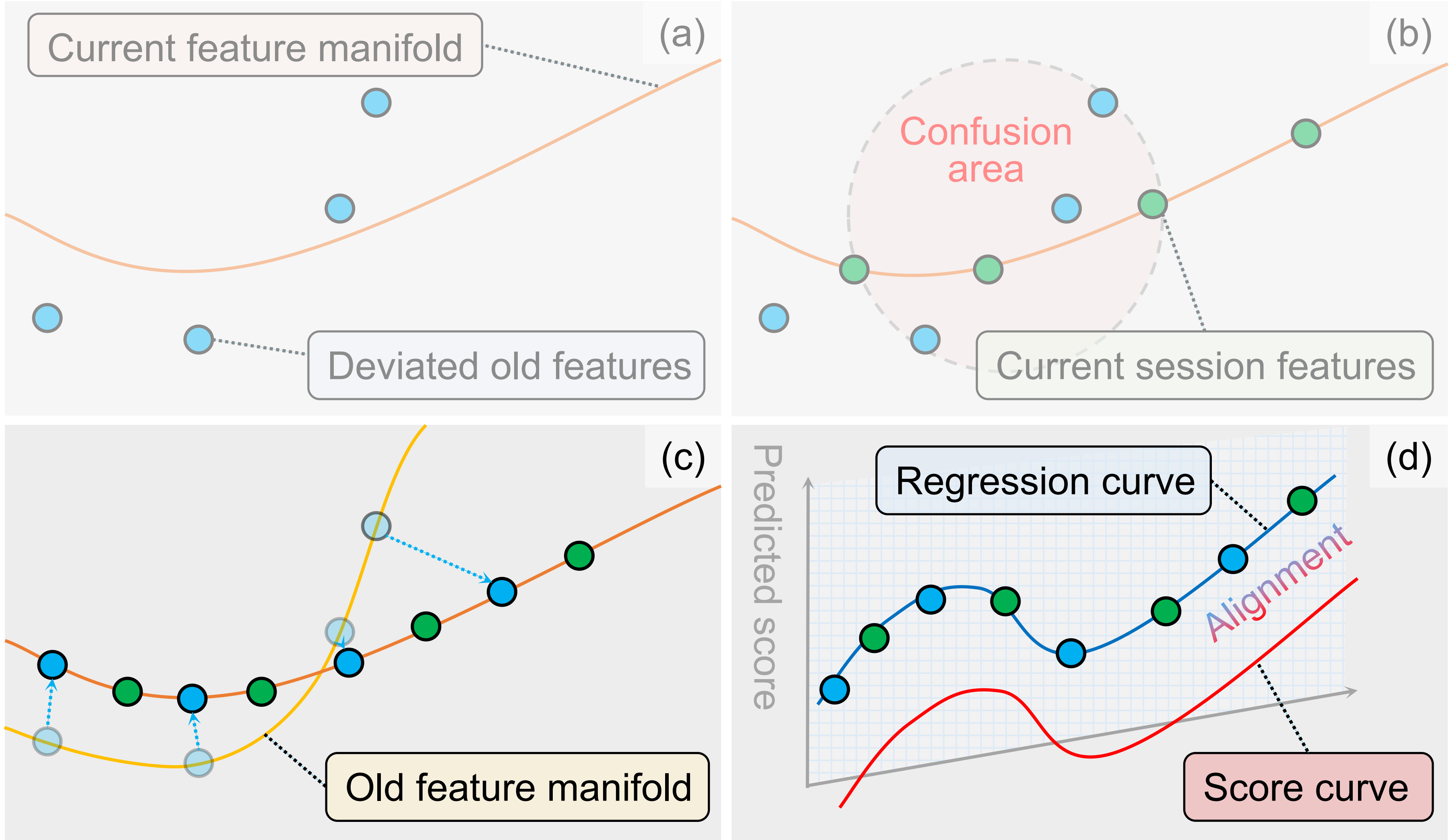


1.4 New Challenges within CAQA: Misalignment



(1) To mitigate the catastrophic forgetting, we adopt feature replay rather than raw data replay to prioritize user privacy; (2) To improve the adaptability, the complexity of AQA requires backbone updating that induces **the misalignment between static old features and dynamically evolving feature manifolds.**

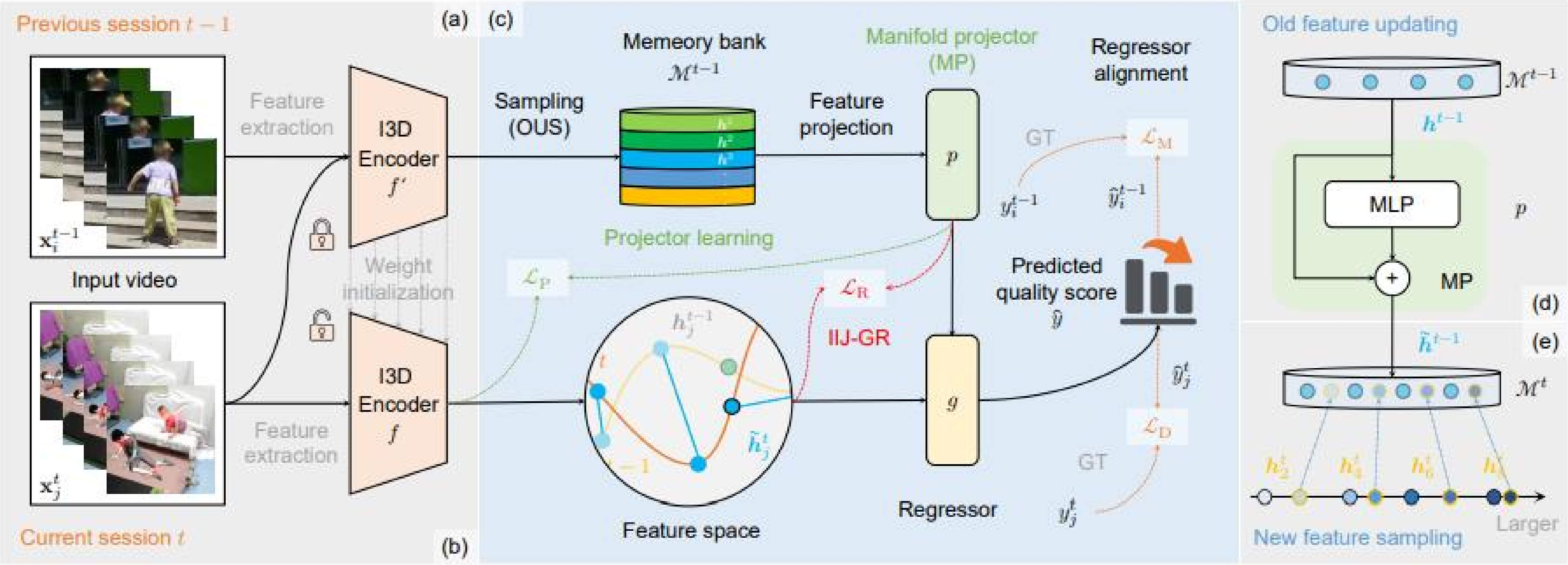
1.5 Core Idea to Address the Misalignment: Two Steps



2. Manifold-Aligned Graph Regularization (MAGR)

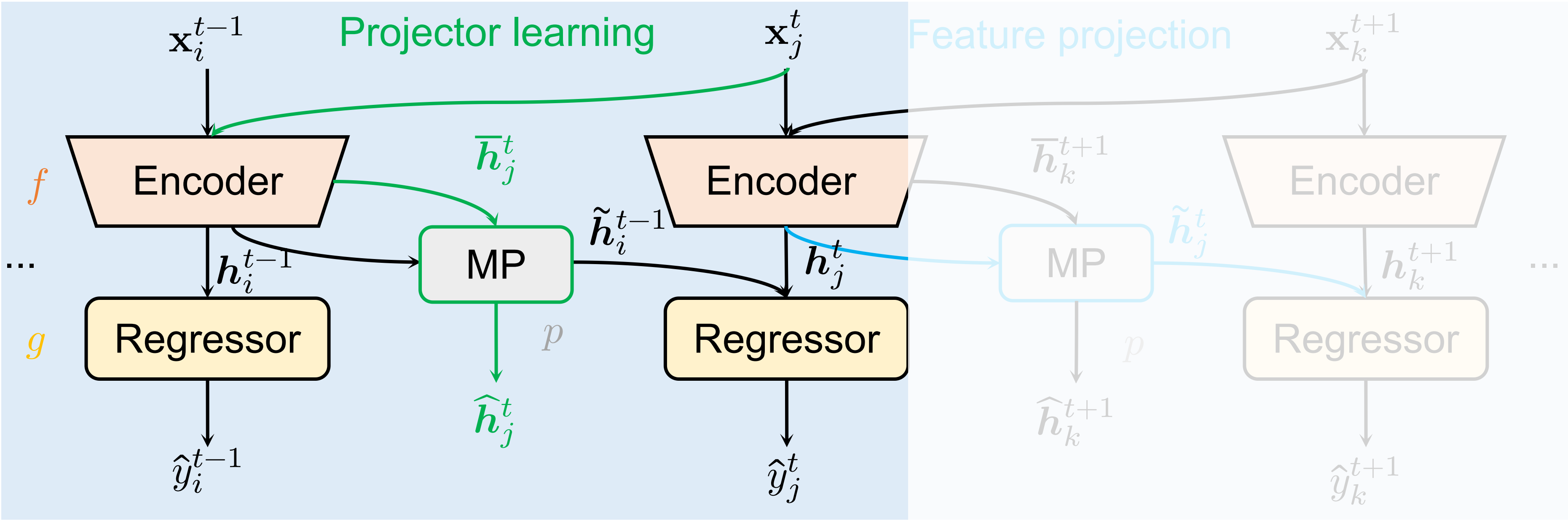
2.1 Framework Overview

- MP translates old features to the current manifold
- IIJ-GR regulates the entire feature space to align with the quality space



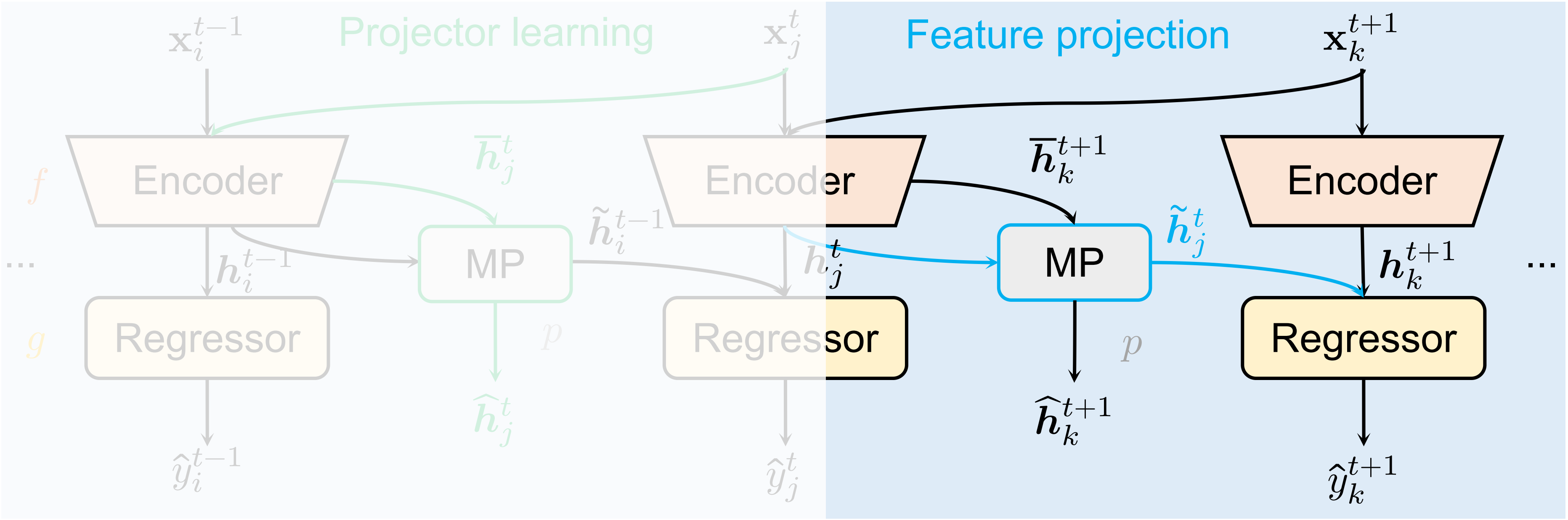
2.2 Manifold Projector: Deviated Feature Translation

- Projector Learning estimates the manifold shift at each model update using dependencies from the current session data.



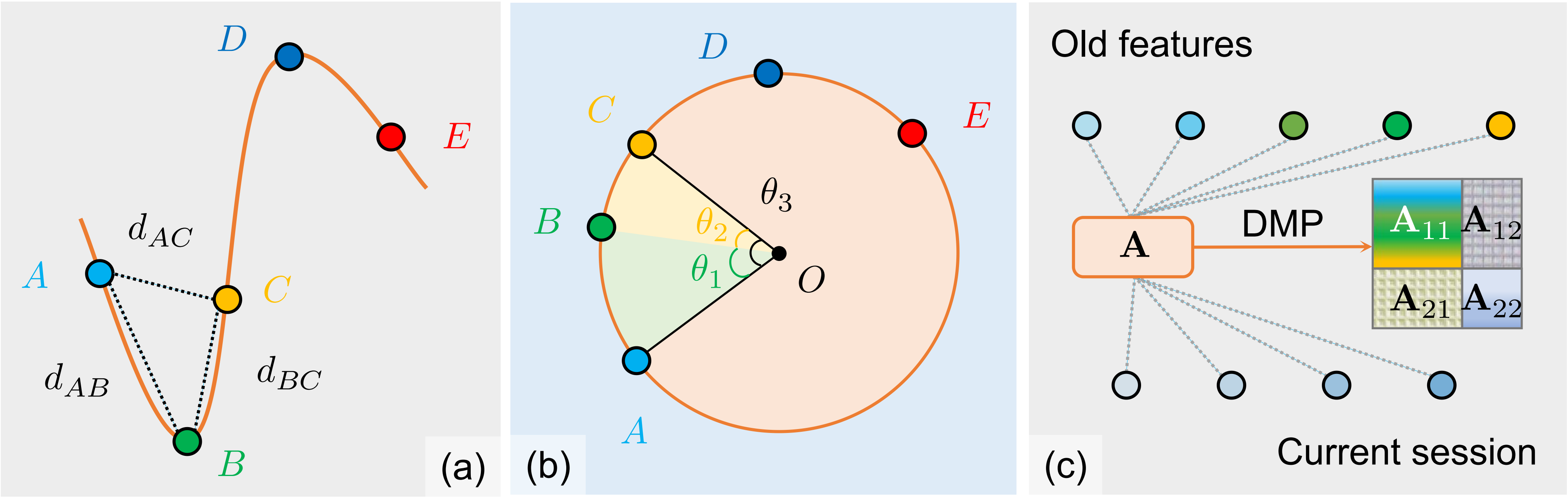
2.2 Manifold Projector: Deviated Feature Translation

- Feature Projection is designed to translate old features to the current manifold.



2.3 Intra-Inter-Joint Graph Regularizer: Feature Distribution Alignment

- We propose IIJ-GR to regulate the feature space for accurate score regression.



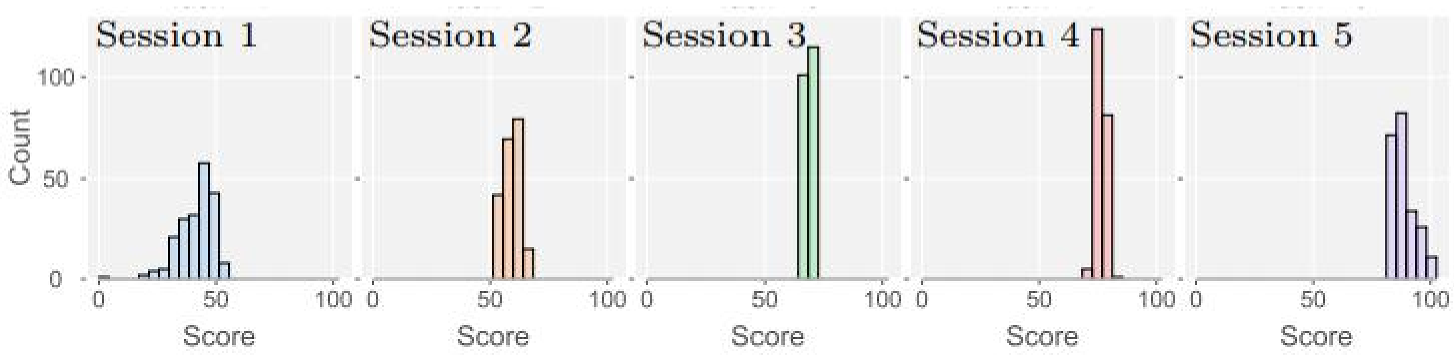
$d(A, C) \neq d(A, B) + d(B, C)$

$\theta_{AOC} = \theta_{AOB} + \theta_{BOC}$

3. Experiments and Results

3.1 Data Split and Metrics

- Discretizing the continuous quality space into distinct intervals corresponding to different action grades and ensuring an equal number of samples in each session, resulting in more challenging score variations.



3.1 Data Split and Metrics

- SRCC, forgetting, forward transfer

$$\rho = \frac{\sum_i (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_i (p_i - \bar{p})^2 \sum_i (q_i - \bar{q})^2}},$$

$$\rho_{\mathrm{aft}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \max_{i,j \in \{1,2,\cdots,T\}} \left(\rho_{i,t} - \rho_{j,t}\right),$$

$$\rho_{\mathrm{fwt}} = \frac{1}{T-1} \sum_{t=2}^{T} \left(\rho_{t-1,t} - \tilde{\rho}_t\right),$$

3.2 Comparison with Recent Strong Baselines

- MAGR outperforms recent strong baselines with up to 6.56%, 5.66%, 15.64%, and 9.05% correlation gains on the MTL-AQA, FineDiving, UNLV-Dive, and JDM-MSA split datasets, respectively.

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 [42]	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 [11]	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 [13]	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 [22]	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++ [3]	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 [26]	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 [12]	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 [43]	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 [39]	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

3.3 Ablation Study

- On the MTL-AQA split dataset

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 residual link	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 loss)	0.8447 $^{\downarrow 6\%}$	0.0265 $^{\uparrow 16\%}$	0.1890 $^{\downarrow 1\%}$
w/o OUS (random sampling)	0.8619 $^{\downarrow 4\%}$	0.0876 $^{\uparrow 293\%}$	0.1027 $^{\downarrow 46\%}$

3.4 Impact of Buffer Size

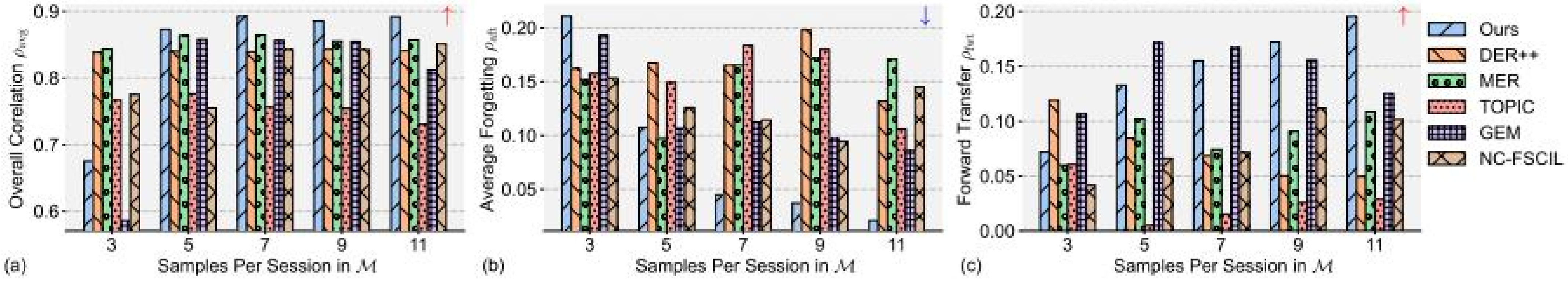


Fig. 7: Memory size comparisons with replay-based methods on MTL-AQA. ↑ indicates that higher values are better for the metric, whereas ↓ indicates the opposite.

3.5 Visualizations

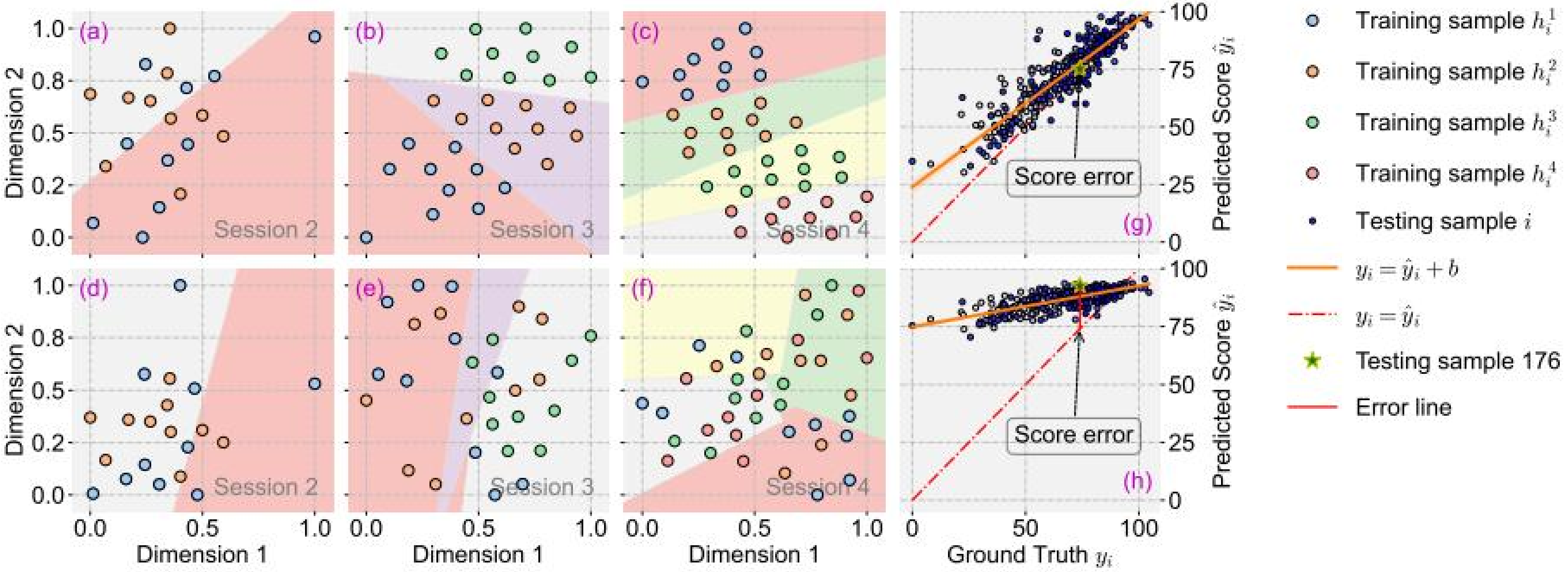


Fig. 9: Visualizations of feature distribution (a-f) and score correlation (g-h): MAGR (top) and Feature MER (bottom). The explicit division of different sessions validates the effectiveness of MAGR in mitigating catastrophic forgetting.

3.6 More Results: Computational Performance

Table S1: Computational performance on the MTL-AQA dataset.

Method	Param. (M)	Training Time (h)	Offline Performance		
			ρ_{avg} (↑)	ρ_{aft} (↓)	ρ_{fwt} (↑)
SLCA [14]	13.62	2.27	0.7223	0.1852	0.1665
NC-FSCIL [12]	12.62	2.33	0.8426	0.1146	0.0718
Feature MER	12.62	2.22	0.7283	0.2255	0.0535
MAGR (Ours)	12.63	2.23	0.8979	0.0223	0.1914

3.6 More Results: Online Performance Comparison

Table S2: Online continual learning (ρ_{avg} is the main metric).

Method	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} (↑)
SLCA [14]	0.4880	0.0430	-0.0282	0.3935	0.3360	0.2346	0.3119	0.1641	-0.3082	0.1726	0.0589	0.0382
NC-FSCIL [12]	0.4971	0.0291	-0.0463	0.3810	0.0079	**0.2518**	0.3136	**0.1282**	-0.4892	0.1540	**0.0355**	0.0378
Feature MER	0.3571	0.1444	**-0.0213**	0.1935	0.0998	0.1559	0.1308	0.2126	-0.4571	0.1699	0.0356	0.0382
MAGR (Ours)	**0.5196**	**0.0269**	-0.0337	**0.4641**	**0.0062**	0.2020	**0.4202**	0.1947	**-0.0499**	**0.2029**	0.0356	**0.0449**

4. Conclusions and Future Work

4.1 Conclusions

- We are the first to introduce CAQA to enable efficient AQA model refinement using sparse new data, addressing the unique challenges versus traditional classification tasks in CL.
- 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, thereby advancing CL and AQA research.

4.2 Future Work

- Advanced network architectures like ViT
- Incorporating prompt-based techniques

Thanks!