OSATTA: One-Shot Automatic Test Time Augmentation for Domain Adaptation

Felix Küper and Sergi Pujades

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

Abstract. Fundamental models (FM) are reshaping the research paradigm by providing ready-to-use solutions to many challenging tasks, such as image classification, registration, or segmentation. Yet, their performance on new dataset cohorts significantly drops, particularly due to domain gaps between the training (source) and testing (target) data. Recently, test-time augmentation strategies aim at finding target-to-source-mappings (t2sm), which improve the performance of the FM on the target dataset by leveraging the FM weights, thus assuming access to them. While this assumption holds for open research models, it does not for commercial ones (e.g., Chat-GPT). These are provided as black boxes; thus, the training data and the model weights are unavailable. In our work, we propose a new generic few-shot method that enables the computation of a target-to-source mapping by only using the black-box model's outputs. We start by defining a parametric family of functions for the t2sm. Using a simple loss function, we optimize the t2sm parameters based on a single labeled image volume. This effectively provides a mapping between the source domain and the target domain. In our experiments, we investigate how to improve the segmentation performance of a given FM (a UNet), and we outperform state-of-the-art accuracy in the 1-shot setting, with further improvement in a few-shot setting. Our approach is invariant to the model architecture as the FM is treated as a black box, which significantly increases our method's practical utility in real-world scenarios. The code is available for reproducibility purposes at https://osatta.gitlabpages.inria.fr/MedAGI.

1 Introduction

Fundamental models (FM) are reshaping the research paradigm by providing ready-to-use solutions to many challenging tasks, such as image classification, registration, or segmentation. The notion of a *fundamental model* is based on the hypothesis that the model is trained with such a high quantity and diversity of data, that it can generalize to any new dataset. In real-world scenarios, though, when dealing with small specific datasets, the performance of FM models severely deteriorates, mainly due to the domain gap between the training data used for the FM (source) and the new dataset (target). These domain gaps naturally arise from several sources. One source is the diversity in the acquisition devices; different vendor scanners or calibration parameters produce images with different biases and noise patterns. A frequent domain gap also appears

due to the differences in the observed patient population. These can be related to gender, age, ethnicity, anatomy, or pathologies. For these reasons, leveraging the potential of FM models on small specific datasets remains challenging to this day.

Interestingly, many different strategies have been studied in the literature to address this problem. They can be structured according to their stated hypothesis with regard to the nature of the FM and the target dataset. These hypotheses can be defined by answering the following questions: i) Is the training source data available? ii) What is the quantity of labeled and unlabeled target data samples? iii) Are the FM model weights available? Different combinations arise, which shape today's state-of-the-art. We start by describing approaches that try to make the FM as generic as possible by using data augmentation and then describe strategies to adapt them to target data.

Domain generalization by data augmentation at training time. The goal here is to build the strongest possible model from the available data to best generalize to new unseen data. One widespread approach is data augmentation. However, choosing suitable image transformations is not trivial. Therefore, Cubuk et al. introduced AutoAugment [1] in which optimal augmentation strategies are automatically found. Since augmenting during training time is a bi-level optimization problem, their solution is highly compute-intensive. Later, Li et al. introduced 'Differentiable Automatic Data Augmentation" [2], in which they use the Gumbel-softmax parametrization trick to enable gradient descent over augmentation strategies. Follow-up work further optimized these methods, especially for the field of medical imaging [3–7].

Domain adaptation. If the FM weights can be adapted using source and target domain data, we are in the field of Domain Adaptation (DA) [8]. If access to source data is not possible, we are in the setting of source-free domain adaptation (SFDA) [9] or source-free unsupervised domain adaptation (SFUDA) [10]. If sufficient labeled target data is available, Continual Learning strategies can be used [11]. These do not only allow to adapt a FM to a new domain, but also to new tasks. However, in a clinical context, the availability of labeled target data is limited, thus special strategies have been proposed. For example, Gaillochet et al. use uncertainty estimates during test-time adaptation to determine the best image to train on next [12], and Xu et al. introduced a gradual multi-stage technique to improve fine-tuning in low-resource scenarios [13]. If labeled target data is available, but data sharing is not possible, federated learning is a possibility. Li et al. introduced a technique that enables multi-site learning while preserving the privacy of the respective datasets [14]. For a more detailed look at this approach, we refer the reader to this survey [15]. Some methods assume access to unlabeled data from both target and source domain. Inspired by works such as CycleGan [16], their approach is to learn the t2sm directly from image pairs. With the recent advent of diffusion models, Gao et al. train a diffusion model to correct corruptions in images [17].

Test-time augmentation (TTA). Recently, TTA has been used to improve the robustness and accuracy of a FM without retraining/finetuning it. Not unlike

data augmentation, the choice and weighting of different augmentation functions are non-trivial. Shanmugam et al. learned weighting functions for different augmentations to achieve new high scores on image-classification task [18]. For an extensive literature review on test-time adaptation strategies, we refer the reader to a recent survey [19]. Kimura et al. established a mathematical framework and showed the theoretical optimum of weighting augmentation functions [20]. In 2022, Tomar et al. introduced OptTTA [21], which showed that TTA can be used for domain adaptation. They exploit the fact that batch-normalization layers store statistics on the mean and variance of input data and feature layers. They use these weights to create an unsupervised loss function, optimizing augmentation strategies for the target dataset to achieve state-of-the-art accuracy. Their method is source-free and does not require access to any label for the target dataset. While they do not retrain the model, they assume access to the weights, which is key in the definition of their unsupervised loss function. In a follow-up work, You et al. introduced SaGTTA [22]. Instead of using the batch-norm layers information, they use saliency maps [23] to guide the optimization of the augmentation strategies. By minimizing the similarity between class saliency maps, they encourage confidence in predictions. Saliency maps also aid the interpretability of their results.

Both of these works [21,22] have inspired the proposed method. They assume no access to source data and optimize directly using unlabeled target data. However, they both assume access to the weights of the FM, which our approach, OSATTA, does not.

Annotating a full cohort of images is a very time-consuming task. However, the assumption that a few images (1-5) can be labeled, is a reasonable investment when dealing with a specific cohort. Our work also differs from OptTTA and SaGTTA in that regard; our approach, OSATTA, leverages a few labeled examples of the target domain in order to increase the performance of the FM in that specific cohort.

In summary, we propose OSATTA—for One Shot Automatic Test Time Augmentation, a one-shot approach that can adapt a black-box FM by finding a target-2-source mapping for the whole target dataset.

2 Method

Problem statement. The input to our method is a fundamental model (FM) f performing a segmentation task in the image domain: $f(I) = S$, where, without loss of generality, $I \in \mathbb{R}^{W \times H}$ ($W, H \in \mathbb{N}$) is considered a 2D intensity image. The segmented image $S \in \mathbb{C}^{W \times H}$, has the same size as I with values spanning the set of possible $C \in \mathbb{N}$ class labels $\mathbb{C} = \{1, \ldots, C\}$. In addition to f we are provided with a small set of pairs of images and their labels, that we note $\mathcal{P} = \{(I_i, S_i)\}$, where $|\mathcal{P}| = N \in \mathbb{N}$ is small, typically $1 \leq N \leq 5$. Given a metric measuring the segmentation quality $D(f(I) = S, S^{GT})$, such as Dice score [24], our goal is to find a target-to-source-mapping (t2sm) $m : \mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$, composed of geometric transformation ($g : \mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$) and style transformations (s :

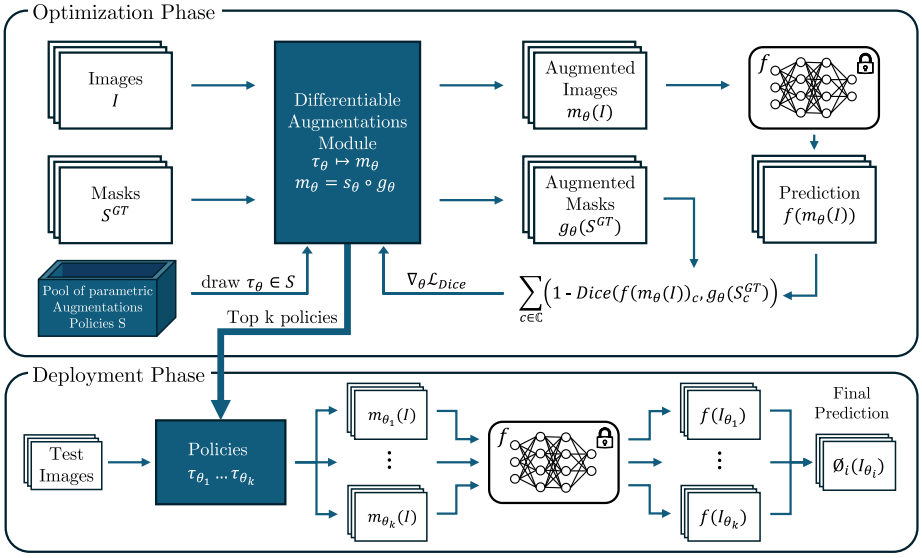

Fig. 1: Overview of our method. All possible policies from the pool get optimized using the available data. The top-k policies get picked to be used during the deployment phase, where we generate the final predictions.

$\mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$) so that $m = g \circ s$ and $D(f(m(I)), g(S^{GT})) \geq D(f(I), S^{GT})$. Where g is an invertible spatial transformation of the image. In practice, g can be effectively written as a product of multiple affine transformations, making it invertible. This invertibility is necessary to obtain the final segmentation mask on the original image frame, i.e., $g^{-1} \circ f \circ m(I) = S^m$. Note that in case g crops the image, then the inverse function has to assume a predetermined value for out-of-bounds values for the segmentation. We discuss this case in Sec. 4. The style function s is defined as a (potentially non-linear) change in the intensity values which does not affect the geometry of the image. Note that, unlike g, which can be applied to an image mask C, it is not possible to apply s to C.

The overview of OSATTA is presented in Fig. 1. In the rest of the section, we describe the set of elementary functions that we consider to build the s, g and thus m functions (Sec. 2.1), we explain how we find/optimize the elementary function combinations (Sec. 2.2), and how these policies are ensembled to obtain our results at test-time (Sec. 2.3).

2.1 T2sm - Augmentation Space

The t2sm could theoretically be any invertible function that projects the image back into the original image space as in $m : \mathbb{R}^{W \times H} \mapsto \mathbb{R}^{W \times H}$. This includes very complex non-linear mappings, such as neural networks, CNNs, diffusion models, etc. In practice, we aim to find the best t2sm from a parametric family

of functions. Importantly, the family of functions needs to be differentiable with regard to the parameters.

In this work, we chose to use the same simple augmentation-based strategy as previous work [21,22], which has two main advantages: comparability to previous work and the interpretability of results (see sec. 3). We build a set S of "augmentation sub-policies" τ, made up of elementary parametric transformations $\mathcal{O}$. This idea was first introduced by Cubuk et al. [1]. Our set of image transformations $\mathcal{O}$ is divided into two sub-sets: style ($\mathcal{O}_s$) and geometric ($\mathcal{O}_g$). The style transformations $\mathcal{O}_s$ are *Identity, Gamma Correction, Gaussian Blur, Contrast modification, Brightness modification* and the geometric transformations $\mathcal{O}_g$ are *Resize Crop, Horizontal Flip, Vertical Flip, and Random Rotation.* An augmentation policy τ_θ is then defined as a selection of N sequentially applied transformations parametrized by θ. For a given number of transformations N, we then generate the potential pool of sub-policies, by building all combinations of transformations, giving us $\binom{9}{N}$ sub-policies to optimize and choose from. Our final policy is then a set of optimized sub-policies. We derive the t2sm $m_\theta = g_\theta \circ s_\theta$ from the sub-policy τ_θ with the parameters θ by sequentially applying the parametric transformations, therefore $\tau_\theta \mapsto m_\theta$

2.2 Supervised Optimization of Augmentations

Differentiable Augmentation Functions. As the image transformations themselves are not directly differentiable, we have to use a "re-parametrization trick": each sub-policy τ is optimized by minimizing the expected loss $\mathcal{L}_\tau$, which is the expectation of the loss over random augmentations of the data. By expressing the augmentation magnitudes in terms of learnable distribution parameters μ^τ and σ^τ, we enable the numeric estimation of gradients for these parameters. For a detailed description, we refer the reader to section 2.2.2 of OptTTA [21].

Training Phase. Our training regime is a modified version of the one used in OptTTA [21]. We next provide an overview, pointing out the key differences. During training, given a number of maximum augmentations, we generate our pool of sub-policies as described in sec. 2.1. For a given number of iterations, we then optimize the parameters θ of each sub-policy in the following way. First, we draw mini-batches of image slices across our entire training set. This is a key difference compared to OptTTA and SaGTTA, which optimize one image volume at a time. By optimizing across multiple images, we aim to generalize better to new unseen images. Then, we augment the mini-batch using the current sub-policy (with optimized parameters θ). Importantly, for all geometric transformations, we use the same seed to augment the matching segmentation masks. As discussed earlier, we do not apply any style transformations to the segmentation masks, as we assume that any intensity change should not affect the segmentation labels. We then generate predictions for our current augmented mini-batch and calculate the loss using the augmented segmentation masks. We use the Dice loss [25]

$$\mathcal{L}(I, S^{GT}; \theta) = \sum_{c \in \mathbb{C}} (1 - Dice(f(m_\theta(I))_c, g_\theta(S_c^{GT}))) \tag{1}$$

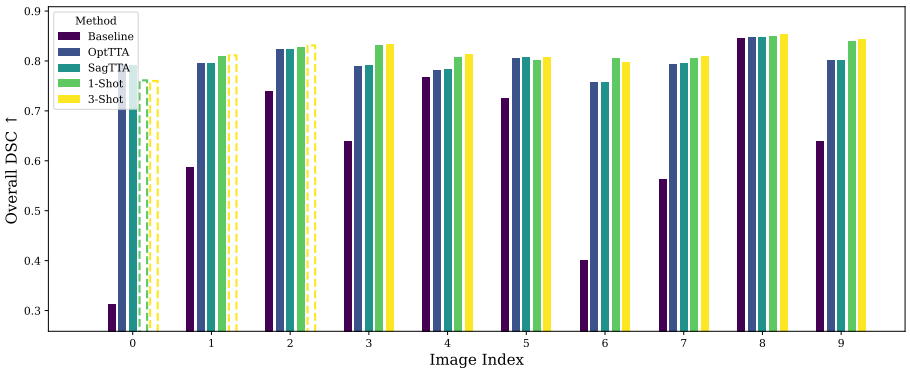

Fig. 2: Individual Dice scores for each image, compared across methods, dotted lines for images used during few-shot training (score provided for completeness)

to optimize the parameters θ of the current sub-policy via gradient descent. We train our optimization policies using AdamW optimizer for 3k iterations at a learning rate of 10^{-3}. We implemented all experiments using PyTorch and ran them on an NVIDIA RTX 2080 Ti.

2.3 Deployment Phase

After optimizing the entire pool of possible augmentation sub-policies, we select the top k sub-policies to use during deployment. The selection criterion here is Dice loss of the training image. To reduce overfitting, we considered using an additional labeled image volume to compute a validation loss as a selection criterion. However, in practice, including this image in the training set and using the training loss as a criterion proved to generalize better to the test set. In contrast to previous (unsupervised) works, we do not fine-tune the top policies on every test image during deployment. In our evaluation, we discuss the performance and accuracy implications of this (see sec. 3). Besides that, the ensembling of the final prediction is done in the same fashion as SaGTTA [22]. We draw M random samples of each sub-policy, average the M results, and obtain an estimate of the expected value for the given sub-policy. We then average over all k sub-policies, obtaining our final prediction for the total policy made up of our top-k optimized sub-policies.

3 Experiments and Results

3.1 Baselines and Dataset

As described in Sec. 1, the two existing works that provide a solution to the considered problem are OptTTA [21] and SagTTA [22]. Although their solutions are fully unsupervised, they provide a baseline to the performance of OSATTA.

Method	FM no adapt.	OptTTA	SaGTTA	SaGTTA no exploit	1-Shot (ours)	3-Shot (ours)
DSC Mean	.5709	.7943	.8016	.7806	<u>.8197</u>	**.8229**
DSC Std	.0200	.0416	.0275	.0601	<u>.0249</u>	**.0209**

Table 1: Comparison of DSC score for different methods on target test data[1]. FM trained on site 1, and t2sm learned with target site 3. Bold indicates the best value, and underline the second best.

To compare to them we follow their experimental protocol. For the dataset, we use the public Spinal Cord Gray Matter Segmentation challenge Dataset [26], consisting of data collected at four different sites, each with different vendors and protocols. The training data has segmentation annotations for gray and white matter, and all images were re-sampled to a common 1mm isotropic resolution using bi-cubic interpolation. For the experimental setting, we focus on adapting the data from site #3 to a model with fixed weights trained on data from site #1. As described in SaGTTA [22], these two sites happen to show the most significant domain gap of the dataset. The learned model, considered as the FM, is a 2D U-net trained with weighted cross entropy loss with the RMSprop optimizer (learning rate 10^{-5}; 250K iterations). The segmentation performance on left-out data is measured using the Dice score (DSC) [24].

3.2 Evaluation

In Table 1, we report the computed mean DSC values between the GT segmentation masks and the predicted ones. Our approach obtains the highest values, with a significant improvement with respect to the FM (no adapt.) in the one-shot scenario. Unlike the baselines, our approach leverages the availability of more labeled training target data, as visible in the 3-shot approach, slightly increasing the performance over the 1-shot. In Fig. 2, we present the performance on the individual target test data, where one can observe how our method bridges the domain gap for the other target images that have not been seen. Interestingly, we outperform SagTTA, even though it was further fine-tuned individually on each image. It is worth noting that SagTTA [22] obtains, via a thorough grid search, a theoretical upper bound of .8036 for the DSC performance. Both our approaches outperform this upper bound with few examples (.8229).

One interesting feature of our approach, inherited from the selection of the set of policies [21], is that their effects can be visually observed. In the top of Fig. 2,

[1] Values for the baselines (Baseline, OptTTA, SaGTTA) were obtained from [22], as these results were averaged over 20 runs. Due to time constraints, we were unable to replicate their extensive number of runs. Nonetheless, our reproduced values are consistent with the literature, falling within a similar range.

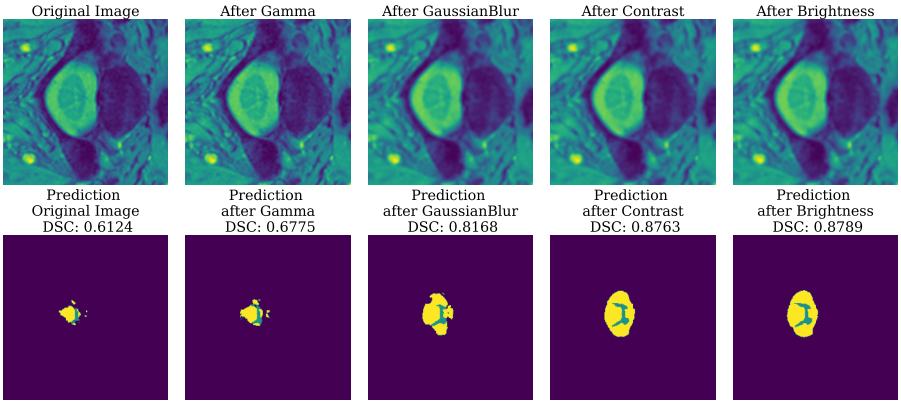

Fig. 3: Visualization of the learned top policy: Top row: Image after augmentation, Bottom row: Prediction of the augmented image above

we present, from left to right, the original image and the different transformations applied until the final image. On the bottom, we observe the applied FM to the corresponding image. It is interesting to note how each transformation further improves the prediction of the FM.

Speed. We further analyze the speed of the application of the approaches. During training, we obtain equal iteration speed to OptTTA [21] with ~ 4 it/s. SaGTTA [22] is one order of magnitude slower with 0.5 it/s. However, as we do not change the ensembling process of the trained augmentation policies for the final prediction, our method is orders of magnitude faster at test time. While the accuracy of previous methods relies on fine-tuning policies for each image, we can skip this step and apply the learned augmentation directly.

4 Conclusion

In this work, we propose OSATTA: One Shot Automatic Test Time Adaptation. OSATTA learns a target-2-souce-mapping (t2sm) to improve the segmentation accuracy of a Fundamental Model (FM) on a new target domain, whithout requiring access to the weights of the FM. By leveraging one labeled sample of the target domain, OSATTA improves the performance of the FM model by over .2 points DICE and also outperforms existing unsupervised state-of-the-art approaches. OSATTA learns a t2sm from a set of simple parametrized augmentation functions. The definition of this set allows to obtain an interpretability of the learned function, and thus visualize the domain gap between the source and target domains.

While OSATTA is presented with a segmentation tasks, it is interesting to note that OSATTA can be easily extended to other tasks, such as registration or classification. To do so, one only needs to replace the output quality mea-

sure function, so that it is aligned with the considered task. We leave these experimentations for future work.

Another lead is to extend the family of elementary functions to optimize for. Learning complex non-linear intensity changes, adding 2D/3D warpings, and parametrizing the current non-learnable geometric transformations are just some of the possibilities that could be explored. Moreover, one could try to learn a full style transfer model, such as a U-Net. Although this seems a difficult task in the low data regime (few-shot), one could potentially go beyond pre-defined functions and learn other patterns between the source and target domain. OSATTA opens a new perspective on how to improve the fundamental models performance in unseen target domains.

Acknowledgements Felix Küper and Sergi Pujades' work was funded by the ANR PRC INORA project.

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

References

1. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: AutoAugment: Learning Augmentation Strategies From Data. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 113–123. IEEE, Long Beach, CA, USA (Jun 2019). https://doi.org/10.1109/CVPR.2019.00020
2. Li, Y., Hu, G., Wang, Y., Hospedales, T., Robertson, N.M., Yang, Y.: Differentiable Automatic Data Augmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 580–595. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58542-635
3. He, W., Liu, M., Tang, Y., Liu, Q., Wang, Y.: Differentiable Automatic Data Augmentation by Proximal Update for Medical Image Segmentation. IEEE/CAA Journal of Automatica Sinica **9**(7), 1315–1318 (Jul 2022). https://doi.org/10.1109/JAS.2022.105701
4. Liu, A., Huang, Z., Huang, Z., Wang, N.: Direct Differentiable Augmentation Search. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12199–12208. IEEE, Montreal, QC, Canada (Oct 2021). https://doi.org/10.1109/ICCV48922.2021.01200
5. Xu, J., Li, M., Zhu, Z.: Automatic Data Augmentation for 3D Medical Image Segmentation (Dec 2020). https://doi.org/10.48550/arXiv.2010.11695, http://arxiv.org/abs/2010.11695, arXiv:2010.11695 [cs, eess]
6. Liu, Z., Lv, Q., Li, Y., Yang, Z., Shen, L.: MedAugment: Universal Automatic Data Augmentation Plug-in for Medical Image Analysis (Nov 2023). https://doi.org/10.48550/arXiv.2306.17466, http://arxiv.org/abs/2306.17466
7. Luo, Y., Wang, Y., Zhang, Z., Liu, M., Tang, Y.: IOADA: An Optimal Automated Augmentation Algorithm for Medical Image Segmentation. In: 2023 China Automation Congress (CAC). pp. 3900–3905 (Nov 2023). https://doi.org/10.1109/CAC59555.2023.10450689, iSSN: 2688-0938
8. Guan, H., Liu, M.: Domain Adaptation for Medical Image Analysis: A Survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (Mar 2022). https://doi.org/10.1109/TBME.2021.3117407, https://ieeexplore.ieee.org/document/9557808/
9. Yu, Z., Li, J., Du, Z., Zhu, L., Shen, H.T.: A Comprehensive Survey on Source-free Domain Adaptation (Feb 2023). https://doi.org/10.48550/arXiv.2302.11803, http://arxiv.org/abs/2302.11803, arXiv:2302.11803 [cs]
10. Fang, Y., Yap, P.T., Lin, W., Zhu, H., Liu, M.: Source-Free Unsupervised Domain Adaptation: A Survey (Jan 2023). https://doi.org/10.48550/arXiv.2301.00265, http://arxiv.org/abs/2301.00265, arXiv:2301.00265 [cs]
11. Wang, L., Zhang, X., Su, H., Zhu, J.: A Comprehensive Survey of Continual Learning: Theory, Method and Application (Feb 2024). https://doi.org/10.48550/arXiv.2302.00487, http://arxiv.org/abs/2302.00487, arXiv:2302.00487 [cs]
12. Gaillochet, M., Desrosiers, C., Lombaert, H.: TAAL: Test-Time Augmentation for Active Learning in Medical Image Segmentation. In: Nguyen, H.V., Huang, S.X., Xue, Y. (eds.) Data Augmentation, Labelling, and Imperfections. pp. 43–53. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-17027-0_5
13. Xu, H., Ebner, S., Yarmohammadi, M., White, A.S., Van Durme, B., Murray, K.: Gradual Fine-Tuning for Low-Resource Domain Adaptation (Sep 2021), http://arxiv.org/abs/2103.02205, arXiv:2103.02205 [cs]

14. Li, X., Gu, Y., Dvornek, N., Staib, L., Ventola, P., Duncan, J.S.: Multi-site fMRI Analysis Using Privacy-preserving Federated Learning and Domain Adaptation: ABIDE Results (Dec 2020), http://arxiv.org/abs/2001.05647
15. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y.: A survey on federated learning. Knowledge-Based Systems **216**, 106775 (Mar 2021). https://doi.org/10.1016/j.knosys.2021.106775, https://www.sciencedirect.com/science/article/pii/S0950705121000381
16. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (Aug 2020), http://arxiv.org/abs/1703.10593, arXiv:1703.10593 [cs]
17. Gao, J., Zhang, J., Liu, X., Darrell, T., Shelhamer, E., Wang, D.: Back to the Source: Diffusion-Driven Adaptation to Test-Time Corruption. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11786–11796. IEEE, Vancouver, BC, Canada (Jun 2023). https://doi.org/10.1109/CVPR52729.2023.01134
18. Shanmugam, D., Blalock, D., Balakrishnan, G., Guttag, J.: Better Aggregation in Test-Time Augmentation (Oct 2021), http://arxiv.org/abs/2011.11156, arXiv:2011.11156 [cs]
19. Liang, J., He, R., Tan, T.: A Comprehensive Survey on Test-Time Adaptation under Distribution Shifts (Mar 2023). https://doi.org/10.48550/arXiv.2303.15361, http://arxiv.org/abs/2303.15361, arXiv:2303.15361 [cs]
20. Kimura, M.: Understanding Test-Time Augmentation (Feb 2024). https://doi.org/10.48550/arXiv.2402.06892, http://arxiv.org/abs/2402.06892
21. Tomar, D., Vray, G., Thiran, J.P., Bozorgtabar, B.: OptTTA: Learnable Test-Time Augmentation for Source-Free Medical Image Segmentation Under Domain Shift. Proceedings of Machine Learning Research, Volume 172: International Conference on Medical Imaging with Deep Learning, 6-8 July 2022, Zurich, Switzerland (2022)
22. You, S., Tomar, D., Bozorgtabar, B., Reyes, M.: SaGTTA: Saliency Guided Test Time Augmentation for Medical Image Segmentation Across Vendor Domain Shift. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–4 (Apr 2023). https://doi.org/10.1109/ISBI53787.2023.10230764, iSSN: 1945-8452
23. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic Attribution for Deep Networks (Jun 2017). https://doi.org/10.48550/arXiv.1703.01365, http://arxiv.org/abs/1703.01365, arXiv:1703.01365 [cs]
24. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology **26**(3), 297–302 (1945). https://doi.org/10.2307/1932409, https://onlinelibrary.wiley.com/doi/abs/10.2307/1932409
25. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. vol. 10553, pp. 240–248 (2017). https://doi.org/10.1007/978-3-319-67558-9_28, http://arxiv.org/abs/1707.03237, arXiv:1707.03237 [cs]
26. Prados, F., Ashburner, J., Blaiotta, C., Brosch, T., Carballido-Gamio, J., Cardoso, M.J., Conrad, B.N., Datta, E., Dávid, G., Leener, B.D., Dupont, S.M., Freund, P., Wheeler-Kingshott, C.A.M.G., Grussu, F., Henry, R., Landman, B.A., Ljungberg, E., Lyttle, B., Ourselin, S., Papinutto, N., Saporito, S., Schlaeger, R., Smith, S.A., Summers, P., Tam, R., Yiannakas, M.C., Zhu, A., Cohen-Adad, J.: Spinal cord grey matter segmentation challenge. NeuroImage **152**, 312–329 (May 2017). https://doi.org/10.1016/j.neuroimage.2017.03.010, https://www.sciencedirect.com/science/article/pii/S1053811917302185