





FOCUS: Feature Replay with Optimized Channel-Consistent Dropout for U-Net Skip-Connections

Simon Johannes Joham¹, Franz Thaler^{1,2}, Arnela Hadzic¹, and Martin Urschler¹

¹ Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria

² Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria
`martin.urschler@medunigraz.at`

Abstract. We address image segmentation in the domain-incremental continual learning scenario, a use-case frequently encountered in medical diagnostics where privacy regulations and storage constraints prevent access to historical data. In this scenario, segmentation models must learn to cope with new domains (e.g., difference in imaging protocols or patient population) while maintaining performance on previously learned domains without full access to past data. Feature-based replay addresses the privacy concerns by only storing latent feature representations instead of original images. However, existing feature replay approaches have a critical limitation: they sacrifice U-Net skip-connections, which are essential for achieving high segmentation accuracy and fast convergence. This limitation significantly impacts clinical viability, especially when alternatives such as full model retraining or maintaining domain-specific models are available. Therefore, we propose feature replay with optimized channel-consistent dropout for U-Net skip-connections (FOCUS). FOCUS enables crucial skip-connections in feature replay while respecting privacy and storage constraints, and integrates recent domain generalization techniques based on data augmentation. Evaluation across two domain-incremental continual MRI segmentation settings demonstrates that FOCUS achieves substantial improvements (up to 21% average DSC) over existing methods, while saving only 0.5% of the original feature information per domain. The code is available at <https://github.com/imigraz/FOCUS/>.

Keywords: continual learning · segmentation · deep learning.

1 Introduction

In domain-incremental continual segmentation, models are sequentially trained on datasets from different domains and need to adapt continuously while also maintaining performance on past domains. Domains in clinical practice can be

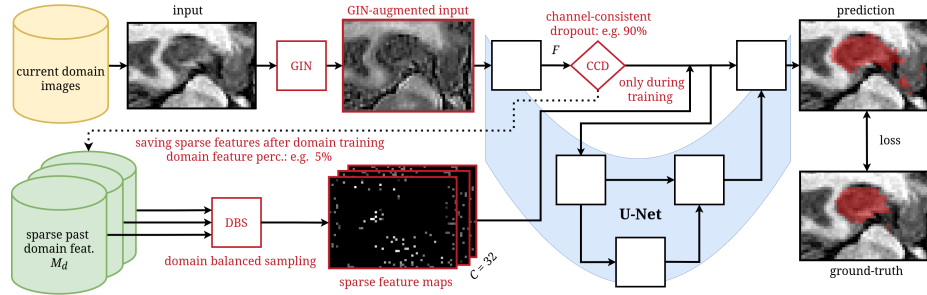


Fig. 1. Overview of FOCUS. Contributions are highlighted in red.

understood as differences in patient populations, imaging protocols, or disease categories [6,20]. However, storing several large medical datasets can be impractical [6], and access to past patient data may be restricted by privacy regulations [5]. In settings where access to past data is limited, models become prone to catastrophic forgetting, where they lose their ability to perform well on previously learned domains when trained on new data. The alternative of maintaining domain-specific models introduces its own challenges, demanding precise domain identification during inference and increased resource requirements. Therefore, it is vital to develop continual learning approaches that allow a single model to adapt to new data, preserve performance on previous tasks, manage computational resources, and adhere to privacy constraints.

Replay-based approaches have proven most successful at mitigating catastrophic forgetting in continual learning, for both natural [23] and medical semantic image segmentation [7]. Exemplar-replay methods preserve and reuse actual image samples [2,6,7,17], which, however, raises privacy and storage concerns [23]. Instead of storing actual samples, generative replay techniques synthesize training samples using generative models trained on previous domains [12]. Lastly, there are methods [11,16] that rely on latent feature representations obtained from the model instead of original data, which are domain-invariant and low-dimensional, thus offering privacy protection. Furthermore, they reduce the computational complexity for generative-replay [11]. However, existing feature replay methods in medical continual segmentation [2,11] do not use U-Net [19] skip-connections, as they rely on the lowest dimensional features. Nevertheless, these are vital for high segmentation performance and convergence speed [24].

To lift this limitation, we propose feature replay with optimized channel-consistent dropout for U-Net skip-connections (FOCUS), as shown in Figure 1. FOCUS boosts segmentation performance by employing feature replay with vital skip-connections and retaining first level encoder features. To comply with privacy and storage constraints, we introduce channel-consistent dropout (CCD), which deletes 90% of the spatial information before saving features, and overall save only 5% of the domain features. Moreover, we propose a domain-stratified feature sampling scheme with domain balanced feature sampling (DBS) to retain good performance on all past domains. Lastly, FOCUS leverages recent ad-

vances in domain generalization [15,22] with global intensity non-linear augmentation (GIN) [15], not investigated in continual learning before. GIN promotes the encoder to learn domain-invariant shape features. Our evaluation using two popular continual learning settings (prostate and hippocampus) demonstrates that FOCUS notably outperforms state-of-the-art domain-incremental continual MRI segmentation methods.

2 Methods

Our method uses direct replay of stored latent domain features. Using the basic framework of [11], we apply feature replay to the established U-Net architecture [19], which contains skip-connections between encoder and decoder. In our setting of sequential domains, we collect features from all patients after the training phase of the current domain d , based on 2D image slices of the original 3D inputs. We store output feature maps from the first encoder level $F \in \mathbb{R}^{H \times W \times C}$, where H and W are the spatial dimensions and C is the number of feature channels, in a domain-specific memory buffer M_d . Storing features from the first encoder level is sufficient for supporting skip-connections, as feature maps from lower levels can be computed. Starting with the second domain, we interleave training between new domain images and our feature replay that samples features F from each M_d , alternating between current domain and feature replay after each batch of input images. Stored features are directly inserted as the output of the first convolution block during replay, as shown in Figure 1.

FOCUS enhances feature replay with skip-connections with three contributions: (1) we delete spatial information from features F adhering to privacy and storage constraints with channel-consistent dropout (CCD), (2) domain balanced feature sampling (DBS) that enhances feature sampling to balance across domains and (3) integration of GIN [15] that lets the model focus on domain-invariant shape features.

Channel-Consistent Dropout. Storing first-level encoder feature maps F introduces privacy and storage concerns as they contain patient-specific information and scale with C . To address this, we introduce channel-consistent dropout (CCD), defined as

$$\text{CCD} : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}^{H \times W \times C} \text{ with } \text{CCD}(F, p)_{h,w,c} := F_{h,w,c} D_{h,w}, \quad (1)$$

where $D \in \{0,1\}^{H \times W}$ with $D_{h,w} \sim \text{Bernoulli}(1-p)$ and p is the dropout probability (e.g., $p=0.9$ for 90% CCD). CCD can be understood as a Hadamard product between F and D , applied channel-wise.

Our approach provides two key advantages: First, it enables efficient feature compression through spatial dropout [21] that is shared over all channels, reducing the storage complexity from scaling with C to being constant for all locations, where dropout is selected. Second, it provides privacy through a multi-layered protection scheme. CCD applies high spatial dropout (e.g., 90% CCD) uniformly across channels and leverages non-invertible transformations from the

first convolution block that includes non-linear activations. Furthermore, we remove information by keeping only a small percentage (e.g., 5%) of overall features F from each domain d (domain feature percentage, DFP). This prevents spatial consistency in 3D that is not needed for effective 2D feature replay. We select the samples to be dropped using patient-stratified random sampling, which preserves balanced patient representation. Overall, this combination of CCD and DFP makes original image reconstruction very challenging, as it requires recovery of spatial information, despite high sparsity, non-linear mappings, and inconsistent 3D patterns.

During training, we apply CCD to 25% of samples within each batch of current domain images. This ensures both encoder and decoder adapt to CCD-induced perturbations that are always present in feature replay while maintaining performance. Importantly, CCD is only applied during training.

Domain Balanced Feature Sampling. Domain balanced feature sampling (DS) is designed to both balance across domains and to effectively sample relevant anatomical structures within each domain d . Instead of randomly sampling over all saved features of all M_d , which would bias toward larger datasets with more available features, we equally weight each d for better generalization across domains. Specifically, given a batch with size B and N domains, we allocate B/N samples per domain. Within each domain’s batch partition, we follow nnU-Net’s sampling strategy where 33.3% of patches are guaranteed to contain foreground classes while the remaining 66.7% are sampled randomly (foreground oversampling was highlighted as missing in the implementation of [11]).

Global Intensity Non-Linear Augmentation. After training the first domain, we freeze the encoder weights, otherwise there is a severe distribution shift in the features, which would render feature replay ineffective [16]. This is a significant shortcoming of feature replay, as it depends on the assumption that the encoder can learn robust domain-invariant features already from the first domain. Therefore, we take advantage of recent advances in domain generalization [15,22] that have shown promising results through strong data augmentation such as global intensity non-linear augmentation (GIN) [15]. To the best of our knowledge, we are the first to adapt this strategy to continual segmentation. GIN consists of a series of convolution layers and is used for data augmentation by randomly re-initializing convolution kernels. During training, GIN augmentation is stochastically applied with 50% probability to a batch of input images from the current domain after standard data augmentation. In this way, GIN promotes the encoder, before being frozen after the first domain, to learn domain-invariant shape features. Please note that GIN is not applied during the feature extraction phase when populating M_d , to keep domain-specific information intact.

2.1 Comparison Methods

We evaluate our exemplar-based feature replay approach against established continual learning methods [4,9] and recent techniques specifically developed for domain-incremental medical image segmentation [11,25]. Elastic weight consolidation (EWC) [9] is a regularization-based approach built on the idea to reg-

ularize the change of weights that are important for past tasks. For knowledge distillation-based approaches, we investigate modeling the background (MiB) [4] and tri-enhanced distillation framework (TED) [25]. While both methods target semantic segmentation, TED is specifically designed for domain-incremental medical image segmentation. Furthermore, we compare our method with double conditioned variational auto-encoder (ccVAE) [11], which employs generative-replay with bottleneck features. Contrary to our method, ccVAE disables skip-connections and is computationally and storage expensive, as there is the need to train and save a large fully connected VAE model per domain. EWC, TED and MiB storage complexity scales with the number of U-Net weights and thus is negligible. All comparison methods are considered privacy-protecting (PP).

3 Experimental Results

We use the continual learning setting of [7], where each model trains sequentially on several domains. Per step, only current domain images and labels are accessible. Still, the model should perform well on all past and current domains.

3.1 Datasets

We followed the dataset configuration from [11] using public domain-incremental MRI datasets. For continual prostate segmentation, we utilized a multi-site dataset [13] with the sequential training order: BIDMC (12 subjects) \rightarrow I2CVB (19 subjects) \rightarrow HK (12 subjects) \rightarrow UCL (13 subjects), where MRI acquisitions alternated between using and not using endorectal coils. For hippocampus segmentation, we employed three datasets in sequence: DecathHip [1] (130 subjects, both healthy and schizophrenia patients), Dryad [10] (25 healthy subjects), and HarP [3] (135 elderly subjects with and without Alzheimer’s disease). While DecathHip provided pre-cropped volumes of interest (VOI) containing either the left or right hippocampus, we extracted VOIs containing both hippocampi from Dryad and HarP to prevent patient overlap between training and testing sets.

3.2 Implementation Details

All experiments use the popular nnU-Net framework [8] that was extended for continual learning by [7]. Same as previous work [11,25], we use a 2D U-Net. Following [11], we split datasets into training (56%), validation (24%), and test (20%) sets. FOCUS uses 90% CCD and 5% DFP. Moreover, for experiments with DFP set to 5%, we average over five runs. First level-encoder features F have 32 channels C . EWC, MiB, and ccVAE use default parameters from [11] given identical settings. For TED, we evaluate the knowledge distillation weighting parameter $\lambda \in \{1, 0.1, 0.001\}$ and only report the best result. Furthermore, we enhanced the loss functions that TED introduces to be compatible with the deep supervision feature of nnU-Net. The baseline consists of naive sequentially trained U-Net (Seq.) and individual domain models (Individ.). Furthermore, we

Table 1. Continual hippocampus test set performance. Bold values indicate the best performing method per column. Underlined values indicate second-best performance. Methods at the bottom and top are complementary and excluded from comparison.

Method	CL metrics \uparrow			DSC (%) after last domain \uparrow		
	AVG	BWT	FWT	DecathHip	Dryad	HarP
Individ. (250 ep.)	–	–	–	89.2 \pm 2	89.8 \pm 1	84.8 \pm 8
Seq.	85.5 \pm 3	-7.4	1.5	77.9 \pm 7	86.8 \pm 3	86.8 \pm 9
EWC [9]	85.2 \pm 5	<u>0.0</u>	-15.9	88.9\pm3	81.3 \pm 1	61.9 \pm 14
MiB [4]	<u>87.8\pm2</u>	<u>-3.1</u>	1.3	84.6 \pm 5	<u>88.8\pm2</u>	86.6 \pm 9
TED [25] ($\lambda=0.001$)	86.3 \pm 3	-6.0	<u>1.6</u>	80.3 \pm 6	87.4 \pm 3	<u>86.9\pm9</u>
ccVAE [11]	79.5 \pm 8	8.7	-14.0	<u>87.7\pm4</u>	87.9 \pm 2	83.2 \pm 8
Seq.+GIN	86.2 \pm 3	-6.5	1.5	82.9 \pm 4	87.2 \pm 3	87.0\pm9
FOCUS	89.5\pm1	-1.6	1.7	88.9\pm3	89.2\pm2	86.8 \pm 9
↳(No Skip, No CCD, 100% DFP)	77.1 \pm 12	10.6	-18.6	88.7 \pm 2	88.4 \pm 1	83.5 \pm 8
↳(Skip, No CCD, 5% DFP)	88.9 \pm 1	-1.3	1.4	88.5 \pm 3	89.2 \pm 2	86.8 \pm 9

also extend Seq. with GIN (Seq.+GIN), which was not done before in continual learning literature. ccVAE experiments ran on a system with dual A100 GPUs (each 80GB VRAM) with 128GB CPU RAM. Other experiments could run on a workstation with GTX 3090 (24GB VRAM, 64GB RAM).

3.3 Evaluation Metrics

Segmentation Metrics. We use Dice Similarity Coefficient (DSC) [2,7,11]. Furthermore, for our ablation study, we adopt a dual-metric approach using both DSC (counting-based) and Mean Average Surface Distance (MASD, distance-based). While our model operates on 2D slices, we employ 3D evaluation metrics for accuracy, utilizing the MetricsReloaded [14] implementation for MASD³.

Continual Learning Metrics. We evaluate continual learning using average backward transfer (BWT) [18] computed over all continual learning time-steps and forward transfer (FWT) as defined in [7]. BWT quantifies how learning new tasks affects performance on past learned tasks. FWT evaluates how knowledge from previously seen domains affects learning of new domains. Additionally, we evaluate the average segmentation performance on previously seen domains over all time-steps (AVG) [18].

3.4 Segmentation Results

Hippocampus Setting. This setting consists of a large collection of small images, e.g. 260 small-resolution images ($36\times50\times35$) for DecathHip, representing a straightforward segmentation task after VOI cropping (Table 1, Figure 2). In the continual learning context, sequential training (Seq.) exhibits substantial catastrophic forgetting (-7.4 BWT). While EWC has neutral BWT, its limited plasticity impairs adaptation to new domains, resulting in lower average

³ <https://github.com/Project-MONAI/MetricsReloaded/> last accessed 03.02.2025

Table 2. Continual prostate setting performance.

Method	CL metrics \uparrow			DSC (%) after last domain \uparrow			
	AVG	BWT	FWT	BIDMC	I2CVB	HK	UCL
Individ. (250 ep.)	—	—	—	67.3 \pm 34	83.2 \pm 2	87.6 \pm 5	80.6 \pm 7
Seq.	57.4 \pm 8	-30.5	<u>2.0</u>	57.3 \pm 20	45.3 \pm 17	42.4 \pm 33	<u>86.4\pm4</u>
EWC [9]	48.9 \pm 17	-21.6	-37.7	53.5 \pm 23	15.7 \pm 14	26.2 \pm 8	46.1 \pm 17
MiB [4]	59.0 \pm 7	-31.7	0.6	55.9 \pm 10	37.2 \pm 11	27.6 \pm 30	83.1 \pm 7
TED [25] ($\lambda=0.1$)	52.7 \pm 9	-23.3	-18.1	31.1 \pm 22	55.4 \pm 14	67.3 \pm 0	75.8 \pm 9
ccVAE [11]	53.8 \pm 15	2.6	-10.7	66.5 \pm 23	37.7 \pm 28	66.1 \pm 16	69.3 \pm 5
Seq.+GIN	<u>76.5\pm11</u>	-20.5	2.8	65.8 \pm 4	79.5 \pm 2	<u>72.6\pm11</u>	86.7\pm3
FOCUS	80.5\pm7	<u>-10.3</u>	0.7	75.7\pm6	81.2\pm3	84.3\pm6	83.3 \pm 5

performance (85.2 \pm 5) compared to Seq. (85.5 \pm 3). Despite incorporating hyperparameter tuning and deep supervision, TED shows only marginal improvement over Seq. (86.3 \pm 3 AVG). ccVAE’s high BWT (8.7) but poor AVG DSC (79.5 \pm 8) stems from its U-Net architecture without skip-connections, leading to slow convergence and compromised FWT (-14.0). MiB emerges as the best performing related work with 87.8 \pm 2 AVG and minimal forgetting (-3.1 BWT). Our GIN integration with Seq. yields modest improvements over default Seq. in both average performance (86.2 \pm 3) and forgetting resistance (-6.5 BWT).

FOCUS substantially outperforms existing methods, achieving 89.5 \pm 1 AVG while storing only 5% of features and using 90% CCD. It demonstrates strong forgetting resistance (-1.6 BWT) and transfer capability (1.7 FWT), enabled by its unique approach of applying CCD exclusively during training. This preserves skip-connection benefits without compromising privacy or storage efficiency and is further highlighted by the poor performance of the skip-connection-free variant (No Skip) without CCD and feature removal. Notably, using high CCD (90%) surpasses using no CCD (89.5 vs. 88.9), likely because the increased training difficulty promotes more robust feature learning when sufficient data is available.

Prostate Setting. This setting presents a challenging scenario with large images (48 \times 384 \times 384) but a limited number of samples (e.g., 12 images overall for BIDMC), characterized by significant domain gaps (Table 2, Figure 2). Traditional continual learning approaches struggle, with sequential training (Seq.) achieving only 57.4 \pm 8 AVG DSC. Even domain-specific models show inconsistent performance, particularly with BIDMC’s image 7. MiB slightly outperforms Seq. (59.0 \pm 7), while ccVAE exhibits strong backwards transfer (2.6 BWT) but poor overall performance (53.8 \pm 15). Notably, our Seq.+GIN, which extends Seq. with stronger data augmentation, shows high effectiveness with 76.5 \pm 11 AVG.

In this difficult setting, FOCUS also shows superior performance, achieving 80.5 \pm 7 AVG while storing only 0.5% of original feature information. However, the limited sample size and consequently fewer features make the model more sensitive to high CCD and overall feature elimination, resulting in some forgetting (-10.3 BWT), but with slightly positive forward transfer (0.7 FWT).

Ablation. Table 3 demonstrates the impact of each proposed component of FOCUS on performance, storage complexity and privacy-protection. Generative

Table 3. Ablation results for the continual prostate setting. CCD and DFP are percentage-based (%). The relative storage factor (RSF) indicates the approximate storage needed compared to storing the original images. Privacy-protecting (PP) specifies if a method applies considerable privacy-protection mechanisms for saved features.

Method	Components					DSC (%) \uparrow			MASD (mm) \downarrow			RSF	PP
	Skip	DBS	GIN	CCD	DFP	AVG	BWT	FWT	AVG	BWT	FWT	\downarrow	
ccVAE	–	–	–	–	–	53.8 \pm 15	<u>2.6</u>	-10.7	10.4 \pm 9	<u>-7.5</u>	2.3	14	\checkmark
No GR				0	100	58.7 \pm 15	6.9	-10.9	8.5 \pm 9	-7.7	1.7	0.1	\checkmark
Ablation	\checkmark			0	100	66.9 \pm 3	-15.8	-1.2	3.3 \pm 1	2.0	0.4	32	
	\checkmark	\checkmark		0	100	68.8 \pm 4	-13.0	-0.3	3.7 \pm 2	2.2	<u>0.2</u>	32	
	\checkmark		\checkmark	0	100	<u>83.4\pm4</u>	-4.9	0.2	<u>2.0\pm1</u>	0.1	0.4	32	
	\checkmark	\checkmark	\checkmark	0	100	85.2\pm2	-1.6	<u>0.6</u>	1.7\pm0	0.2	<u>0.2</u>	32	
	\checkmark	\checkmark		90	100	70.2 \pm 6	-12.6	<u>0.0</u>	6.8 \pm 5	7.2	0.0	3.2	\checkmark
	\checkmark	\checkmark	\checkmark	90	100	81.5 \pm 6	-8.1	0.0	2.5 \pm 1	1.8	0.0	3.2	\checkmark
	\checkmark	\checkmark	\checkmark	0	5	81.7 \pm 7	-7.7	0.7	2.2 \pm 1	1.0	<u>0.2</u>	1.6	
	\checkmark	\checkmark	\checkmark	95	100	81.0 \pm 8	-9.1	0.5	2.6 \pm 1	1.7	<u>0.3</u>	1.6	\checkmark
FOCUS	\checkmark	\checkmark	\checkmark	90	5	80.5 \pm 7	-10.3	0.7	2.5 \pm 1	1.9	0.4	<u>0.2</u>	\checkmark

replay (GR) as used in ccVAE ensures privacy-protection, but decreases performance. Additionally, using fully-connected VAE models take up more space than the investigated MRI datasets. Skip-free architectures achieve high BWT but suffer from poor FWT and AVG DSC, making them impractical for clinical applications. GIN provides substantial performance improvements by letting the model focus on domain-invariant shape features, thereby mitigating the impact of the frozen encoder by design. Interestingly, using GIN improves intra-domain performance even exceeding the domain model results especially on BIDMC. Furthermore, domain balanced sampling (DBS) slightly boosts performance with and without GIN. Even though channel-consistent dropout (CCD) reduces performance in the continual prostate setting, except for FWT, it enables vital privacy protection and storage efficiency of saved domain features. This is prominently demonstrated by the low relative storage factor (0.2 RSF) of FOCUS.

4 Conclusion

We propose FOCUS, a novel approach for domain-incremental continual MRI segmentation that substantially outperforms the related work. FOCUS enables the use of crucial U-Net skip-connections in feature replay while also maintaining strict privacy constraints. Moreover, our approach stores only 0.5% of original feature information per domain. Our integration of data augmentation techniques from domain generalization literature substantially enhances continual learning performance. Evaluation on two continual MRI datasets demonstrates that FOCUS improves average DSC by up to 21% compared to state-of-the-art methods, moving continual learning closer to a clinically viable solution for privacy-compliant medical image segmentation across changing domains.

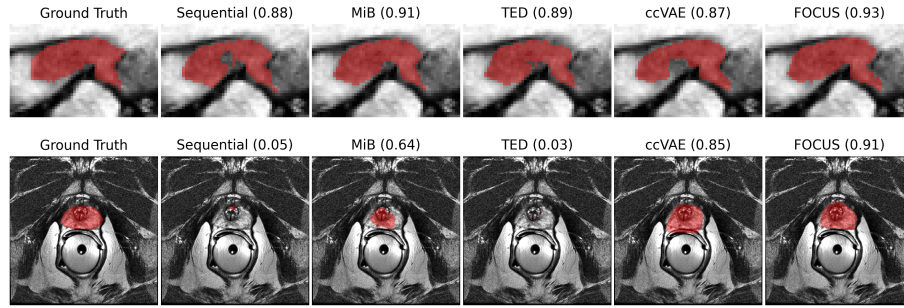


Fig. 2. Qualitative results of both continual learning settings, using models that have seen all domains, evaluated on the first domain. First row shows images and predictions from DecathHip, second row the same for BIDMC. Caption shows DSC performance of shown slice. FOCUS uses 90% CCD and 5% DFP.

Acknowledgments. We thank Mateusz Kozinski for fruitful discussions during the course of this project. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/PAT1748423.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. *Nature Communications* **13**(1), 4128 (2022). <https://doi.org/10.1038/s41467-022-30695-9>
2. Bera, S., Ummadi, V., Sen, D., Mandal, S., Biswas, P.K.: Memory replay for continual medical image segmentation through atypical sample selection. In: *Medical Image Computing and Computer Assisted Intervention*. pp. 513–522. Cham (2023). https://doi.org/10.1007/978-3-031-43901-8_49
3. Boccardi, M., Bocchetta, M., Morency, F.C., Collins, D.L., Nishikawa, M., Ganzola, R., Grothe, M.J., Wolf, D., Redolfi, A., Pievani, M., Antelmi, L., Fellgiebel, A., Matsuda, H., Teipel, S., Duchesne, S., Jack Jr, C.R., Frisoni, G.B., EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation, Alzheimer’s Disease Neuroimaging Initiative: Training labels for hippocampal segmentation based on the EADC-ADNI harmonized hippocampal protocol. *Alzheimer’s & Dementia* **11**(2), 175–183 (2015). <https://doi.org/10.1016/j.jalz.2014.12.002>
4. Cermelli, F., Mancini, M., Rota Bulò, S., Ricci, E., Caputo, B.: Modeling the background for incremental learning in semantic segmentation. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9230–9239 (2020). <https://doi.org/10.1109/CVPR42600.2020.00925>
5. Chen, L., Song, L., Feng, H., Zeru, R.T., Chai, S., Zhu, E.: Privacy-SF: An encoding-based privacy-preserving segmentation framework for medical images.

- Image and Vision Computing **151**, 105246 (2024). <https://doi.org/10.1016/j.imavis.2024.105246>
6. Chou, Y.C., Zhou, Z., Yuille, A.: Embracing massive medical data. In: Medical Image Computing and Computer Assisted Intervention. pp. 24–35 (2024). https://doi.org/10.1007/978-3-031-72378-0_3
 7. González, C., Ranem, A., Pinto dos Santos, D., Othman, A., Mukhopadhyay, A.: Lifelong nnU-Net: a framework for standardized medical continual learning. Scientific Reports **13**(1), 9381 (2023). <https://doi.org/10.1038/s41598-023-34484-2>
 8. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021). <https://doi.org/10.1038/s41592-020-01008-z>
 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. National Academy of Sciences **114**(13), 3521–3526 (2017). <https://doi.org/10.1073/pnas.1611835114>
 10. Kulaga-Yoskovitz, J., Bernhardt, B.C., Hong, S.J., Mansi, T., Liang, K.E., van der Kouwe, A.J.W., Smallwood, J., Bernasconi, A., Bernasconi, N.: Multi-contrast submillimetric 3 Tesla hippocampal subfield segmentation protocol and dataset. Scientific Data **2**(1), 150059 (2015). <https://doi.org/10.1038/sdata.2015.59>
 11. Lemke, N., González, C., Mukhopadhyay, A., Mundt, M.: Distribution-aware replay for continual MRI segmentation. In: Artificial Intelligence in Pancreatic Disease Detection and Diagnosis, and Personalized Incremental Learning in Medicine. pp. 73–85. Cham (2024). https://doi.org/10.1007/978-3-031-73483-0_7
 12. Liu, M., Xiao, L., Jiang, H., He, Q.: A new generative replay approach for incremental class learning of medical image for semantic segmentation. In: International Conference on Intelligent Medicine and Health. p. 51–56 (2022). <https://doi.org/10.1145/3560071.3560080>
 13. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1013–1023 (2021)
 14. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M.D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., et al.: Metrics reloaded: Recommendations for image analysis validation. Nature Methods **21**(2), 195–212 (2024). <https://doi.org/10.1038/s41592-023-02151-z>
 15. Ouyang, C., Chen, C., Li, S., Li, Z., Qin, C., Bai, W., Rueckert, D.: Causality-inspired single-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(4), 1095–1106 (2023). <https://doi.org/10.1109/TMI.2022.3224067>
 16. Pellegrini, L., Graffieti, G., Lomonaco, V., Maltoni, D.: Latent replay for real-time continual learning. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 10203–10209 (2020). <https://doi.org/10.1109/IROS45743.2020.9341460>
 17. Perkonig, M., Hofmanninger, J., Herold, C.J., Brink, J.A., Pianykh, O., Prosch, H., Langs, G.: Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging. Nature Communications **12**(1), 5678 (2021). <https://doi.org/10.1038/s41467-021-25858-z>

18. Rodríguez, N.D., Lomonaco, V., Filliat, D., Maltoni, D.: Don't forget, there is more than forgetting: New metrics for continual learning. arXiv:1810.13166 (2018), <http://arxiv.org/abs/1810.13166>
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
20. Sahiner, B., Chen, W., Samala, R.K., Petrick, N.: Data drift in medical machine learning: implications and potential remedies. *British Journal of Radiology* **96**(1150), 20220878 (2023). <https://doi.org/10.1259/bjr.20220878>
21. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**(1), 1929–1958 (2014)
22. Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M.: Robust and generalizable visual representation learning via random convolutions. In: International Conference on Learning Representations (2021)
23. Yuan, B., Zhao, D.: A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **46**(12), 10891–10910 (2024). <https://doi.org/10.1109/TPAMI.2024.3446949>
24. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging* **39**(6), 1856–1867 (2020). <https://doi.org/10.1109/TMI.2019.2959609>
25. Zhu, Z., Ma, X., Wang, W., Dong, S., Wang, K., Wu, L., Luo, G., Wang, G., Li, S.: Boosting knowledge diversity, accuracy, and stability via tri-enhanced distillation for domain continual medical image segmentation. *Medical Image Analysis* **94**, 103112 (2024). <https://doi.org/10.1016/j.media.2024.103112>