UNSUPERVISED DOMAIN ADAPTATION FOR BRAIN VESSEL SEGMENTATION THROUGH TRANSWARP CONTRASTIVE LEARNING

*Fengming Lin[1] *, Yan Xia[1] *, Michael MacRaild[1], Yash Deo[1], Haoran Dou[1], Qiongyao Liu[1], Kun Wu[1], Nishant Ravikumar[1] †, Alejandro F. Frangi[1 2] †*

[1] University of Leeds, [2] University of Manchester https://github.com/fmlinks/domain

ABSTRACT

Unsupervised domain adaptation (UDA) aims to align the labelled source distribution with the unlabelled target distribution to obtain domain-invariant predictive models. Since cross-modality medical data exhibit significant intra and inter-domain shifts and most are unlabelled, UDA is more important while challenging in medical image analysis. This paper proposes a simple yet potent contrastive learning framework for UDA to narrow the inter-domain gap between labelled source and unlabelled target distribution. Our method is validated on cerebral vessel datasets. Experimental results show that our approach can learn latent features from labelled 3DRA modality data and improve vessel segmentation performance in unlabelled MRA modality data.

***Index Terms*—** Unsupervised Domain Adaptation, Contrastive Learning, Cerebral Vessel Segmentation

1. INTRODUCTION

Cerebral vessel segmentation offers invaluable assistance in precise diagnosis, surgical planning, monitoring disease progression, and evaluating treatment outcomes. In silico trials [1] using segmentation-derived vascular models enable the simulation of biological behaviours and support drug development efforts, ultimately enhancing the efficacy of personalised treatment strategies. Traditional manual segmentation methods, while highly accurate, are labour-intensive and time-consuming. This has motivated the exploration of automated segmentation techniques [2] as a more efficient alternative. However, the transition to computerised methods poses a new challenge: the reliance on fully labelled data for training.

In scenarios where clinical vessel annotations in 3D Rotational Angiography (3DRA) [3] modalities are ambiguous, semi-supervised learning methods [4] with teacher-student structure [5] leveraging both labelled and unlabeled patches have been proposed to address the issue. However, when the challenge escalates further, with the network being fed unlabelled Magnetic Resonance Angiography (MRA) [6, 7] data as input, semi-supervised methods become inadequate for supervision. To address the challenge of data of different modes without labelled data, unsupervised domain adaptation (UDA) techniques, such as FDA [8], DAFormer [9], HRDA [10], MIC [11], and MSCDA [12], have emerged. These methods transfer knowledge from well-annotated source domains to unlabelled target domains.

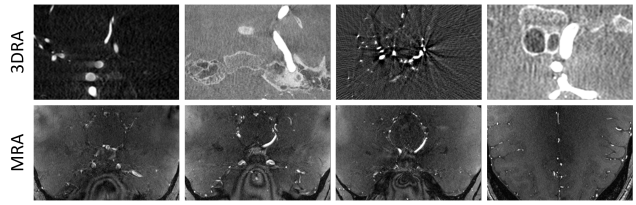

Fig. 1. Visualization of 3DRA and MRA data reveals significant intra- and inter- domain shifts.

These methods address the domain shift between source and target data. However, the domain shift is considerable, even from inner source data in the context of cerebral vascular images. Fig.1 shows that 3DRA and MRA exhibit a substantial domain shift, with a significant domain shift even within the 3DRA modality when clinical data from different data centres are scanned. Mitigating domain shifts for segmentation across other modalities is more challenging due to the intricate nature of blood vessels, individual variabilities, and inherent noise and artefacts in imaging techniques.

Our work introduces a unique method integrating insights from unsupervised domain adaptation, semi-supervised learning, and contrastive learning to address significant domain shifts in cerebral vascular images. The key innovations are as follows:

- This work presents an innovative symmetric adaptation network tailored for cross-modality segmentation of brain vasculature. This represents the first study to apply UDA techniques to the segmentation of the cerebral vessels from 3DRA to MRA.

* Contribute equally to this work

† Joint last authors

† AFF is supported by the Royal Academy of Engineering INSILEX Chair (CiET1919/19), UKRI Frontier Research Guarantee INSILICO (EP/Y030494/1), and EC Sixth Framework Programme @neurIST (FP6-2004-IST-4-027703).

- We introduce transwarp contrastive learning, a method that investigates features in the time and frequency domains within the latent space to achieve the alignment of content and style.
- Furthermore, we design a new homocentric squares Fourier domain adaptation to handle cross-domain nuisance variability without explicit feature alignment.

2. METHODS

This section presents our novel unsupervised domain adaptation approach (see Fig.2) to learning instance-specific and domain-invariant representations.

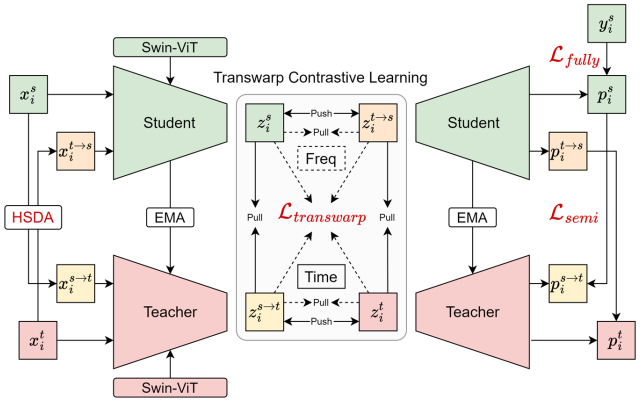

Fig. 2. Schematic of the proposed method. The method utilizes a composite loss function incorporating fully supervised, semi-supervised, and transwarp Contrastive Learning.

2.1. Method Overview and Problem Formulation

For image style transfer, FDA has demonstrated that low-frequency components of images represent style features [8]. Therefore, we utilise the low-frequency components of latent features to extract style features from different inputs. This novel paradigm in transwarp contrastive learning utilises a student-teacher network architecture, incorporating both content and style features using Fourier transform with a low-frequency mask, with the ultimate aim of narrowing the gap caused by different data modalities and magnifying the invariant feature extraction capabilities of the model.

As shown in Fig. 2, the student receives two inputs: labelled source domain data (x_i^s) and unlabelled target domain data undergone style transfer ($x_i^{t\to s}$). On the contrary, the teacher processes unlabelled target domain data (x_i^t) and style-transferred labelled source domain data ($x_i^{s\to t}$). At the same time, we undertake the extraction of both content and style features. From the student and teacher networks, the characteristics of the content in the time domain are defined as z_i^s, $z_i^{t\to s}$, z_i^t and $z_i^{s\to t}$. In parallel, we extract style features in the frequency domain, capturing the Fourier low-frequency attributes. These style features are expressed as s_i^s, $s_i^{t\to s}$, s_i^t, and $s_i^{s\to t}$.

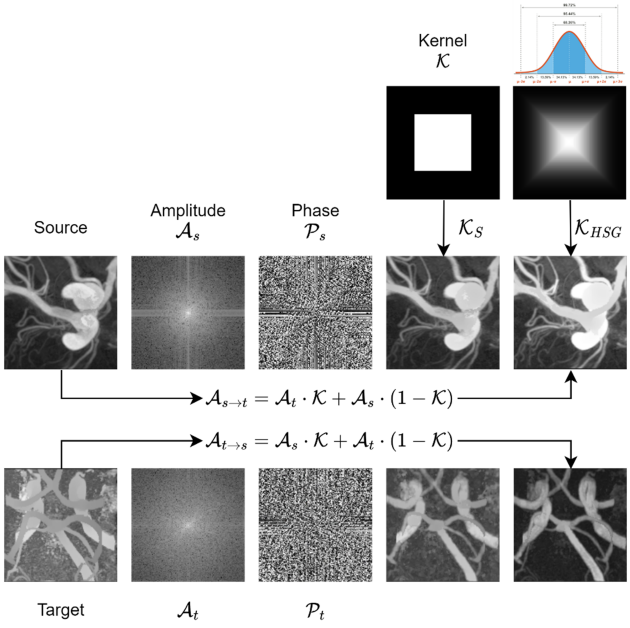

Fig. 3. Homocentric squares Gaussian kernel $\mathcal{K}_{HSG}$ for image adaptation on 3DRA (source) and MRA (target) vessel patch.

The prediction of the student network on the source domain and the target on the source domain is denoted as p_i^s and $p_i^{t\to s}$. We also denote the prediction of the teacher network in the target domain and the source in the target domain as p_i^t and $p_i^{s\to t}$. Our goal is to learn a task-specific network using a labelled source data set $\{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and an unlabelled target data set $\{(x_i^t)\}_{i=1}^{N_t}$ to predict labels on test data from the target domain accurately.

2.2. Homocentric Squares Domain Adaptation (HSDA)

The purpose of pre-processing is to remove noise and facilitate the model extracting features specific to the vessels. FDA proposes that the style of an image can be migrated by removing the low-frequency amplitude between the target image and the source image [8]. However, as observed in [8] [13], the generated images consist of incoherent dark patches caused by abrupt changes in amplitude around the rectangular mask. Also, considering that the distribution of the spectrum is square instead of circles, we propose a 3D homocentric square Gaussian mask in Fig. 3 to exchange the target and source amplitude to make a smooth transition in their style.

Let $\mathcal{F}_A(\cdot)$ and $\mathcal{F}_P(\cdot)$ be the amplitude and phase spectrum in frequency space of a 3D image, and $\mathcal{F}^{-1}$ indicate the inverse Fourier transform. We define a 3D homocentric

square Gaussian mask $\mathcal{K}_{HSG}$ of the same size as $\mathcal{F}_A$, with σ being the standard deviation. Given two patches x_i^s and x_i^t, our proposed HSDA can be formulated as:

$$x_i^{s\to t} = \mathcal{F}^{-1}[\mathcal{F}_P(x_i^s), \mathcal{F}_A(x_i^t)\cdot\mathcal{K}_{HSG}+\mathcal{F}_A(x_i^s)\cdot(1-\mathcal{K}_{HSG})] \tag{1}$$

2.3. Fully-supervised Learning

In the student network, labelled source samples $\{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ are passed through the task-specific segmentation network to minimise the supervised loss $\mathcal{L}_{fully}$, which includes Dice similarity coefficient loss and cross-entropy:

$$\mathcal{L}_{fully} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(1 - \frac{2\left|p_i^s \cap y_i^s\right|}{|p_i^s| + |y_i^s|} - y_i^s \log\left(p_i^s\right)\right) \tag{2}$$

2.4. Semi-supervised Learning

We employ semi-supervised learning, as the source data is labelled while the target data lacks labels. This involves using the predictions of the student network as pseudo-labels to supervise the output of the teacher network.

To perform alignment at the instance level, for the source patches, there are two predictions p_i^s and $p_i^{s\to t}$. For the target patches, there are two predictions p_i^t and $p_i^{t\to s}$. First, for the student network, the model's ability to segment the source style is improved in fully supervised learning, where p_i^s is supervised from y_i^s. After style migration, the student network also improves the segmentation performance of $p_i^{t\to s}$. We then use the output of the student network as a pseudo-label to supervise the output of the teacher network. Specifically, we employ the mean square error defined in Eq. 3 to reduce the discrepancy between the two predictions of the same patch.

$$\mathcal{L}_{semi} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(p_i^s - p_i^{s\to t}\right)^2 + \frac{1}{N_t}\sum_{i=1}^{N_t}\left(p_i^{t\to s} - p_i^t\right)^2 \tag{3}$$

2.5. Transwarp contrastive learning

Our objective in the time domain (content) is to enable the network to learn instance-specific features. First, we aim to achieve positive content pairs (see Eq. 5) by reducing the distance between content features derived from similar patches. The cosine similarity is computed to pull z_i^s closer to $z_i^{s\to t}$ and $z_i^{t\to s}$ closer to z_i^t. As these pairs originate from identical patches and have only undergone processing through different networks, their latent features should closely align. On the contrary, we set negative content pairs (see Eq. 6), derived from different modalities and patches, to maximise their inherent distance. This involves using the cosine similarity to push apart c_i^s from $c_i^{t\to s}$ and similarly, distance $c_i^{s\to t}$ from c_i^t.

Venturing into the frequency domain (style), we aim for the network to become predominantly agnostic to modality regarding feature style. The fundamental objective is for the network to encapsulate the inherent vasculature data style, surpassing the intricacies of specific patches or modalities. The post-encoding outcome should display a more consistent style representation irrespective of the data's source, whether from various patches or modalities.

To achieve this goal, we take advantage of the low-frequency component of the latent feature as the style feature, aiming to ensure stylistic consistency across all segments. To accomplish this, we calculate the upbeat style pairs as outlined in Eq. 7, aligning s_i^s, $s_i^{s\to t}$, $s_i^{t\to s}$, and s_i^t to be near each other. Furthermore, to preserve the integrity of the latent feature and mitigate excessive distortion, we restrict the number of encoder blocks to three, allowing for just two layers of max-pooling.

$$h(u, v) = \frac{u^T v}{\|u\|_2\|v\|_2} \tag{4}$$

$$pos_i^c = h\left(z_i^s, z_i^{s\to t}\right) + h\left(z_i^{t\to s}, z_i^t\right) \tag{5}$$

$$neg_i^c = h\left(z_i^s, z_i^{t\to s}\right) + h\left(z_i^{s\to t}, z_i^t\right) \tag{6}$$

$$pos_i^s = h\left(s_i^s, s_i^{t\to s}\right) + h\left(s_i^s, s_i^t\right) + h\left(s_i^{s\to t}, s_i^t\right) + h\left(s_i^{s\to t}, s_i^{t\to s}\right) \tag{7}$$

$$\mathcal{L}_{transwarp} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{(e^{pos_i^c} + e^{pos_i^s})/\tau}{e^{pos_i^c} + e^{pos_i^s} + e^{neg_i^c}} \tag{8}$$

In summary, transwarp contrastive learning leverages Fourier transformation to extract temporal components as content features while utilising low-frequency components as style features. This approach enhances contrast in content aspects while reducing contrast in style aspects and enhancing network adaptation capabilities across domains.

2.6. Overall framework and training objective

The weight ratio of fully supervised loss is higher than that of semi-supervised loss and contrastive loss of the transwarp (8:1:1). We prioritise the fully supervised loss to ensure training robustness and prevent the network from becoming overly confident and introducing noise during the initialisation stage.

$$\mathcal{L} = \lambda_1\cdot\mathcal{L}_{fully} + \lambda_2\cdot\mathcal{L}_{semi} + \lambda_3\cdot\mathcal{L}_{transwarp} \tag{9}$$

3. EXPERIMENTS

3.1. Datasets

Our study trained the network using labelled 3DRA and unlabelled MRA from the Aneurist data set and tested it with

the SMILE data set. **Aneurist-3DRA** [14]: This dataset includes 223 partially annotated 3D brain vessel images from four different centres, each using unique scanners and imaging protocols. **Aneurist-MRA** [14]: Consisting of 207 high-quality clinical MRA data, this dataset lacks vascular segmentation labels and is not paired with Aneurist-3DRA. **SMILE-UHURA** [15]: This dataset offers 14 fully labelled cerebral vessel cases in MRA. ToF angiography images were initially pre-segmented automatically and underwent extensive manual refinement to ensure accuracy.

3.2. Experimental Setup

Our experiments used an NVIDIA GeForce RTX 3090 GPU (24GB VRAM) with 128GB RAM. We use Swin-UNet [16] for both student and teacher networks, using patch-based learning with patch size [128, 128, 128]. We aim to have 100 training epochs with optimisation using the Adam optimiser. Data augmentation techniques, including random rotational adjustments and flipping, were applied. The initial learning rate is 0.001 and decreases by a factor of 0.1 every ten epochs. Teacher network parameters θ_{tea} received standard updates, while student network parameters θ_{stu} were updated using the exponential moving average approach in Eq. 10.

$$\theta_{stu} = \alpha \cdot \theta_{stu} + (1 - \alpha) \cdot \theta_{tea} \tag{10}$$

$$\alpha = \min\left(1 - \frac{1}{iter + 1}, decay\right) \tag{11}$$

3.3. Quantitative Results

In Table 1, we present a summary of segmentation performance compared to state-of-the-art (SOTA) techniques, which encompasses various scenarios: (1) Source-only learning on the target image, (2) Different UDA methods, and (3) Fully supervised learning with labelled target data. MSCDA is designed for medical images in UDA methods, and DAFormer, MIC, and HRDA are natural image UDA methods. Our approach excels in critical metrics, including the Dice similarity coefficient (DSC), Sensitivity (Sen), Jaccard index (Jac) and Volume similarity (VS), highlighting its effectiveness in detecting positive instances, achieving overall agreement, and ensuring volume matching accuracy. While we outperform other domain adaptation methods and the source-only learning baseline, there is room for further enhancement in our unsupervised approach compared to fully supervised methods.

Table 2 presents an ablation study that builds upon the foundation of the FDA and progressively integrates various components into our approach. As we observe improvements in the experimental results, it becomes evident that each component in our method contributes positively to the outcome.

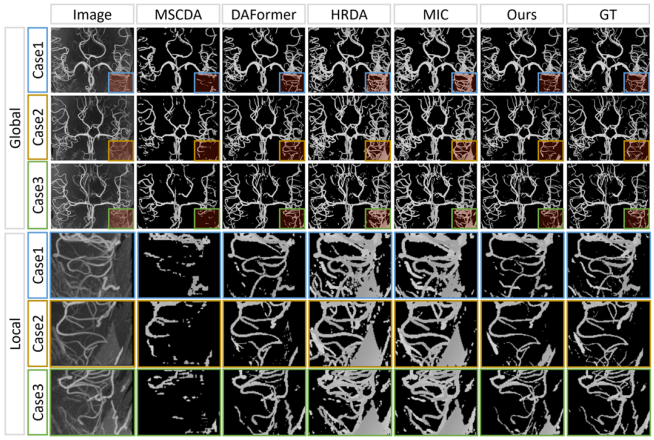

Fig. 4. Visualisation comparison on MIP maps. Ours shows less over-segmentation on local area.

Table 1. Comparison of Segmentation Performance with UDA SOTAs and different training strategy.* indicates $p < 0.05$ in t-test.

Methods	DSC (%) ↑	Sen (%) ↑	Jac (%) ↑	VS (%) ↑
$\mathcal{S} \rightarrow \mathcal{T}$ [2]	31.48 ± 6.76	18.89 ± 5.00	18.88 ± 5.00	31.52 ± 6.75
MSCDA [12]	41.18 ± 4.70	27.57 ± 4.96	26.04 ± 3.84	49.12 ± 8.69
DAFormer [9]	57.75 ± 6.35	42.84 ± 8.07	40.89 ± 6.52	63.37 ± 9.70
MIC [11]	67.16 ± 2.02	59.07 ± 7.16	50.59 ± 2.27	84.18 ± 9.49
HRDA [10]	68.35 ± 2.74	60.03 ± 8.57	51.98 ± 3.14	83.31 ± 9.68
Ours	**72.65 ± 6.65** *	**64.75 ± 8.06** *	**57.46 ± 7.80** *	**85.47 ± 9.65** *
$\mathcal{T} \rightarrow \mathcal{T}$ [16]	79.76 ± 1.92	74.61 ± 7.77	66.37 ± 2.69	90.06 ± 5.74

Table 2. Ablation Study: Gradual Addition of Components from Top to Bottom.

Components	DSC (%) ↑	Sen (%) ↑	Jac (%) ↑	VS (%) ↑
$\mathcal{L}_{fully}$	61.84 ± 7.08	46.29 ± 8.48	45.16 ± 7.77	64.88 ± 8.47
$\mathcal{L}_{semi}$	64.60 ± 7.36	49.08 ± 8.75	48.00 ± 8.17	67.48 ± 8.42
$\mathcal{L}_{transwarp}$	67.55 ± 6.81	52.75 ± 8.65	50.95 ± 7.67	72.16 ± 8.47
Ours HSDA	72.65 ± 6.65	64.75 ± 8.06	57.46 ± 7.80	85.47 ± 9.65

3.4. Visual Inspection

In addition to conducting quantitative comparisons, we also performed visual comparisons of results with similar numerical values in Table 1. In Fig. 4, we visually compared maximum intensity projection (MIP) maps between our method and the top-performing algorithm MIC and HRDA from UDA-SOTAs. It is evident from the image that within the red bounding box, MIC and HRDA tends to exhibit over-segmentation. Its capacity to distinguish between vessels and brain tissue noise is significantly inferior to our approach.

4. CONCLUSION

In conclusion, our novel transwarp contrastive learning framework represents a significant advancement in the field of UDA in the segmentation of brain vessels. We have effectively managed the challenges posed by diverse data modali-

ties by utilising content pairs from the time domain and style pairs from the frequency domain. This innovative approach not only facilitates knowledge transfer from source domains with domain shift but also ensures exceptional precision in feature extraction for application in target domains. Future work will focus on enhancing domain generalisation capabilities.

5. REFERENCES

[1] Qiongyao Liu, Ali Sarrami-Foroushani, Yongxing Wang, Michael MacRaild, Christopher Kelly, Fengming Lin, Yan Xia, Shuang Song, Nishant Ravikumar, Tufail Patankar, Z Taylor, Toni Lassila, and Alejandro F. Frang, "Hemodynamics of thrombus formation in intracranial aneurysms: an in-silico observational study," *APL Bioengineering*, 2023.

[2] Fengming Lin, Yan Xia, Shuang Song, Nishant Ravikumar, and Alejandro F Frangi, "High-throughput 3dra segmentation of brain vasculature and aneurysms using deep learning," *Computer Methods and Programs in Biomedicine*, vol. 230, pp. 107355, 2023.

[3] Hrvoje Bogunović, José María Pozo, María Cruz Villa-Uriol, Charles BLM Majoie, Rene van den Berg, Hugo AF Gratama van Andel, Juan M Macho, Jordi Blasco, Luis San Román, and Alejandro F Frangi, "Automated segmentation of cerebral vasculature with aneurysms in 3dra and tof-mra using geodesic active regions: an evaluation study," *Medical physics*, vol. 38, no. 1, pp. 210–222, 2011.

[4] Fengming Lin, Yan Xia, Nishant Ravikumar, Qiongyao Liu, Michael MacRaild, and Alejandro F Frangi, "Adaptive semi-supervised segmentation of brain vessels with ambiguous labels," *arXiv preprint arXiv:2308.03613*, 2023.

[5] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 605–613.

[6] Yash Deo, Rodrigo Bonazzola, Haoran Dou, Yan Xia, Tianyou Wei, Nishant Ravikumar, Alejandro F Frangi, and Toni Lassila, "Learned local attention maps for synthesising vessel segmentations from t2 mri," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2023, pp. 32–41.

[7] Yash Deo, Haoran Dou, Nishant Ravikumar, Alejandro F Frangi, and Toni Lassila, "Shape-guided conditional latent diffusion models for synthesising brain vasculature," *arXiv preprint arXiv:2308.06781*, 2023.

[8] Yanchao Yang and Stefano Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4085–4095.

[9] Lukas Hoyer, Dengxin Dai, and Luc Van Gool, "Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9924–9935.

[10] Lukas Hoyer, Dengxin Dai, and Luc Van Gool, "Hrda: Context-aware high-resolution domain-adaptive semantic segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 372–391.

[11] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool, "Mic: Masked image consistency for context-enhanced domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11721–11732.

[12] Sheng Kuang, Henry C Woodruff, Renee Granzier, Thiemo JA van Nijnatten, Marc BI Lobbes, Marjolein L Smidt, Philippe Lambin, and Siamak Mehrkanoon, "Mscda: Multi-level semantic-guided contrast improves unsupervised domain adaptation for breast mri segmentation in small datasets," *Neural Networks*, vol. 165, pp. 119–134, 2023.

[13] Hritam Basak and Zhaozheng Yin, "Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning," *arXiv preprint arXiv:2307.02798*, 2023.

[14] Siegfried Benkner, Antonio Arbona, Guntram Berti, Alessandro Chiarini, Robert Dunlop, Gerhard Engelbrecht, Alejandro F Frangi, Christoph M Friedrich, Susanne Hanser, Peer Hasselmeyer, et al., "@ neurist: infrastructure for advanced disease management through integration of heterogeneous data, computing, and complex processing services," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 6, pp. 1365–1377, 2010.

[15] S. Chatterjee, H. Mattern, F. Dubost, S. Schreiber, A. Nürnberger, and O. Speck, "Smile-uhura challenge 2023," *https://doi.org/10.7303/syn47164761*, 2023.

[16] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.