Source-free Domain Adaptation for Multi-site and Lifespan Brain Skull Stripping

Yunxiang Li¹, Ruilong Dan¹, Shuai Wang², Yifan Cao¹, Xiangde Luo³, Chenghao Tan¹, Gangyong Jia¹, Huiyu Zhou⁴, Yaqi Wang¹,⁵(✉), Li Wang⁶(✉)

¹ School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
² School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China
³ School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
⁴ School of Computing and Mathematical Sciences, University of Leicester, UK
⁵ College of Media Engineering, Communication University of Zhejiang, Hangzhou, China
⁶ Developing Brain Computing Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA

wangyaqi@cuz.edu.cn
li_wang@med.unc.edu

Abstract. Skull stripping is a crucial prerequisite step in the analysis of brain magnetic resonance (MR) images. Although many excellent works or tools have been proposed, they suffer from low generalization capability. For instance, the model trained on a dataset with specific imaging parameters (source domain) cannot be well applied to other datasets with different imaging parameters (target domain). Especially, for the lifespan datasets, the model trained on an adult dataset is not applicable to an infant dataset due to the large domain difference. To address this issue, numerous domain adaptation (DA) methods have been proposed to align the extracted features between the source and target domains, requiring concurrent access to the input images of both domains. Unfortunately, it is problematic to share the images due to privacy. In this paper, we design a source-free domain adaptation framework (SDAF) for multi-site and lifespan skull stripping that can accomplish domain adaptation without access to source domain images. Our method only needs to share the source labels as shape dictionaries and the weights trained on the source data, without disclosing private information from source domain subjects. To deal with the domain shift between multi-site lifespan datasets, we take advantage of the brain shape prior which is invariant to imaging parameters and ages. Experiments demonstrate that our framework can significantly outperform the state-of-the-art methods on multi-site lifespan datasets.

Keywords: Transformer · Skull stripping · Domain adaptation · Shape dictionary · Lifespan brain
1 Introduction

Skull stripping, the separation of brain tissue from non-brain tissue, is a critical preprocessing step for the characterization of brain MRIs. Plenty of skull stripping methods have been proposed, e.g., morphology-based method: Brain Surface Extractor (BSE) [1], surface-based method: Brain Extraction Tool (BET) [2], and meta-method: Brain MAPS [3]. Compared with traditional skull stripping tools, deep learning has recently been proven more suitable for skull stripping, where 3D U-Net is the most popular backbone [4,5]. However, the high performance of deep learning-based methods requires that the training and testing datasets share a similar data distribution, which is hardly met due to a variety of device manufacturers, magnetic field strength, and acquisition protocols. Moreover, there is also a substantial data distribution difference across lifespan, e.g., the adult and infant brain MRIs in Fig. 1, where the infant’s brain is undergoing myelination and maturation.

Fig. 1. A schematic overview of our proposed source-free domain adaptation framework.

Many efforts have been devoted to addressing the domain shift problem. Among them, the most widely used method is domain adaptation to align the latent feature distributions of the two domains [6,7,8]. Unfortunately, one limitation is that it requires concurrent access to the input images of both the source and target domains. It is well known that, compared with natural images, medical images are difficult to share due to privacy. To effectively enhance the segmentation performance in different domains, we propose a novel source-free domain adaptation method in this work to deal with the absence of concurrent access to the source and target domain images. The main contributions of our method are three-fold: 1) A novel source-free domain adaptation framework is designed, without disclosing the privacy. 2) A shape dictionary based on the
Fourier Descriptors is proposed to fully utilize the anatomical prior knowledge of the brain shape. 3) To better model the overall shape information, we designed a Shape AutoEncoder (SAE) based on Shuffle Transformer.

2 Method

2.1 Overall Architecture:

The overall framework is illustrated in Fig. 1. Specifically, we can arbitrarily take an off-the-shelf model trained on the source domain as the segmentation network, and share the source labels (excluding the privacy information) with the target domain. Fourier Descriptors [9] of the source labels are computed to build a shape dictionary. Then, the target domain image is directly input into the shared model, and Fourier Descriptors of the segmentation results are computed. Through it, we can retrieve a label with the closest shape from the shape dictionary. Finally, the segmentation results together with the retrieved labels serve as inputs into Shape AutoEncoder to further refine the segmentation results.

2.2 Shape Dictionary

To make full use of the anatomy prior knowledge of brain shape, we calculate the corresponding Fourier Descriptors for each subject according to the source labels and store them in the dictionary. Assuming source labels \( L_s = \{l_s^i\}_{i=1}^{L_s^i} \), the constructing process of shape dictionary \( D_s \) is defined as Eq. (1)

\[
D_s = \{d_s^i = F(l_s^i)\}_{i=1}^{L_s^i}
\] (1)

where \( F \) is the Fourier Descriptor for a quantitative representation of closed shapes independent of their starting point, scale, location, and rotation. The whole process of computing Fourier Descriptors consists of three steps: (1) Establish a coordinate system in the upper left corner of the boundary, and the coordinate axis is tangent to the boundary. (2) Take the two axes as real and imaginary numbers respectively, and the coordinates of points \((x_m, y_m)\) on the boundary are expressed in complex numbers \(z(m) = x_m + j y_m\). (3) The discrete Fourier transform (DFT) is applied to the above coordinates to obtain the Fourier Descriptor of the boundary shape, which is defined as \(Z(k)\) in Eq. (2).

\[
Z(k) = \frac{1}{N} \sum_{m=0}^{N-1} z(m)e^{-j2\pi mk/N}, k = 0, 1, 2, \ldots, N - 1
\] (2)

where \(N\) is the amount of the boundary points. Specifically, we choose 10 low-frequency coefficients as the final Fourier Descriptors, which is sufficient to achieve the required accuracy for retrieving [10].
2.3 Shape AutoEncoder (SAE)

The Shape AutoEncoder is designed to further refine the segmentation results based on both segmentation maps from the segmentation network and shape reference from the shape dictionary, with the detailed structure shown in Fig. 2. To better extract global information such as overall shape, Transformer-based methods are good candidates. Nevertheless, most 3D Transformer-based networks face cubic computational complexity. There are two solutions to reduce the computational complexity. The first one, which is represented by ViT [11], is to take a patch with a certain size as an element, but it is in the image level and more suitable for classification tasks. The second one is Swin Transformer [12][13][14], which performs pixel-level operations in a small local window, and gradually obtains global information via deepening the network, but it does not fully utilize Transformer’s global modeling capability. Therefore, we design a novel model so-called Shuffle Transformer to capture global capability with low computational complexity.

**Shuffle Transformer:** Our shuffle Transformer mixes the voxel features regularly, then evenly divides them into non-overlapping groups. Specifically, Shuffle Module takes the feature $X$ as input and outputs shuffled blocks $X^b$ thus it can be formulated through Eq. (3).

$$X^b = \{x_i^b \mid x_i^b = \text{Split}(\text{Shuffle}(\text{Group}(X, i))), x_i^b \in \mathbb{R}^{D_{n_1} \times H_{n_2} \times W_{n_3}} \}_{i=1}^{(n_1 \times n_2 \times n_3)}$$

(3)

where the details of Group, Shuffle, and Split are depicted in Fig. 2. After these operations we obtain a total number of $(n_1 \times n_2 \times n_3)$ shuffled blocks $x^b$, each with a size of $D_{n_1} \times H_{n_2} \times W_{n_3}$. Subsequently, each group is computed by multi-head self-attention. Since our shuffle operation keeps the relative position of each element in the shuffled blocks as the same as the original feature block, our position encoding method is based on relative-distance-aware position encoding [15][16][17]. The query-key-value (QKV) attention [18] in each small block $x^b$ can
be computed by Eq. (4).

\[
\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}} + B)V
\]  

(4)

where \(Q\), \(K\), and \(V\) stand the query, key, and value matrices of dimension \(d_k\) respectively, and \(B\) denotes the position biases matrix.

**Self-supervised Training of Shape AutoEncoder:** As for the training process of SAE, it is based on self-supervision, meaning the supervisory signals are generated by the input, and we use the shared source labels as the training images of SAE. Each image is processed by two kinds of random spatial transformations \((T_1, T_2)\). Assume we have images \(Y = \{y_i\}_{i=1}^\eta\) where \(\eta\) denotes the number of images, thus transformed images are generated via Eq. (5).

\[
\begin{align*}
Y^{T_1} &= \{y_i^{T_1} \mid y_i^{T_1} = T_1(y_i)\}_{i=1}^\eta \\
Y^{T_2} &= \{y_i^{T_2} \mid y_i^{T_2} = T_2(y_i)\}_{i=1}^\eta
\end{align*}
\]  

(5)

To mimic the unreliable segmentation results caused by domain shift, random noises \(RN\) are added to \(T_2\) through placing random false positive and false negative stain to generate \(Y^{T_2\circ RN}\) formulated by Eq. (6).

\[
Y^{T_2\circ RN} = RN(Y^{T_2})
\]  

(6)

Let \(SAE(y_i^{T_1}, y_i^{T_2\circ RN}; \theta)\) be the Shape AutoEncoder, thus refined outputs \(\hat{Y}\) are defined via Eq. (7).

\[
\hat{Y} = SAE(Y^{T_1}, Y^{T_2\circ RN}; \theta)
\]  

(7)

\(\hat{Y}\) and \(\theta\) represent the prediction output and the trainable parameters of SAE respectively, and the self-supervised loss function of the SAE is defined by Eq. (8).

\[
L_{SAE}(\hat{Y}, Y^{T_2}) = -\frac{1}{\eta} \sum_{i=1}^{\eta} (T_2(y_i) \times \ln(SAE(T_1(y_i), RN(T_2(y_i)); \theta)) \\
+ (1 - T_2(y_i)) \times (1 - \ln(SAE(T_1(y_i), RN(T_2(y_i)); \theta))))
\]  

(8)

In this way, it can not only combine segmentation results and shape reference but also denoise and refine the unreliable segmentation results automatically through learning shape prior knowledge from the labels.

### 3 Implementation and Experiments

#### 3.1 Datasets and Evaluation Metrics

We evaluated the proposed method on the publicly available dataset, where the source domain is from Neurofeedback Skull-stripped (NFBS) [19], and the target
domains are from Alzheimer’s Disease Neuroimaging Initiative (ADNI) [20] and Developing Human Connectome Project (dHCP) [21]. Note that subjects from NFDS are young adults from 21 to 45 years old and ADNI are elder adults from 55 and 90 years old, and dHCP are newborns. After resampling and padding, the size of the individual scan is $256 \times 256 \times 256$ and each voxel size is $1 \times 1 \times 1 \text{mm}^3$. We selected 25 subjects from NFBS with manual labels as the training dataset and 10 subjects from ADNI and dHCP as the testing dataset, and 3-fold cross-validation is used. It is worth noting that there is no available publicly manual label of dHCP, thus we only compare the results by visual inspection. Due to the limited GPU memory, a sub-volume of size $64 \times 64 \times 64$ is used as the first stage segmentation network input. In order to better capture the overall shape, the input size of the Shape AutoEncoder is set to $8 \times 256 \times 256$, and the middle layer slice with the size of $256 \times 256$ is used to compute the Fourier Descriptors and retrieve the most similar shape from the dictionary. For all experiments described below, the Average Surface Distance (ASD), the Dice Coefficients (DICE), the sensitivity (SEN) and specificity (SPE) are chosen for evaluation metrics.

3.2 Implementation Details

Our all experiments were based on the PyTorch framework and carried out on 4 Nvidia RTX 2080Ti GPUs. We trained our network from scratch for a total of 10000 iterators, and the parameters were updated by the Adam algorithm (momentum = 0.97, weight decay = $5 \times 10^{-4}$). We adopt a batch size of 4 and set the learning rate as $2 \times 10^{-4}$.

| Method   | Adaptation               | ASD (mm)      | DICE (%)    | SPE (%)     | SEN (%)     |
|----------|--------------------------|---------------|-------------|-------------|-------------|
| 3D U-Net | Train on Source Only     | 11.57±9.83    | 88.30±3.53  | 99.51±0.68  | 82.69±3.10  |
| CycleGAN | Appearance Translation   | 18.96±7.88    | 86.27±2.85  | 98.72±0.51  | 85.19±2.58  |
| EMNet    | Feature Alignment        | 8.67±9.86     | 91.49±2.39  | 99.37±0.49  | 85.32±2.10  |
| Tent     | Minimize Entropy (Source-free) | 7.19±0.24 | 90.52±1.70  | 99.64±0.28  | 85.60±2.18  |
| SDAF (ours) | Shape Prior (Source-free) | 4.63±0.96    | 91.57±1.38  | 99.55±0.15  | 88.09±2.67  |

3.3 Experimental Results on Cross-site Dataset

Our results are presented in Table 1 from which it can be observed that our method achieves better performance than 3D U-Net. Moreover, even without concurrent access to the source and target domain images, our method outperforms the state-of-the-art domain adaptation method, i.e., EMNet. Interestingly, the ASD achieved by other methods was unexpectedly far higher than ours. This may be due to the fact that the training data from the source domain do not have the shoulder part, but the testing data in the target domain do have the shoulder, which is illustrated in Fig. 3. Our method is capable to identify the segmentation result of the shoulder as unreliable results and exclude it from
the brain tissues. We can also notice that the output of CycleGAN mistakenly identifies many non-brain regions as the brain. Consequently, its performance is unexpectedly poorer than those obtained from the source model without domain adaptation.

![Fig. 3. Visualization of segmentation results on ADNI.](image)

### 3.4 Visualization Results on Newborns

We qualitatively compare the segmentation outputs of the proposed SDAF and the 3D U-Net on the newborn. Although the infant shares a similar shape and basic structure with the adult, the infant’s brain develops rapidly throughout the first year of life, resulting in huge appearance differences with the adult. As shown in Fig. 4, 3D U-Net cannot achieve accurate brain tissues with fuzzy brain boundaries, while our method is able to generate smooth and reasonable segmentations.

![Fig. 4. Qualitative comparison of segmentation results on newborn brain MRIs from dHCP.](image)
3.5 Ablation Study

In order to verify the effectiveness of the main components of our proposed SDAF, we conducted the following ablation study: a) implement our SDAF without Shape AutoEncoder, referred to as SDAF w/o SAE. b) implement our SDAF without shape dictionary, that is, only the unrefined segmentation results are input to Shape AutoEncoder, referred to as SDAF w/o SD. c) implement our SDAF without shuffle Transformer, that is, replace the shuffle Transformer block with the basic convolution block, denoted as SDAF w/o ST. Quantitative comparison results of the three variants along with the SDAF are illustrated in Table 2. The proposed SDAF achieves improved performance, especially in terms of important ASD and DICE metrics.

| Method       | ASD (mm)   | DICE (%) | SPE (%) | SEN (%) |
|--------------|------------|----------|---------|---------|
| SDAF w/o SAE | 11.57±9.83 | 88.30±1.53 | 99.51±0.68 | 82.69±1.10 |
| SDAF w/o SD  | 6.86±3.53  | 89.73±1.42 | 99.41±0.13 | 86.04±2.21 |
| SDAF w/o ST  | 6.39±4.24  | 90.62±1.40 | **99.76±0.13** | 84.79±2.66 |
| SDAF (ours)  | 4.63±0.98  | **91.57±1.38** | 99.55±0.15 | **88.09±2.67** |

3.6 Comparison with the Traditional Tools

As skull stripping is typically an initial step of most brain MRI studies, plenty of pipelines/tools were developed previously. To demonstrate the advantage of the proposed framework, we further make comparisons with BET, one of the most widely used tools. As illustrated in Fig. 5, we can observe that BET always under-segments the brain tissues from the infant dataset, while our method can achieve reasonable results. Furthermore, we also apply our framework to refine the unreasonable segmentation by BET through our Shape AutoEncoder, and we find the BET results are improved by visual inspection, indicating that our method has a wide application.

![Fig. 5. Qualitative results of comparison and combination between traditional tool BET and our method on dHCP.](image)
4 Conclusion

In summary, we presented a source-free domain adaptation method and successfully applied it to the skull stripping task on multi-site lifespan datasets without the need for any annotations on the target domain or concurrent access to the input images of both the source and target domains. Our method consists of a segmentation network, a shape dictionary, and a Shape AutoEncoder. With the assistance of the shape dictionary, the Shape AutoEncoder makes full use of the anatomical prior knowledge to refine the unreliable segmentation results on the target domain. Experimental results demonstrated that the proposed method achieves better performance than the state-of-the-art unsupervised domain adaptation methods, and the proposed Shape AutoEncoder can further enhance traditional skull stripping tools.

References

1. David W Shattuck, Stephanie R Sandor-Leahy, Kirt A Schaper, David A Rottenberg, and Richard M Leahy. Magnetic resonance image tissue classification using a partial volume model. NeuroImage, 13(5):856–876, 2001.
2. Stephen M Smith. Fast robust automated brain extraction. Human brain mapping, 17(3):143–155, 2002.
3. Kelvin K Leung, Josephine Barnes, Marc Modat, Gerard R Ridgway, Jonathan W Bartlett, Nick C Fox, Sébastien Ourselin, Alzheimer’s Disease Neuroimaging Initiative, et al. Brain maps: an automated, accurate and robust brain extraction technique using a template library. Neuroimage, 55(3):1091–1108, 2011.
4. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention, pages 424–432. Springer, 2016.
5. Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021, pages 424–432, 2021.
6. Yunxiang Li, Jingxiong Li, Ruilong Dan, Shuai Wang, Kai Jin, Guodong Zeng, Jun Wang, Xiangji Pan, Qianni Zhang, Huiyu Zhou, et al. Dispensed transformer network for unsupervised domain adaptation. arXiv preprint arXiv:2110.14944, 2021.
7. Guodong Zeng, Florian Schmaranzer, Till D Lerch, Adam Boschung, Guoyan Zheng, Jürgen Burger, Kate Gerber, Moritz Tannast, Klaus Siebenrock, Young-Jo Kim, et al. Entropy guided unsupervised domain adaptation for cross-center hip cartilage segmentation from mri. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 447–456. Springer, 2020.
8. Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiaihai Zhuang, and Pheng-Ann Heng. Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access, 7:99065–99076, 2019.
9. Mark S. Nixon and Alberto S. Aguado. Chapter 7 - object description. In Mark S. Nixon and Alberto S. Aguado, editors, *Feature Extraction & Image Processing for Computer Vision (Third Edition)*, pages 343–397. Academic Press, Oxford, third edition edition, 2012.

10. Christoph Dalitz, Christian Brandt, Steffen Goebbels, and David Kolanus. Fourier descriptors for broken shapes. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–11, 2013.

11. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

12. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

13. Yuxiang Li, Guodong Zeng, Yifan Zhang, Jun Wang, Qin Jin, Lingling Sun, Qianli Zhang, Qisi Lian, Guiping Qian, Neng Xia, et al. Agmb-transformer: Anatomy-guided multi-branch transformer network for automated evaluation of root canal therapy. *IEEE Journal of Biomedical and Health Informatics*, 2021.

14. Yuxiang Li, Shuai Wang, Jun Wang, Guodong Zeng, Wenjun Liu, Qianli Zhang, Qin Jin, and Yaqi Wang. Gt u-net: A u-net like group transformer network for tooth root segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 386–395. Springer, 2021.

15. Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 464–468, 2018.

16. Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jon Shlens. Stand-alone self-attention in vision models. *Advances in Neural Information Processing Systems*, 32, 2019.

17. Irwan Bello, Barret Zoph, Ashish Vaswani, Jonathon Shlens, and Quoc V. Le. Attention augmented convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

18. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

19. Simon F Eskildsen, Pierrick Coupé, Vladimir Fonov, José V Manjón, Kelvin K Leung, Nicolas Guizard, Shafik N Wassef, Lasse Riis Östergaard, D Louis Collins, Alzheimer's Disease Neuroimaging Initiative, et al. Beast: brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012.

20. Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

21. Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.
22. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232, 2017.

23. Yue Sun, Kun Gao, Zhengwang Wu, Guannan Li, Xiaopeng Zong, Zhihao Lei, Ying Wei, Jun Ma, Xiaoping Yang, Xue Feng, et al. Multi-site infant brain segmentation algorithms: The iseg-2019 challenge. IEEE Transactions on Medical Imaging, 40(5):1363–1376, 2021.

24. Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In International Conference on Learning Representations, 2021.