Graph-guided Sequential Transfer for Medical Image Segmentation

Jingyun Yang, Guoqing Zhang, Jingge Wang, Yang Li*
Shenzhen Key Laboratory of Ubiquitous Data Enabling
Tsinghua Shenzhen International Graduate School
Tsinghua University
Shenzhen, China
yangli@sz.tsinghua.edu.cn

***Abstract*—The medical image processing field often encounters the critical issue of scarce annotated data. Transfer learning has emerged as a solution, yet how to select an adequate source task and effectively transfer the knowledge to the target task remains challenging. To address this, we propose a novel source selection framework designed to identify the landmark source with an effective sequential transfer path for the given target task. Specifically, we first assess the relatedness among source tasks, estimated by our task affinity metric. Considering the characteristics of medical image segmentation tasks, we analyze the image and label similarity between tasks and compute the task affinity score. Following this, we construct a comprehensive source graph and combine the informativeness and representativeness of each node to identify the landmark source for the target. To ensure a positive transfer, we pinpoint a sequential transfer path to the target by minimizing both transfer and search costs. We gradually narrow the domain discrepancy and consequently improve the transfer performance on the target task by incorporating intermediate source tasks. Extensive experiments on three brain MRI medical datasets demonstrate the efficacy of the proposed framework in finding the best source sequence. The results show that our method outperforms other transfer learning approaches by a considerable margin, improving state-of-the-art performance by 6.61% for FeTS 2022, 0.66% for iSeg-2019, and 1.70% for WMH in terms of segmentation Dice score. Code is available at the git repository: SeqTransferLM.**

***Index Terms*—source selection, sequential transfer learning, transferability estimation, medical image analysis.**

I. Introduction

Advances in deep learning have led to rapid developments in medical image processing. As training from scratch is not a scalable solution in medical image analysis tasks for insufficient annotated data, transfer learning (TL) has become a critical technique in training deep neural networks to address the problem [1], [2]. To ensure robust representational capabilities, this paradigm requires pre-training on adequate source datasets and then we can fine-tune the model on the desired target task where only a small amount of annotated data is available. A phenomenon known as negative transfer [3] happens when the knowledge is transferred from a less related source task, which may inversely hurt the performance on the target task. Therefore, the selection of appropriate source tasks is crucial for TL.

*Corresponding author.

Although existing studies have made the stride in estimating the transferability for source selection [4]–[6], there remains a gap in applying these findings to the field of medical image processing as they haven't fully investigated the characteristics of medical image segmentation tasks. Several experimental studies tried to figure out what factors affect the transfer performance for medical image analysis. Cheng and Lam [7] explored the efficacy of pre-training with different datasets in improving lung ultrasound image segmentation performance and Wen et al. [8] analyzed the advantages and disadvantages of model-integration-based transfer learning strategies for medical image analysis. However, these methods haven't explored the properties of medical images themselves and answered the key question of how to select the best source task.

Recent TL studies on source selection have incorporated various methods to estimate the knowledge transferability between the source and target tasks for natural images, such as H-score [9], OTCE [6], and GBC [5]. Nguyen et al. [10] introduced a metric named LEEP to utilize the log-likelihood between the target labels and the predictions from the source model. LogME [11] computed the maximum evidence of model fit based on the assumption of linear parameters, focusing on the compatibility between features and labels. Nevertheless, these metrics are designed for classification and regression tasks that can use a single n-dimensional feature vector to represent each image while it's difficult for segmentation tasks to extract a global semantic representation and directly estimate the transferability [12]. Moreover, they rely on features extracted by pre-trained models which results in significant computational cost and they just focus on the relationship between the embeddings and target labels without exploring the properties of the medical images themselves, such as RoI shape similarity between the source and target tasks and modality difference [13]. For example, the T1ce scan is found to be more useful in displaying the enhanced tumor than other modalities [14]. If the knowledge of these representative modalities can be successfully identified and transferred, the model can produce promising results on the desired target [15] while aiding clinicians in choosing relevant scans. However, few of the aforementioned methods have

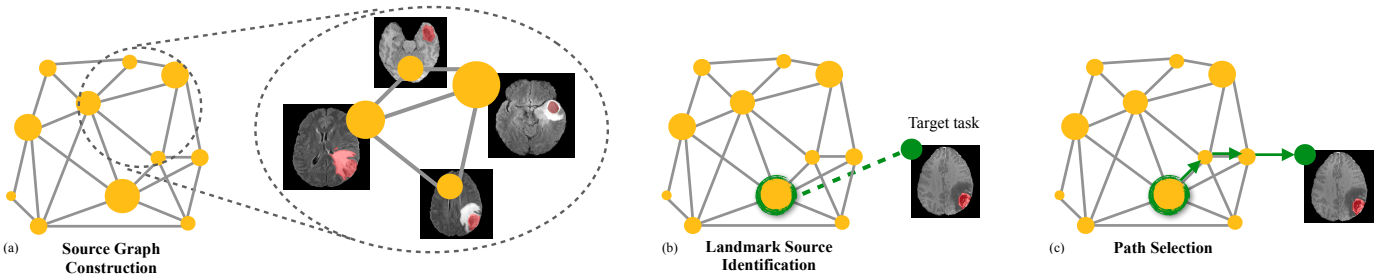

Fig. 1: Illustration of the proposed landmark source sequential transfer framework. (a) shows the graph we construct on source tasks. The edges connecting the nodes signify affinity between these medical image processing tasks. (b) depicts the landmark source we identify for the target task, represented by the circled node. (c) illustrates the most effective sequential transfer path selected for the target task, indicated by the green arrows.

considered the representativeness of source tasks and utilized such rich information in the medical image domain.

Aiming at these issues, in this work, we propose to select the landmark source for a given target task by leveraging the latent information of medical image segmentation tasks and estimating their informativeness and representativeness. First, we propose a simple yet effective task affinity metric by calculating the Wasserstein distance of low dimensional image features and structural similarity (SSIM) score of labels to assess the relatedness between tasks. In addition, we estimate the dataset diversity, density, and objective segmentation performance of source tasks as indicators of their informativeness and representativeness. Taking into account all these factors, we identify the landmark source for a given target task. Meanwhile, we've observed that there may exist less ideal datasets that closely align with the target dataset in terms of pathological features but fall short in volume, due to the high cost of labeling disease-specific instances. Training a model on a generalizable source dataset could ensure robust representation capabilities while learning on a similar source dataset is found to be useful to enhance the model's capacity for accurate image reconstruction [16]. In [1], sequential knowledge acquisition was proposed to facilitate the learning process of target tasks within the realm of continual learning. In light of these insights, to ensure a positive and effective transfer, we explore beneficial intermediate domains and find the sequential transfer path to fine-tune the model in a few-shot setting to get close to the target.

This systematic approach, as shown in Fig 1, ensures the enhanced performance of specific target tasks while adapting to various medical image segmentation tasks. Cause we focus on the target task performance instead of preserving the ability in every step in the source learning in continual learning, we propose a sequential transfer learning strategy to help improve the target task performance by optimizing the use of available rare source medical image data. We pre-train on a designated source task and across a spectrum of intermediate source tasks and sequentially transfer to the target. Compared to multi-source transfer [17], [18] and multi-task learning methods [19], [20], sequential transfer provides a solution for target-focused optimization with few-shot training requirements on source tasks and lower risks of negative transfer.

In summary, our main contributions are:

- **A novel graph-based sequential transfer framework.** We successfully apply a graph-guided sequential transfer learning pipeline in the medical image processing field to enhance target performance: a 6.61% gain for FeTS 2022, a 0.66% gain for iSeg-2019, and a 1.70% gain for WMH in terms of segmentation Dice score compared to state-of-the-art transfer methods.
- **Landmark source identification.** By fully exploring the characteristics of medical image segmentation tasks, we identify the most informative and representative source task for the given target task.
- **Sequential transfer path selection.** Based on the graph, we select the optimal sequential transfer path for the target task, driven by task affinity estimation, to ensure an effective transition sequence in the learning process.

II. Methodology

A. Problem Definition

Suppose we have a set of source tasks $\mathcal{S} = \{s_1, ..., s_N\}$ and a target task t. For $s_i \in \mathcal{S}$ and t, their data are $D_{s_i} = \{(x_j^{s_i}, y_j^{s_i})\}_{j=1}^{n_i} \sim P_{s_i}(x, y)$ and $D_t = \{(x_i^t, y_i^t)\}_{i=1}^{m} \sim P_t(x, y)$, respectively. In the context of transfer learning, given a source model parameterized by θ_s we fine-tune its decoder on the target task and obtain transfer accuracy, which can be measured by a segmentation metric (e.g., Dice score), denoted by $\mathcal{A}_{s \to t}$. The goal is to select the best landmark source s^* with an effective sequential transfer path $\mathcal{P}^\star$ for t, iteratively fine-tuning the model on each subsequent intermediate source task in the path, to achieve the best transfer accuracy $\mathcal{A}^\star$.

B. Source Graph Construction

First, we construct the graph $G = (V, E)$ on source tasks, as shown in Fig 2, where $V \subseteq \mathcal{S}$ represents the set of vertices corresponding to the source tasks, and $E \subseteq V \times V$ is the set of edges, estimated by the task affinity metric. We estimate the task affinity with the consideration of both images and labels of medical segmentation tasks. Accordingly, we filter out some edges to reduce the path search cost, detailed in Section III-A.

1) Image Similarity Analysis: given a pair of tasks (i, j) with sample sizes (N_i, N_j), the image similarity $\mathcal{H}(i, j)$ is measured by the Wasserstein distance [21] for its stability and ability to handle shifts in data distributions [12]:

$$\mathcal{H}(i,j) \triangleq \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \mathcal{W}(\hat{P}_k, \hat{P}_l), \tag{1}$$

where $(\hat{P}_k, \hat{P}_l)$ are distributions of images $(\hat{x}_k, \hat{x}_l)$ after dimension reduction using principal components analysis. And the data-pair Wasserstein distance $\mathcal{W}(\hat{P}_k, \hat{P}_l)$ is defined as:

$$\mathcal{W}(\hat{P}_k, \hat{P}_l) = \inf_{\gamma \in \Pi(\hat{P}_k, \hat{P}_l)} \mathbb{E}_{(x,y)\sim\gamma} \|x - y\|. \tag{2}$$

2) Label Similarity Analysis: we propose to use the structural similarity (SSIM) index [22] to quantify the similarity of task objectives. The label similarity $\mathcal{R}(i, j)$ is denoted as:

$$\mathcal{R}(i,j) \triangleq \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} SSIM(y_k, y_l). \tag{3}$$

SSIM is often used to evaluate the visual similarity between two images. The idea is that natural images often contain highly structural information, i.e., neighboring pixels in natural images have a strong correlation. Given two voxels (p, q), the data-pair SSIM is:

$$SSIM(p,q) = \frac{(2\mu_p\mu_q + C_1)(2\sigma_{pq} + C_2)}{(\mu_p^2 + \mu_q^2 + C_1)(\sigma_p^2 + \sigma_q^2 + C_2)}, \tag{4}$$

where μ_p is the average of p and μ_q is the average of q. σ_p^2 is the variance of p, σ_q^2 is the variance of q, and σ_{pq} is the covariance of p and q. C_1 and C_2 are constants for maintaining stability.

3) Task Affinity Estimation: for the pair of tasks (i,j), we calculate the task affinity metric $\mathcal{T}_{ij}$ and set the edges accordingly:

$$\omega(i,j) = \mathcal{T}_{ij} = \alpha\mathcal{H}(i,j) + \beta\mathcal{R}(i,j), \tag{5}$$

where the hyper-parameters set $\{\alpha, \beta\}$ are determined through Bayesian Optimization (BO). To decrease the exhaustive traversal of all path combinations while preserving the effectiveness of sequential transfer, we filter out edges between tasks with neither modality nor segmentation objective in common as transferring from source tasks with different modalities or limited region of interest (RoI) shape similarity to the target task has been found to be less effective [13].

C. Landmark Source Identification

To achieve good performance on the target in TL, the source data should not only have a similar distribution to the target but also provide sufficient information to enable the model to learn robust representation capabilities. Thus we combine the meta-information of each source to identify the landmark source.

For a source task s_i, we quantify the informativeness based on the amount of sample size it holds to ensure sufficient size to capture the complexities and variabilities in the data, denoted as ζ_i. Meanwhile, we calculate the representativeness, denoted as γ_i, based on its density ρ_i and objective segmentation performance, estimated by the dice score using the source model $F(\cdot; \theta_{s_i})$ parameterized by θ_{s_i}:

$$\begin{aligned} \gamma_i &= \rho_i \mathbf{O}(s_i) \\ &= \sum_{s_j \in \mathcal{S}, j \neq i} e^{-\left(\frac{\omega(s_i, s_j)}{\omega_c}\right)^2} \sum_{k=1}^{K_i} d(F(x_k^{s_i}; \theta_{s_i}), y_k^{s_i}), \end{aligned} \tag{6}$$

where ω_c is the neighborhood distance threshold and $d(\cdot, \cdot)$ is the function to calculate the dice score on the test set $\{(x_k^{s_i}, y_k^{s_i})\}_{k=1}^{K_i} \subset D_{s_i}$.

In the graph context, given a target task, we define it as a new node v_t. The identification of the landmark source s_{lm}^* is formulated as:

$$s_{lm}^* = v_{lm}^* = \arg\min_{v_i \in V} \frac{\omega(v_i, v_t)}{\zeta_i \gamma_i}. \tag{7}$$

For a transfer to be successful, we propose a sequential transfer strategy that incorporates the latent beneficial intermediate source nodes to identify an effective transition from the landmark source node v_{lm} to the target node v_t.

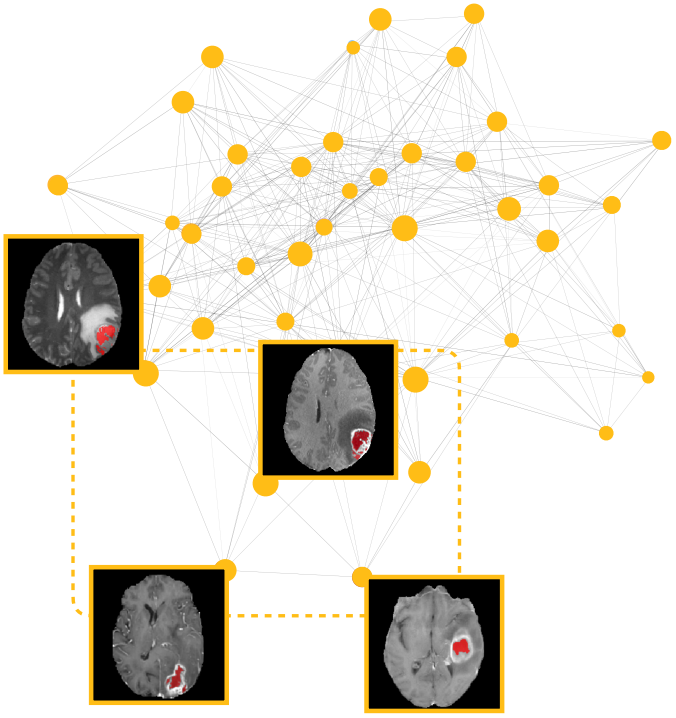

Fig. 2: Illustration of the source graph we construct. The edges connecting the nodes signify affinity between tasks. In order to display the nodes and edges clearly, we present 1/2 of the source tasks and depict four tasks.

D. Sequential Transfer Path Selection

To select an effective sequential transfer path $\mathcal{P} = \{v_{lm}^* \rightarrow v_1^p \rightarrow \ldots \rightarrow v_{l-1}^p \rightarrow v_t\} = \{v_0^p \rightarrow v_1^p \rightarrow \ldots \rightarrow v_{l-1}^p \rightarrow$

$v_l^p\}$ whose length is l, we formulate the objective function for ***O**ptimal **S**equential **T**ransfer* problem as follows:

$$L(\mathcal{P}) = \lambda \frac{1}{l} \sum_{i=0}^{l-1} \omega(v_i^p, v_{i+1}^p) + (1-\lambda) l C_{search}, \quad (8)$$

where a search cost C_{search} is added to limit the search steps, excluding all transfer paths longer than 5 as they offer minor improvements, keeping the total cost manageable, while λ is used to balance the exploration and exploitation, with detailed settings analyzed in Section III-D. The goal is to transfer from the landmark source to the target in as few steps as possible in as small a stride as possible.

The optimization problem is formulated as:

$$\begin{aligned} \min_{v_i^p \in V, l \in \mathbb{N}^+} \quad & L(\mathcal{P}), \\ \text{s.t.} \quad & \omega(v_i^p, v_{i+1}^p) < \omega(v_0^p, v_l^p), 0 \le i \le l-1, \\ & v_0^p = v_{lm}^*, v_l^p = v_t. \end{aligned} \quad (9)$$

We use the Dijkstra algorithm to solve the OST problem. In scenarios where the landmark source is sufficiently ideal, the solution could be a direct transfer, namely $\mathcal{P}^* = \{v_{lm}^* \to v_t\}$. The source graph is pre-constructed and prepared for arbitrary target tasks. When the target is given, it only requires the computation of edges between the target and source tasks.

III. Experiments and Results

A. Datasets

Three publicly available brain MRI segmentation datasets are used in our work: FeTS 2022 [23]–[25], iSeg-2019 [26], and WMH [27]. For the FeTS 2022 dataset, we use MRI volumes across T1, T2, FLAIR, and T1ce modalities, segmenting for enhancing tumor (ET), edema (ED), and necrotic core (NCR), with a resolution of $240 \times 240 \times 155$. This dataset is split into 22 partitions by the provider, according to different institutions and information extracted from images. Thus, each partition can be seen as an individual domain. We select datasets from 8 institutions (01, 04, 06, 13, 16, 18, 20, and 21), each with a sample size exceeding 30, and reorganize the datasets into a collection of binary segmentation tasks on every available modality. In total, we select 7x4x3 source tasks and 1x4x3 target tasks (from Institution 16). The iSeg-2019 dataset includes T1, T2 modalities and segments for white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), with image dimensions of $144 \times 192 \times 256$. The White Matter Hyperintensity(WMH) dataset focuses on FLAIR modality for white matter hyperintensities from three institutions, namely, VU Amsterdam (A), NUHS Singapore (S), and UMC Utrecht (U), sized $132 \times 256 \times 83$, $256 \times 232 \times 48$, and $240 \times 240 \times 48$, respectively. Thus 1x2x3 and 3x1x1 tasks are also used to perform experiments as source or target. For each dataset, we use the other two datasets to construct the source graph. We denote a task as "institute-modality-segmentation objective" and sort the dataset accordingly. For example, the WMH dataset includes scans from 3 institutes of 1 modality with 1 label, so we sort it into 3x1x1 tasks.

B. Training Setup

In the experiment setting, we filter out edges between nodes that don't share any modalities or similar segmentation objectives as transferring from tasks with the same modality or stronger RoI shape similarity to the target task has been found to be more effective [13]. We keep the same hyper-parameters for different settings. We use the same nnU-Net model architecture [28] for all experiments and follow the prevalent transfer training fashion, which is pre-training the model on a source task and fine-tuning it on the next task. During fine-tuning, the encoder is frozen and the parameters of the decoder are updated while we add EWC regularization [29] to protect the parameters with high Fisher information to stay close to the values needed for the previous task. We pre-train on the landmark source using 60 samples, based on the minimum size of the landmark source dataset. For fine-tuning, we experimented with 10, 5, 3, and 2 samples, and found that using 3 samples yields satisfactory results in sequential transfer settings where each subsequent domain is similar, while also reducing the size requirements of the auxiliary dataset. Therefore, we fine-tune the model with 3 samples for each sequential transfer step, including the final step on the target task. Training and test samples are consistently used in compared methods. Due to the large image size and memory constraints, we set a batch size of 2. We crop all data to the region of nonzero values in the same size. We use the Adam optimizer with an initial learning rate of 0.01 and set it to decrease periodically if the losses do not improve enough. To avoid overfitting, we utilize a large variety of data augmentation methods on the fly during training: random rotations, random scaling, and random elastic deformations. All MRI images used in experiments are preprocessed via a standard pipeline: registration, skull stripping, and bias field correction. All experiments are conducted on a CentOS 7.6.1810 system with one GeForce RTX 3090 GPU.

C. Performance Evaluation

We evaluate our proposed sequential transfer from the landmark source framework in comparison with state-of-the-art transfer learning methods, including single-source selection and multi-source adaptation. The baseline methods involve LEEP [10], LogME [11], a multi-source transfer (MST) framework using a fusion layer [18], and a mixed-batch multi-task learning (MTL) model [19], where we use the 3D UNet encoder in place of ResNet34 used in [19] to maintain consistency with the experiment settings. All results are averages from three random sets of 10 test samples. A quantitative analysis of transfer performance on 21 target tasks from three datasets is detailed in Table I.

Our framework surpasses all the other methods in the average Dice score. The existing source selection methods are inferior to ours because they are not designed for medical image segmentation tasks with not enough representative feature vectors for transferability estimation. Then we compare our method with multi-source and multi-task learning methods to prove that the enhancing performance on the target task

TABLE I: Model performance comparison of segmentation Dice score on three datasets. **Bold** marks the best-performing method.

Target	Method					Target	Method				
	LogME	LEEP	MST	MTL	Ours		LogME	LEEP	MST	MTL	Ours
16-Fl-ET	0.4976	0.5482	0.6137	0.5433	**0.6801**	iS-T1-GM	0.8909	0.8951	0.8924	0.8913	**0.9032**
16-Fl-ED	0.8770	0.8918	0.8993	0.8889	**0.9165**	iS-T1-WM	0.8824	**0.8973**	0.8940	0.8904	0.8966
16-Fl-NCR	0.2672	0.3465	0.3499	0.3254	**0.4136**	iS-T1-CSF	0.9285	0.9340	0.9322	**0.9397**	0.9351
16-T1-ET	0.4886	0.5081	0.5320	0.5049	**0.5902**	iS-T2-GM	0.8799	0.8831	0.8870	0.8881	**0.8894**
16-T1-ED	0.7657	0.7708	0.7698	0.7665	**0.7773**	iS-T2-WM	0.8634	0.8751	0.8676	0.8663	**0.8789**
16-T1-NCR	0.3299	0.3548	**0.3695**	0.3500	0.3691	iS-T2-CSF	0.8944	0.9001	**0.9028**	0.8983	0.9007
16-T1c-ET	0.8445	**0.8911**	0.8860	0.8813	0.8845	Avg	0.8899	0.8975	0.8960	0.8957	**0.9007**
16-T1c-ED	0.7769	0.8060	0.8211	0.8178	**0.8342**	-	-	-	-	-	-
16-T1c-NCR	0.7937	0.8190	0.8165	0.8122	**0.8190**	-	-	-	-	-	-
16-T2-ET	0.5418	0.5813	0.6184	0.5576	**0.6330**	S-Fl-WMH	0.8313	0.8386	0.8397	0.8346	**0.8426**
16-T2-ED	0.8856	0.8874	0.8877	0.8825	**0.8909**	U-Fl-WMH	0.7712	0.7925	0.8084	0.7945	**0.8141**
16-T2-NCR	0.3517	0.3946	0.3930	0.3888	**0.4209**	A-Fl-WMH	0.6576	0.6857	0.6863	0.6796	**0.6872**
Avg	0.6184	0.6500	0.6631	0.6433	**0.6858**	Avg	0.7534	0.7723	0.7781	0.7696	**0.7813**

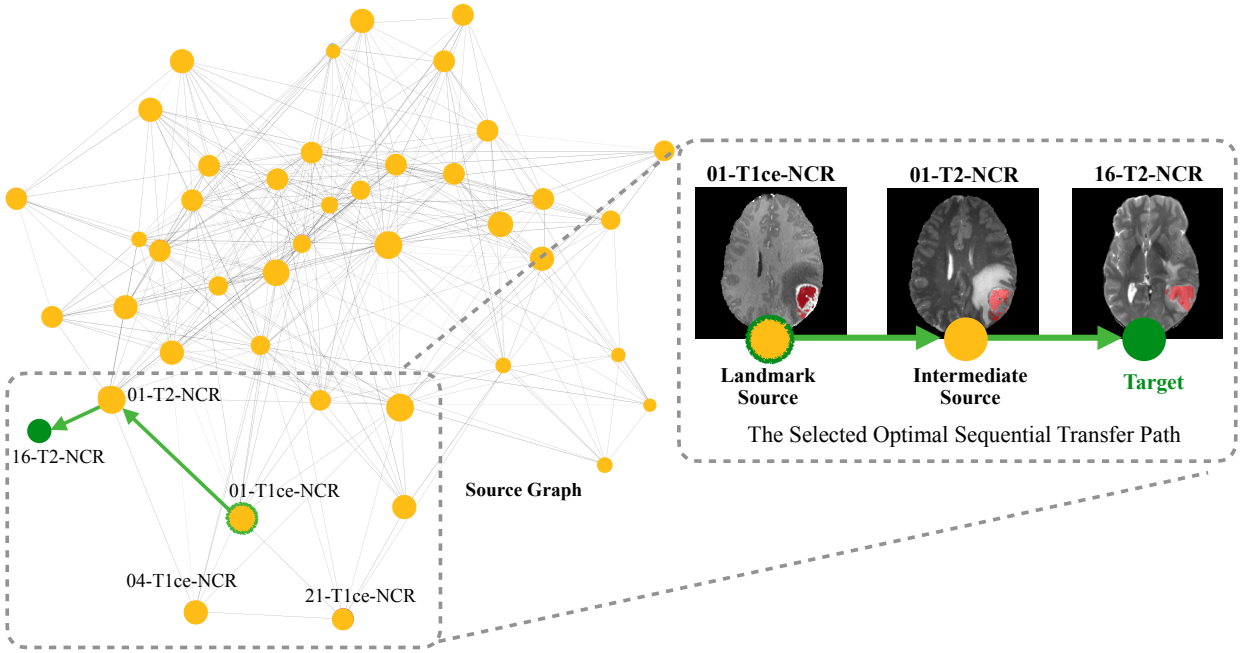

Fig. 3: Illustration of the selected optimal sequential transfer (OST) path for a specific target task, 16-T2-NCR. To ensure clarity in displaying the nodes and edges, we present 1/2 of the source tasks.

is not merely a result of increased training sample size. The results show that the sequential transfer framework provides a reasonable learning sequence for knowledge transfer, which is superior to learning all domains simultaneously. A closer look at the 16-T2-NCR target task, where our method achieves a Dice score of 0.4209, surpassing the MTL method by 8.26%, offers an illustrative example of the sequential transfer mechanism, shown in Fig 3. In the prior studies [15], we learned that NCR segmentation's best results are on the T1ce modality. Interestingly, the OST path chosen for 16-T2-NCR is 01-T1ce-NCR→01-T2-NCR. This suggests that the model first learns NCR detection well on the T1ce modality. After that, it strategically shifts to the T2 modality, getting it closer to the target. By following this step-by-step knowledge acquisition, the model achieves better transfer performance, showing the strength of landmark source and sequential transfer learning. Moreover, even when target tasks are quite different from source tasks, like brain tissue segmentation on iSeg-2019, our method still identifies a beneficial landmark source for learning how to extract features and an effective sequential transfer path

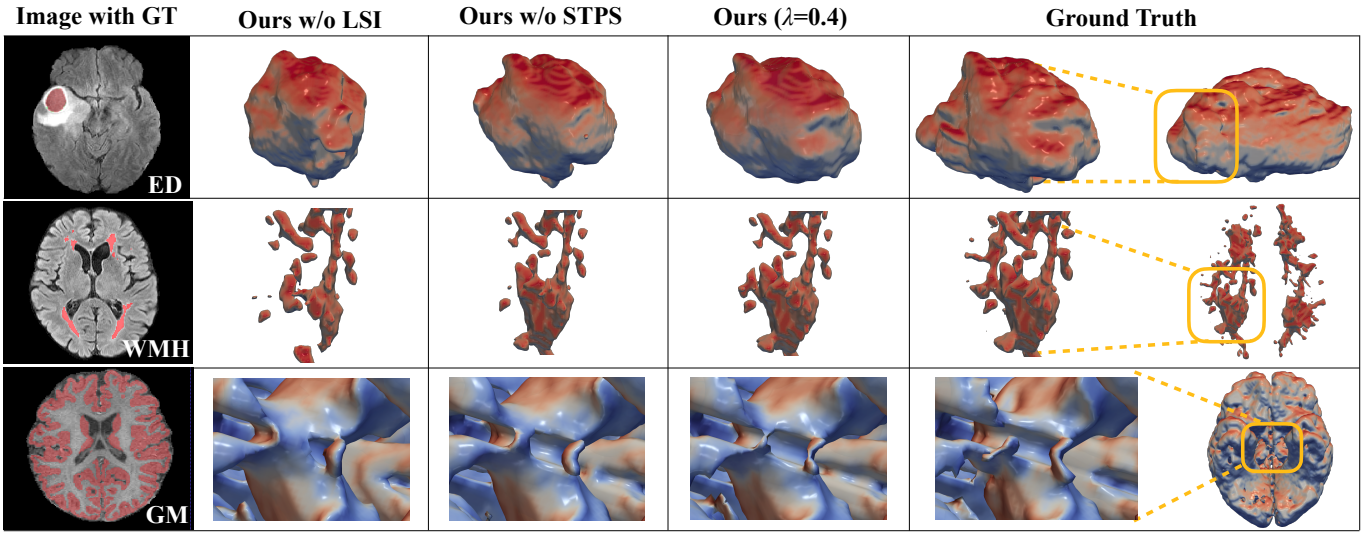

Fig. 4: A 3D visualization of transfer methods performance. From left to right: 1) image with ground truth, 2) sequentially transfer from the random initial source, i.e., w/o LSI, 3) directly transfer from the landmark source to the target, i.e., w/o STPS, 4) the proposed method, 5) ground truth.

to narrow the domain discrepancy.

D. Ablation Study

The results demonstrated in Table II underscore the importance of each part of our framework. First, we analyze the effect of the landmark source identification (LSI) and the sequential transfer path selection (STPS). In the "w/o LSI" experiments, we determine the initial source node for the sequential transfer path by (1) randomly choosing one and calculating the average score from 10 such selections, and (2) employing LogME and LEEP for source selection. Sequential transfer relies on the accurate selection of source models—a process that, if not executed properly, could lead to suboptimal transfer paths and a decrease in performance gains. The enhanced results show the superiority of the proposed landmark source selection. In the "w/o STPS" experiment, we directly transfer from the landmark source to the target to prove that our sequential transfer framework can identify an effective transition sequence in the learning process to improve the target performance.

We explore the trade-off between exploration and exploitation and show the results in Table III. A lower value of λ makes the framework tend to select shorter paths. We can observe that the results are less than ideal when $\lambda = 1.0$, indicating that the task affinity metric is not universally applicable to paths of all lengths and the extended paths may weaken the transfer performance. The best results at $\lambda = 0.4$ suggest this parameter setting effectively balances the trade-off between transfer and search costs. We present visualizations of the segmentation results of three target tasks from different medical datasets in Fig 4, clearly demonstrating the enhancements our method brings to various medical image segmentation tasks.

TABLE II: Ablation study on the effectiveness of different parts in our strategy. **Bold** marks the best Dice score.

Method	Target			Avg
	FeTS2022	iSeg-2019	WMH	
STPS	0.6109	0.8759	0.7476	0.7448
LogME + STPS	0.6184	0.8899	0.7534	0.7539
LEEP + STPS	0.6622	0.8996	0.7768	0.7795
LSI	0.6613	0.9002	0.7756	0.7790
LSI + STPS	**0.6858**	**0.9007**	**0.7813**	**0.7893**

TABLE III: Ablation study on different settings of λ. **Bold** marks the best Dice score.

Method	Target			Avg
	FeTS2022	iSeg-2019	WMH	
$\lambda = 1.0$	0.6820	0.8911	0.7805	0.7845
$\lambda = 0.5$	0.6833	0.8911	0.7809	0.7851
$\lambda = 0.4$	**0.6858**	0.9007	**0.7813**	**0.7893**
$\lambda = 0.3$	0.6827	**0.9008**	0.7801	0.7878
$\lambda = 0.0$	0.6613	0.9002	0.7756	0.7790

IV. Discussion

In this study, we propose a novel landmark source sequential transfer learning framework to select the landmark source task and sequentially transfer the knowledge to the target task through an effective transfer path. The framework identifies the most beneficial source task well while the stepwise learning process of sequential transfer notably improves the target task performance. Experiments on three benchmark medical datasets show that the proposed method achieves state-of-the-

art performance in what and how to transfer problems within the realm of medical image processing. In the future, we will extend our exploration to include datasets with a broader range of anatomical regions and more medical imaging modalities, such as CT, X-ray, and PET, among others. To assess more complex datasets, we are working on learning task embeddings and computing their Wasserstein distance for multi-modal multi-class medical image segmentation tasks. Another important research direction is to better align with real-world medical scenarios with various regulatory standards and ethical considerations. In current work, we address scenarios where certain public datasets are accessible to aid targets. Should ideal datasets be restricted by regulations like HIPAA [30], we plan to combine transferability metrics, e.g., LEEP and OTCE, to select the source model with our sequential transfer strategy to leverage auxiliary data.

ACKNOWLEDGMENT

This work is supported in part by the Natural Science Foundation of China (Grant 62371270) and Shenzhen Key Laboratory of Ubiquitous Data Enabling (No.ZDSYS20220527171406015).

REFERENCES

[1] Chaitanya Baweja, Ben Glocker, and Konstantinos Kamnitsas, "Towards continual learning in medical imaging," *arXiv preprint arXiv:1811.02496*, 2018.

[2] Jian Wang, Hengde Zhu, Shui-Hua Wang, and Yu-Dong Zhang, "A review of deep learning on medical image analysis," *Mobile Networks and Applications*, vol. 26, pp. 351–380, 2021.

[3] Zirui Wang, Zihang Dai, Barnabás Póczos, and Jaime Carbonell, "Characterizing and avoiding negative transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11293–11302.

[4] Kshitij Dwivedi, Jiahui Huang, Radoslaw Martin Cichy, and Gemma Roig, "Duality diagram similarity: a generic framework for initialization selection in task transfer learning," in *European Conference on Computer Vision*. Springer, 2020, pp. 497–513.

[5] Michal Pándy, Andrea Agostinelli, Jasper Uijlings, Vittorio Ferrari, and Thomas Mensink, "Transferability estimation using bhattacharyya class separability," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9172–9182.

[6] Yang Tan, Yang Li, and Shao-Lun Huang, "Otce: A transferability metric for cross-domain cross-task representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15779–15788.

[7] Dorothy Cheng and Edmund Y Lam, "Transfer learning u-net deep learning for lung ultrasound segmentation," *arXiv preprint arXiv:2110.02196*, 2021.

[8] Yang Wen, Leiting Chen, Chuan Zhou, Yu Deng, Huiru Zeng, Shuo Xi, and Rui Guo, "On the effective transfer learning strategy for medical image analysis in deep learning," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 827–834.

[9] Yajie Bao, Yang Li, Shao-Lun Huang, Lin Zhang, Lizhong Zheng, Amir Zamir, and Leonidas Guibas, "An information-theoretic approach to transferability in task transfer learning," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2309–2313.

[10] Cuong Nguyen, Tal Hassner, Matthias Seeger, and Cedric Archambeau, "Leep: A new measure to evaluate transferability of learned representations," in *International Conference on Machine Learning*. PMLR, 2020, pp. 7294–7305.

[11] Kaichao You, Yong Liu, Jianmin Wang, and Mingsheng Long, "Logme: Practical assessment of pre-trained models for transfer learning," in *International Conference on Machine Learning*. PMLR, 2021, pp. 12133–12143.

[12] Yuncheng Yang, Meng Wei, Junjun He, Jie Yang, Jin Ye, and Yun Gu, "Pick the best pre-trained model: Towards transferability estimation for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 674–683.

[13] Yicong Li, Yang Tan, Jingyun Yang, Yang Li, and Xiao-Ping Zhang, "Finding the most transferable tasks for brain image segmentation," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 1620–1625.

[14] Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng, "Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*. Springer, 2019, pp. 447–456.

[15] Hu Wang, Congbo Ma, Jianpeng Zhang, Yuan Zhang, Jodie Avery, Louise Hull, and Gustavo Carneiro, "Learnable cross-modal knowledge distillation for multi-modal learning with missing modality," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 216–226.

[16] Simon Kornblith, Jonathon Shlens, and Quoc V Le, "Do better imagenet models transfer better?," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2661–2671.

[17] Qingshan She, Yinhao Cai, Shengzhi Du, and Yun Chen, "Multi-source manifold feature transfer learning with domain selection for brain-computer interfaces," *Neurocomputing*, vol. 514, pp. 313–327, 2022.

[18] Yifu Zhang, Hongru Li, Tao Yang, Rui Tao, Zhengyuan Liu, Shimeng Shi, Jiansong Zhang, Ning Ma, Wujin Feng, Zhanhu Zhang, et al., "Multi-source adversarial transfer learning for ultrasound image segmentation with limited similarity," *arXiv preprint arXiv:2305.19069*, 2023.

[19] Simon Graham, Quoc Dang Vu, Mostafa Jahanifar, Shan E Ahmed Raza, Fayyaz Minhas, David Snead, and Nasir Rajpoot, "One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification," *Medical Image Analysis*, p. 102685, 2022.

[20] David Tellez, Diederik Höppener, Cornelis Verhoef, Dirk Grünhagen, Pieter Nierop, Michal Drozdzal, Jeroen Laak, and Francesco Ciompi, "Extending unsupervised neural image compression with supervised multitask learning," in *Medical Imaging with Deep Learning*. PMLR, 2020, pp. 770–783.

[21] Victor M Panaretos and Yoav Zemel, "Statistical aspects of wasserstein distances," *Annual review of statistics and its application*, vol. 6, pp. 405–431, 2019.

[22] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[23] Sarthak Pati, Ujjwal Baid, Maximilian Zenk, Brandon Edwards, Micah Sheller, G Anthony Reina, Patrick Foley, Alexey Gruzdev, Jason Martin, Shadi Albarqouni, et al., "The federated tumor segmentation (fets) challenge," *arXiv preprint arXiv:2105.05874*, 2021.

[24] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al., "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:2107.02314*, 2021.

[25] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[26] 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*, vol. 40, no. 5, pp. 1363–1376, 2021.

[27] Hugo J Kuijf, J Matthijs Biesbroek, Jeroen De Bresser, Rutger Heinen, Simon Andermatt, Mariana Bento, Matt Berseth, Mikhail Belyaev, M Jorge Cardoso, Adria Casamitjana, et al., "Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge," *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2556–2568, 2019.

[28] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al., "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[29] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al., "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[30] U.S. Department of Health and Human Services, "Health insurance portability and accountability act of 1996 (hipaa)," 1996, Accessed: 2024-06-20.