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], and Xiaohui Liang[1,5(✉)]

[1] State Key Laboratory of VR Technology and Systems, Beihang University
[2] Department of Computer Science and Technology, Institute for AI, BNRist Center, Tsinghua-Bosch Joint ML Center, THBI Lab, Tsinghua University
[3] Department of Computer Science, Durham University
[4] Children's Hospital of Capital Institute of Pediatrics
[5] Zhongguancun Laboratory
liang_xiaohui@buaa.edu.cn

Abstract. Action Quality Assessment (AQA) evaluates diverse skills but models struggle with non-stationary data. We propose Continual AQA (CAQA) to refine models using sparse new data. Feature replay preserves memory without storing raw inputs. However, the misalignment between static old features and the dynamically changing feature manifold causes severe catastrophic forgetting. To address this novel problem, we propose Manifold-Aligned Graph Regularization (MAGR), which first aligns deviated old features to the current feature manifold, ensuring representation consistency. It then constructs a graph jointly arranging old and new features aligned with quality scores. Experiments show 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. This validates MAGR for continual assessment challenges arising from non-stationary skill variations. Code is available at https://github.com/ZhouKanglei/MAGR_CAQA.

Keywords: Continual Learning · Action Quality Assessment

1 Introduction

Action Quality Assessment (AQA) evaluates performance beyond recognition [46]. Traditional AQA methods trained on small static datasets [16, 17, 45] struggle with dynamically changing skills in sports [35, 42] and rehabilitation [12, 37, 44], requiring updated standards. Continual Learning (CL) provides promising solutions to non-stationarity [30, 41], yet faces challenges like catastrophic forgetting in sequential training [29, 30]. By enabling continuous learning with memory preservation, CL facilitates lifelong adaptation and stability.

While CL offers promising solutions for dynamic skill assessments, prior work has scarcely explored its application to AQA, which poses a unique CL challenge through its reliance on subtle quality regression over evolving feature manifolds, defining a novel problem of continual assessment. To address this issue, we introduce Continual AQA (CAQA) to seamlessly refine AQA models using sparse

new data (see Fig. 1) without catastrophic forgetting. Unlike traditional CL research focused on discrete classifications [5, 26, 28], CAQA involves continuous quality score regression requiring adaptation to changing quality score distributions characterizing how skills evolve over time. To advance CAQA research, we propose the novel task of incrementally refining AQA models over sequential sessions using only a few arriving samples, fulfilling real-world needs while posing distinct challenges compared to traditional classification tasks in CL.

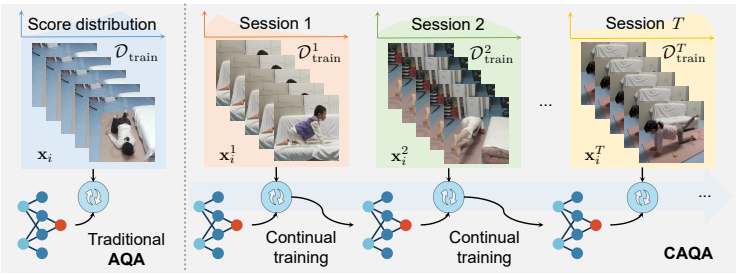

Fig. 1: Traditional AQA vs CAQA: CAQA refines AQA from a few sequentially arrived instances without exhaustive retraining, which advances CL beyond classification.

A key challenge in CAQA implementation is the misalignment between static old features and the evolving feature manifold, resulting in catastrophic forgetting. Effectively retaining and utilizing past knowledge is essential across sequential sessions. While experience replay methods like those in [10, 20] are common in CL, they raise privacy concerns, especially in sensitive AQA domains like sports training or medical care. Feature replay methods, such as those proposed in [36, 40], offer privacy advantages but struggle with adapting to new data distributions. The dilemma of using a fixed backbone limits model adaptability, while updating the backbone risks severe catastrophic forgetting by misaligning old features with the evolving feature manifold. Addressing this specific misalignment issue and updating backbones are critical for CAQA's adaptability.

To mitigate the misalignment, we propose Manifold-Aligned Graph Regularization (MAGR), a novel feature replay method consisting of two essential steps. Firstly, MAGR iteratively learns the manifold shift between sessions using only the current session's raw data due to privacy constraints, ensuring accurate alignment of old features to the current feature manifold. Secondly, MAGR focuses on readjusting the feature distribution with the quality score distribution from local and global perspectives to eliminate confusion of features across sessions and within the same session, thereby improving assessment accuracy.

To stimulate research in this important yet understudied field, we establish a comprehensive CAQA benchmark study. This includes splitting multiple AQA datasets, defining custom evaluation protocol and metrics, and incorporating recent strong baselines. Experiments demonstrate that MAGR outperforms recent strong baselines by a substantial margin, achieving correlation gains of up to 6.56%, 5.66%, 15.64%, and 9.05% on the MTL-AQA, FineDiving, UNLV-Dive, and JDM-MSA split datasets, respectively. Our main contributions are:

- To the best of our knowledge, 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.
- 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, thereby advancing CL and AQA research.

2 Related Work

Action Quality Assessment (AQA) evaluates the quantitative performance of performed actions in various areas [32], such as sports [46], medical rehabilitation [7, 44], and skill assessment [21, 33]. Earlier methods depended heavily on hand-crafted features and heuristics, revealing certain limitations. By integrating deep learning, various models [17, 22, 38] have shown improved performance. Due to label scarcity, existing AQA datasets [32] are relatively small, risking over-fitting. To mitigate this, pre-trained backbones such as C3D [24], I3D [4], and VST [14] are commonly employed. As every frame may contain essential AQA cues [18], videos are often segmented into clips for separate processing due to the limited computation resources, which hinders a complete understanding of the action. To address this, a GCN-based method [46] has been proposed to eliminate semantic ambiguities. Despite the success, the ever-evolving nature of individual skills and environments requires lifelong adaptation in AQA. To address this practical problem, we seamlessly incorporate CL paradigms into AQA systems to ensure efficient adaptation.

Continual Learning (CL), also known as incremental or lifelong learning [30], aims to train a model over a sequence of tasks, ensuring the retention of learned tasks and mitigating catastrophic forgetting. Since AQA is significantly restricted by relatively small datasets, we primarily focus on few-shot CL [41, 43]. which suffers from severe over-fitting. Recent strategies to combat this issue include preserving representation topology [23], constructing exemplar relation graphs for knowledge distillation [8], and using generative replays of earlier data distributions [13]. These advanced CL strategies often rely on the experience replay of old samples [31], raising legitimate privacy concerns [41]. This is particularly critical in sensitive AQA domains such as proprietary sports training and medical rehabilitation. Feature replay methods [26, 36, 40] offer privacy benefits and have demonstrated strong performance. However, generative replay [25, 27] faces limitations due to the variable quality of the generated data, which can impact its effectiveness in real-world scenarios. Moreover, a fixed backbone [36] simplifies the model but limits its adaptability to evolving real-world scenarios, while an updated backbone [40] risks misalignment with old features, deteriorating the forgetting issue. Our work is dedicated to addressing the misalignment with backbone updates in the context of CAQA.

3 Continual Action Quality Assessment (CAQA)

We propose the novel task of CAQA to adapt evolving individual skills or health conditions over time. For instance, in athlete rehabilitation, movement quality evolves with recovery stages, challenging traditional AQA systems trained on static datasets. Unlike traditional CL methods focused on classification [30, 43], CAQA involves continuous quality score regression, which is crucial for accurately capturing subtle variations in performance. This aspect is especially relevant in real-world scenarios where subtle variations in quality assessment can have significant implications, presenting a unique and pressing challenge in the field. By introducing CAQA, we aim to address this critical gap and provide a robust framework for continual refinement of AQA models in dynamic environments. The following defines AQA and CAQA.

AQA Task aims to assign a quality score $\hat{y} \in \mathcal{Y}$ (typically $y \in \mathbb{R}$) to an input video $\mathbf{x} \in \mathcal{X}$ (typically $\mathbb{R}^{L\times H\times W\times 3}$), where L, H, W and 3 represent length, height, width, and channels of inputs, respectively. The goal is to learn mappings $\hat{y} = g_{\theta_g}(\boldsymbol{h})$ and $\boldsymbol{h} = f_{\theta_f}(\mathbf{x})$ between $\mathcal{X}$ and $\mathcal{Y}$ from data $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_n, y_n)\}_{n=1}^{N}$ using encoders $f(\cdot)$ and regressors $g(\cdot)$, where $\boldsymbol{h}$ denotes the latent feature.

CAQA Task aims to seamlessly integrate CL into AQA to enable continuous adaptation of assessment capabilities. It 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:

$$\min_{\theta_f,\theta_g} \ \mathcal{L}_{\text{D}} + \mathcal{L}_{\text{M}}, \tag{1}$$

where $\mathcal{L}_{\text{D}}$ and $\mathcal{L}_{\text{M}}$ are score regression losses on current data $\mathcal{D}^t_{\text{train}}$ and memory bank $\mathcal{M}^t$, enabling incremental refinement without exhaustive retraining.

4 Manifold-Aligned Graph Regularization (MAGR)

We first motivate MAGR and then explain its major novel technical components.

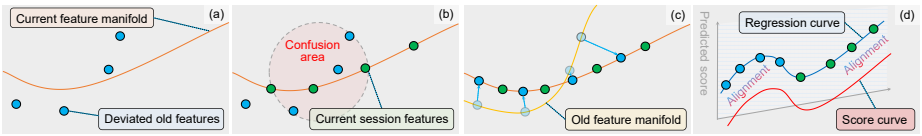

Fig. 2: Our core idea: (a) Deviation of old features (blue circles) from the current manifold (orange curve) caused by the manifold shift; (b) Potential confusion for score regression due to the mixture of old features and current session features (green circles); (c) Translation of old features from the previous manifold (yellow curve) to the current; (d) Readjustment of the feature distribution to align with the quality score distribution.

4.1 Motivation and Pipeline of MAGR

Motivation. We propose MAGR as a solution to address the key challenge of CAQA. Its motivation is threefold: (1) To mitigate the catastrophic forgetting, we adopt feature replay rather than raw data replay to prioritize user privacy which is crucial for sensitive AQA domains; (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 (see Figs. 2(a) and 2(b)); and (3) To tackle the misalignment, MAGR adopts a two-step alignment process by dynamically translating deviated features (see Fig. 2(c)) and aligning the feature distribution (see Fig. 2(d)).

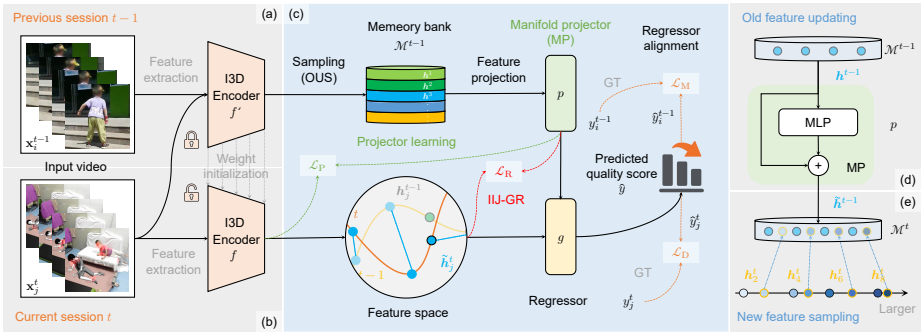

Fig. 3: MAGR framework: (a) At the end of session $t-1$, representative samples are chosen and stored in the memory bank $\mathcal{M}^{t-1}$, and the feature extractor f' is frozen. (b) Throughout session t, MP translates old features to the current manifold, while IIJ-GR regulates the entire feature 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.

Pipeline and Insight. Fig. 3 depicts the MAGR framework, where we consider two consecutive sessions $t-1$ and t. At the end of session $t-1$, we employ Ordered Uniform sampling (OUS) to select representative features that are then stored in a memory bank $\mathcal{M}^{t-1}$ to log the old distribution. OUS involves sorting the entire training set before performing uniform sampling, ensuring coverage across the whole range. In comparison to DER [3], OUS maximally preserves the old quality score distribution, thereby improving CAQA performance (refer to Tab. 2). During session t, one branch is dedicated to refining AQA by learning new data, while the other retrieves a mini-batch of old features from the memory bank $\mathcal{M}^{t-1}$ to maintain the memory stability. Old features are refined using the Manifold Projector (MP, see Sec. 4.2), while the Intra-Inter-Joint Regularizer (IIJ-GR, see Sec. 4.3) aligns both old and new features with the quality score distribution. Together, the objective in Eq. (1) can be reformulated as:

$$\begin{aligned} \min_{\Theta} \ & \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\mathrm{M}} + \lambda_{\mathrm{P}}\mathcal{L}_{\mathrm{P}} + \lambda_{\mathrm{R}}\mathcal{L}_{\mathrm{R}}, \\ \text{s.t.} \ & \hat{y}_i^s = g_{\theta_g}(p_{\theta_p}(\boldsymbol{h}_i^s)), \ (\boldsymbol{h}_i^s, y_i^s) \in \mathcal{M}^t, \\ & \hat{y}_j^t = g_{\theta_g}(f_{\theta_f}(\mathbf{x}_j^t)), \ (\mathbf{x}_j^t, y_j^t) \in \mathcal{D}_{\mathrm{train}}^t, \end{aligned} \tag{2}$$

where $\boldsymbol{h}_i^s = f_{\theta_f}(\mathbf{x}_i^s)$ denotes old features, $\Theta = \{\theta_f, \theta_p, \theta_g\}$ is the parameter set, and $\mathcal{L}_{\mathrm{P}}$ and $\mathcal{L}_{\mathrm{R}}$ encourage correcting deviated features and regulating the feature space, respectively. λ_{P}^t and λ_{R} balance the two constraints. In addition, the training process is detailed in Sec. 4.4.

4.2 Manifold Projector: Deviated Feature Translation

MP is designed to learn a mapping from the previous manifold to the current one. Fig. 4 highlights two sub-steps of MP across three consecutive sessions: projector learning and feature projection.

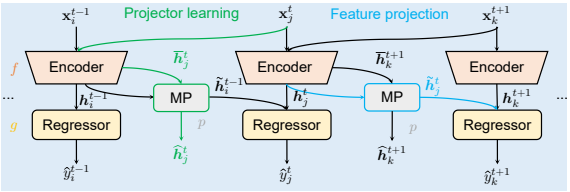

Fig. 4: Illustrations of MP: Projector learning estimates the manifold shift, and feature projection translates old features to the current manifold.

Projector Learning is designed to estimate the manifold shift at each model update using dependencies from the current session data. We replicate and freeze the encoder $f'(\cdot)$ from the previous session at the beginning of session t. This allows us to obtain the initial feature $\bar{\boldsymbol{h}}_j^t$ for the current session data $\mathbf{x}_j^t$ as a base to estimate manifold shifts. For simplicity, we will omit arguments in the lower-right corners of functions in subsequent discussions. Next, $\bar{\boldsymbol{h}}_j^t$ is passed through the projector p to obtain the predicted updated feature $\hat{\boldsymbol{h}}_j^t$, which is:

$$\hat{\boldsymbol{h}}_j^t = \bar{\boldsymbol{h}}_j^t + p(\bar{\boldsymbol{h}}_j^t), \text{ where } \bar{\boldsymbol{h}}_j^t = f'(\mathbf{x}_j^t). \tag{3}$$

Here, we observed that the inclusion of a residual link in feature projection enhances effectiveness in learning through updates (refer to the results in Tab. 2).

The learning process of MP is supervised by minimizing the difference between the predicted and actual updated features of the current session data:

$$\mathcal{L}_{\mathrm{P}} = \frac{1}{|\mathcal{D}_{\mathrm{train}}^t|} \sum_j \|\boldsymbol{h}_j^t - \hat{\boldsymbol{h}}_j^t\|_2^2, \tag{4}$$

where $|\cdot|$ indicates the set size, and $\boldsymbol{h}_j^t = f(\mathbf{x}_j^t)$ is the actual updated feature.

Feature Projection is designed to translate old features to the current manifold. For each old feature $\tilde{\boldsymbol{h}}_i^s$ $(i = 1, 2, \cdots, |\mathcal{M}^{t-1}|)$ from the memory bank, its corrected feature can be calculated by:

$$\tilde{\boldsymbol{h}}_i^s = \tilde{\boldsymbol{h}}_i^s + p(\tilde{\boldsymbol{h}}_i^s), \text{ where } s = 1, 2, \cdots, t-1. \tag{5}$$

Benefit of MP. By leveraging dependencies from the current session to correct deviated old features, MP plays a crucial role in addressing the manifold shift in feature-replay methods [36, 40]. Its novelty is its ability to adaptively

learn the manifold shift between sessions without needing access to raw old inputs. This sets it apart from traditional experience replay methods involving raw data [20]. MP offers both privacy and resource-efficiency advantages, making it particularly valuable in sensitive AQA domains. In addition, MP overcomes the restriction of fixing backbones found in [36], providing high adaptability to real-world complexities. Utilizing an MLP with residual links, MP presents a simple yet effective solution for addressing misalignment in feature replay.

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

While MP has translated deviated old features with the current feature manifold, it may not ensure alignment between the feature distribution and its quality score distribution. In Fig. 2(b), it is difficult to discern the relative relationships of quality scores between inter-session features (across different sessions) and intra-session features (within the same session), posing challenges for quality score regression and ultimately impacting AQA performance. To this end, we propose IIJ-GR to regulate the feature space for accurate score regression.

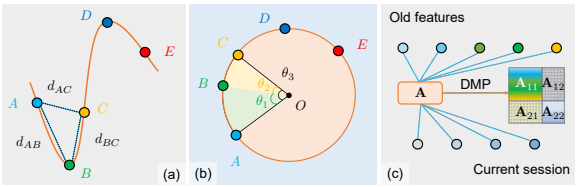

Fig. 5: Illustrations of IIJ-GR: (a) Euclidean distance, (b) Angular distance, and (c) Distance Matrix Partitioning (DMP).

Fig. 5 depicts the fundamental concept behind our proposed IIJ-GR module. We identify a significant limitation in graph-based CL methods [8, 23], where Euclidean distance is commonly utilized to measure point separation within a manifold. Indeed, the quality score relationship inherent in the data particularly meets the geodesic property. However, using the Euclidean distance often fails to satisfy such relationship $d_{AC} = d_{AB} + d_{BC}$ (see Fig. 5(a)). To address this problem, we leverage the insight that geodesic distance is proportional to angular differences [2]. Consequently, we normalize features to fit a unit sphere and utilize these angular differences for distance estimation (see Fig. 5(b)). Additionally, to effectively regulate the feature space, we propose partitioning the distance matrix (see Fig. 5(c)). This matrix encapsulates dependencies from both previous and current sessions. Our proposed Distance Matrix Partitioning (DMP) first divides the matrix into block matrices, which are then collaboratively optimized to ensure a quality-aware manifold from both local and global perspectives.

Distance Matrix Partitioning. Let's consider a mini-batch b_1 from previous sessions, symbolized by $\boldsymbol{h}_i^s$, and a batch b_2 from the current session, represented by $\boldsymbol{h}_j^t$. Concatenating them results in the feature matrix $\mathbf{H} = [\boldsymbol{h}_1^s, \boldsymbol{h}_2^s, \cdots, \boldsymbol{h}_{b_1}^s, \boldsymbol{h}_1^t, \boldsymbol{h}_2^t, \cdots, \boldsymbol{h}_{b_2}^t] \in \mathbb{R}^{(b_1+b_2)\times D}$. Similarly, we can obtain the corresponding score vector $\boldsymbol{y} \in \mathbb{R}^{(b_1+b_2)\times 1}$. In this way, the distance matrix $\mathbf{A} \in \mathbb{R}^{(b_1+b_2)\times(b_1+b_2)}$ can be calculated as:

$$\mathbf{A} = \arccos(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\top), \text{where } \tilde{\mathbf{H}} = \mathbf{H}/\|\mathbf{H}\|. \tag{6}$$

Thereafter, $\mathbf{A}$ can be divided into four sub-matrices: $\mathbf{A}_{11} \in \mathbb{R}^{b_1 \times b_1}$ captures relationships within the previous sessions, $\mathbf{A}_{12} \in \mathbb{R}^{b_1 \times b_2}$ and $\mathbf{A}_{21} \in \mathbb{R}^{b_2 \times b_1}$ characterize relationships between the previous and current sessions, $\mathbf{A}_{22} \in \mathbb{R}^{b_2 \times b_2}$ encapsulates relationships within the current session, and $\mathbf{A}$ as a whole signifies the integrated relationships spanning all observed sessions.

Graph Regularization. Essentially, this step involves creating a quality score distance matrix $\mathbf{S} = \boldsymbol{y} - \boldsymbol{y}^\top$, which acts as supervision for $\mathbf{A}$. To achieve this, we first define a loss function to gauge the distribution discrepancy between two matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{N \times N}$ as:

$$L(\mathbf{P}, \mathbf{Q}) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{KL}(\sigma(\mathbf{P}_i), \sigma(\mathbf{Q}_i)), \tag{7}$$

where $\mathrm{KL}(\cdot)$ represents the Kullback–Leibler (KL) divergence, and $\sigma(\cdot)$ denotes the softmax function. Here, the KL divergence offers a looser and more holistic constraint compared to MSE, aligning well with the correlation evaluation metric and ultimately leading to better performance (refer to Tab. 2). Then, the total regularization loss can be represented as:

$$\mathcal{L}_{\mathrm{R}} = L(\mathbf{A}, \mathbf{S}) + \sum_{i=1}^{2} \sum_{j=1}^{2} L(\mathbf{A}_{ij}, \mathbf{S}_{ij}), \tag{8}$$

where the partition of $\mathbf{S}$ is the same as that of $\mathbf{A}$. This aids in closely aligning the underlying feature space with the actual quality distribution of the data.

Benefit of IIJ-GR. IIJ-GR offers several advantages over manual pair selection in contrastive loss [1]. It effectively captures complex feature relationships from both local and global perspectives, which helps mitigate catastrophic forgetting and ensures assessment improvement across different sessions. While kernel alignment [6] aims to maximize the alignment between kernels in a static feature space, IIJ-GR explicitly learns to align the raw feature space itself with the quality score space in a dynamic manner. Compared to graph-based CL methods [8, 23], DMP considers both inter-session and intra-session constraints, enhancing feature representation fidelity and aligning it with the quality space.

4.4 Training Procedure

In Algorithm 1, the training begins with careful parameter initialization. The encoder leverages pre-trained weights from [46], providing domain knowledge, while other components are randomly initialized to adapt to CAQA sessions.

For each session, we maintain a previous encoder copy to compute the relative manifold shift for learning the MP. During training, for each batch, we extract features using the encoder and predict scores through a regressor. If memory is non-empty, we employ feature replay to mitigate catastrophic forgetting. MAGR incorporates MP and IIJ-GR for feature space alignment. We then replay a mini-batch of these corrected features. Model parameter optimization is iterative until convergence. At the end of the session, old features are first updated, and then representative features are drawn and stored in the memory bank.

Algorithm 1: The training process of MAGR.

Input: Training datasets $\mathcal{D}_{\text{train}}^t, t \in \{1, 2, \cdots, T\}$, and the network parameter $\Theta = \{\theta_f, \theta_g, \theta_p\}$.

Output: The trained model with the optimal parameter Θ.

Initialize θ_f with the pre-trained I3D weight, randomly initialize θ_p and θ_g, and initialize $\mathcal{M}$ with $\varnothing$;

for $t \leftarrow 1, 2, \cdots, T$ **do**

  $f' \leftarrow \text{copy}(f)$; `// copy and fix the previous encoder`

  while *not converged* **do**

    $\hat{y}_i^t \leftarrow g(\boldsymbol{h}_i^t),\ \boldsymbol{h}_i^t \leftarrow f(\mathbf{x}_i^t)$; `//` $i = 1, 2, \cdots, b_2$

    $\mathcal{L}_{\text{D}} \leftarrow 1/b_2 \sum_i (\hat{y}_i^t - y_i^t)^2$; `//` $(\mathbf{x}_i^t, y_i^t) \in \mathcal{D}_{\text{train}}^t$

    if $\mathcal{M}^t \neq \varnothing$ **then**

      $\hat{\boldsymbol{h}}_i^t \leftarrow f'(\mathbf{x}_i^t) + p(f'(\mathbf{x}_i^t))$; `// feature projection in Eq. (3)`

      $\mathcal{L}_{\text{P}} \leftarrow 1/b_2 \sum_i^{b_2} \|\boldsymbol{h}_i^t - \hat{\boldsymbol{h}}_i^t\|_2^2$; `// feature learning in Eq. (4)`

      Calculate $\mathcal{L}_{\text{R}}$ using Eq. (8); `// IIJ-GR in Sec. 4.3`

      $\tilde{\boldsymbol{h}}_i^s \leftarrow \tilde{\boldsymbol{h}}_i^s + p(\tilde{\boldsymbol{h}}_i^s)$; `//` $s < t$, `Eq. (5)`

      $\mathcal{L}_{\text{M}} \leftarrow 1/b_1 \sum_i^{b_1} (\hat{y}_i^s - y_i^s)^2,\ \hat{y}_i^s \leftarrow g(\tilde{\boldsymbol{h}}_i^s)$; `// regressor alignment`

    Update Θ by optimizing Eq. (2); `// backward`

  Update old features $\tilde{\boldsymbol{h}}^s$ from $\mathcal{M}^{t-1}$ to $\mathcal{M}^t$;

  Draw representative features $\boldsymbol{h}^t$ to $\mathcal{M}^t$ using the OUS strategy;

5 The CAQA Benchmark

We construct a comprehensive benchmark study for advancing CAQA research.

5.1 Datasets

We choose four datasets, namely MTL-AQA [16], FineDiving [35], UNLV-Dive [18], and MSA-JDM [44], to ensure a holistic evaluation across diverse domains and scenarios, each with varying sample sizes. Leveraging representative AQA datasets spanning sports and medical care domains allows us to address privacy concerns and ensure the generalization of CAQA models. For more details about each dataset, please refer to the supplementary material.

5.2 Experiment Protocol

To simulate the real-world skill variations, we propose a novel grade-incremental setting for CAQA, characterized by the challenges of both regression and classification tasks (see Fig. 6). This is achieved by 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. Unlike the uniform discrete class semantic space in traditional class-incremental tasks [41], our setting involves contextual relationships between adjacent grades, and data samples in the same grade may present quality variations of actions, posing a new challenge for lifelong learning in preserving these dependencies to mitigate catastrophic forgetting.

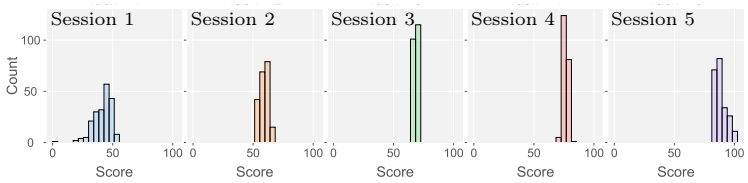

Fig. 6: Illustration of the grade-incremental setting for CAQA: Our setting is characterized by challenges of both classification and regression tasks. The interdependencies of non-stationary grades pose a new CL challenge for memorizing such dependencies.

For MTL-AQA, FineDiving, and UNLV-Dive, we partition them into five subsets, ensuring each subset has an equal number of samples according to their label distribution. JDM-MSA, with a limited number of actions, is divided into three subsets. This split creates a challenging protocol for CAQA models to handle the variability in quality scores, allowing us to evaluate their adaptability to real-world scenarios. To address timely updates and labeling scarcity, we train models with a few samples per session, reserving the remaining samples for base session fine-tuning. During inference, all samples in their subset are evaluated for a comprehensive assessment. Additionally, we assess the effectiveness and generalization of CAQA models in learning from limited instances by varying the number of training samples per session (refer to Fig. 8).

5.3 Evaluation Metrics

The Spearman's rank correlation coefficient, denoted as ρ quantifies the correlation between the ground truth $\boldsymbol{y}$, and its predicted score $\hat{\boldsymbol{y}}$. Given that $\boldsymbol{p}$ and $\boldsymbol{q}$ represent the rank vectors of $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$, ρ is defined as:

$$\rho = \frac{\sum_i (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_i (p_i - \bar{p})^2 \sum_i (q_i - \bar{q})^2}}, \tag{9}$$

where $\bar{p}$ and $\bar{q}$ denote the average values of $\boldsymbol{p}$ and $\boldsymbol{q}$.

To comprehensively assess the efficacy of CAQA models, we have adopted three CL metrics as outlined in [30]. We utilize the overall correlation, denoted as ρ_{avg}, as a metric to gauge total performance. This approach differs from previous methodologies [23,36], which tend to compute individual metrics in isolation. By aggregating samples from all preceding sessions, we aptly address the rank correlation's sensitivity to sample size variations. Furthermore, to quantify memory stability and learning plasticity, we employ the average forgetting ρ_{aft} and the forward transfer ρ_{fwt}, respectively. ρ_{aft} and ρ_{fwt} can be defined as:

$$\rho_{\text{aft}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \max_{i,j \in \{1,2,\cdots,T\}} \left(\rho_{i,t} - \rho_{j,t}\right), \tag{10}$$

$$\rho_{\text{fwt}} = \frac{1}{T-1} \sum_{t=2}^{T} \left(\rho_{t-1,t} - \tilde{\rho}_t\right), \tag{11}$$

where $\rho_{i,j}$ represents the correlation evaluated on the test set of the j-th task after incremental learning of the i-th task ($j \leq i$), and $\tilde{\rho}_t$ denotes the correlation of a randomly initialized reference model evaluated on that of the t-th task.

6 Experimental Results

Using PyTorch, all models are trained with two NVIDIA RTX 3090 GPUs. We adopted Adam with a learning rate and weight decay both set to 10^{-4}. Each training session spans a maximum of 50 epochs. We set the batch size b_1 to 5 and the mini-batch size b_2 to 3 across all models. We have frozen the batch normalization layer within the backbone to counterbalance the influence of batch size. In addition, we employed two MLP layers for the MP module, and both loss weight parameters, λ_{P} and λ_{R}, are set to 1. We present the main results here, with supplementary material offering additional details.

Comparison with Recent Strong Baselines. We compared MAGR against a series of baselines, including both memory-free methods [3, 10, 19, 23, 36] and memory-based approaches [9, 11, 39, 40]. Detailed results are provided in Tab. 1.

Table 1: 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.

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 [39]	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 [9]	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 [11]	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 [19]	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 [23]	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 [10]	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 [40]	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 [36]	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

Joint training represents the upper performance bound for CL. In contrast, using sequential fine-tuning, the lower bound leads to a significant drop in performance across all datasets, such as a 41.69% correlation decline on the MTL-AQA dataset. This verifies the challenge of catastrophic forgetting in the CAQA setting. While memory-free models make some effort, their performance pales in comparison to raw data replay and feature replay.

The raw data replay is effective but raises privacy concerns, where MER [19] achieves the best performance among them. Although feature replay can handle privacy concerns, it encounters challenges when the backbone undergoes fine-tuning in new sessions. This can lead to feature deviation from the current data manifold, impacting assessment performance. This issue can be confirmed when comparing the Feature MER (feature replay adaptation of MER) with that of MER, where the former performs poorly compared to the latter. NC-FSCIL [36] adopts a fixed backbone approach to avoid deviations but sacrifices adaptability to real-world complexity, resulting in lower performance than MER. Our method is specifically engineered to address feature deviation and thus achieves the best performance, underscoring its superior design and efficacy in CAQA.

We have noted that recent prompt-based approaches [26] have excelled in CL on pre-trained models. However, due to the distinctive nature of AQA pre-training in terms of model architectures and datasets, the use of single-task (ViT+prompt) remains unexplored. Recently, SLCA [40] has demonstrated superior performance to classical prompt-based approaches like DualPrompt [34]. Therefore, we compare MAGR with SLCA, which performs even worse than NC-FSCIL due to its consistently lower quality of generative features. This unstable generation affects the CAQA performance in mitigating catastrophic forgetting. This emphasizes the clear advantages of our method.

Ablation Study. In Tab. 2, we have conducted an ablation study on the MTL-AQA dataset. The first row represents the performance of our MAGR model. Each subsequent row delineates the performance by removing a core component from MAGR. From Tab. 2, removing the MP module results in the most significant drop in performance, underlining its central role in rectifying feature deviations and ensuring that old features align with the evolving data manifold. The profound impact on results without the IIJ-GR emphasizes its essentiality in preserving regressor alignment across sessions. Separately removing Intra-Inter Graph Regularizer (II-GR) and Joint Graph Regularizer (J-GR) clarifies that both local and global regularizations are vital for realizing CAQA. Each independently enhances performance over the scenario where both are excluded, proving the distinct contributions of both local and global components. While the performance drop of removing OUS is not as pronounced upon its exclusion, the role of OUS is beyond mere metrics. It embodies the efficiency of MAGR, ensuring a smart and compact replay by selecting the most representative exemplars, thus economizing storage. In sum, the ablation study verifies the collective importance of each module. Each component is not just an additive piece but brings a unique aspect.

Table 2: Ablation results on the MTL-AQA 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 res. 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)	0.8447 $^{\downarrow 6\%}$	0.0265 $^{\uparrow 16\%}$	0.1890 $^{\downarrow 1\%}$
w/o OUS (random)	0.8619 $^{\downarrow 4\%}$	0.0876 $^{\uparrow 293\%}$	0.1027 $^{\downarrow 46\%}$

Impact of Memory Size. The memory bank requires extra storage space to retain old features, constrained by limited resources. To trade off the performance and memory size, we varied the number of replayed samples per session in Fig. 7. MAGR demonstrates resilience within the range of 5 to 11 features per session, owing to the integration of OUS to sample representative features. However, when the sample number is 3, MAGR's performance degrades. This is likely because OUS may sample extreme boundary samples that fail to maintain old distributions effectively, whereas random sampling may perform better in such scenarios. We maintained 10 samples per session for a fair comparison in Tab. 1.

Robustness to Label Scarcity, Noises and Severe Deviations. Unlike classification, obtaining reliable quality scores for actions typically requires domain experts and specialized annotation procedures. This scarcity of labeled

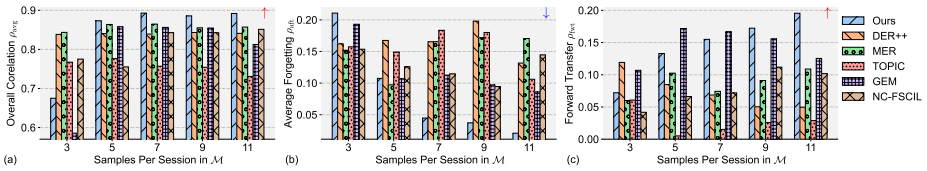

Fig. 7: Memory size comparisons with replay-based methods on MTL-AQA. ↑ indicates higher values are better for the metric, whereas ↓ indicates the opposite.

AQA data is a major challenge not addressed by most CL methods tested on plentiful classification benchmarks. We evaluated MAGR's performance under different label scarcity conditions (see Fig. 8(a)) and examined its robustness to noise (see Fig. 8(b)) on the MTL-AQA dataset. Different levels of label noise were only introduced to the training data for comparison against recent strong baselines [23, 36, 40], showcasing MAGR's effectiveness in learning from fewer labeled examples and robustness to noise. Additionally, we have visualized the correlation plots at noise intensity 9 to intuitively compare a recent baseline [36] (see Fig. 8(c)) and MAGR (see Fig. 8(d)), demonstrating MAGR's resilience to varying noise intensities. We further quantify the feature deviations between pre-trained and fine-tuned features of Joint Training in Tab. 3 (JDM-AQA is not considered due to the different feature scales). Our correlation gains compared to the strongest baseline increase significantly with feature deviations, indicating MAGR's robustness in handling severe deviations. This is due to the graph regularization's ability to preserve the feature space structure.

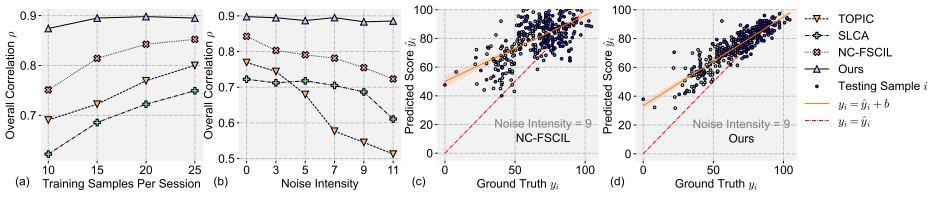

Fig. 8: Illustrations of label scarcity and noise robustness: (a) Training samples per session plot, (b) Noise intensity plot, (c) Correlation plot of NC-FSCIL at noise intensity 9, and (d) Correlation plot of MAGR at noise intensity 9.

Table 3: Statistics of feature deviations and correlation gains.

Dataset	FineDiving	MTL-AQA	UNLV-Dive
Degree of Feature Deviations (MSE)	26.85	35.28	51.75
Overall Correlation Gains (%)	5.66	6.56	15.64

Visualization of Mitigating Catastrophic Forgetting. We employed t-SNE [15] to project the features into the 2D space on the MTL-AQA dataset. Fig. 9(c) highlights the efficacy of MAGR in organizing samples across various sessions coherently, while Feature MER struggles with feature deviation (see Fig. 9(f)). We further showcase correlation plots at the end of the last session in the last column. From Fig. 9(g) and Fig. 9(h), it is evident to demonstrate our

superior correlation between the predicted scores and the actual ground truth, evaluating the effectiveness of maintaining feature alignment across sessions. For example, the ground truth score of sample #176 (see Fig. 10) is 73.85. Our method's prediction remains within a margin of 2 points from the ground truth score, indicating a closer alignment compared to Feature MER.

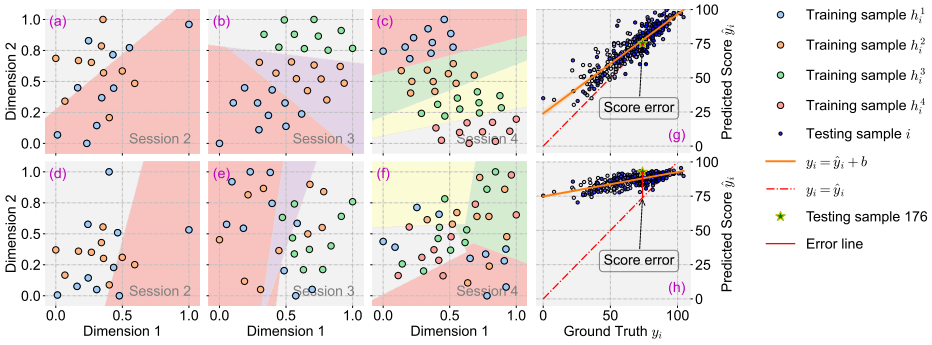

Fig. 9: Visualizations of feature distribution (a-f) and score correlation (g-h): MAGR (top) and Feature MER (bottom). The explicit division of different sessions validates the effectiveness of MAGR in mitigating catastrophic forgetting.

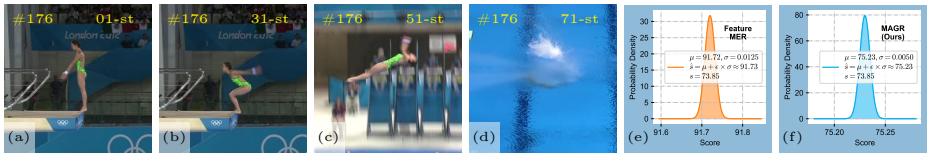

Fig. 10: Qualitative assessment comparison of sample #176 (a-d) using Feature MER (e) and MAGR (f) on the MTL-AQA dataset. We adopted the reparameterization trick [46] to obtain the predicted score, where μ and σ represent the mean and the standard deviation (please see our supplementary material for details).

7 Conclusions and Future Work

This work proposes the novel task of CAQA to accommodate real-world complexities. To mitigate catastrophic forgetting while prioritizing user privacy, we propose MAGR as a solution to address the misalignment issue due to backbone updates. By integrating MP and IIJ-GR, MAGR iteratively refines deviated old features and regulates the feature space across incremental sessions. Experiments on three AQA datasets show the superiority of MAGR compared to recent strong baselines. We believe this to enhance AQA systems in real-world applications, offering improved capabilities to serve our human beings. Future research should explore more robust designs capable of handling complex data. This may involve optimizing the number of layers and investigating advanced network architectures like ViT. In this way, incorporating prompt-based techniques could be considered to ensure parameter-efficient tuning and enhance adaptability.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Project 62272019, and also in part by the International Joint Doctoral Education Fund of Beihang University.

References

1. Bai, Y., Zhou, D., Zhang, S., Wang, J., Ding, E., Guan, Y., Long, Y., Wang, J.: Action quality assessment with temporal parsing transformer. In: European Conference on Computer Vision. pp. 422–438. Springer (2022)
2. Bao, Y., Lu, F.: Pcfgaze: Physics-consistent feature for appearance-based gaze estimation. arXiv preprint arXiv:2309.02165 (2023)
3. Buzzega, P., Boschini, M., Porrello, A., Abati, D., Calderara, S.: Dark experience for general continual learning: a strong, simple baseline. Advances in Neural Information Processing Systems **33**, 15920–15930 (2020)
4. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017)
5. Chi, Z., Gu, L., Liu, H., Wang, Y., Yu, Y., Tang, J.: Metafscil: A meta-learning approach for few-shot class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14166–14175 (2022)
6. Cortes, C., Mohri, M., Rostamizadeh, A.: Algorithms for learning kernels based on centered alignment. The Journal of Machine Learning Research **13**(1), 795–828 (2012)
7. Ding, X., Xu, X., Li, X.: Sedskill: Surgical events driven method for skill assessment from thoracoscopic surgical videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–45. Springer (2023)
8. Dong, S., Hong, X., Tao, X., Chang, X., Wei, X., Gong, Y.: Few-shot class-incremental learning via relation knowledge distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 1255–1263 (2021)
9. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., Hadsell, R.: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences **114**(13), 3521–3526 (2017). https://doi.org/10.1073/pnas.1611835114
10. Kukleva, A., Kuehne, H., Schiele, B.: Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9020–9029 (2021)
11. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(12), 2935–2947 (2017)
12. Liao, Y., Vakanski, A., Xian, M.: A deep learning framework for assessing physical rehabilitation exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering **28**(2), 468–477 (2020)
13. Liu, H., Gu, L., Chi, Z., Wang, Y., Yu, Y., Chen, J., Tang, J.: Few-shot class-incremental learning via entropy-regularized data-free replay. In: European Conference on Computer Vision. pp. 146–162. Springer (2022)
14. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3202–3211 (2022)

15. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(11) (2008)
16. Parmar, P., Morris, B.: Action quality assessment across multiple actions. In: 2019 IEEE Winter Conference on Applications of Computer Vision. pp. 1468–1476. IEEE (2019)
17. Parmar, P., Morris, B.T.: What and how well you performed? a multitask learning approach to action quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 304–313 (2019)
18. Parmar, P., Tran Morris, B.: Learning to score olympic events. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 20–28 (2017)
19. Riemer, M., Cases, I., Ajemian, R., Liu, M., Rish, I., Tu, Y., Tesauro, G.: Learning to learn without forgetting by maximizing transfer and minimizing interference. In: ICLR (2019)
20. Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., Wayne, G.: Experience replay for continual learning. Advances in Neural Information Processing Systems **32** (2019)
21. Smith, E.M., Williamson, M., Shuster, K., Weston, J., Boureau, Y.L.: Can you put it all together: Evaluating conversational agents' ability to blend skills. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 2021–2030. Association for Computational Linguistics (2020)
22. Tang, Y., Ni, Z., Zhou, J., Zhang, D., Lu, J., Wu, Y., Zhou, J.: Uncertainty-aware score distribution learning for action quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9839–9848 (2020)
23. Tao, X., Hong, X., Chang, X., Dong, S., Wei, X., Gong, Y.: Few-shot class-incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12183–12192 (2020)
24. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4489–4497 (2015)
25. Wang, L., Lei, B., Li, Q., Su, H., Zhu, J., Zhong, Y.: Triple-memory networks: A brain-inspired method for continual learning. IEEE Transactions on Neural Networks and Learning Systems **33**(5), 1925–1934 (2021)
26. Wang, L., Xie, J., Zhang, X., Huang, M., Su, H., Zhu, J.: Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality. Advances in Neural Information Processing Systems **36** (2024)
27. Wang, L., Yang, K., Li, C., Hong, L., Li, Z., Zhu, J.: Ordisco: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5383–5392 (2021)
28. Wang, L., Zhang, M., Jia, Z., Li, Q., Bao, C., Ma, K., Zhu, J., Zhong, Y.: Afec: Active forgetting of negative transfer in continual learning. Advances in Neural Information Processing Systems **34**, 22379–22391 (2021)
29. Wang, L., Zhang, X., Li, Q., Zhang, M., Su, H., Zhu, J., Zhong, Y.: Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nature Machine Intelligence **5**(12), 1356–1368 (2023)
30. Wang, L., Zhang, X., Su, H., Zhu, J.: A comprehensive survey of continual learning: Theory, method and application. arXiv preprint arXiv:2302.00487 (2023)
31. Wang, L., Zhang, X., Yang, K., Yu, L., Li, C., Hong, L., Zhang, S., Li, Z., Zhong, Y., Zhu, J.: Memory replay with data compression for continual learning. In: International Conference on Learning Representations (2021)

32. Wang, S., Yang, D., Zhai, P., Yu, Q., Suo, T., Sun, Z., Li, K., Zhang, L.: A survey of video-based action quality assessment. In: 2021 International Conference on Networking Systems of AI (INSAI). pp. 1–9. IEEE (2021)
33. Wang, T., Wang, Y., Li, M.: Towards accurate and interpretable surgical skill assessment: A video-based method incorporating recognized surgical gestures and skill levels. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 668–678. Springer (2020)
34. Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.Y., Ren, X., Su, G., Perot, V., Dy, J., et al.: Dualprompt: Complementary prompting for rehearsal-free continual learning. In: European Conference on Computer Vision. pp. 631–648. Springer (2022)
35. Xu, J., Rao, Y., Yu, X., Chen, G., Zhou, J., Lu, J.: Finediving: A fine-grained dataset for procedure-aware action quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2949–2958 (2022)
36. Yang, Y., Yuan, H., Li, X., Lin, Z., Torr, P., Tao, D.: Neural collapse inspired feature-classifier alignment for few-shot class-incremental learning. In: ICLR (2023)
37. Yao, L., Lei, Q., Zhang, H., Du, J., Gao, S.: A contrastive learning network for performance metric and assessment of physical rehabilitation exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2023)
38. Yu, X., Rao, Y., Zhao, W., Lu, J., Zhou, J.: Group-aware contrastive regression for action quality assessment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7919–7928 (2021)
39. Zenke, F., Poole, B., Ganguli, S.: Continual learning through synaptic intelligence. In: International Conference on Machine Learning. pp. 3987–3995. PMLR (2017)
40. Zhang, G., Wang, L., Kang, G., Chen, L., Wei, Y.: Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. arXiv preprint arXiv:2303.05118 (2023)
41. Zhang, J., Liu, L., Silven, O., Pietikäinen, M., Hu, D.: Few-shot class-incremental learning: A survey. arXiv preprint arXiv:2308.06764 (2023)
42. Zhang, S., Dai, W., Wang, S., Shen, X., Lu, J., Zhou, J., Tang, Y.: Logo: A long-form video dataset for group action quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2405–2414 (2023)
43. Zhao, L., Lu, J., Xu, Y., Cheng, Z., Guo, D., Niu, Y., Fang, X.: Few-shot class-incremental learning via class-aware bilateral distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11838–11847 (2023)
44. Zhou, K., Cai, R., Ma, Y., Tan, Q., Wang, X., Li, J., Shum, H.P., Li, F.W., Jin, S., Liang, X.: A video-based augmented reality system for human-in-the-loop muscle strength assessment of juvenile dermatomyositis. IEEE Transactions on Visualization and Computer Graphics **29**(5), 2456–2466 (2023)
45. Zhou, K., Li, J., Cai, R., Wang, L., Zhang, X., Liang, X.: Cofinal: Enhancing action quality assessment with coarse-to-fine instruction alignment. arXiv preprint arXiv:2404.13999 (2024)
46. Zhou, K., Ma, Y., Shum, H.P., Liang, X.: Hierarchical graph convolutional networks for action quality assessment. IEEE Transactions on Circuits and Systems for Video Technology (2023)