Supplementary Material
NCAdapt: Dynamic adaptation with domain-specific Neural Cellular Automata for continual hippocampus segmentation

Amin Ranem[1] John Kalkhof[1] Anirban Mukhopadhyay[1]

[1] Technical University of Darmstadt, amin.ranem@gris.tu-darmstadt.de

1. Cardiac Datasets

We focus on the continual segmentation of the left ventricle (LV) in cardiac MRIs, utilizing data from the Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) [1], Table 1. This dataset includes 75 labeled cases acquired with Siemens scanners and 75 cases acquired with Philips scanners. Although the original dataset includes annotations for the left and right ventricles as well as the myocardium, our study specifically targets LV segmentation.

Table 1. Image and label characteristics of the used cardiac datasets.

Dataset	Siemens	Philips
# Cases	75	75
Resolution	[12 256 256]	[12 256 256]

The multi-class nature of the original problem, which involves different anatomical structures, allows us to investigate how segmentation performance varies depending on the shape and size of the region of interest. However, by focusing solely on the LV, we concentrate on the challenges posed by this particular structure while ensuring that the method is robust to variations across different scanners and patient populations. For each dataset, we maintain a consistent split by using 20% of the data for testing purposes across all experiments.
This targeted approach allows us to effectively evaluate the adaptability and performance of our method in a controlled yet challenging scenario, demonstrating the potential benefits of continual learning (CL) in hippocampus and cardiac image segmentation.

2. CL performance for LV segmentation

In this section, we evaluate the CL performance of our method specifically for left ventricle (LV) segmentation in cardiac MRIs. The emphasis is on evaluating how effectively the model generalizes and adapts to new tasks while preserving knowledge from previous tasks beyond the initial hippocampus segmentation task from the main manuscript. We utilize backward transfer (BWT) and forward transfer (FWT) metrics defined in the main manuscript to measure the model's ability to learn incrementally and avoid catastrophic forgetting.

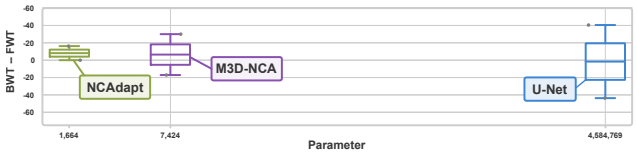

Figure 1. Comparison of BWT and FWT performance against number of parameters for NCAdapt and sequential U-Net trained on the Cardiac datasets; smaller boxes indicate superior performance. Note: Values above zero on the y-axis represents FWT and values below BWT.

To provide a comprehensive evaluation of continual learning (CL) performance, Figure 1 presents a comparison of backward transfer (BWT) and forward transfer (FWT) metrics across different models, including NCAdapt, sequential U-Net, and TransU-Net, all trained on the Cardiac datasets. The figure highlights the relationship between the number of parameters and the CL performance, with smaller boxes indicating superior performance. This visual comparison helps illustrate how each model balances the trade-off between model complexity and learning efficiency. Notably, positive values on the y-axis represent FWT, while negative values represent BWT, offering insights into how well the models generalize to new tasks while retaining knowledge from previous ones.
In addition, Table 2 summarizes the overall CL performance of the final models on the Cardiac datasets, including mean Dice scores, BWT and FWT metrics, along with the total number of trainable parameters, training runtime, and in-

Table 2. CL performance of the final model on the Cardiac datasets; mean Dice, BWT and FWT over all tasks including standard deviation, total amount of trainable parameters, training runtime and inference time in seconds; best values for sequential setup are marked in bold. Methods marked with * are rehearsal-based methods.

Method	Fixed param	Tuned param	Dice ↑ [%]	BWT ↑ [%]	FWT ↑ [%]	# Parameters (train) ↓	Runtime [sec] ↓	GPU sec ↓
$\text{Sequential}_{\text{U-Net}}$	–	–	57.63 ± 33.24	-43.73	48.97	4,584,769	$\mathbf{8.63}$	1.05
$\text{Sequential}_{\text{NCA}}$			72.67 ± 15.76	-17.10	-29.98	7,424	24.0	1.20
$\text{Sequential}_{\text{NCAdapt}}$			71.85 ± 1.83	$\mathbf{-0.001}$	-16.22	**1,664**	19.16	$\mathbf{0.123}$
EWC_{NCA}	–	$\lambda = 0.4$	71.28 ± 17.29	-18.60	-26.24	7,424	23.08	1.15
$\text{RWalk}_{\text{NCA}}$	$\alpha = 0.9$	$\lambda = 0.4$	67.58 ± 6.15	0.919	-23.90	7,424	23.59	1.30
$\text{SI}^{*}_{\text{U-Net}}$	$c = 0.4$	–	56.30 ± 33.42	-45.13	-39.07	4,584,769	8.80	1.11
$\text{FDR}^{*}_{\text{U-Net}}$	–		58.29 ± 31.75	-47.24	-45.28		9.54	1.14
$\text{DER}^{*}_{\text{U-Net}}$	$\alpha = 0.4$		59.70 ± 28.94	-33.15	-39.33		10.04	1.00
$\text{AGem}^{*}_{\text{U-Net}}$	–		59.33 ± 29.71	-33.70	-49.41		10.15	1.00

ference time. Since the setup involves two datasets, only a single BWT and FWT value for each model can be reported, i.e. with standard deviation 0%. This table provides a detailed comparison of the models, with the best values for the sequential setup highlighted in bold. Methods marked with an asterisk (*) are rehearsal-based, indicating that they leverage previously seen data during training to improve CL performance. This table serves as a key reference for understanding the trade-offs between performance, computational cost, and memory efficiency in the context of LV segmentation.

References

[1] Victor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martin-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021. 1