FR^{2}Seg: Continual Segmentation Across Multiple Sites via Fourier Style Replay and Adaptive Consistency Regularization

Cheng Xu[1, 2], Weiwen Zhang[1], Hongrui Zhang[1], Xuemiao Xu[1, 3, 4, 5, 6*], Huaidong Zhang[1], Jing Zou[2], Jing Qin[2]

[1]South China University of Technology, China
[2]The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]Guangdong Engineering Center for Large Model and GenAI Technology
[4]State Key Laboratory of Subtropical Building Science
[5]Ministry of Education Key Laboratory of Big Data and Intelligent Robot
[6]Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information
cschengxu@gmail.com, csvv@mail.scut.edu.cn, xhdr196884@gmail.com, {xuemx, huaidongz}@scut.edu.cn, {zoujing.zou, harry.qin}@polyu.edu.hk

Abstract

In clinical imaging, medical segmentation networks typically require continually adapting to new data from multiple sites over time, as aggregating all data for learning at once can be impractical due to storage limitations and privacy concerns. However, existing methods basically overlook domain-specific characteristics and fall short of adequately capturing domain-invariant knowledge during continual learning, leading to undesired catastrophic forgetting of previous sites and inferior generalization to new sites. To tackle this issue, this paper introduces FR^{2}Seg, to sufficiently exploit both domain-specific and domain-invariant knowledge for efficient continual learning with the aid of low-frequency cues. For the former aspect, we propose a Fourier style replay module to synthesize pseudo images with old-site styles for data augmentation during new-site training, effectively preventing catastrophic forgetting without sacrificing data privacy. For the latter, we present a Fourier adaptive consistency regularization to identify and constrain the optimization of domain-invariant parameters with explicit awareness of knowledge transferability across sites, ensuring excellent generalizability to new sites. Experimental results on two public datasets confirm our method's superiority over existing state-of-the-art continual learning methods.

Code — https://github.com/cschengxu/FR2Seg

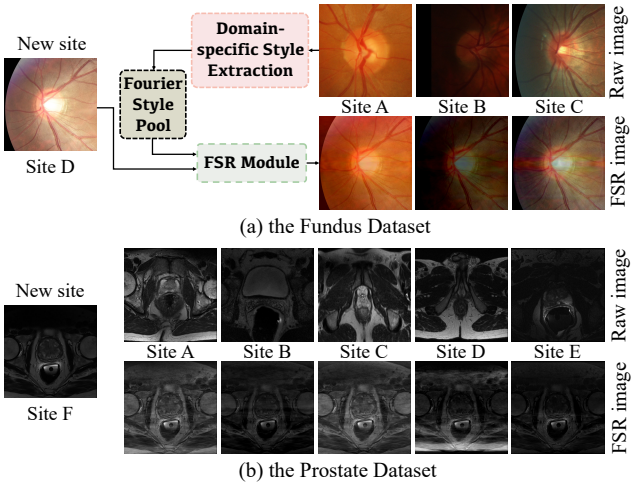

Figure 1: Visualization of Fourier Style Replay (FSR) images. By swapping the low-frequency components of new-site images with those from raw images of old sites, the resulting FSR images maintain the content of the new site while exhibiting the styles of the old sites. We strategically leverage these low-frequency cues to enhance both domain-specific and domain-invariant learning, achieving effective continual learning without compromising privacy.

Introduction

Training a robust medical image segmentation model usually requires extensive training data, which are often gathered across multiple sites (Glocker et al. 2019). In clinical practice, it is impractical to aggregate multi-site data as a consolidated set due to high storage costs and potential data privacy leakage. A sequenced training of multi-site data is preferred for efficiency and privacy. However, simply fine-tuning with newly arrived data often leads to catastrophic forgetting (McCloskey and Cohen 1989), due to the large data discrepancy across multiple sites. It is crucial to continually adapt the model to new sites without compromising its performance on previous sites.

Prior works tackle this problem by mainly focusing on mitigating the catastrophic forgetting issue. Replay-based methods (Ranem, González, and Mukhopadhyay 2022; Hayes, Cahill, and Kanan 2019) store a subset of old data for rehearsal training, which compromises data privacy. Parameter isolation-based methods (Rao et al. 2019; Rusu et al.

*Corresponding author.

2016; Zhao et al. 2022) assign different parameters to each task such that learning of new tasks will not interfere with the knowledge of old tasks. However, their performance is heavily limited by model capacity. Other methods either distill the knowledge from previous models to relieve forgetting (Li and Hoiem 2017; Douillard et al. 2020, 2021; Zhang et al. 2022), or identify important weights and penalize optimization of specific parameters (Kirkpatrick et al. 2017; Özgün et al. 2020; Zhang et al. 2021; Zhang and Gu 2023). Nevertheless, these approaches generally overlook domain-specific information and struggle with distilling comprehensive domain-invariant knowledge across different sites, leading to inferior segmentation performance.

Indeed, domain-specific and domain-invariant knowledge are both indispensable for efficient continual learning. On the one hand, domain-specific characteristics provide abundant memorability of previous data, significantly alleviating catastrophic forgetting. On the other hand, thoroughly exploring domain-invariant knowledge helps identify commonalities and patterns across sites, effectively bolstering the generalizability to new sites. To fully exploit the merits from both aspects, our key idea is to leverage low-frequency cues to simulate images from previous sites, enhancing new-site training and identifying domain-invariant knowledge. Particularly, the amplitude of low-frequency signals capture intensity and texture information (Yang and Soatto 2020; Xu et al. 2021), which represent crucial domain-specific characteristics (see Fig. 1). These cues can be effectively used for both domain-specific and domain-invariant learning through data augmentation and perturbations.

Based on the above observation, we propose a novel Fourier-based replay and regularized framework, dubbed FR^{2}Seg, for continual segmentation across multiple sites. We introduce a Fourier-Style Replay module designed to retain the styles of previously sites. This module serves a dual purpose: first, by augmenting the current site data with styles from previous sites, it effectively mitigates catastrophic forgetting. Second, leveraging these stored styles, we develop a Fourier Consistency Regularization technique that simulates domain shifts across different sites. This process enables the network to accurately identify and maintain domain-invariant parameters, thereby preserving knowledge that is consistent across domains. Moreover, considering the inherent transferability of knowledge from previous sites to the current one, we customize an Adaptive Inter-Domain Transferability Weighting Mechanism, which preserves domain-invariant knowledge more accurately and prevents potential negative transfer, further enhancing the performance of continual learning. Extensive experiments demonstrate our method show significant advantanges over state-of-the-art continual learning techniques.

In summary, our main contributions are three-fold:

- We introduce FR^{2}Seg, an innovative approach for domain continual learning. To our knowledge, this is the first initiative leveraging low frequency cues to exploit reciprocal strengths of domain-specific and domain-invariant learning for domain continual learning, effectively preventing catastrophic forgetting and boosting new-site learning, while ensuring data privacy.
- We present a Fourier adaptive consistency regularization scheme to promote domain-invariant learning. On the one hand, we design a Fourier intra-domain consistency regularization (FCR) to precisely identify parameters important for domain-invariant representations and penalize their optimization, thereby encouraging domain-invariant knowledge learning. On the other hand, an adaptive inter-domain transferability weighting mechanism is tailored to inject explicit cross-site transferability into FCR, enhancing domain-invariant generalization.
- Extensive experiments on two public benchmarks confirm our method's superiority over state-of-the-art continual learning approaches.

Related Work

Continual learning (CL) aims to continuously update models with new tasks and datasets while avoiding catastrophic forgetting (French 1999). Existing methods for continual learning can be generally divided into four categories: Rehearsal-based methods (Ranem, González, and Mukhopadhyay 2022; Hayes, Cahill, and Kanan 2019) mitigate the forgetting of previously learned knowledge by retaining samples, synthesized images, or features from prior tasks. While effective in reducing catastrophic forgetting, these methods often pose privacy risks, particularly in scenarios where access to sensitive data, such as medical scans from multiple hospitals, is restricted. Parameter-isolated methods (Rao et al. 2019; Rusu et al. 2016; Zhao et al. 2022) address different tasks by employing task-specific parameters, which can be achieved through network expansion or the addition of task-specific subnetworks. However, these approaches typically require increasingly larger storage resources, and their effectiveness may still be constrained. Distillation-based methods (Michieli and Zanuttigh 2019; Li and Hoiem 2017) transfer knowledge from previous tasks to alleviate catastrophic forgetting, yet they are susceptible to accumulated errors over extended periods of learning. Regularization-based methods (Aljundi et al. 2018; Zenke, Poole, and Ganguli 2017; Kirkpatrick et al. 2017; Zhang and Gu 2023) preserve previously learned knowledge by identifying important parameters in the network and applying penalties to their updates, thus reducing catastrophic forgetting. However, this strategy can also hinder the network's ability to effectively learn new knowledge. Diverging from conventional continual learning approaches, we propose a novel scheme that introduces low-frequency cues and integrates the complementary strengths of both rehearsal-based and regularization-based methods. This approach enables the preservation of both domain-specific and domain-invariant knowledge, leading to effective continual learning while preserving data privacy.

Domain Continual Semantic Segmentation has received considerable attention with the advancement of deep learning (Chen et al. 2023; Yu et al. 2024; Yang et al. 2024; Huang et al. 2024) in recent years. This is particularly edivent in medical image segmentation, where substantial domain shifts between data from different hospitals, coupled

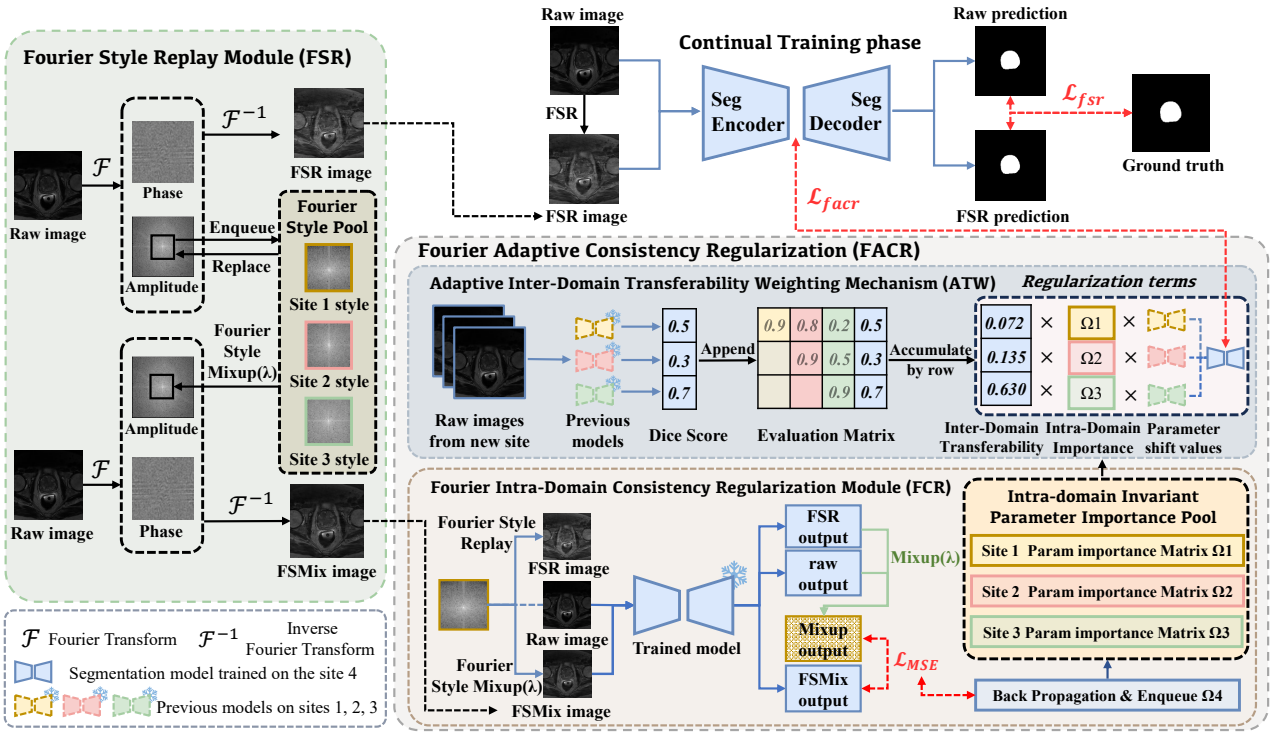

Figure 2: Overview of FR^{2}Seg. Our framework consists of the segmentation backbone network and two innovative modules: (1) Fourier Style Replay (FSR) module extracts and memorizes domain-specific characteristics for subsequent data augmentation and perturbations; (2) Fourier Adaptive Consistency Regularization (FACR) identifies the domain-invariant parameters and constrains their optimization, in which, the adaptive inter-domain transferability weighting (ATW) mechanism estimates the transferability across sites and further regularizes the domain-invariant learning.

with the inability to access all data simultaneously, pose significant challenges to effective continual learning. To address these challenges, MAS-LR (Özgün et al. 2020) introduces a learning rate regularization strategy that directly limits the adaptation of important network parameters, thereby preventing catastrophic forgetting. Zhang et al. (Zhang et al. 2021; Zhang and Gu 2023) propose a comprehensive importance-based selective regularization strategy that preserves shape priors and semantic information from previously learned sites. Ranem et al. (Ranem, González, and Mukhopadhyay 2022) combine global knowledge with sequential input data using a transformer-based framework. Zhu et al. (Zhu et al. 2024) develop a tri-enhanced distillation network that mitigates catastrophic forgetting by improving distillation performance across different stages. However, existing methods struggle to retain both domain-specific and domain-invariant knowledge during the continual segmentation learning process. In contrast, our approach introduces Fourier-style replay and mixing, which facilitate the accurate retention of both domain-specific and domain-invariant knowledge during learning, thereby significantly enhancing the performance of continual segmentation.

Method

Our goal is to fully explore low-frequency cues for achieving domain-specific and domain-invariant learning in continual segmentation without compromising data privacy, thereby mitigating catastrophic forgetting on previous sites and enhancing generalizability to new sites. We follow the problem formulation in CISR (Zhang et al. 2021). Given a sequential stream of medical images from T sites, in step $t \in [1, T]$ of the continual learning process, only the images and annotated labels $\{(x_n, y_n)\}_{n=1}^{N}$ of site D_t are available. Fig. 2 illustrates the overview of our method. which comprises a segmentation backbone network for continual learning, a Fourier Style Replay (FSR) module for storing domain-specific characteristics to boost domain-specific learning, and a Fourier Adaptive Consistency Regularization (FACR) module for identifying and constraining updates of domain-invariant parameters. In the following sections, we elaborate on each component in detail.

Fourier Style Replay

To prevent catastrophic forgetting without sacrificing data privacy, we devise a Fourier Style Replay (FSR) module to memorize domain-specific characteristics of previous sites in a Fourier style pool. These characteristics can be used for faithful synthesis of pseudo images with previous-site styles for continual learning on new sites, enabling the model to maintain previous-site knowledge without assessing previous raw data.

Fourier Domain-specific Style Extraction. To extract domain-specific characteristics, we first use Fourier transform $\mathcal{F} = (\mathcal{A}(X), \mathcal{P}(X))$ to convert the image x to the frequency space X, where $\mathcal{A}(X)$ and $\mathcal{P}(X)$ indicate the amplitude and phase, respectively. Then, we move the low-frequency component to the centre of the spectrum. As mentioned in previous studies (Yang and Soatto 2020; Xu et al. 2021; Zhou, Qi, and Shi 2022), amplitude indicates the intensity and texture distribution of the image, characterizing critical style information. Therefore, after training on a site, we introduce a binary mask $\mathcal{M}_\alpha = \mathbb{I}_{(h,w)\in[-\alpha:\alpha,-\alpha:\alpha]}$ to extract low-frequency components of the amplitude from images on a site. We average these components for comprehensive characteristics acquisition, preventing data privacy leakage. Therefore, we obtain a Fourier domain-specific style $S = \frac{1}{N}\sum_{n=1}^{N} \mathcal{M}_\alpha \mathcal{A}(X_n)$, which is then sent to the Fourier Style Pool $\mathbb{S} = \{S_1, S_2, \ldots, S_i\}$ for subsequent style replay, where i indicates the index of a previous site.

Fourier Domain-specific Style Replay. When training on a new site, we replace the low-frequency part of amplitude $\mathcal{A}(X)$ of a raw image with randomly chosen previous Fourier style S_i, and then we can synthesize a new amplitude spectrum: $(1 - \mathcal{M}_\alpha)\mathcal{A}(X) + S_i$. The new amplitude spectrum is combined with the original phase spectrum $\mathcal{P}(X)$ and undergoes the inverse Fourier Transform $\mathcal{F}^{-1}$, rendering a pseudo image with previous style S_i. This pseudo image preserves the content of the new site image while exhibiting style of the previous site. This Fourier Domain-specific Style Replay (FSR) process can be formulated as:

$$\mathrm{FSR}(x, S_i) = \mathcal{F}^{-1}((1 - \mathcal{M}_\alpha)\mathcal{A}(X) + S_i, \mathcal{P}(X)). \quad (1)$$

During training, we randomly choose a Style S_i from $\mathbb{S}$ and utilize the FSR module to synthesize images for a mini batch training $\{x_n, y_n\}$. A segmentation loss $\mathcal{L}_{seg}$ (including the standard CE loss (Wang et al. 2024a,b; Ren et al. 2021; Xu et al. 2023) and dice loss) is applied to both of the raw images x and synthesized images $\mathrm{FSR}\left(x_n, S_i\right)$. The Fourier Style Replay loss $\mathcal{L}_{fsr}$ is as follows:

$$\mathcal{L}_{fsr} = \frac{1}{2N}\sum_{n=1}^{N}(\mathcal{L}_{seg}(x_n, y_n) + \mathcal{L}_{seg}(\mathrm{FSR}\left(x_n, S_i\right), y_n)). \quad (2)$$

By learning from rich data with previous-site distribution via FSR, the model effectively preserves previous domain-specific knowledge, thus resisting catastrophic forgetting.

Fourier Adaptive Consistency Regularization

Apart from domain-specific memorization, preserving domain-invariant knowledge is critical to continual learning. To this end, we propose a Fourier Adaptive Consistency Regularization, consisting of a Fourier Intra-Domain Consistency Regularization (FCR) for identifying domain-invariant parameters and constraining their optimization, and an Adaptive Inter-Domain Transferability Weighting mechanism (ATW) for further enhancing FCR with cross-site transferability guidance.

Fourier Intra-Domain Consistency Regularization. To distill domain invariant knowledge, we propose to identify domain-invariant parameters and penalize their updates. To achieve this, we perform a mixup operation, namely Fourier Style Mixup (FSMix), between a randomly selected Fourier style S_i from Fourier style pool $\mathbb{S}$ and low-frequency part of amplitude spectrum $\mathcal{A}(X)$ from a new-site image, yielding an augmented image that simulates diverse domain shifts:

$$\mathrm{FSMix}_\lambda(x, S_i) = \mathcal{F}^{-1}((1 - \lambda\mathcal{M}_\alpha)\mathcal{A}(X) + \lambda S_i, \mathcal{P}(X)). \quad (3)$$

Here, λ is the mixing factor. Considering the segmentation of the new-site image and its augmented counterpart should be consistent, we tailor a Fourier Style Mixup Consistency loss $\mathcal{L}_{mc}$ to identify domain-invariant parameters as follows:

$$\begin{aligned}\mathcal{L}_{mc} = \mathrm{MSE}(&F_\theta\left(\mathrm{FSMix}_\lambda\left(x_n, S_j\right)\right), \\ &\mathrm{Mix}_\lambda\left(F_\theta(x_n), F_\theta\left(\mathrm{FSR}\left(x_n, S_j\right)\right)\right)).\end{aligned} \quad (4)$$

Here, $\mathrm{Mix}_\lambda(a, b) = \lambda a + (1 - \lambda)b$. λ is uniformly sampled from $[0.1, 0.9]$. S_j is randomly drawed from $\mathbb{S}$. The same λ and S_j are input to both FSMix and FSR modules for a mini batch. Specifically, the mean square error loss (MSE) is adopted to measure the consistency between the model output from input FSMix-based mixup $F_\theta\left(\mathrm{FSMix}_\lambda\left(x_n, S_j\right)\right)$ and the mixup of raw output $F_\theta(x_n)$ with FSR-based output $F_\theta\left(\mathrm{FSR}\left(x_n, S_j\right)\right)$.

Subsequently, we assess the influence of a slight change in each parameter θ with respect to loss $\mathcal{L}_{mc}$. Thus, Ω_i denotes the importance of parameter θ_i in maintaining domain-invariant representations. To penalize optimization of the domain-invariant parameters during training, we introduce a Fourier Intra-domain Consistency Regularization term $\mathcal{L}_{fcr}$:

$$\begin{aligned}\Omega_i &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial \left\|\mathcal{L}_{mc}\right\|_2^2}{\partial\theta_i}, \\ \mathcal{L}_{fcr}(\theta, t) &= \sum_i \Omega_i \left(\theta_i - \theta_{i,t}\right)^2.\end{aligned} \quad (5)$$

Here, t denotes the index of previous site D_t. $\theta_{i,t}$ represents the value of the i-th parameter trained on site D_t. θ_i is the corresponding parameter trained on the current site. By enforcing this regularization, the domain-invariant parameters can be accurately captured. The domain-invariant knowledge can thus be strategically retained by constraining the optimization of these parameters.

Adaptive Inter-Domain Transferability Weighting Mechanism (ATW). Given the unique correlations between knowledge from various previous sites and the current site, it is crucial to consider these relationships to better preserve domain-invariant knowledge and avoid potential negative transfer. To achieve this, we assign different weights for regularization terms from different sites according to the inter-domain transferability between a previous site D_t and the new site D_s, denoted as $Tr_{t,s}$. In particular, upon the arrival of data $\{x_{D_s}, y_{D_s}\}$ from a new site D_s, we feed it to all trained models of previous sites and compute the Dice Score (DSC) output by each model.

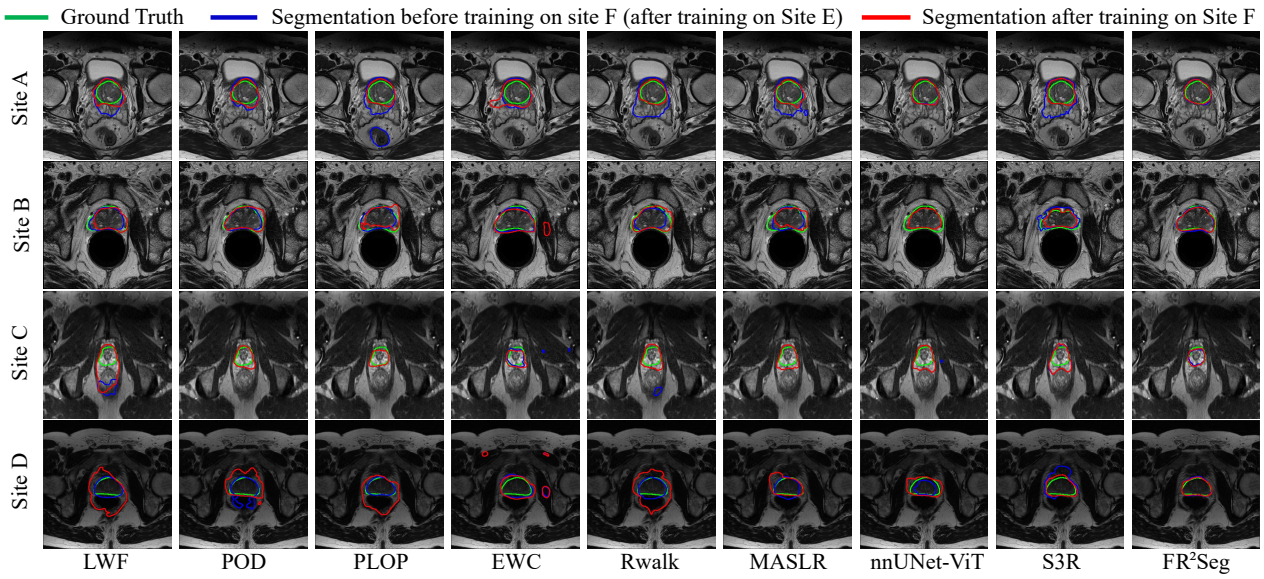

Figure 3: Visualization of segmentation on the Prostate dataset before and after training on site F (learning in order).

The transferability is then defined as the cumulative product of the DSC obtained by the model F_t trained on site D_t, when evaluated on the training sets from site D_t to site D_s (see the blue part of Fig. 2). The calculation process is as follows:

$$Tr_{t,s} = \prod_{m=t}^{s} \mathrm{DSC}\left(F_t\left(x_{D_m}\right), y_{D_m}\right) \tag{6}$$

$$= Tr_{t,s-1} \times \mathrm{DSC}\left(F_t\left(x_{D_s}\right), y_{D_s}\right). \tag{7}$$

Thus far, the original Fourier consistency regularization can be enhanced by the adaptive transferability weighting mechanism, rendering the final Fourier adaptive consistency regularization term $\mathcal{L}_{facr}$. From previous sites during continual learning on current site s, $\mathcal{L}_{facr}$ can be defined as follows:

$$\mathcal{L}_{facr}(\theta) = \sum_{t=1}^{s-1} Tr_{t,s} \times \mathcal{L}_{fcr}(\theta, t). \tag{8}$$

During new-site training, the total loss for our framework is:

$$\mathcal{L}_{total} = \mathcal{L}_{fsr} + \beta \mathcal{L}_{facr}, \tag{9}$$

where β is the weighting parameter of the regularization terms in the total loss. By combining the reciprocal strengths of the Fourier style replay loss and the Fourier adaptive consistency regularization, domain-specific information and domain-invariant knowledge are sufficiently explored during the continual training process, granting our method superior capability to prevent catastrophic forgetting on previous sites and excellent generalizability to new sites.

Experiments

Experimental Settings

Datasets. We evaluate our approach on two medical image segmentation tasks: **1) Prostate MRI segmentation** includes prostate T2-weighted MRI from 6 different data sources (Liu, Dou, and Heng 2020). Among these data, samples of Sites A and B are from the NCI-ISBI13 dataset (Bloch and et al 2015). Samples of Site C are from the I2CVB dataset (Lemaître et al. 2015). Samples of Sites D, E, F are from the PROMISE12 dataset (Litjens et al. 2014). Following CISR (Zhang et al. 2021), we organize this multi-site dataset in a sequential stream ordered by Site A→B→C→D→E→F; **2) Fundus image segmentation** includes retinal fundus images from four medical institutions out of three public datasets. Samples of sites A and B are from REFUGE (Orlando et al. 2020), samples of site C are from RIM-ONE-r3 (Fumero et al. 2011), and samples of site D are from Drishti-GS (Sivaswamy et al. 2015). We average the performance of optic disc and cup segmentation for

Methods	Prostate				Fundus			
	in-order		in-reverse		in-order		in-reverse	
	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑
FineTune	66.82	-25.26	58.62	-24.42	76.34	-12.50	79.92	-7.99
JointTrain	90.38	0.03	88.57	-0.63	89.86	0.24	89.46	0.31
LWF	73.96	-18.21	65.57	-25.10	81.91	-7.87	81.74	-6.98
POD	77.46	-14.69	71.56	-18.39	81.68	-8.42	83.72	-5.91
PLOP	75.68	-15.06	71.72	-17.78	82.24	-7.55	82.93	-6.11
MASLR	74.89	-16.55	68.35	-21.41	81.39	-8.08	81.51	-7.81
EWC	75.06	-8.18	71.91	-12.85	82.20	-5.94	83.45	-4.81
RWalk	78.68	-13.58	72.56	-18.42	82.49	-6.27	82.95	-6.56
nnUNet-ViT	83.26	-5.47	74.91	-6.26	83.37	-5.38	84.23	-5.52
S3R	85.14	-2.23	76.80	-5.93	84.08	-2.62	84.20	-3.37
FR²Seg	**88.30**	**-1.43**	**79.68**	**-3.01**	**85.18**	**-1.11**	**85.38**	**-3.11**

Table 1: Quantitative comparison on the Prostate dataset and the Fundus dataset. Values are in %.

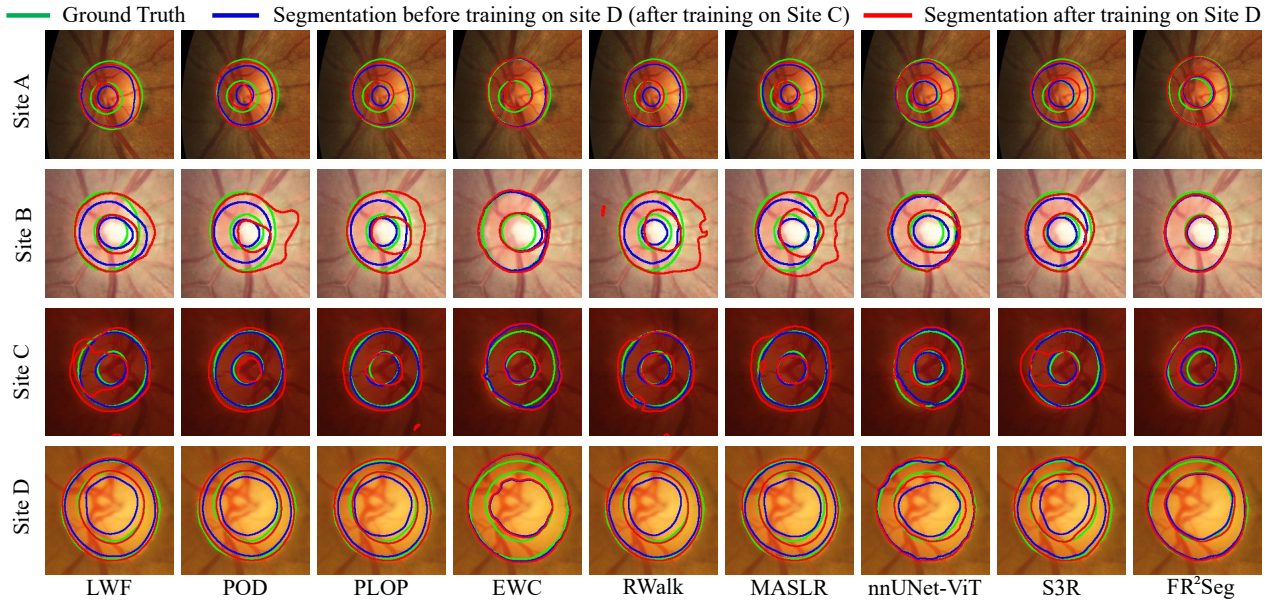

Figure 4: Visualization of segmentation on the Fundus dataset before and after training on site D (learning in order).

evaluation. This dataset is organized in a sequential stream ordered by Site A→B→C→D.

Implementation Details. We adopt 2D-nnUNet (Isensee et al. 2021) as the network backbone. Input images are uniformly resized to 384×384 and undergo preprocess pipeline of nnUNet. Cases of each domain are randomly divided into five folds, among which four folds are used for training and the rest for testing. The batch size is set as 8 and number of training epochs is set as 100. The initial learning rate is 0.01, with exponential decay. Empirically, α is set to 11 for both datasets. λ is set to 100 for the Prostate dataset and 500 for the Fundus dataset. All experiments are conducted on 2 NVIDIA GeForce RTX 3090 GPUs.

Metrics. We adopt two metrics for evaluation, including AVG and BWT (Díaz-Rodríguez et al. 2018). Seen Average (**AVG**) is the average performance of the network on seen domains, which is the most widely used metric to evaluate continual learning performance. Backward Transfer (**BWT**) measures the forgetting degree of the network on previous domains. These two metrics can be formulated as follows:

$$AVG = \frac{2\sum_{i=1}^{D}\sum_{j=1}^{i} R_{i,j}}{D(D+1)}, \tag{10}$$

$$BWT = \frac{2\sum_{i=2}^{D}\sum_{j=1}^{i-1} (R_{i,j} - R_{j,j})}{D(D-1)}, \tag{11}$$

where $R \in \mathbb{R}^{D\times D}$ is the train-test-accuracy-matrix, where D is the number of domains. Each entry $R_{i,j}$ is the mean Dice score (DSC) of the model on domain j after sequential learning from site 1 to site i.

Methods	Prostate				Fundus			
	in-order		in-reverse		in-order		in-reverse	
	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑
FineTune	66.82	-25.26	58.62	-24.42	76.34	-12.50	79.92	-7.99
JointTrain	90.38	0.03	88.57	-0.63	89.86	0.24	89.46	0.31
w/o FSR	85.45	-4.43	75.49	-8.47	84.75	-2.43	84.67	-3.59
w/o FACR	81.61	-10.59	74.84	-14.63	84.14	-4.01	85.10	-4.77
w/o ATW	87.01	-1.98	78.67	-5.51	84.73	-1.70	85.21	-4.37
FR²Seg	**88.30**	**-1.43**	**79.68**	**-3.01**	**85.18**	**-1.11**	**85.38**	**-3.11**

Table 2: Ablation study on the Prostate dataset and the Fundus dataset. Values are in %.

Comparison with State-of-the-Arts

Baselines. We compare to several state-of-the-art continual learning methods, including three distillation-based methods: LWF (Li and Hoiem 2017), POD (Douillard et al. 2020), and PLOP (Douillard et al. 2021); along with four regularization-based methods: EWC (Kirkpatrick et al. 2017), RWalk (Chaudhry et al. 2018), MASLR (Özgün et al. 2020), and S3R (Zhang and Gu 2023). nnUNet-ViT (Ranem and et al. 2022) is also involved for comparison. As mentioned in (Ranem and et al. 2022), we choose EWC loss to train the nnUNet-ViT model for optimal performance. Additionally, the FineTune scheme is incorporated as a baseline. A model trained on a collective dataset of all sites (joint-training scheme) is referenced as the upper bound. All the methods, including ours, are implemented on the lifelong nnUNet framework (González et al. 2023).

Quantitative comparison. Table 1 compares different methods across two datasets in original and reverse orders.

Number of Latest Models	in-order		in-reverse	
	AVG↑	BWT↑	AVG↑	BWT↑
1	86.74	-3.82	78.85	-5.43
2	87.24	-2.65	78.95	-4.24
3	87.64	-2.29	79.39	-3.42
6 (all)	**88.30**	**-1.43**	**79.68**	**-3.01**

Table 3: Performance Comparion of different numbers of previous models on the Prostate Dataset. Values are in %.

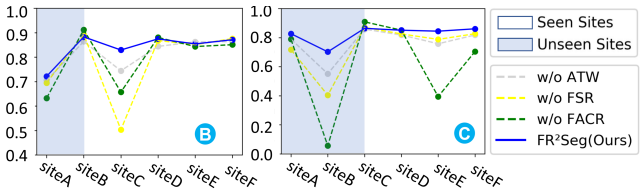

Figure 5: Tendency of segmentation performance (Y-axis) on site B and Site C during continual learning from site A to site F (X-axis) on the Prostate dataset.

Dataset	Prostate				Fundus			
Orders	CAEDBF		BEAFCD		CADB		BDAC	
Methods	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑	AVG↑	BWT↑
FineTune	60.03	-23.17	62.12	-25.72	79.89	-8.76	79.77	-9.70
JointTrain	90.52	-0.04	89.53	-1.55	89.17	-0.01	90.02	-0.30
LWF	69.68	-22.51	68.39	-23.79	81.47	-7.45	82.44	-8.82
POD	74.46	-16.93	69.51	-22.76	82.63	-6.71	82.76	-8.11
PLOP	71.08	-20.62	66.49	-24.39	83.08	-6.03	81.98	-8.72
MASLR	71.72	-20.06	74.29	-17.91	82.23	-6.99	81.90	-8.83
EWC	77.01	-5.56	71.02	-12.71	81.22	-3.85	82.20	-4.08
RWalk	75.02	-16.51	73.07	-18.60	83.61	-6.22	83.28	-7.61
nnUNet-ViT	82.39	-6.92	80.66	-6.75	84.44	-3.98	84.28	-5.77
S3R	82.64	-2.82	81.00	-3.67	84.77	-2.59	85.00	-4.13
FR^{2}Seg	**84.13**	**-2.08**	**83.11**	**-3.18**	**85.66**	**-2.30**	**86.33**	**-1.47**

Table 4: Quantitative results under different continual learning site orders on the Prostate dataset and the Fundus dataset. Values are in %.

Distillation-based methods (LWF, POD, PLOP) and several prior-based methods (RWalk, MASLR) struggle with domain-invariant representations learning, leading to catastrophic forgetting with low BWT values. While S3R and nnUNet-ViT achieve better BWT and AVG than the above-mentioned methods, they neglect domain-specific knowledge during continual training, and thus, there exists a notable gap between their AVG and our approach.

Qualitative comparison. Figs. 3 and 4 show segmentation results on the Prostate and Fundus datasets, respectively. Our method maintains the highest overlap with ground truth on slices from site A to site D (B) before and after training on site E (C) for the Prostate (Fundus) dataset, demonstrating its superiority in mitigating catastrophic forgetting.

Ablation Study

Here we provide in-depth analysis of efficacy of our main proposals. Moreover, we also explore the impacts of number of saved previous models, the window size in FSR, and our framework's robustness to different learning orders.

Effectiveness of the main components. As shown in Fig. 5, when the FSR or FACR is removed, the performance of site B and site C significantly decreases after training on site C and Site E, respectively. Accordingly, a significant drop in AVG is also observed in Table 2. This indicates that both domain-specific characteristics extracted by FSR and domain-invariant knowledge distilled by FACR are indispensable for effective continual segmentation. Furthermore, thanks to sufficient consideration of the transferability across sites, ATW effectively encourages domain-invariant knowledge maintenance and alleviates the negative transfer, thus contributing to further improvement in AVG.

Impacts of number of saved previous models. Table 3 shows the impact of different number of preserved previous models. It can be observed that the segmentation performance generally improves as the number of preserved models increases, reaching its peak when all six models are retained. However, even when only the most recent model is preserved, our method also achieves state-of-the-art performance, indicating that the Fourier intra-domain consistency regularization is effective in accurately identifying domain-consistent parameters and retaining cross-domain invariant knowledge. Considering that saving six models can significantly improve model performance while incurring only acceptable storage overhead (see Table 1 in the Appendix), we save the six most recent models as our default setting.

Robustness to learning orders. We also examined the impact of different learning orders on continual segmentation. Beyond the original and reverse orders in Table 1, we conducted experiments using two random orders for each dataset. The results, summarized in Table 4, demonstrate that our method exhibits significant robustness to varying learning orders and consistently achieves the highest average performance among all the methods compared.

Conclusion

In this paper, we propose FR^{2}Seg for multi-site continual segmentation with the aid of low-frequency cues. The Fourier-based replay module memorizes domain-specific characteristics and synthesizes faithful pseudo images of previous-site styles for domain-specific learning, mitigating catastrophic forgetting. The Fourier adaptive consistency regularization module encompasses an intra-domain consistency regularization and an adaptive inter-domain transferability weighting mechanism, effectively capturing domain-invariant knowledge and ensuring excellent model generalizability. Our method demonstrates superiority on two datasets in efficient continual segmentation, paving the way for storage-efficient and privacy-preserved continual learning applications in clinical practice.

Acknowledgments

This work is supported by Key-Area Research and Development Program of Guangzhou City (No. 2023B01J0022), NSFC Key Project (No. U23A20391), China National Key R&D Program (No. 2023YFE0202700), Guangdong Provincial Natural Science Foundation for Outstanding Youth Team Project (No. 2024B1515040010), a General Research Fund under Hong Kong Research Grants Council (No. 15218521), and a Shenzhen-Hong Kong-Macao Science and Technology Plan Project (Category C Project) under Shenzhen Municipal Science and Technology Innovation Commission (No. SGDX20230821092359002).

References

Aljundi, R.; Babiloni, F.; Elhoseiny, M.; Rohrbach, M.; and Tuytelaars, T. 2018. Memory aware synapses: Learning what (not) to forget. In *ECCV*, 139–154.

Bloch, N.; and et al. 2015. NCI-ISBI 2013 challenge: automated segmentation of prostate structures. Cancer Imaging Arch. *Cancer Imaging Arch*.

Chaudhry, A.; Dokania, P. K.; Ajanthan, T.; and Torr, P. H. 2018. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *ECCV*, 532–547.

Chen, Y.; He, S.; Jin, Y.; and Qin, J. 2023. Surgical activity triplet recognition via triplet disentanglement. In *MICCAI*, 451–461.

Díaz-Rodríguez, N.; Lomonaco, V.; Filliat, D.; and Maltoni, D. 2018. Don't forget, there is more than forgetting: new metrics for Continual Learning. *arXiv preprint arXiv:1810.13166*.

Douillard, A.; Chen, Y.; Dapogny, A.; and Cord, M. 2021. Plop: Learning without forgetting for continual semantic segmentation. In *CVPR*, 4040–4050.

Douillard, A.; Cord, M.; Ollion, C.; Robert, T.; and Valle, E. 2020. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *ECCV*, 86–102.

French, R. M. 1999. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4): 128–135.

Fumero, F.; Alayón, S.; Sanchez, J. L.; Sigut, J.; and Gonzalez-Hernandez, M. 2011. RIM-ONE: An open retinal image database for optic nerve evaluation. In *international symposium on computer-based medical systems (CBMS)*, 1–6.

Glocker, B.; Robinson, R.; Castro, D. C.; Dou, Q.; and Konukoglu, E. 2019. Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. *arXiv preprint arXiv:1910.04597*.

González, C.; Ranem, A.; Pinto dos Santos, D.; Othman, A.; and Mukhopadhyay, A. 2023. Lifelong nnU-Net: a framework for standardized medical continual learning. *Scientific Reports*, 13(1): 9381.

Hayes, T. L.; Cahill, N. D.; and Kanan, C. 2019. Memory efficient experience replay for streaming learning. In *2019 International Conference on Robotics and Automation (ICRA)*, 9769–9776.

Huang, Z.; Xu, X.; Xu, C.; Zhang, H.; Zheng, C.; Qin, J.; and He, S. 2024. Beat-It: Beat-Synchronized Multi-Condition 3D Dance Generation. In *ECCV*, 273–290.

Isensee, F.; Jaeger, P. F.; Kohl, S. A.; Petersen, J.; and Maier-Hein, K. H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2): 203–211.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13): 3521–3526.

Lemaître, G.; Martí, R.; Freixenet, J.; Vilanova, J. C.; Walker, P. M.; and Meriaudeau, F. 2015. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Computers in biology and medicine*, 60: 8–31.

Li, Z.; and Hoiem, D. 2017. Learning without forgetting. *IEEE TPAMI*, 40(12): 2935–2947.

Litjens, G.; Toth, R.; Van De Ven, W.; Hoeks, C.; Kerkstra, S.; van Ginneken, B.; Vincent, G.; Guillard, G.; Birbeck, N.; Zhang, J.; et al. 2014. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Medical Image Analysis*, 18(2): 359–373.

Liu, Q.; Dou, Q.; and Heng, P.-A. 2020. Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In *MICCAI*, 475–485.

McCloskey, M.; and Cohen, N. J. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, 109–165.

Michieli, U.; and Zanuttigh, P. 2019. Incremental learning techniques for semantic segmentation. In *ICCV workshops*, 0–0.

Orlando, J. I.; Fu, H.; Breda, J. B.; Van Keer, K.; Bathula, D. R.; Diaz-Pinto, A.; Fang, R.; Heng, P.-A.; Kim, J.; Lee, J.; et al. 2020. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59: 101570.

Özgün, S.; Rickmann, A.-M.; Roy, A. G.; and Wachinger, C. 2020. Importance driven continual learning for segmentation across domains. In *Machine Learning in Medical Imaging*, 423–433.

Ranem, A.; and et al. 2022. Continual hippocampus segmentation with transformers. In *CVPR*, 3711–3720.

Ranem, A.; González, C.; and Mukhopadhyay, A. 2022. Continual hippocampus segmentation with transformers. In *CVPR*, 3711–3720.

Rao, D.; Visin, F.; Rusu, A.; Pascanu, R.; Teh, Y. W.; and Hadsell, R. 2019. Continual unsupervised representation learning. *NeurIPS*, 32.

Ren, J.; Hu, X.; Zhu, L.; Xu, X.; Xu, Y.; Wang, W.; Deng, Z.; and Heng, P.-A. 2021. Deep texture-aware features for camouflaged object detection. *TCSVT*, 33(3): 1157–1167.

Rusu, A. A.; Rabinowitz, N. C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; and Hadsell, R. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

Sivaswamy, J.; Krishnadas, S.; Chakravarty, A.; Joshi, G.; Tabish, A. S.; et al. 2015. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, 2(1): 1004.

Wang, K.; Zheng, F.; Cheng, L.; Dai, H.-N.; Dou, Q.; and Qin, J. 2024a. Breast Cancer Classification from Digital Pathology Images via Connectivity-aware Graph Transformer. *IEEE TMI*.

Wang, K.; Zheng, F.; Guan, D.; Liu, J.; and Qin, J. 2024b. Distilling heterogeneous knowledge with aligned biological entities for histological image classification. *Pattern Recognition*, 111173.

Xu, C.; Chen, Z.; Mai, J.; Xu, X.; and He, S. 2023. Pose-and attribute-consistent person image synthesis. *ACM TOMM*, 19(2s): 1–21.

Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; and Tian, Q. 2021. A fourier-based framework for domain generalization. In *CVPR*, 14383–14392.

Yang, H.; Xu, X.; Xu, C.; Zhang, H.; Qin, J.; Wang, Y.; Heng, P.-A.; and He, S. 2024. G 2 Face: High-Fidelity Reversible Face Anonymization via Generative and Geometric Priors. *IEEE Transactions on Information Forensics and Security*.

Yang, Y.; and Soatto, S. 2020. Fda: Fourier domain adaptation for semantic segmentation. In *CVPR*, 4085–4095.

Yu, Y.; Liu, B.; Zheng, C.; Xu, X.; Zhang, H.; and He, S. 2024. Beyond textual constraints: Learning novel diffusion conditions with fewer examples. In *CVPR*, 7109–7118.

Zenke, F.; Poole, B.; and Ganguli, S. 2017. Continual learning through synaptic intelligence. In *ICML*, 3987–3995.

Zhang, C.-B.; Xiao, J.-W.; Liu, X.; Chen, Y.-C.; and Cheng, M.-M. 2022. Representation compensation networks for continual semantic segmentation. In *CVPR*, 7053–7064.

Zhang, J.; and Gu, R. 2023. S3R: Shape and Semantics-based Selective Regularization for Explainable Continual Segmentation across Multiple Sites. *IEEE TMI*.

Zhang, J.; Gu, R.; Wang, G.; and Gu, L. 2021. Comprehensive importance-based selective regularization for continual segmentation across multiple sites. In *MICCAI*, 389–399.

Zhao, T.; Wang, Z.; Masoomi, A.; and Dy, J. 2022. Deep bayesian unsupervised lifelong learning. *Neural Networks*, 149: 95–106.

Zhou, Z.; Qi, L.; and Shi, Y. 2022. Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In *ECCV*, 420–436.

Zhu, Z.; Ma, X.; Wang, W.; Dong, S.; Wang, K.; Wu, L.; Luo, G.; Wang, G.; and Li, S. 2024. Boosting knowledge diversity, accuracy, and stability via tri-enhanced distillation for domain continual medical image segmentation. *Medical Image Analysis*, 94: 103112.