Autoencoder-Based Hybrid Replay for Class-Incremental Learning

Milad Khademi Nori [1] Il-Min Kim [2] Guanghui Wang [1]

[1]Toronto Metropolitan University, Toronto, Canada [2]Queen's University, Kingston, Canada

Hybrid Replay Minimizes Memory & Achieves SOTA

Table 1. Hybrid replay strategy versus baseline strategies.

CIL Strategies	Memory	Compute	Performance
Generative Replay	$\mathcal{O}(cte)$	$\mathcal{O}(t)$	non-SOTA
Generative Classifier	$\mathcal{O}(t)$	$\mathcal{O}(cte)$	non-SOTA
Exemplar Replay	$\mathcal{O}(t)$	$\mathcal{O}(t)$	SOTA
Hybrid Replay (ours)	$\mathcal{O}(0.1t)$	$\mathcal{O}(t)$	SOTA

Autoencoder-Based Hybrid Replay (AHR)

Challenge: Current class-incremental learning methods either store raw exemplars with growing memory complexity $\mathcal{O}(t)$ or rely on generative replay that uses constant memory but suffers from poor-quality pseudo-data and catastrophic forgetting as tasks increase.

- **Theoretical Foundation:** We introduce a hybrid autoencoder (HAE) that jointly minimizes reconstruction error and enforces class-wise clustering via a charged-particle system energy minimization framework, guaranteeing compressed embeddings with bounded fidelity loss.
- **Unified Approach:** AHR blends exemplar and generative replay by storing compact latent codes (memory $\mathcal{O}(0.1t)$) and decoding them on demand for rehearsal, leveraging the decoder's memorization for faithful data recovery.
- **Novel Autoencoder:** HAE employs Coulomb-inspired energy equations and a Repulsive Force Algorithm (RFA) in the latent space to systematically disperse class centroids, enabling effective classification via Euclidean distance without expanding the architecture.
- **State-of-the-Art Results:** Across five benchmarks and ten baselines, AHR achieves SOTA accuracy using the same compute budget ($\mathcal{O}(t)$) while reducing memory footprint by an order of magnitude, demonstrating superior performance and efficiency.

Reconstruction Repulsion: Core AHR Equations

We jointly minimize reconstruction error and latent clustering via

$$L(x,\hat{x},z) = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{K_j^i} \|x_{j,k}^i - \hat{x}_{j,k}^i\|_2^2 + \lambda \|z_{j,k}^i - p_j^i\|_2^2, \quad (1)$$

which compresses inputs while clustering same-class embeddings. Class centroids (CCEs) are treated as charged particles whose potential energy is

$$U = \sum_{i,j} \frac{(q_j^i)^2}{2} \sum_{\substack{i',j' \\ (i',j')\neq(i,j)}} \frac{1}{\|p_{j'}^{i'} - p_j^i\|}, \quad (2)$$

and optimized via Coulomb-inspired repulsive forces to evenly disperse centroids.

From Pixels to Latents and Back

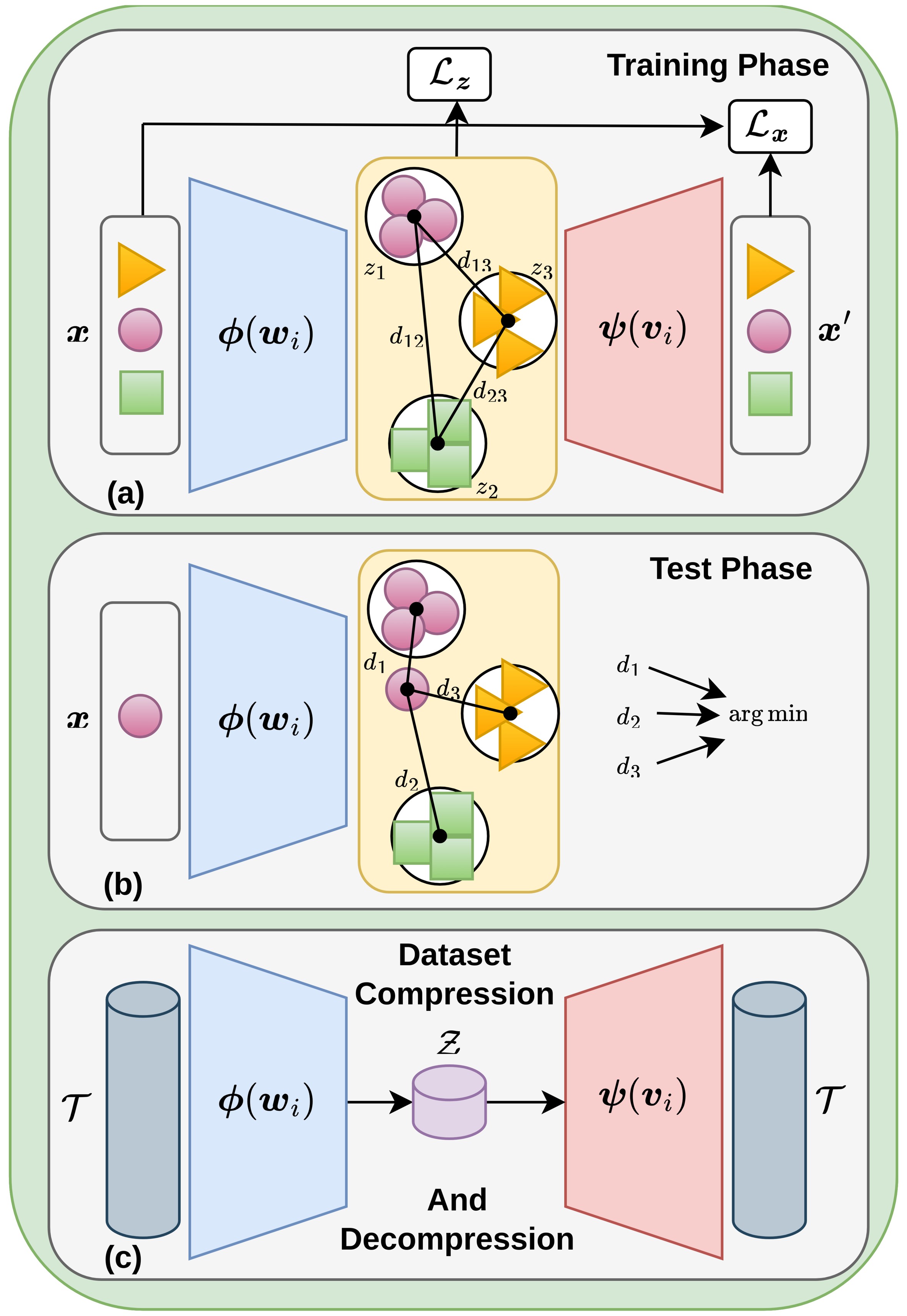


Figure 1. HR workflow. Task 1: Train autoencoder, store M_1. Tasks $h > 1$: Model Update Phase: Decode $M_{1:h-1}$, interleave with D_h, update model. Memory Update Phase: Compute and store $M_{1:h}$.

Workflow Overview: We train a hybrid autoencoder with charged-particle energy minimization to learn compact, class-aware latents and a decoder for faithful reconstruction.

Key Phases:

- **Training:** Minimize
$$\mathcal{L}_x = \|x - \psi(\phi(x))\|_2^2, \quad \mathcal{L}_z = \lambda \sum_{i,j} \|z_j^i - p_j^i\|_2^2.$$
- **Inference:** Classify by $\hat{y} = \arg\min_j \|\phi(x) - p_j\|_2$.
- **Compression:** Store latent buffer $\mathcal{M} = \{\phi(x)\}$ instead of raw images.
- **Decompression:** Reconstruct exemplars via $x' = \psi(m),\ \forall m \in \mathcal{M}$, for replay.

Experiments: Memory & Compute Efficiency

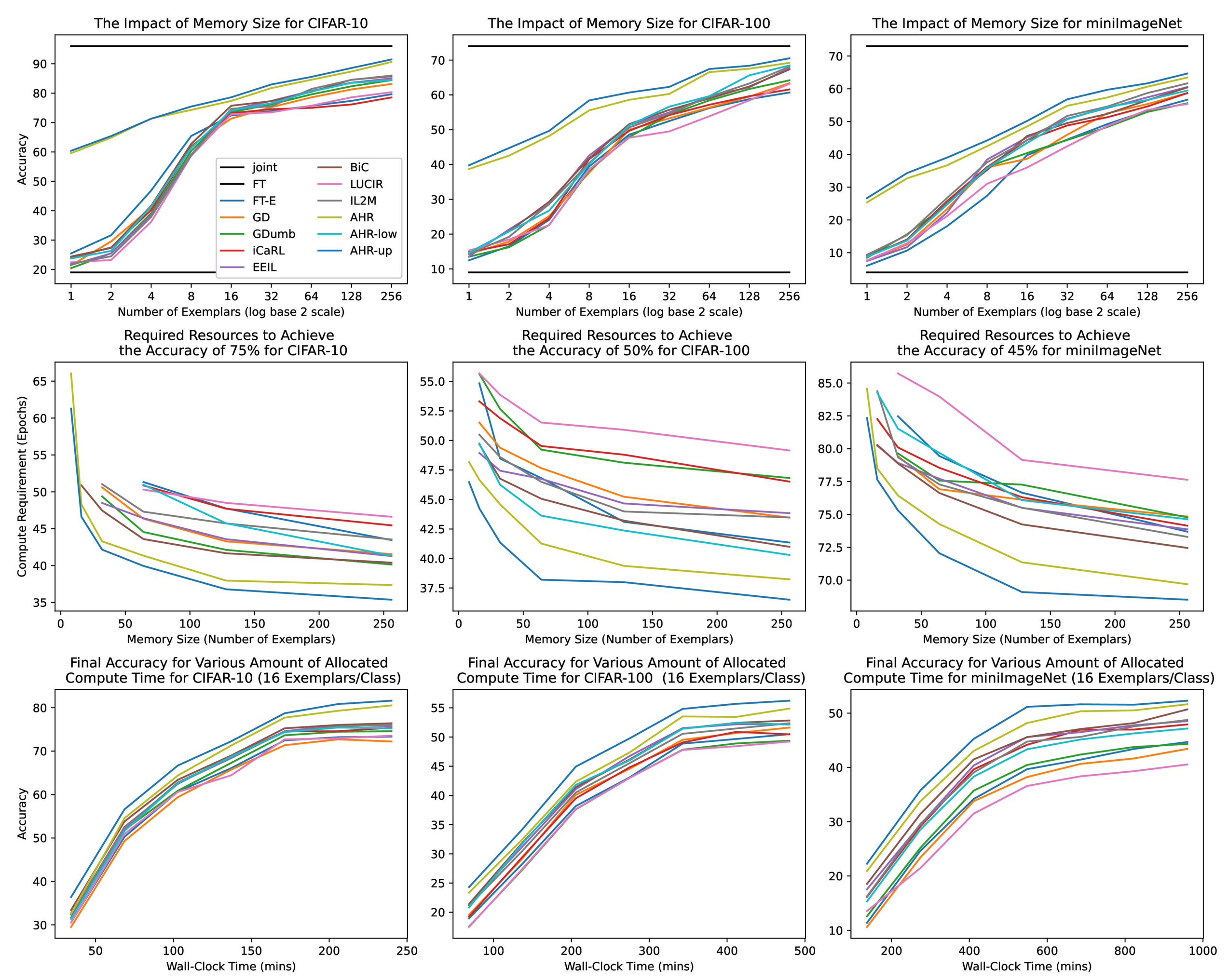


Figure 2. The impact of the memory size (first row). The required resources to achieve a target performance (second row). The achieved performance for a given compute time (third row).

The performance improvement of hybrid approaches, compared to exemplar rehearsal-based strategies, can be attributed to exemplar diversity (availability of more exemplars). The effectiveness of the availability of more exemplars, exemplar diversity, is well-documented in the literature.

Distortion of the Latent Exemplars at Each Increment

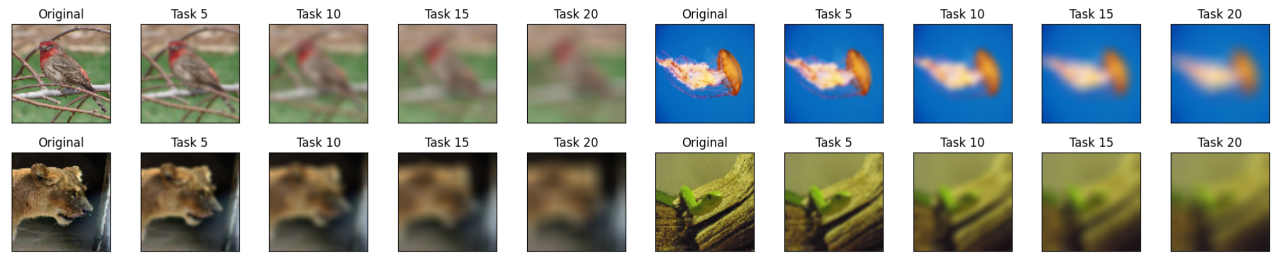
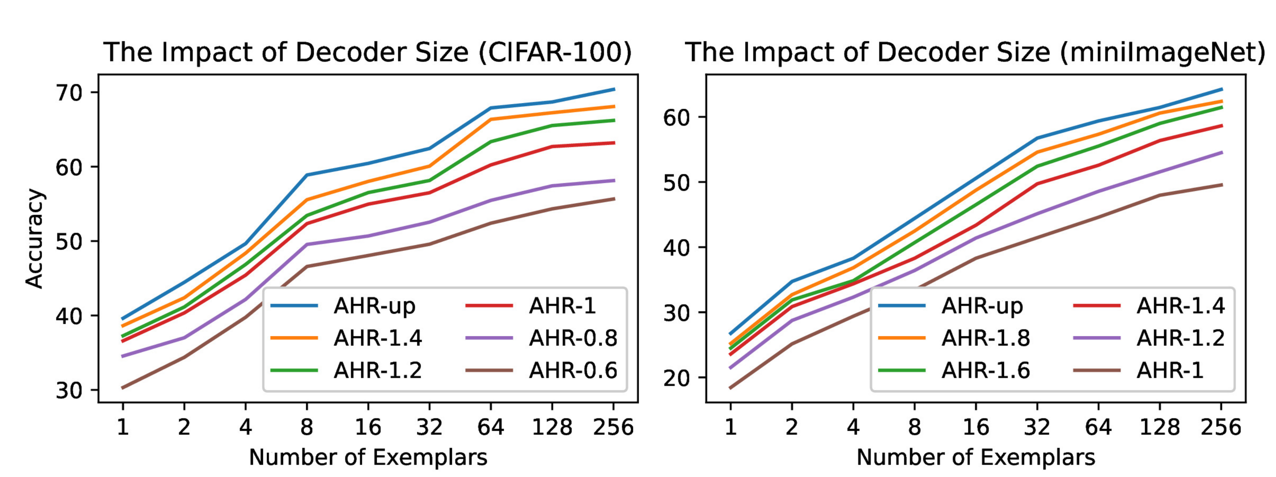


Figure 3. Images produced by the decoder at different tasks (decoder size of 1.8M).

Performances for various decoder/memory sizes



Performances for fixed memory (both the decoder and exemplars) and compute

Table 2. Performances for fixed memory (both the decoder and exemplars) and compute budgets.

Strategies	Benchmarks	# Exemplars	Memory		# Epochs	Wall-Clock Time	Performance
			Decoder	Exemplar			
AHR	CIFAR-100(10/10)	150 (latent)	1.4M	4.6M	50	462min	**54.43** ±0.93
	miniImageNet(20/5)	190 (latent)	1.8M	40.54M	70	842min	**48.09** ±0.64
BiC	CIFAR-100(10/10)	20 (raw)	-	6M	60	473min	52.12 ±0.91
	miniImageNet(20/5)	20 (raw)	-	42.34M	80	837min	45.23 ±0.62
IL2M	CIFAR-100(10/10)	20 (raw)	-	6M	60	455min	50.81 ±0.74
	miniImageNet(20/5)	20 (raw)	-	42.34M	80	861min	44.67 ±0.63
EEIL	CIFAR-100(10/10)	20 (raw)	-	6M	60	478min	51.70 ±0.79
	miniImageNet(20/5)	20 (raw)	-	42.34M	80	859min	41.83 ±0.55

AHR slashes the combined decoder + exemplar footprint to under 6 M parameters—yet still tops CIFAR-100 (54.43% vs. BiC's 52.12%) and miniImageNet (48.09% vs. BiC's 45.23%) under the same 50 – 70 epoch and ∼460 – 840 min budgets.

References

Generative Classifier Strategies

- Hayes, T. L. & Kanan, C. *Lifelong Machine Learning with Deep Streaming Linear Discriminant Analysis*. In *Proc. CVPR Workshops*, 2020.
- Van de Ven, G. M., Li, Z. & Tolias, A. S. *Class-incremental Learning with Generative Classifiers*. In *Proc. CVPR Workshops*, 2021.
- Zając, M., Tuytelaars, T. & Van de Ven, G. M. *Prediction Error-based Classification for Class-Incremental Learning*. arXiv:2305.18806, 2023.

Generative Replay Strategies

- Shin, H., Lee, J. K., Kim, J. & Kim, J. *Continual Learning with Deep Generative Replay*. NeurIPS Workshop on Continual Learning, 2017.
- Wu, C., Herranz, L., Liu, X., van de Weijer, J., Raducanu, B. & Tuytelaars, T. *Memory Replay GANs: Learning to Generate New Categories without Forgetting*. In *NeurIPS*, 2018.
- Van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. *Brain-inspired Replay for Continual Learning with Artificial Neural Networks*. *Nature Communications*, 11(1):4069, 2020.

Exemplar Replay Strategies

- Rebuffi, S.-A., Kolesnikov, A., Sperl, G. & Lampert, C. H. *iCaRL: Incremental Classifier and Representation Learning*. In *Proc. CVPR*, pp. 2001–2010, 2017.
- Castro, F. M., Marín-Jiménez, M. J., Guil, N., Schmid, C. & Alahari, K. *End-to-End Incremental Learning*. In *ECCV*, pp. 233–248, 2018.
- Lee, K., Lee, K., Shin, J. & Lee, H. *Overcoming Catastrophic Forgetting with Unlabeled Data in the Wild*. In *ICCV*, pp. 312–321, 2019.
- Prabhu, A., Torr, P. H. S. & Dokania, P. K. *GDumb: A Simple Approach that Questions Our Progress in Continual Learning*. In *ECCV*, pp. 524–540, 2020.
- Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y. & Fu, Y. *Large Scale Incremental Learning*. In *Proc. CVPR*, pp. 374–382, 2019.
- Hou, S., Pan, X., Loy, C. C., Wang, Z. & Lin, D. *Learning a Unified Classifier Incrementally via Rebalancing (LUCIR)*. In NeurIPS Workshop on Continual Learning, 2019.
- Belouadah, E. & Popescu, A. *IL2M: Class Incremental Learning with Dual Memory*. In *Proc. CVPR*, pp. 583–592, 2019.

Hybrid Replay Strategies

- Tong, S., Dai, X., Wu, Z., Li, M., Yi, B. & Ma, Y. *Incremental Learning of Structured Memory via Closed-Loop Transcription*. arXiv:2202.05411, 2022.
- Hayes, T. L., Kafle, K., Shrestha, R., Acharya, M. & Kanan, C. *REMIND: Remind Your Neural Network to Prevent Catastrophic Forgetting*. In *ECCV*, pp. 466–483, 2020.
- Wang, K., van de Weijer, J. & Herranz, L. *ACAE-REMIND for Online Continual Learning with Compressed Feature Replay*. *Pattern Recognition Letters*, 150:122–129, 2021.

arXiv:2505.05926
350 Victoria St, Toronto, ON M5B 2K3

miladkhademinori@gmail.com