SALIENCY-DRIVEN EXPERIENCE REPLAY FOR CONTINUAL LEARNING

G. Bellitto[1], F. Proietto Salanitri[1], M. Pennisi[1], M. Boschini[2], L. Bonicelli[2], A. Porrello[2], S. Calderara[2], S. Palazzo[1], C. Spampinato[1]

[1] University of Catania, Italy

[2] University of Modena and Reggio Emilia, Italy

Problem Background

Machine Learning models struggles with **Continual Learning** (CL) – the ability to learn new information without forgetting previously acquired knowledge.

Traditional models face **Catastrophic Forgetting** (CF) when exposed to non-stationary data streams, leading to a decrease in accuracy on previously learned tasks.

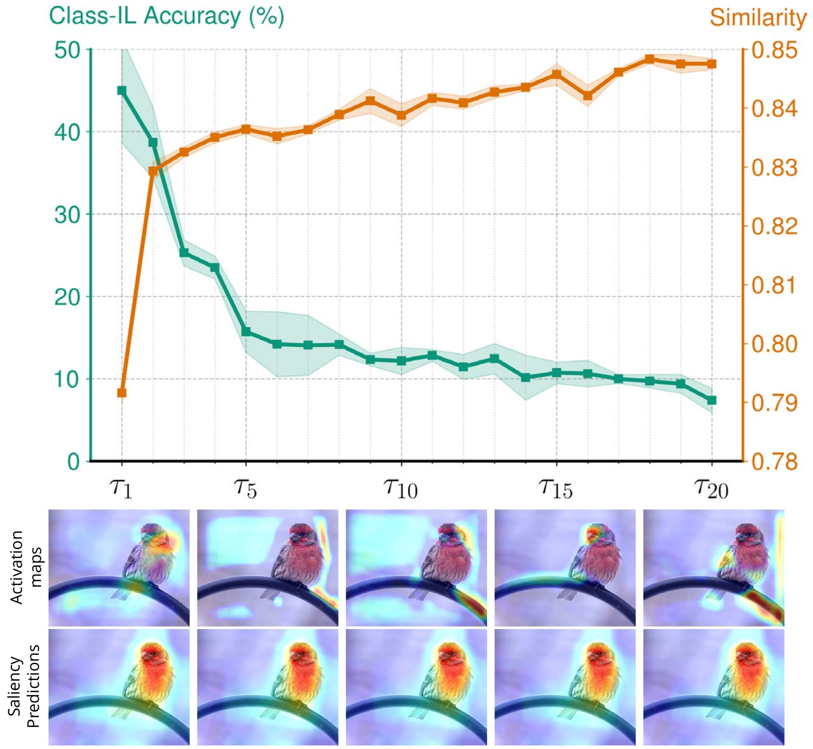
Our goal is to find a biologically inspired method to make CL more effective, reducing CF and making models more stable over time.

Human-Learning insight

- ➢In humans, the Visual System prioritizes salient information of the visual scene (e.g., movements, contrast, …)

- ➢**Selective Attention** retains an ancestral saliency bias, highlighting a stable, inherited visual processing trait that resists forgetting over time [1].

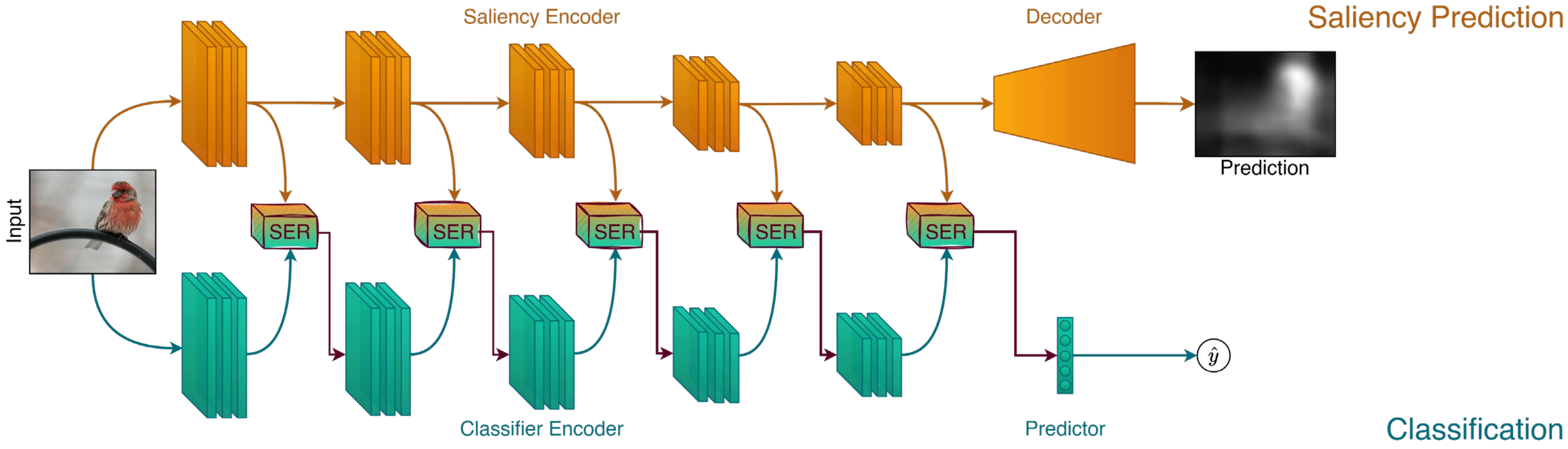
- ➢Visual features from a Saliency Predictor are **highly robust** to CF.



[1] J. New et al. “Category-specific attention for animals reflects ancestral priorities, not expertise”. National Academy of Sciences, 2007.

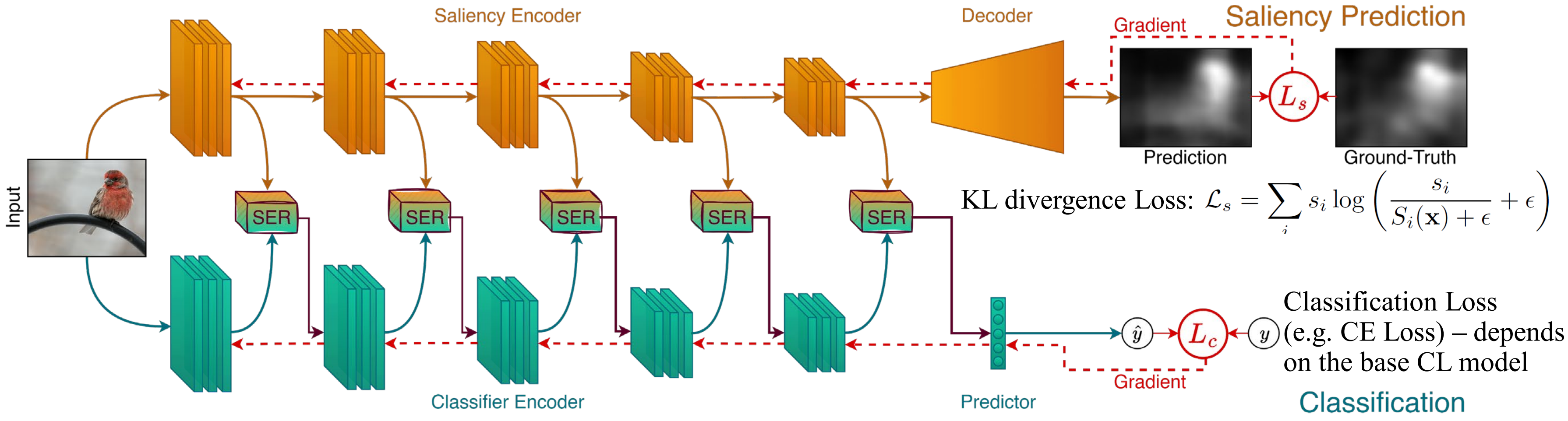
SER: Saliency-based Experience Replay

- A dual-branch architecture, with a Saliency Predictor **S** and a Continual Classifier **C**.
- **S** modulates the feature learning of **C** by emphasizing “salient” features at intermediate representations.



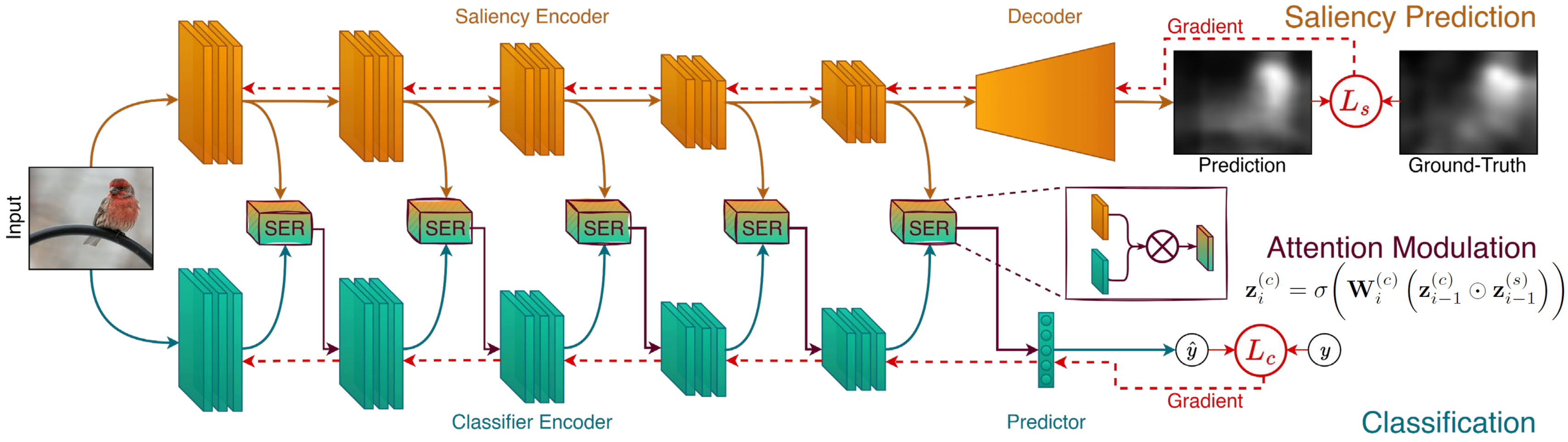
SER: Saliency-based Experience Replay

➢ **S** and **C** observe the same data stream, but they are trained with different objective functions.



SER: Saliency-based Experience Replay

- **S** and **C** observe the same data stream, but they are trained with different objective functions.
- Saliency modulation is performed through a Hadamard product between corresponding features.



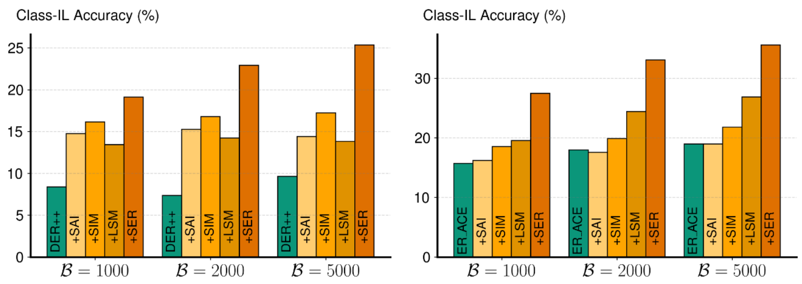
Performance Comparison

Table 1: **Class-Incremental accuracy of SOTA rehearsal-based methods** with and without SER.

Model	**Split Mini-ImageNet**			**Split FG-ImageNet**		
Joint	14.79±1.17			9.06±1.07		
Fine-tune	3.43±0.35			2.43±0.81		
Buffer size	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**
DER++	14.95±3.11	12.82±4.97	14.58±2.55	8.08±1.54	8.27±1.72	9.20±0.86
↪**SER**	**19.13**±1.62	**22.92**±2.25	**25.35**±2.56	**11.71**±2.36	**12.97**±1.62	**13.73**±1.95
ER-ACE	20.86±3.69	24.93±3.20	26.31±5.22	14.28±0.96	16.45±1.24	18.21±3.45
↪**SER**	**27.48**±2.83	**33.09**±1.28	**35.58**±1.79	**20.03**±3.13	**23.80**±2.11	**28.68**±0.50
CoPE	21.58±1.60	23.58±4.39	24.77±3.56	16.45±1.38	16.81±0.83	17.77±2.02
↪**SER**	**26.66**±2.22	**33.35**±4.67	**45.04**±2.44	**18.17**±2.79	**27.14**±1.62	**34.34**±3.51
	Dual-branch methods					
TwF	23.78±1.67	29.05±2.02	–	15.32±2.59	18.72±1.75	–
↪**SER**	**28.36**±3.72	**35.55**±0.61	–	**20.04**±1.63	**22.54**±2.20	–
DualNet	20.57±0.91	27.41±1.79	32.08±1.55	15.62±1.54	21.04±1.08	22.07±2.08
↪**SER**	**28.58**±1.40	**33.76**±1.21	**36.44**±0.77	**19.48**±0.59	**22.53**±1.56	**24.83**±2.01

DER++ : P. Buzzega et al. “Dark Experience for General Continual Learning”. NeurIPS 2020.
Er-ACE: L. Caccia et al. “New Insights on Reducing Abrupt Representation Change in Online Continual Learning”. ICLRW 2022.
CoPE: M. De Lange and T. Tuytelaars. Continual prototype evolution: Learning online from nonstationary data streams”. ICCV 2021.
TwF: M. Boschini et al. ”Transfer without forgetting”. ECCV 2022.
DualNet: Q. Pham et al. “Dualnet: Continual learning, fast and slow”. NeurIPS 2021.

Assessing saliency integration strategies



	Split Mini-ImageNet		Split FG-ImageNet	
SER Scheme	**DER++**	**ER-ACE**	**DER++**	**ER-ACE**
1 1 1 0 0	12.97±2.62	23.72±0.77	6.54±0.67	18.08±0.96
1 1 1 1 0	17.46±1.02	26.44±2.33	8.77±1.45	16.55±2.55
1 1 1 1 1	**22.92**±2.25	**33.09**±1.28	**12.97**±1.62	**23.80**±2.11

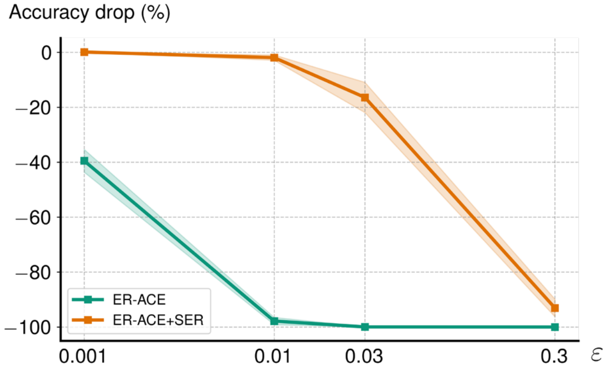
Alternative saliency integration methods evaluated:

- SAI: Saliency as additional input
- SIM: Saliency-based input modulation
- LSM: Learning saliency-based modulation

Selective-driven modulation applied across the entire network flow yields the best results, aligning with neurophysiological insights [2, 3]

[2] S. Treue and J. C. nez Trujillo. "Feature-based attention influences motion processing gain in macaque visual cortex". Nature, Jun 1999.
[3] J. C. Martinez-Trujillo and S. Treue. Feature-based attention increases the selectivity of population responses in primate visual cortex". Curr Biol, May 2004

Effects of saliency features on Model robustness



Method	Class-IL	Task-IL
ER-ACE	50.07±3.88	86.77±1.63
ER-ACE$^{\mathcal{SF}}$	28.46±3.46	74.40±4.37
$\hookrightarrow$**SER**	**44.08**±3.67	**83.04**±3.06

- In case of adversarial input space perturbations (PGD attack [4]), SER significantly improves model stability by reducing performance degradation through saliency-based feature regularization.
- Testing on an ad-hoc benchmark, SER recovers almost all the performance lost due to spurious features, making the model more stable and adaptable across tasks.

[3] A Madryet al. "Towards deep learning models resistant to adversarial attacks." ICLR 2018.

Conclusions

SER, a biologically-plausible approach based on replicating human visual saliency to enhance classification models in CL.

By incorporating saliency-driven modulation, SER improves state-of-the-art CL methods, reducing forgetting.

The saliency-based modulation significantly enhance robustness to adversarial attacks

SER highlights the potential of integrating neurophysiological principles to advance CL in AI systems.

Università di Catania

PeRCeiVe Lab

NEURAL INFORMATION PROCESSING SYSTEMS

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

THANK YOU FOR WATCHING!

Paper and code are available here:

SCAN ME

giovanni.bellitto@unict.it