Comprehensive Importance-Based Selective Regularization for Continual Segmentation Across Multiple Sites

Jingyang Zhang[1,2,4], Ran Gu[3,4], Guotai Wang[3], and Lixu Gu[1,2,4](✉)

[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai,, China
gulixu@sjtu.edu.cn
[2] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
[3] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
[4] SenseTime Research, Shanghai, China

Abstract. In clinical practice, a desirable medical image segmentation model should be able to learn from sequential training data from multiple sites, as collecting these data together could be difficult due to the storage cost and privacy restriction. However, existing methods often suffer from catastrophic forgetting problem for previous sites when learning from images from a new site. In this paper, we propose a novel comprehensive importance-based selective regularization method for continual segmentation, aiming to mitigate model forgetting by maintaining both shape and reliable semantic knowledge for previous sites. Specifically, we define a comprehensive importance weight for each model parameter, which consists of shape-aware importance and uncertainty-guided semantics-aware importance, by measuring how a segmentation's shape and reliable semantic information is sensitive to the parameter. When training model on a new site, we adopt a selective regularization scheme that penalizes changes of parameters with high comprehensive importance, avoiding the shape knowledge and reliable semantics related to previous sites being forgotten. We evaluate our method on prostate MRI data sequentially acquired from six institutes. Results show that our method outperforms many continual learning methods for relieving model forgetting issue. Code is available at https://github.com/jingyzhang/CISR.

Keywords: Continual learning · Multi-site segmentation · Comprehensive importance · Selective regularization

1 Introduction

Convolutional neural networks have achieved remarkable performance in medical image segmentation [9]. These architectures require a large number of training images, which are commonly acquired from multiple sites (or hospitals), to

J. Zhang and R. Gu—The authors contributed equally to this work.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12901, pp. 389–399, 2021.
https://doi.org/10.1007/978-3-030-87193-2_37

improve model generalization capability. However, it is impractical to aggregate together such large multi-site datasets due to the expensive storage cost and the privacy restriction across institutes. An alternative way is to train a model with a sequential stream of multi-site data rather than a consolidated set, where data of different sites arrives in sequence without storing and access to old data of previous sites. In this setting, a naive continuous model fine-tuning scheme concerning only the new incoming site would cause considerable performance degradation on previously learned sites, called *catastrophic forgetting* [15], due to the data distribution discrepancy across multiple sites with different acquisition protocols. It is desired yet challenging to enable a model to continually segment on a new site without sacrificing the performance on previous sites.

Much effort has been directed at continual learning for mitigating model forgetting[18]. For example, experience replay methods are proposed to strengthen memory of old knowledge by explicitly storing old raw data [13,19] or implicitly training generative models [21], which yet requires additional replay storage and selection criterion. Besides, dynamically expandable networks [23] augment architecture with new modules (e.g., gating autoencoders [2] and batch normalization layers [5]) to accommodate new knowledge, contributing to zero forgetting yet causing quadratic parameter increase and requiring task label for each sample at test time. In a task-agnostic manner with fixed network architecture, selective regularization methods [1,7,25] explore model parameters that are important for preserving old knowledge, and then minimize their alterations when learning new knowledge. Despite of their success in image-level classification, a naive translation of these methods to continual segmentation would yield sub-optimal performance [4] for two-fold reasons. First, the selected important parameters are not aware of shape information that is abundant in structural dense segmentation predictions, aggravating model forgetting especially for shape knowledge. Second, the segmentation reliability is ignored, which misguides the parameter selection to fit and even remember semantic noise in the segmentation results.

In this paper, we propose a novel comprehensive importance-based selective regularization (CISR) method for continual multi-site segmentation, which mitigates model forgetting by simultaneously preserving shape information and reliable semantics for previously learned sites. To prioritize parameter usage in the model related to shape and reliable semantic information, we propose a comprehensive importance (CI) weight for each parameter that consists of shape-aware importance (SpAI) and uncertainty-guided semantics-aware importance (USmAI). Concretely, SpAI is measured by the parameter sensitivity to shape-relevant predictions (i.e., a level set representation and a segmentation embedding), accounting for the complementary shape information with local boundary and global topology. USmAI is estimated for each parameter based on the sensitivity to only confident segmentation predictions with reliable semantic information instead of uncertain ones with potential noise, by exploiting uncertainty estimation with Monte Carlo Dropout. Finally, when training model on a new site, we utilize a selective regularization scheme that penalizes changes of

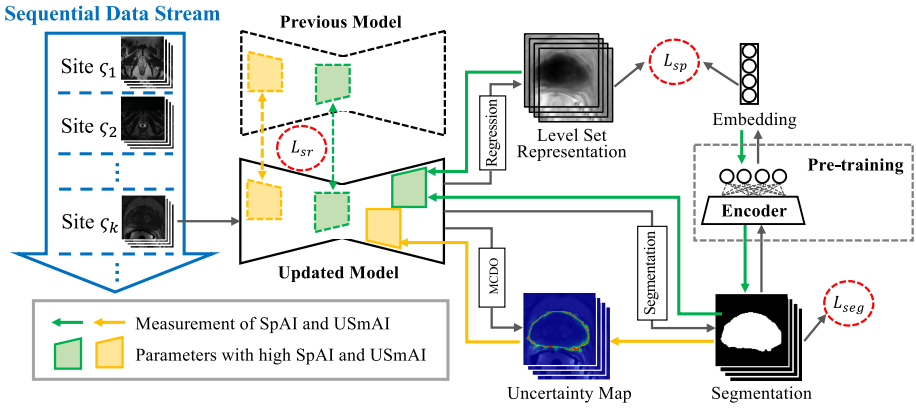

Fig. 1. Overview of our framework. For each model parameter, we define a comprehensive importance (CI) consisting of: (1) shape-aware importance (SpAI) considering complementary shape information by a level set representation and a segmentation embedding (Sect. 2.1); and (2) uncertainty-guided semantics-aware importance (USmAI) considering reliable semantic information without uncertain segmentation results (Sect. 2.2). When fine-tuning the model on a new incoming site in the sequential data stream, a selective regularization loss L_{sr} penalizes changes of important parameters with high accumulated CI, mitigating model forgetting for previously learned sites (Sect. 2.3).

important parameters with high CI, preventing both shape and semantic knowledge for previous sites being overwritten and forgotten. We have evaluated our method with the application of prostate MRI segmentation, using a sequential stream of public datasets acquired from six institutes with different acquisition protocols. The results validate that our method effectively alleviates model forgetting issue and outperforms many state-of-the-art continual learning methods.

2 Methods

In our problem setting, we are given a sequential stream of images from K sites, which are sequentially used to train a segmentation model. In round $k \in [1, K]$ of this continual learning procedure, we can only obtain images and ground truths $\{(x_n, y_n)\}_{n=1}^{N}$ from a new incoming site ζ_k without access to old data from previous sites $\{\zeta_i\}_{i=1}^{k-1}$. Figure 1 illustrates our proposed comprehensive importance-based selective regularization (CISR) method, with an objective to consecutively learn on a new site without sacrificing the performance on previous sites.

2.1 Shape-Aware Importance (SpAI)

Different from the classification task considering sample-wise isolated accuracy, the segmentation task requires structural dense predictions with abundant shape

information, which would complicate the model forgetting problem. Therefore, we first augment the vanilla segmentation model with new shape-relevant outputs to explicitly exploit the shape characteristics of segmentation object. Then we measure how these shape-relevant outputs are sensitive to each model parameter, reflecting the importance of this parameter for preserving the explicit shape information, i.e., the shape-aware importance (SpAI).

Complementary Shape-Relevant Outputs. Typically, shape information could be divided into two complementary aspects [16], i.e., local boundary and global topology. To exploit both shape aspects, we augment the network backbone with an additional regression head and also attach an autoencoder to the vanilla segmentation head. Specifically, the regression head learns from the level set representation of ground truth with signed distance transform [14], providing rich boundary delineation. The autoencoder on the top of segmentation head has encoder-decoder components, learning an intermediate representation from which the input segmentation can be reconstructed. Internally, encoder component is designed to be undercomplete [17] with a last fully-connected layer. It, pretrained by ground truth[1], can compress segmentation result into a compact embedding with highly reduced dimension, thus encoding global topology in it.

Formally, based on the mean squared error L_{mse}, we define a joint shape loss L_{sp} for shape-relevant outputs with trade-off parameter α_r and α_e:

$$L_{sp} = \alpha_r L_r + \alpha_e L_e, \text{ with } L_r = L_{mse}\left(r_n, \mathcal{T}_r(y_n)\right), \; L_e = L_{mse}\left(\mathcal{T}_e(s_n), \mathcal{T}_e(y_n)\right), \tag{1}$$

where $\mathcal{T}_r(y_n)$ denotes the level set representation of ground truth y_n, as defined by a signed distance transform [14]. Loss L_r encourages the regression output r_n to be a predicted level set representation of the segmentation target in image x_n, which delineates the detailed object boundary. Besides, $\mathcal{T}_e(s_n)$ and $\mathcal{T}_e(y_n)$ are the segmentation embedding and ground truth embedding, respectively, by passing the segmentation result s_n and ground truth y_n through the encoder component of autoencoder. Loss L_e enables $\mathcal{T}_e(s_n)$ to exploit the global topology of segmentation target in x_n. Therefore, after convergence, r_n and $\mathcal{T}_e(s_n)$ are regarded as shape-relevant outputs, characterizing complementary shape information.

Measurement of SpAI. To mitigate model forgetting for shape information, shape-relevant outputs are the targets that need to be preserved when learning on a new site. Therefore, motivated by [1], we measure the sensitivity of shape-relevant outputs r_n and $\mathcal{T}_e(s_n)$ with respect to a change of each model parameter:

$$\Omega_{ij}^{sp} = \frac{1}{N}\sum_{n=1}^{N} \frac{\beta_r \partial \left\|r_n\right\|_2^2 + \beta_e \partial \left\|\mathcal{T}_e(s_n)\right\|_2^2}{\partial \theta_{ij}}, \tag{2}$$

[1] Before each round of continual learning, the encoder component is pretrained and consecutively fine-tuned with the coupled decoder component, by minimizing a reconstruction loss with ground truth mask inputs. It should be frozen [24] in the later to avoid being corrupted by incomplete shape predictions due to model forgetting.

where $\frac{\partial \|r_n\|_2^2}{\partial \theta_{ij}}$ and $\frac{\partial \|\mathcal{T}_e(s_n)\|_2^2}{\partial \theta_{ij}}$ denote the gradient of squared L_2 norm of r_n and $\mathcal{T}_e(s_n)$ with respect to parameter θ_{ij}. β_r and β_e are trade-off parameters. Sensitivity Ω_{ij}^{sp} is obtained by averaging gradients over all N images. Intuitively, it reflects that how much a small perturbation to θ_{ij} would change the shape-relevant outputs r_n and $\mathcal{T}_e(s_n)$. Therefore, sensitivity Ω_{ij}^{sp} can also be regarded as SpAI, measuring parameter importance for preserving shape knowledge. Parameters with high SpAI should be unchanged to avoid forgetting shape knowledge when training on subsequent sites, while parameters with small SpAI can be updated without constraints since they slightly affect the shape-relevant outputs.

2.2 Uncertainty-Guided Semantics-Aware Importance (USmAI)

Besides shape knowledge, segmentation semantics is also crucial since it accounts for the pixel-wise predictions with inherent image property. However, considering the low contrast and inhomogeneous appearance of medical images [22], segmentation results may be noisy and unreliable, misguiding the selective regularization with important parameters for semantic noise that is commonly around segmentation boundary. To solve this problem, we estimate uncertainty for each segmentation prediction, and then propose a uncertainty-guided scheme to measure the parameter importance regarding only confident segmentation results with reliable semantic information, i.e., the uncertainty-guided semantics-aware importance (USmAI).

Uncertainty Estimation. Given image x_n and ground truth y_n, the network learns to predict semantic segmentation via segmentation head with loss:

$$L_{seg} = L_{ce}(s_n, y_n) + L_{dice}(s_n, y_n), \tag{3}$$

where s_n is the segmentation result of x_n. L_{ce} and L_{dice} denote the cross-entropy loss and dice loss, respectively. At the same time, we estimate uncertainty for s_n by Monte Carlo Dropout (MCDO) [6]. Specifically, given the same input x_n, we perform D times forward passes through the network backbone and segmentation head with the activation of dropout operations, leading to D-fold segmentation predictions $\{\tilde{s}_n^d\}_{d=1}^D$. The average of them is denoted by μ_n, and then used to calculate the entropy as the estimated uncertainty u_n:

$$\mu_n = \frac{1}{D}\sum\nolimits_{d=1}^{D} \tilde{s}_n^d, \qquad \text{and} \qquad u_n = -\mu_n \log \mu_n. \tag{4}$$

Measurement of USmAI. Under the guidance of uncertainty u_n, we select only confident predictions from s_n and then measure their sensitivity concerning a change of each model parameter, which is defined as USmAI for this parameter:

$$\Omega_{ij}^{sm} = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial \left\| \mathbb{I}\left(u_n < T\right) s_n \right\|_2^2}{\partial \theta_{ij}}, \tag{5}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. T is a threshold for u_n to select confident targets (low uncertainty) in s_n with reliable semantics. Ω_{ij}^{sm} is measured by averaging the gradients of these confident targets concerning parameter θ_{ij}. A higher Ω_{ij}^{sm} indicates that even a small perturbation for θ_{ij} would largely change reliable segmentation results, implying a higher importance of θ_{ij} for preserving meaningful semantic information. Therefore, the changes of parameters with high USmAI should be penalized to overcome forgetting for reliable semantics.

2.3 Comprehensive Importance-Based Selective Regularization

SpAI and USmAI are combined into a comprehensive importance (CI), prioritizing the parameter usage for keeping shape knowledge and reliable semantics:

$$\Omega_{ij}^{c} = \Omega_{ij}^{sp} + \Omega_{ij}^{sm}. \tag{6}$$

When training model on a new site, in addition to the inter-site supervised loss $L_{sp} + L_{seg}$ for this site by Eq. (1) and Eq. (3), we design a selective regularization loss L_{sr} that penalizes changes of parameters with high accumulated CI to avoid forgetting shape knowledge and reliable semantics for previous sites:

$$L = L_{sp} + L_{seg} + \lambda L_{sr}, \quad \text{with} \quad L_{sr} = \sum\nolimits_{i,j} \Omega_{ij}^{c^*} (\theta_{ij} - \theta_{ij}^*)^2, \tag{7}$$

where λ is a trade-off parameter. Notably, Ω_{ij}^{c} is computed in each learning round for a specific incoming site, and accumulated over all previously learned sites by moving average, as denoted by $\Omega_{ij}^{c^*}$. It is used as a weight for the change between current parameter θ_{ij} and old parameter θ_{ij}^* (as determined by optimizing Eq. (7) for the previous site in the sequence), formulating the selective regularization loss L_{sr} to avoid changing parameters with high $\Omega_{ij}^{c^*}$ that are important for previous sites. In this way, both shape and reliable semantic knowledge can be effectively preserved, and thus catastrophic forgetting problem would be mitigated.

3 Experiments

Dataset. We employed a well-established multi-site prostate T2-weighted MRI dataset [12], including 30 cases with in/through plane resolution 0.6–0.625/3.6–4 mm from RUNMC [3] (Site A), 30 cases with resolution 0.4/3mm from BMC [3] (Site B), 19 cases with resolution 0.67–0.79/1.25 mm from HCRUDB [8] (Site C), 13 cases with resolution 0.325–0.625/3-3.6 mm from UCL [10] (Site D), 12 cases with resolution 0.25/2.2–3 mm from BIDMC [10] (Site E), and 12 cases with resolution 0.625/3.6 mm from HK [10] (Site F). We organized this multi-site dataset in a sequential stream ordered by Site A→B→C→D→E→F. For pre-processing, we resized all images to size 384×384 in the axial plane and normalized them to zero mean and unit variance. For each site, we used images from 60%, 15% and 25% of cases for training, validation and testing.

Implementation. We adopted 2D-UNet [20] as network backbone due to the large variance on through-plane resolution among different sites [11]. Weight λ was empirically set as a large value 10^5 [18]. Parameter α_r, α_e, β_r and β_e were set as 0.001, 0.1, 0.1, 0.001 for suitable trade-off. The autoencoder was designed with mirrored encoder-decoder components, where the encoder component was used as an embedding network containing three cascaded blocks with Conv 3×3 (16, 32, 64 kernels) using stride 2 and 1 in each, followed by a flatten operation and a fully-connected layer with 64 hidden units. In each learning round with data from a new site, we optimized the objective function Eq. (7) by AdamOptimizer with learning rate 5×10^{-4}, batch size 5 and epoch number 200.

Evaluation Metrics. We evaluate the segmentation performance by dice similarity coefficient (DSC) and average symmetric surface distance (ASD). After the model finishes continual learning on the last site ζ_K, we compute DSC and ASD on all sites $\{\zeta_i\}_{i=1}^K$ (including the current and previous sites), leading to two sets of results $\{D_{K,i}\}_{i=1}^K$ and $\{A_{K,i}\}_{i=1}^K$, respectively. Entry $D_{K,i}$ and $A_{K,i}$ denote the test DSC and ASD on site ζ_i after learning on the last site ζ_K. Based on them, we define several specialized metrics for continual learning, i.e., the *average* of DSC and ASD ($\text{DSC}^{ave}=\sum_{i=1}^K D_{K,i}/K$, $\text{ASD}^{ave}=\sum_{i=1}^K A_{K,i}/K$) for the generic evaluation; the *backward transfer* [13] of DSC and ASD ($\text{DSC}^{bwt}=\sum_{i=1}^{K-1}(D_{K,i}-D_{i,i})/(K-1)$, $\text{ASD}^{bwt}=\sum_{i=1}^{K-1}(A_{K,i}-A_{i,i})/(K-1)$) that particularly reflect model forgetting for previous sites. Notably, an advanced continual learning method should have a high DSC^{ave} and DSC^{bwt} with a low ASD^{ave} and ASD^{bwt}.

Comparison with State-of-the-art Continual Learning Methods. We compare our CISR with several state-of-the-art continual learning methods, including elastic weight consolidation (EWC) method [7] that preserves old knowledge based on the Fisher information, synaptic intelligence (SI) method [25] that updates network memory in an online manner, and memory-aware synapses (MAS) method [1] considering plain semantics without the uncertainty-guided scheme. We also implement a naive continuous fine-tuning (FT) scheme as baseline method with only the inter-site supervised loss considered in Eq. (7).

Table 1 lists the quantitative evaluation after continual learning finished on the last site F. FT suffers from the severe model forgetting problem (the worst *bwt* measures of DSC^{bwt} and ASD^{bwt}) and the poor segmentation results (the worst *ave* measures of DSC^{ave} and ASD^{ave}). Classical methods of EWC, SI and MAS improves the continual learning performance over FT, while their advantages are still relatively limited. Importantly, our proposed CISR outperforms all these methods, e.g., an advantage over MAS by 3.13% DSC^{ave}, 5.23% DSC^{bwt}, 0.22 mm ASD^{ave}, 0.37 mm ASD^{bwt}, indicating its superiority for continual learning without sacrificing performance on previous sites.

We also report in Fig. 2 how the test performance on site A changes during the entire continual learning. Once finishing learning on new sites, FT immediately forgets previously learned knowledge for site A, leading to evident performance decrease on it. EWC, SI and MAS mitigate this problem to some extent, yet still

Table 1. Evaluation after the model finishes continual learning on the last site F.

Method	$DSC^{ave}(\%)\uparrow$	$DSC^{bwt}(\%)\uparrow$	$ASD^{ave}(mm)\downarrow$	$ASD^{bwt}(mm)\downarrow$
FT (baseline)	40.78 ± 28.26	–50.54 ± 21.58	13.66 ± 8.67	13.75 ± 6.52
EWC [7]	68.83 ± 17.51	–23.90 ± 14.09	4.43 ± 3.34	3.94±3.14
SI [25]	75.97 ± 14.11	–15.27 ± 10.05	3.87 ± 2.40	2.97 ± 2.96
MAS [1]	76.81 ± 11.18	–13.23 ± 8.24	3.64 ± 3.10	2.28 ± 3.32
USmAI	78.11 ± 10.60	–11.42 ± 7.92	3.49 ± 2.80	2.06 ± 3.02
$SpAI_r$	77.85 ± 10.96	–12.10 ± 8.07	3.51 ± 2.75	2.16 ± 2.76
$SpAI_e$	77.53 ± 11.04	–12.74 ± 8.34	3.57 ± 2.85	2.36 ± 2.69
SpAI	79.31 ± 8.86	–9.89 ± 6.19	3.44 ± 2.89	2.11 ± 2.89
CISR (**Ours**)	**79.94 ± 7.71**	**–8.00 ± 6.15**	**3.42 ± 2.83**	**1.91 ± 2.86**

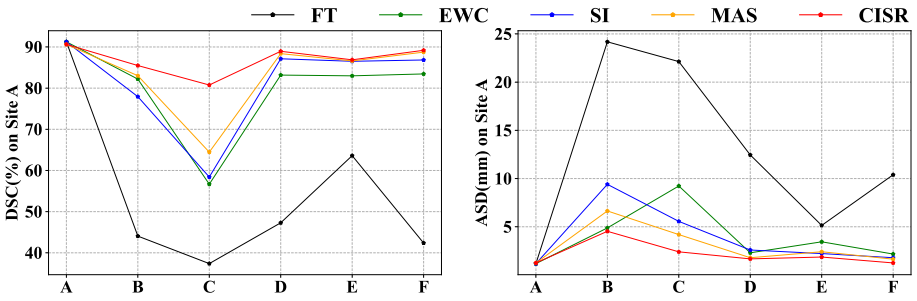

Fig. 2. Changes of test performance on Site A by different methods during continual learning, where the model is sequentially trained on sites A→B→C→D→E→F.

challenged by model forgetting especially after training on site C, since most images from this site contain prostate cancer [8] with visible appearance difference for previous site A. Our CISR maintains consistently the least performance decrease on site A, showing its best capability for reducing model forgetting.

Furthermore, we visualize in Fig. 3 the segmentation results on site A, D and F before and after the model finishes learning on site F. By training on site F that is newly incoming in sequence, all methods achieve highly improved performance on this site (the third row), indicating their sufficient adaptation capability for new site. However, FT leads to severe performance degradation on previously learned site A and D. Among all methods, our CISR achieves the most accurate results on site A and D after training on site F, and maintains the highest overlap ratio with previously obtained results (the first two rows). Besides, high uncertainty focuses on object ambiguous boundaries with semantic noise, explaining the feasibility of our uncertainty guidance for filtering out them.

Ablation Study. We validate the role of SpAI (including $SpAI_r$ using only r_n for boundary delineation and $SpAI_e$ using only $\mathcal{T}_e(s_n)$ for topology awareness)

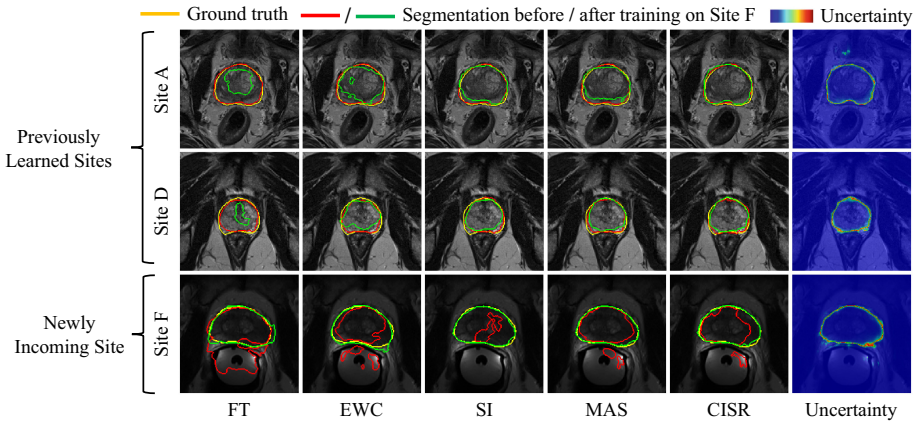

Fig. 3. Visualization of segmentation examples on site A, D and F before (i.e., after training on site A, D and E, respectively) and after training on the newly incoming site F. Also, the uncertainty maps of our CISR is visualized in the last column.

and USmAI for reliable semantics in our method. Table 1 shows that USmAI outperforms MAS owing to our uncertainty-guided scheme for filtering out semantic noise. Preserving shape knowledge by SpAI_r and SpAI_e has an advantage over MAS, and a combination of them in SpAI even improves performance over each single component. Beside, incorporating SpAI and USmAI in our CISR further facilitates more advantages owing to the jointly preserved shape and semantics.

4 Conclusion

This paper proposes a comprehensive importance-based selective regularization method for continual multi-site segmentation. We propose to reduce model forgetting by strengthening network memory for both shape knowledge (with complementary boundary and topology cues) and reliable semantics (with only confident predictions). Experiments show the effectiveness on prostate segmentation with sequential multi-site data. In the further, it would be of interest to study the application for other continual segmentation tasks with longer site sequences.

Acknowledgments. This research is partially supported by the National Key research and development program (No. 2016YFC0106200), Beijing Natural Science Foundation-Haidian Original Innovation Collaborative Fund (No. L192006), and the funding from Institute of Medical Robotics of Shanghai Jiao Tong University as well as the 863 national research fund (No. 2015AA043203).

References

1. Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., Tuytelaars, T.: Memory aware synapses: learning what (not) to forget. In: Proceedings of the European Conference on Computer Vision, pp. 139–154 (2018)

2. Aljundi, R., Chakravarty, P., Tuytelaars, T.: Expert gate: lifelong learning with a network of experts. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3366–3375 (2017)
3. Bloch, N., et al.: NCI-ISBI 2013 challenge: automated segmentation of prostate structures. Cancer Imaging Arch. (2015)
4. Douillard, A., Chen, Y., Dapogny, A., Cord, M.: PLOP: learning without forgetting for continual semantic segmentation. arXiv preprint arXiv:2011.11390 (2020)
5. Karani, N., Chaitanya, K., Baumgartner, C., Konukoglu, E.: A lifelong learning approach to brain MR segmentation across scanners and protocols. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 476–484. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_54
6. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems, pp. 5574–5584 (2017)
7. Kirkpatrick, J., et al.: Overcoming catastrophic forgetting in neural networks. Proc. Natl. Acad. Sci. **114**(13), 3521–3526 (2017)
8. Lemaître, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M., Meriaudeau, F.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. Comput. Biol. Med. **60**, 8–31 (2015)
9. Litjens, G., Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
10. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med. Image Anal. **18**(2), 359–373 (2014)
11. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans. Med. Imaging **39**(9), 2713–2724 (2020)
12. Liu, Q., Dou, Q., Heng, P.-A.: Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 475–485. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_46
13. Lopez-Paz, D., Ranzato, M.: Gradient episodic memory for continual learning. In: Advances in Neural Information Processing Systems, pp. 6467–6476 (2017)
14. Ma, J., He, J., Yang, X.: Learning geodesic active contours for embedding object global information in segmentation CNNs. IEEE Trans. Med. Imaging **40**(1), 93–104 (2021)
15. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: the sequential learning problem. In: Psychology of Learning and Motivation, vol. 24, pp. 109–165. Elsevier (1989)
16. Navarro, F., et al.: Shape-aware complementary-task learning for multi-organ segmentation. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) MLMI 2019. LNCS, vol. 11861, pp. 620–627. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_71
17. Oktay, O., et al.: Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging **37**(2), 384–395 (2017)
18. Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., Wermter, S.: Continual lifelong learning with neural networks: a review. Neural Netw. **113**, 54–71 (2019)
19. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: ICARL: incremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2001–2010 (2017)

20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Shin, H., Lee, J.K., Kim, J., Kim, J.: Continual learning with deep generative replay. In: Advances in Neural Information Processing Systems, pp. 2990–2999 (2017)
22. Wang, G., et al.: Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans. Med. Imaging **37**(7), 1562–1573 (2018)
23. Yoon, J., Yang, E., Lee, J., Hwang, S.J.: Lifelong learning with dynamically expandable networks. arXiv preprint arXiv:1708.01547 (2017)
24. Yue, Q., Luo, X., Ye, Q., Xu, L., Zhuang, X.: Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. In: Medical Image Computing and Computer-Assisted Intervention, pp. 559–567 (2019)
25. Zenke, F., Poole, B., Ganguli, S.: Continual learning through synaptic intelligence. Proc. Mach. Learn. Res. **70**, 3987 (2017)