Transferability-Guided Multi-source Model Adaptation for Medical Image Segmentation

Chen Yang[1], Yifan Liu[2], and Yixuan Yuan[2(✉)]

[1] Department of Electrical Engineering, City University of Hong Kong, Hong Kong, SAR, China
[2] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China
yxyuan@ee.cuhk.edu.hk

Abstract. Unsupervised domain adaptation has drawn sustained attentions in medical image segmentation by transferring knowledge from labeled source data to unlabeled target domain. However, most existing approaches assume the source data are collected from a single client, which cannot be successfully applied to explore complementary transferable knowledge from multiple source domains with large distribution discrepancy. Moreover, they require access to source data during training, which is inefficient and unpractical due to privacy preservation and memory storage. To address these challenges, we study a novel and practical problem, named multi-source model adaptation (MSMA), which aims to transfer multiple source models to the unlabeled target domain without any source data. Since no target label and source data is provided to evaluate the transferability of each source model or domain gap between the source and the target domain, we may encounter negative transfer by those less related source domains, thus hurting target performance. To solve this problem, we propose a transferability-guided model adaptation (TGMA) framework to eliminate negative transfer. Specifically, 1) A label-free transferability metric (LFTM) is designed to evaluate transferability of source models without target annotations for the first time. 2) Based on the designed metric, we compute instance-level transferability matrix (ITM) for target pseudo label correction and domain-level transferability matrix (DTM) to achieve model selection for better target model initialization. Extensive experiments on multi-site prostate segmentation dataset demonstrate the superiority of our framework.

Keywords: Source-free Domain Adaptation · Multi-source · Label-free transferability metric

1 Introduction

Deep neural networks have greatly advanced medical image analysis in recent years [12]. However, a large amount of annotated data is required for training,

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14221, pp. 703–712, 2023.
https://doi.org/10.1007/978-3-031-43895-0_66

which is time-consuming and error-prone, especially in medical image segmentation task that needs pixel-wise annotations. Moreover, a segmentation model trained on one clinical centre (source domain) often fails to generalize well when deployed in a new centre (target domain) due to the discrepancy in the data distribution [2,9,16]. Unsupervised domain adaptation (UDA) [5,14,17] seeks to tackle this dilemma by transferring the knowledge from label-rich source domain to label-rare target domain. However, the source data may become inaccessible due to storage and privacy concerns in medical settings, which hinders the wide applications of domain adaptation. Towards this obstacle, great interests have been invoked to explore source-free domain adaptation (SFDA) [2,6,8,9,16], where a model pre-trained on the labeled source data are adapted to the unlabeled target domain without accessing source data. Though great successes, how to achieve adaptation to the unlabeled target domain with the knowledge from multiple source domains under privacy protection is still an open question to be solved.

To this end, we study a practical and challenging domain adaptation problem which explores transferable knowledge from multiple source domains to target domain with only pre-trained source models rather than the source data, namely *multi-source model adaptation* (MSMA). Although MSMA methods [1,3,7] have made great progress for natural object recognition, there is still a blank in the multi-source-free domain adaptive medical image segmentation. Directly applying existing MSMA methods on medical image segmentation by optimizing all source segmentation models are time-consuming and inefficient due to larger model capacity of segmentation model than classification model. Another trivial solutions to tackle MSMA via SFDA methods [2,6,16,18] are to adapt each source model individually and simply take an average prediction of adapted models. However, this strategy does not take into account the varying contributions of different source models to the target domain, which can result in negative transfer from less related source domains. To rank pre-trained models, transferability metrics [10,13,20] have been widely applied to measure the domain relevance or task relevance for transfer learning, but all of them need target annotations, which is not accessible for multi-source model adaptation. Automatically select an optimal subset of the source models without requiring source data and target annotations in an unsupervised fashion is of far-reaching significance for MSMA.

To address this problem, we develop a novel Transferability-Guided Model Adaptation (TGMA) model, which represents the first attempt to solve MSMA in medical image segmentation. Specifically, a label-free transferability metric (LFTM) is designed to evaluate the relevance between source and target domain without access to the source data. Based on the designed LFTM, we can compute instance-level transferability matrix (ITM) to achieve pseudo-label correction for precise supervision, and domain-level transferability matrix (DTM) to accomplish model selection for better target initialization. To this end, we can achieve adaptation to unlabeled target domain with clean pseudo label and proper model initialization. The main contributions are summarized as:

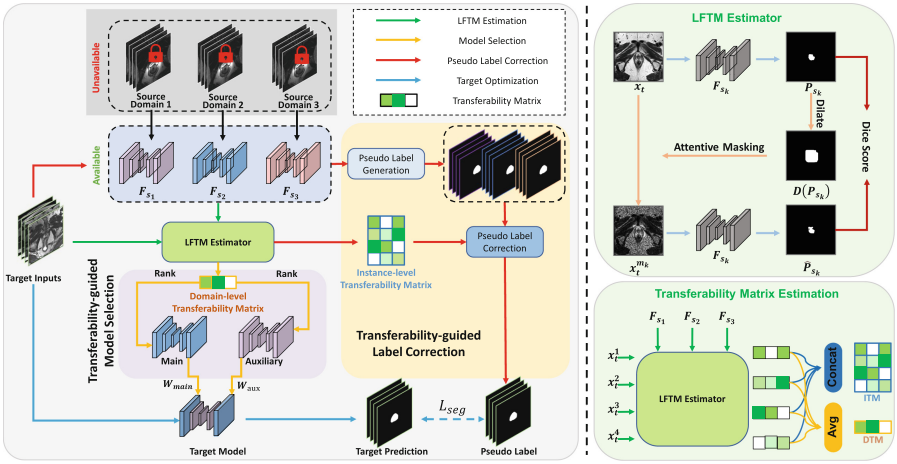

Fig. 1. Illustration of Transferability-Guided multi-source Model Adaptation (TGMA) framework, including (a) label-free transferability metric (LFTM) estimator, (b) transferability-guided model selection and (c) transferability-guided label correction.

- We present the first work that studies the practical domain adaptation problem of transferring knowledge from multiple source segmentation models rather than the source data to unlabeled target domain.
- We design a novel label-free transferability metric (LFTM) based on attentive masking consistency to evaluate the domain relevance for the first time.
- Based on the LFTM, we propose a transferability-guided model adaptation (TGMA) framework including pseudo-label correction by instance-level transferability matrix (ITM) and model selection by domain-level transferability matrix (DTM).
- Extensive experiments on the multi-site prostate segmentation dataset demonstrate the superiority of our TGMA compared with state-of-the-art domain adaptation methods.

2 Method

In MSMA scenario, we address the problem of jointly adapting multiple segmentation models, trained on a variety domains, to a new unlabeled target domain. Formally, let us consider we have a set of source models $\{F_{s_j}\}_{j=1}^{M}$, where the j^{th} model $\{F_{s_j}\}$ is a segmentation model learned using the source dataset $\mathcal{D}_s^j = \{x_{s_j}^i, y_{s_j}^i\}_{i=1}^{N_j}$, with N_j data points, where $x_{s_j}^i$ and $y_{s_j}^i$ denote the i-th source image and the corresponding segmentation label respectively. Now, given a target unlabeled dataset $\mathcal{D}_t = \{x_{t^i}\}_{i=1}^{N_t}$, the problem is to learn a segmentation model F_t, using only the learned source models, without any access to the source dataset. Figure 1 gives an overview of our proposed TGMA framework.

To eliminate negative transfer by domain-dissimilar source models, we design a label-free transferability metric to evaluate the transferability of source models in an unsupervised manner for the first time. Before target training, an instance-level transferability matrix (ITM) is computed to rectify target pseudo labels, and a domain-level transferability matrix (DTM) is calculated to achieve model selection for better model initialization. Based on the rectified pseudo labels and selected models, target segmentation model is trained with dice loss to achieve model adaptation.

2.1 Label-Free Transferability Metric

Most of multi-source model adaptation approaches [1,7] treat all source models equally, leading to negative transfer from irrelevant source domains. To avoid this type of negative transfer, it is important to critically evaluate the relevance of prior knowledge from each source domain to the target domain, and to focus on the most relevant source domains for learning in the target domain. However, it's challenging to evaluate the domain relevance in the absence of source data and target ground truths. To identify the transferability of source models, we develop a label-free transferability metric (LFTM) on the basis of attentive masking consistency to prevent negative transfer for the first time. Our metric is designed based on two assumptions: 1) *Sample relevance*: similar samples should hold identical predictions; 2) *Model stability*: if a source model makes accurate decision on this sample, little permutation on irrelevant regions will not influence the prediction. We follow these two assumptions to construct augmented sample by attentive masking, and compute the consistency as the transferability.

Given unlabeled target data $\mathcal{D}_t = \{x_t^i\}_{i=1}^N$ and a pre-trained source segmentation model F_{s_k}, we import them to the LFTM estimator and compute the transferability metric $LFTM(x_t, F_{s_k})$ with only twice forwards as shown in Fig. 1. In the first forward process, the original target sample x_t is passed into the source model F_{s_k} to generate segmentation map $P_{s_k} = F_{s_k}(x_t)$. Based on the assumption that masking the normal regions from the diseased image will not affect the lesion regions, we preserve the segmentation region of the original image and randomly mask the other regions to generate masked image $x_t^{m_k}$. Since the segmentation results may be affected by receptive field, we enlarge the segmentation map P_{s_k} to $D(P_{s_k})$ by dilation. Then masked image $x_t^{m_t}$ is generated by combination of enlarged lesion regions and masked normal regions:

$$x_t^{m_k} = M(x_t) * (1 - D(P_{s_k})) + x_t * D(P_{s_k}), \tag{1}$$

where $M(x_t)$ is the masking operation to randomly remove pixels. In the second forward process, the masked target sample $x_t^{m_k}$ is passed into the source model F_{s_k} to generate segmentation map $\hat{P}_{s_k} = F_{s_k}(x_t^{m_k})$. Then we calculate the dice score between these two predictions as transferability metric:

$$LFTM(x_t, F_{s_k}) = 2 * \frac{P_{s_k} \cap \hat{P}_{s_k}}{P_{s_k} + \hat{P}_{s_k}}. \tag{2}$$

The larger the LFTM is, the more stable the source model is on the target sample. With M source models and N_t target samples, we can compute the instance-level transferability matrix (ITM) $T_{instance} \in \mathbb{R}^{M \times N_t}$, which can be utilized to correct target pseudo labels. Averaging $T_{instance}$ on the domain-space can generate domain-level transferability matrix (DTM) $T_{domain} \in \mathbb{R}^{M \times 1}$, which represents the contribution of each source model to the target domain. The detailed process is illustrated in Transferability Matrix Estimation of Fig. 1.

2.2 Transferability-Guided Model Adaptation

The basic pipeline for target training needs accurate pseudo labels and suitable model initialization. While there are multiple pseudo labels and source models, simply averaging them as target supervision and model initialization is trivial solution, which ignores the contribution differences of these source domains. To tackle this problem, we propose a transferability-guided model adaptation (TGMA) framework on the basis of LFTM, which consists of two modules: Label Correction and Model Selection. Based on the instance-level transferability matrix $T_{instance}$, we re-weight the pseudo labels generated by multiple source models to achieve pseudo label correction. With the domain-level transferability matrix T_{domain}, we select the most portable source model as the main model initialization and make full use of other source models by weighted optimization strategy.

Transferability-Guided Label Correction. In MSMA, we generate pseudo labels as supervision because no target ground truth is available. However, with multiple pseudo labels predicted by source models for a target sample, prior works [1,7] typically average these labels equally to obtain the final pseudo label. However, negative source models that are poorly suited to the target domain may generate inaccurate pseudo labels, resulting in noisy or unreliable training data. To eliminate negative transfer and improve pseudo-label correction, we can re-weight model predictions from all source models using the calculated instance-level transferability matrix $T_{instance}$.

Taking a target sample x_t for example, we pass this sample to source models $\{F_{s_1}, F_{s_2}, ..., F_{s_M}\}$ to obtain corresponding predictions $\{P_{s_1}, P_{s_2}, ..., P_{s_M}\}$. We take argmax operation on these predictions to generate one-hot pseudo labels $\{y_{s_1}, y_{s_2}, ..., y_{s_M}\}$, where $y = argmax(P)$. The instance-level transferability matrix $T_{instance}$ is applied on these pseudo labels to achieve noise correction by contribution re-weighting:

$$y_t = argmax(\sum_{i=1}^{M} LFTM(x_t, F_{s_i}) * y_{s_i}), \tag{3}$$

where each pseudo label is weighted by the corresponding LFTM score for better combination. This strategy largely prevents the negative transfer problem caused by noisy labels of those domain-irrelevant source models.

Transferability-Guided Model Selection. Previous MSMA methods [1,7] usually treat all models equally and optimize all source models parameters to achieve adaptation to the target domain. On the one hand, they ignore the negative transfer problem led by some less related domains. On the other hand, optimizing all source parameters is time-consuming and inefficient. To better make full use of the source models, we utilize the calculated domain-level transferability matrix T_{domain} to rank all source models.

With T_{domain} representing the transferability of source models, we choose the best source model as main network F_{main} and the second best model as auxiliary network F_{aux}. Only initialing the target model from F_{main} may ignore complementary knowledge of other source models, while optimizing all source models are inefficient. To obtain a compromise solution, we take the second model as auxiliary parameter knowledge. Then a weighted optimization strategy is utilized on the best model and the auxiliary model with weight W_{main} and W_{aul} respectively:

$$F_t = \min_{F \cap W} \mathcal{L}_{dice}(y_t, W_{main} * F_{main}(x_t) + W_{aux} * F_{aux}(x_t)), \tag{4}$$

where $\mathcal{L}_{dice}$ is calculated on the combined target prediction and corresponding pseudo label. This loss optimizes model parameter $W_{main} * F_{main} + W_{aux} * F_{aux}$. The model selection strategy choose optimal source model while makes full use of those sub-optimal source models for better model initialization, thus avoiding the negative transfer by those domain-irrelevant domains.

3 Experiment

3.1 Dataset

Extensive experiments are conducted to verify the effectiveness of our proposed framework on Prostate MR (**PMR**) dataset which is collected and labeled from six different public data sources for prostate segmentation [15]. All of the MRI images have been re-sampled to the same spacing and center-cropped with the size of 384×384. We divide them into six sites, each of which contains $\{261, 384, 158, 468, 421, 175\}$ slices. We denote these six sites as $\{A, B, C, D, E, F\}$ for convenience. At each adaptation process, five sites are selected as source domains and the rest one is set as the target domain. We conduct leave-one-domain-out experiments by selecting one domain to hold out as the target. For example, $\rightarrow A$ denotes adapting source models from $\{B, C, D, E, F\}$ to unlabeled images of A.

3.2 Implementation Details and Evaluation Metrics

The framework is implemented with Pytorch 1.7.0 using an NVIDIA RTX 2080Ti GPU. Following [15], we adopt UNet as our segmentation backbone. We train the target model for 200 epochs with the batch size of 6. Adam optimizer is adopted with the momentum of 0.9 and 0.999, and the learning rate is set to 0.001. We adopt the well-known metrics Dice score for segmentation evaluation.

Table 1. Comparison with state-of-the-art domain adaptation approaches on PMR dataset, measured by dice score.

Type	Method	Source Data	→ A	→ B	→ C	→ D	→ E	→ F	Average
SFDA	SHOT (20') [6]	✗	28.76	34.32	50.57	31.55	33.60	28.70	34.58
	NRC (21') [18]	✗	32.44	38.05	59.39	28.03	40.47	27.35	37.62
	FSM (22') [16]	✗	33.57	38.56	70.72	29.40	36.76	32.52	40.25
MSDA	KD3A (21') [4]	✓	38.05	55.78	64.16	31.77	37.69	49.36	46.13
	CWAN (21') [19]	✓	51.18	61.96	71.62	75.45	55.88	60.11	62.69
	PTMDA (22') [11]	✓	65.16	68.37	79.05	77.23	61.58	69.42	70.13
MSMA	Source only	✗	31.26	39.80	66.95	9.86	14.93	32.77	32.59
	DECISION (22') [1]	✗	48.03	60.72	69.85	71.34	52.94	63.16	61.01
	DINE (22') [7]	✗	54.20	62.82	74.11	72.59	53.46	64.78	63.66
	TGMA (Ours)	✗	62.76	65.73	76.14	75.10	58.59	65.63	67.32
	Ours w/o ITM	✗	56.41	60.65	72.83	71.29	54.42	62.04	62.94
	Ours w/o DTM	✗	59.48	61.56	73.16	73.97	56.85	63.24	64.71

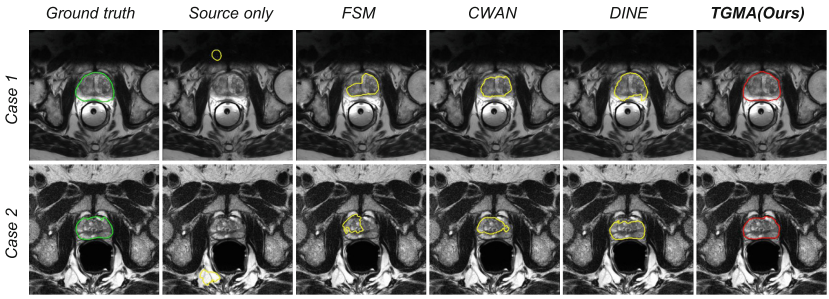

Fig. 2. Qualitative comparison on the PMR dataset of different DA methods.

3.3 Comparison with State-of-the-Arts

We compare our methods to several domain adaptation frameworks, including the single source-free domain adaptation (SFDA) [6,16,18], multi-source domain adaptation (MSDA) [4,11,21] and multi-source model adaptation (MSMA) [1,7] methods. For implementation, as most of these methods are originally designed for the image classification task, we try out best to keep their design principle and adapt them to our image segmentation task. Specifically, SFDA methods are performed on each source model and averaging the adapted model predictions as the final results. The results on prostate segmentation is listed in Table 1. As observed, MSDA methods shows superior performance than MSMA approaches due to access to the source data. Notably, compared with SFDA and MSMA approaches, our TGMA achieves higher performance on nearly all metrics with 67.32% on Average Dice. These clear improvements benefit from our LFTM metric which considers the different contributions of each source model, and largely eliminate negative transfer from the perspective of pseudo label generation and model initialization. Without the rectification by instance-level transferability

matrix (Ours w/o ITM), the pseudo labels are simply generated by average combination of predictions from source models. The significant decrease in performance by 4.38% on Average Dice highlights the criticality of weighting pseudo labels with scores that reflect the relevance of the source domains. Without the model selection by domain-level transferability matrix (Ours w/o DTM), the target models are initialized from each source pre-trained network and trained separately, leading to 2.61% performance drop on Average Dice. It demonstrates that model initialization is also essential to the transfer learning. Moreover, Fig. 2 shows the segmentation results of different methods on two typical cases. We observe that our model with transferability guidance can well eliminate the negative transfer interference by some domain-irrelevant domains.

Table 2. Comparison with different unsupervised metrics on PMR dataset.

Method	Source Data	→ A	→ B	→ C	→ D	→ E	→ F	Average
Entropy	✗	57.84	59.03	74.29	72.56	52.37	65.20	63.54
Rotation	✗	60.46	61.35	73.74	71.82	53.92	62.13	63.90
Cropping	✗	61.14	62.37	74.59	73.83	55.44	63.05	65.07
LFTM	✗	62.76	65.73	76.14	75.10	58.59	65.63	67.32
LFTM w/o Dilation	✗	61.69	63.08	73.95	74.17	56.22	64.54	65.61

3.4 Ablation Analysis

The performance improvement mainly comes from our designed LFTM to detect negative transfer. There are some other unsupervised metrics that can evaluate model stability, such as entropy, rotation-consistency and crop-consistency. To better evaluate the effectiveness of LFTM, we apply these unsupervised metrics to estimate ITM and DTM for label correction and model selection. The comparison results are shown in Table 2. It's obvious that our proposed LFTM outperforms other unsupervised metrics with a large margin. Entropy may make overconfident decisions on model predictions, thus leading to high transferability on those domain-irrelevant source models. Rotation and Cropping are simple data augmentation methods, which can only evaluate the model stability. Our proposed LFTM makes full use of the segmentation mask to construct feature-nearest sample, thus applying sample relevance to evaluate model transferability. Removing dilation operation leads to 1.71% performance degradation on Average Dice, revealing the effect of receptive field.

4 Conclusion

In this paper, we study a practical domain adaptation problem, named multi-source model adaptation where only multiple pre-trained source segmentation

models rather than the source data are provided for adaptation to unlabeled target domain. To eliminate the negative transfer by domain-dissimilar source models, we design a label-free transferability metric based on the attentive masking consistency to evaluate the transferability of each source segmentation model with only target images. Using this metric, we calculate two types of transferability matrices: an instance-level matrix to adjust the target pseudo label, and a domain-level matrix to choose an optimal subset for improved model initialization.

Acknowledgements. This work was supported by National Natural Science Foundation of China 62001410, Hong Kong Research Grants Council (RGC) Early Career Scheme grant 21207420, General Research Fund 11211221.

References

1. Ahmed, S.M., Raychaudhuri, D.S., Paul, S., Oymak, S., Roy-Chowdhury, A.K.: Unsupervised multi-source domain adaptation without access to source data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10103–10112 (2021)
2. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Source-free domain adaptation for image segmentation. Med. Image Anal. **82**, 102617 (2022)
3. Dong, J., Fang, Z., Liu, A., Sun, G., Liu, T.: Confident anchor-induced multi-source free domain adaptation. Adv. Neural. Inf. Process. Syst. **34**, 2848–2860 (2021)
4. Feng, H., et al.: KD3A: Unsupervised multi-source decentralized domain adaptation via knowledge distillation. In: ICML, pp. 3274–3283 (2021)
5. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**(1), 2030–2096 (2016)
6. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning, pp. 6028–6039. PMLR (2020)
7. Liang, J., Hu, D., Feng, J., He, R.: Dine: Domain adaptation from single and multiple black-box predictors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8003–8013 (2022)
8. Liu, X., Yuan, Y.: A source-free domain adaptive polyp detection framework with style diversification flow. IEEE Trans. Med. Imaging **41**(7), 1897–1908 (2022)
9. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1215–1224 (2021)
10. Nguyen, C., Hassner, T., Seeger, M., Archambeau, C.: Leep: A new measure to evaluate transferability of learned representations. In: International Conference on Machine Learning, pp. 7294–7305. PMLR (2020)
11. Ren, C.X., Liu, Y.H., Zhang, X.W., Huang, K.K.: Multi-source unsupervised domain adaptation via pseudo target domain. IEEE Trans. Image Process. **31**, 2122–2135 (2022)
12. 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

13. Tran, A.T., Nguyen, C.V., Hassner, T.: Transferability and hardness of supervised classification tasks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1395–1405 (2019)
14. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR, pp. 7167–7176 (2017)
15. Wang, J., Jin, Y., Wang, L.: Personalizing federated medical image segmentation via local calibration. In: Computer Vision-ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI. pp. 456–472. Springer (2022). https://doi.org/10.1007/978-3-031-19803-8_27
16. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with Fourier style mining. Med. Image Anal. **79**, 102457 (2022)
17. Yang, C., Guo, X., Zhu, M., Ibragimov, B., Yuan, Y.: Mutual-prototype adaptation for cross-domain polyp segmentation. IEEE J. Biomed. Health Inform. **25**(10), 3886–3897 (2021). https://doi.org/10.1109/JBHI.2021.3077271
18. Yang, S., van de Weijer, J., Herranz, L., Jui, S., et al.: Exploiting the intrinsic neighborhood structure for source-free domain adaptation. Adv. Neural. Inf. Process. Syst. **34**, 29393–29405 (2021)
19. Yao, Y., Li, X., Zhang, Y., Ye, Y.: Multisource heterogeneous domain adaptation with conditional weighting adversarial network. IEEE Trans. Neural Netw. Learn. Syst. (2021)
20. You, K., Liu, Y., Wang, J., Long, M.: Logme: practical assessment of pre-trained models for transfer learning. In: International Conference on Machine Learning, pp. 12133–12143. PMLR (2021)
21. Zhao, S., et al.: Multi-source distilling domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12975–12983 (2020)