Ask and Remember: A Questions-Only Replay Strategy for Continual Visual Question Answering

Imad Eddine Marouf [1], Enzo Tartaglione [1], Stéphane Lathuilière [1,3], Joost van de Weijer [2]

[1]LTCI, Télécom-Paris, France [2]Computer Vision Center, Universitat Autónoma de Barcelona, Spain [3]Inria at University Grenoble Alpes, LJK, France

ICCV OCT 19-23, 2025 HONOLULU HAWAII

TL-DR; No images, no problem – QUAD reduces catastrophic forgetting with question-only replay, achieving robust learning without image storage.

Motivation and related work

- **Continual Visual Question Answering (VQACL)** aims to train models on a stream of visual-question tasks without forgetting previous knowledge.

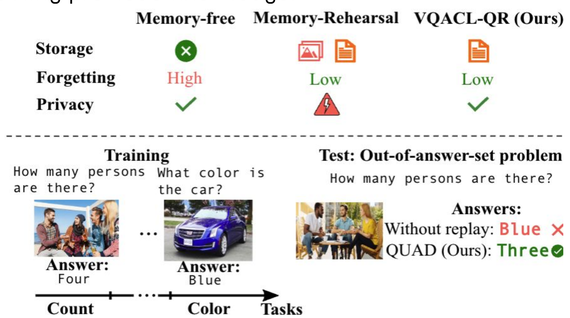

Figure 1: Comparison of continual Visual Question Answering (VQA) methods.

- **Memory-free** methods **forget too much** .
- **Rehearsal methods** require storing visual data → privacy issues.

  Can we preserve knowledge using only textual questions?

- **Out-of-answer set problem** occurs in continual VQA when a model forgets previously seen answers and starts predicting only from the current task's answer space, leading to biased or invalid responses.

Method: QUAD

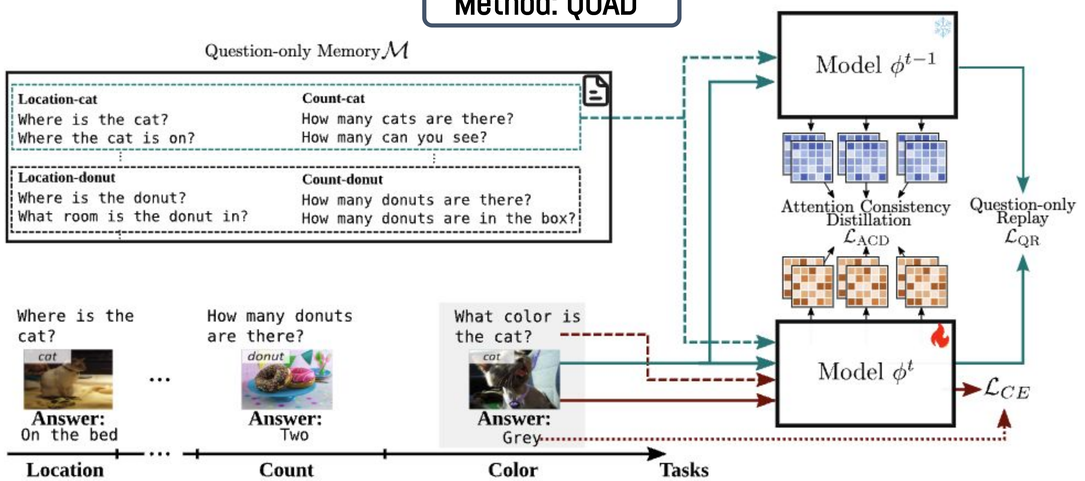

Figure 1: Overview of our QUAD method.

- **Question-Only Memory** 🧠: Stores only previous questions.

- **Question-Only Replay** 🔁: Pairs new images with old questions and uses soft labels from previous model.

$$\mathcal{L}_{\text{QR}} = \mathbb{E}_{x^t \sim \mathcal{T}^t} \mathbb{E}_{q^m \sim \mathcal{M}} \mathcal{L}_{\text{CE}} \left[\phi^t(x^t, q^m), \phi^{t-1}(x^t, q^m) \right]$$

- **Attention Consistency Distillation** 🔍: Aligns attention maps across tasks to preserve visual-linguistic focus.

$$\mathcal{L}_{\text{ACD}} = \mathbb{E}_{x^t \sim \mathcal{T}^t} \mathbb{E}_{q^m \sim \mathcal{M}} \mathbb{E}_{k \sim \mathcal{K}_\phi} \mathcal{L}_{\text{CE}} \left[A_k^t(x^t, q^m), A_k^{t-1}(x^t, q^m) \right]$$

- **Question Selection** 🎯: pick the right past questions to anchor the model without confusing it.

Result

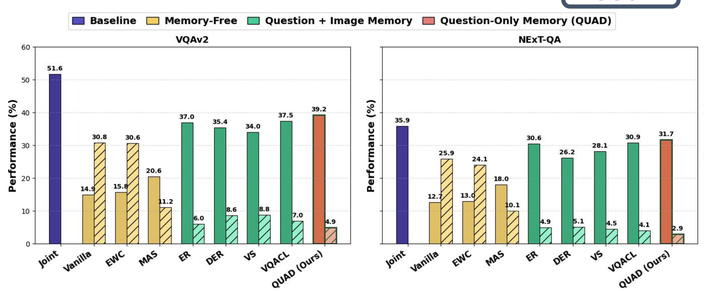

Figure 2: State-of-the-art comparison on VQAv2, and NExT-QA.

- QUAD achieves the best balance between accuracy and forgetting across both VQAv2 and NExT-QA, surpassing rehearsal-based methods without storing images.

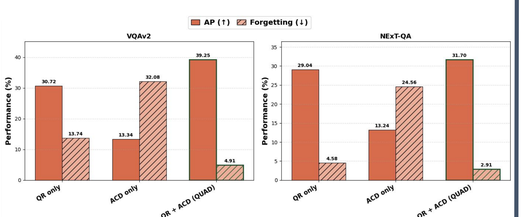

Table 2: Ablation study of QUAD components.

- Combination of replay and attention distillation (QUAD) achieves both high AP and minimal forgetting.

Takeaways

✅ Question-only replay is surprisingly effective.
✅ No visual data stored → privacy preserved.
✅ Attention distillation stabilizes model focus.

Acknowledgment. This work was supported by the French National Research Agency (ANR) JCJC program, the EU's Horizon Europe (ELIAS, No. 101120237), Hi!PARIS Center, and HPC resources of IDRIS under AD011013860.project.