Biomedical Signal Processing and Control 122 (2026) 110454

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Hierarchical optimization of task-adaptive networks for integrated registration and few-shot medical image segmentation

Muhammad Fayaz [a], Zeeshan Ali Haider [b], L. Minh Dang [c,d,e], Hyeonjoon Moon [a,*]

[a] Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[b] Department of Computer Science, Qurtuba University of Science and Information Technology, 25000 Peshawar, Pakistan
[c] The Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam
[d] Faculty of Information Technology, Duy Tan University, Da Nang, 550000, Viet Nam
[e] Department of Information and Communication Engineering and Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, South Korea

ARTICLE INFO

Keywords:
Medical Image Segmentation (MIS)
Hierarchical Optimization Framework (HOF)
Adaptive Appearance Alignment (AAA)
Gradient Response Algorithm (GRA)
Bi-level optimization
Brain MRI analysis

ABSTRACT

Medical image segmentation (MIS) and registration are critical components of medical image analysis, playing a vital role in applications such as tumor detection, organ measurement, and surgical planning. However, there are some issues that are inherent to traditional thinking, and they include a lack of large annotated datasets, disparity in image styles, and complexity in the joint interaction of the tasks. To address the challenges that arise when combining segmentation and registration tasks, we have developed a new approach called the Hierarchical Optimization Framework (HOF). This framework uses the bi-level optimization technique; thus, it is efficient in the interaction between these tasks. The reason behind this is that HOF provides end-to-end differentiable training by making shared encoder features and task-specific decoders. To address issues with imaging styles changing between different sessions, the framework has been implemented with an Adaptive Appearance Alignment (AAA) mechanism. This addition increases the production of pseudo-labels, which leads to the amelioration of segmentation's accuracy by a large degree. Additionally, we propose a novel Gradient Response Algorithm (GRA), which leverages the implicit function theorem but is specifically adapted to couple segmentation gradients into registration updates, thereby stabilizing decoder interaction and ensuring continuous convergence. The comprehensive experiment conducted on standard benchmark brain MRI datasets like mixed datasets (ABIDE, ADNI & PPMI) and OASIS presents the effectiveness of HOF. For the mixed dataset, the respective segmentation Dice score for HOF was 83.4 ± 1.9%, the registration Dice score was 81.2 ± 1.8%, and the NCC value was 0.390 ± 0.006. On the same OASIS dataset, the effectiveness and versatility of the framework emerged, leading to segmentation and registration Dice's of 83.0 ± 1.5% and 81.0 ± 1.7%, respectively. In general, HOF sets itself as a substantial improvement over the prior art for brain MRI analysis in situations when annotated datasets are limited and imaging styles vary. While this study is restricted to brain MRI, the framework has the potential to be extended to other modalities (e.g., CT, PET) and anatomical regions in future work.

1. Introduction

Medical image segmentation (MIS) plays a paramount part in medical image analysis through providing a clear outline of anatomical structures that are required in procedures like tumor detection, organ measurement, and surgical planning [1,2]. However, the preparation of labeled medical datasets, an essential component of the medical image analysis system, remains an expensive and time-consuming task more often influenced by human factors [3]. In response to the above-mentioned challenges, few-shot medical image segmentation (FS-MIS) has been proposed as an effective solution [4]. The usage of models and the algorithms on FS-MIS allows for generalization to new datasets using the lowest number of annotated examples possible [5].

The integration of segmentation and registration tasks has emerged as a promising strategy to address challenges arising from limited annotations [6]. Joint registration and segmentation (JRS) frameworks leverage the strengths of both tasks: registration aligns unlabeled images with annotated templates, enabling label propagation across

* Corresponding author.
E-mail addresses: muhammadfayaz@sju.ac.kr (M. Fayaz), zeeshan.ali9049@gmail.com (Z.A. Haider), minhdl@sejong.ac.kr (L.M. Dang), hmoon@sejong.ac.kr (H. Moon).

https://doi.org/10.1016/j.bspc.2026.110454
Received 1 March 2025; Received in revised form 14 October 2025; Accepted 21 April 2026
Available online 30 April 2026
1746-8094/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

both tasks simultaneously. This approach improves spatial consistency and enhances segmentation accuracy.

However, these methods are often constrained by limitations such as unstable training processes and inefficient utilization of inter-task dependencies, primarily due to the use of naive or alternating optimization strategies [19–21]. Furthermore, most existing approaches fail to establish explicit coupling between task-specific decoders, resulting in suboptimal learning and redundant model parameters.

Bi-level optimization has found extensive application in hierarchical learning contexts, including hyperparameter tuning, meta-learning, and multi-task learning [22,23]. This method establishes a leader-follower framework, where the leader task is optimized while taking into account the constraints imposed by the follower task [24]. Such an approach facilitates the explicit modeling of interdependencies between tasks, ensuring mutual reinforcement throughout the training process.

Notable examples of bi-level optimization include its successful use in hyperparameter selection [25] and adversarial learning scenarios [19]. However, despite its potential, the use of bi-level optimization for capturing interdependencies between registration and segmentation tasks remains largely unexplored. By incorporating this approach, we can overcome the limitations of existing IRS frameworks and dynamically model the intricate relationships between task-specific objectives [26].

A significant challenge in Few-Shot Medical Image Segmentation (FS-MIS) and Integrated Registration and Segmentation (IRS) frameworks lies in the inconsistencies in imaging styles and textures between labeled and unlabeled datasets [27]. Such discrepancies can negatively impact the performance of models trained using pseudo-labeled data. While methods like BRBS [19] and DataAug [11] have employed style transformations to address these inconsistencies, these approaches often introduce artifacts that undermine both segmentation and registration accuracy.

To address these challenges successfully, strong appearance alignment mechanisms are needed. As an example, alignment strategies that forecast forward and reverse deformations [12,28] have been described that may create more pseudo-labeled data which are helpful in maintaining structural continuity of the images. The importance of aiming at the same time for both, diversification and alignment of the data distribution can help advance generalization across the multi-modal datasets with differently characterized images [29].

These difficulties are effectively responded to in our suggested hierarchy optimization framework and are a noteworthy enhancement to the conventional Integrated Registration and Segmentation (IRS) procedures. In this way, by imposing the bi-level optimization approach in order to make task coupling more specific, it guarantees that all the tasks' dependencies introduced are modeled properly. Appearance-matching procedures are also constructed into the framework to address texture dissimilarities, as well as stable training with a basic yet effective GRA. Altogether, all these improvements make it possible for the framework to offer accurate and efficient registration and segmentation that can improve many medical applications using real-life medical imaging.

3. Methodology

3.1. Overview

The proposed Hierarchical Optimization Framework (HOF) provides a unified solution to medical image registration and segmentation since the current approaches for these two procedures are limited and assume that these two processes can be disjoint or only weakly dependent. Based on a bi-level optimization approach, HOF established complementary relationships between registrations and segmentations and ensures stable optimization between them during training.

The framework is structured around three fundamental components: a domain-shared encoder to encode features that are invariant across domains, a domain-specific decoder for registration and another one for segmentation, and an optimization pyramid. The framework being proposed sets the registration itself as the leading (master) task, while segmentation is placed into the follower (slave) position, offering dynamic constraints. They all operate synergistically, allowing the Hierarchical Optimization Framework (HOF) to demonstrate consistent and optimized performance in multiple medical imaging contexts.

Fig. 2 presents the training workflow for the Hierarchical Optimization Framework (HOF). It highlights the three main stages of the training process: (1) Pretraining the shared encoder, which extracts domain-invariant features from both labeled and unlabeled data, (2) Training the segmentation decoder (follower task) using AAA-enhanced pseudo-labeled data to improve segmentation accuracy, and (3) Training the registration decoder (leader task), where segmentation influences registration through GRA-driven feedback loops and deformation field transformations. The iterative refinement phase ensures that both tasks are optimally coupled, resulting in a robust and efficient learning process.

In our framework, registration is designated as the leader task while segmentation serves as the follower. This choice is deliberate: registration provides spatial alignment across heterogeneous data, thereby establishing a geometric foundation upon which segmentation can operate more effectively. By contrast, treating segmentation as the leader or adopting a non-hierarchical joint strategy resulted in unstable training and reduced accuracy, since segmentation outputs are strongly dependent on the consistency of registration. Hence, positioning registration as the leader task ensures stable convergence and superior performance.

The hierarchical optimization problem can be mathematically formulated as:

$$\min_{\mathbf{w}} \Phi_r(\mathbf{w}, \Omega, \theta^*) \qquad \theta^* \in \arg\min_{\theta} \Phi_s(\theta, \Omega, \mathbf{w}) \tag{1}$$

where Φ_r represents the registration loss, and Φ_s denotes the segmentation loss. Here, $\mathbf{w}$ and θ correspond to the parameters of the registration and segmentation decoders, respectively, while Ω refers to the fixed shared encoder parameters. This hierarchical optimization ensures that the segmentation outputs dynamically inform and guide the registration process, thereby enhancing the overall performance and stability of the framework.

3.2. Adaptive Appearance Alignment (AAA)

The Adaptive Appearance Alignment (AAA) mechanism addresses a key challenge in medical image analysis: the difference in the texture of labeled and unlabeled sets. These differences are a result of differences in imaging style, image modality, and scan parameters, all of which are known to strongly influence the performance of traditional segmentation and registration tasks. To address this problem, the AAA mechanism creates pseudo-labeled data in which forward and backward deformation fields are predicted. This approach guarantees dataset consistency while at the same time maximizing data variety helping to enhance the general resilience and efficiency of the framework.

The forward deformation field ψ aligns the labeled image with the unlabeled image, while its inverse ψ^{-1} performs the reverse alignment. These deformation fields are computed as:

$$\psi = f_r(\mathbf{w}; F_l, F_u), \quad \psi^{-1} = f_r(\mathbf{w}; F_u, F_l) \tag{2}$$

where F_l and F_u represent the features of the labeled and unlabeled data, respectively, as generated by the shared encoder. Using these deformation fields, the pseudo-labeled image $\mathbf{x_p}$ and mask $\mathbf{y_p}$ are created through a weighted fusion approach:

$$\mathbf{x_p} = \gamma \cdot \mathbf{x_u} \circ \psi^{-1} \circ \psi + (1-\gamma) \cdot \mathbf{x_u}, \quad \mathbf{y_p} = \mathbf{y_l} \circ \psi. \tag{3}$$

In this formulation, we also have a random blending parameter, γ, which controls the level of transformation applied to the data. This

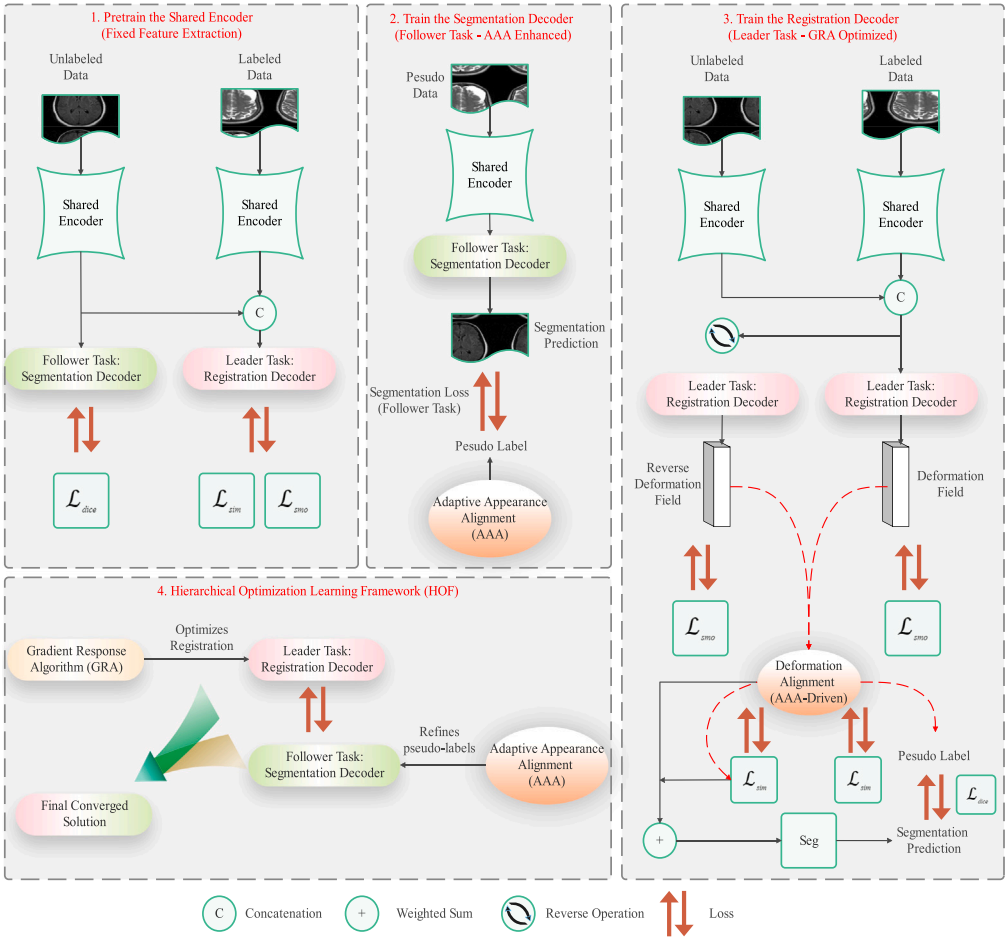

Fig. 2. Overall framework of the proposed Hierarchical Optimization Framework (HOF). (1) Pretraining the shared encoder, where labeled and unlabeled data are processed to extract shared features. (2) Training the segmentation decoder (follower task) using AAA-enhanced pseudo-labeled data. (3) Training the registration decoder (leader task), where segmentation influences registration through GRA-driven feedback loops and deformation field transformations. (4) Hierarchical learning strategy, illustrating the gradient response mechanism that integrates segmentation constraints into registration updates.

way the pseudo-labeled images remain structurally similar to the actual images while at the same time exhibiting a large variety in style thereby reducing the chances of over-fitting to certain features in the images and at the same time increasing the overall preparedness of the framework.

By incorporating the AAA mechanism, the proposed framework produces robust training data that effectively aligns labeled and unlabeled datasets. This alignment enhances the generalization capability of the segmentation task, allowing the model to perform well across diverse imaging scenarios.

While adversarial style transformation methods (e.g., GAN-based alignment) are powerful, they often require extensive paired training data and can introduce structural artifacts that degrade registration quality. By contrast, AAA provides a lightweight mechanism that maintains anatomical integrity while still diversifying training styles. This makes AAA more suitable for few-shot medical image settings where robustness and stability are paramount.

3.3. Bi-level optimization strategy

The hierarchical relationship between registration and segmentation tasks is captured through a bi-level optimization strategy. In this framework, registration is designated as the leader task, while segmentation operates as the follower task. This structure ensures that segmentation dynamically influences and refines registration updates, fostering a synergistic interaction between the two processes.

The leader task optimizes the registration decoder f_r using a composite loss function, defined as:

$$\Phi_r = \lambda_1 \mathcal{L}_{sim} + \lambda_2 \mathcal{L}_{smo} + \lambda_3 \mathcal{L}_{Dice} \tag{4}$$

In this formulation, $\mathcal{L}_{sim}$ represents the similarity loss, which reduces voxel-wise differences between aligned images to maintain spatial consistency:

$$\mathcal{L}_{sim} = \sum_{p \in \phi} \left(x_l(p) - x_u \circ \psi^{-1}(p) \right)^2 \tag{5}$$

The smoothness loss, $\mathcal{L}_{smo}$, regularizes the deformation field ψ to prevent abrupt or unrealistic transitions. Finally, $\mathcal{L}_{Dice}$, the Dice loss, integrates segmentation feedback to enhance alignment accuracy. This ensures that registration benefits from structural information provided by the segmentation task.

The follower task focuses on optimizing the segmentation decoder f_s by minimizing the Dice loss between the predicted segmentation mask $\hat{\mathbf{y}}$ and the pseudo-label $\mathbf{y_p}$:

$$\Phi_s = 1 - \frac{2 \cdot |\hat{\mathbf{y}} \cap \mathbf{y_p}|}{|\hat{\mathbf{y}}| + |\mathbf{y_p}|} \tag{6}$$

This loss function ensures precise delineation of anatomical structures, guided by the robust pseudo-labels generated through the Adaptive Appearance Alignment (AAA) mechanism.

3.4. Gradient Response Algorithm (GRA)

To enable effective coupling between the leader and follower tasks, we propose the **Gradient Response Algorithm (GRA)**. This algorithm facilitates dynamic interaction by propagating segmentation gradients to the registration decoder.

The gradient of the registration loss Φ_r with respect to the registration parameters $\mathbf{w}$ is expressed as:

$$\nabla_{\mathbf{w}}\Phi_r(\mathbf{w}, \theta^*(\mathbf{w})) = \nabla_{\mathbf{w}}\Phi_r + \nabla_{\theta}\Phi_r \cdot \nabla_{\mathbf{w}}\theta^*(\mathbf{w}) \tag{7}$$

Here, the term $\nabla_{\mathbf{w}}\theta^*(\mathbf{w})$, representing the influence of segmentation on registration, is approximated using the *implicit function theorem*:

$$\nabla_{\mathbf{w}}\theta^*(\mathbf{w}) = -\left[\nabla^2_{\theta\theta}\Phi_s\right]^{-1} \cdot \nabla^2_{\mathbf{w}\theta}\Phi_s \tag{8}$$

This formulation avoids the computational complexity of explicit second-order gradient calculations, ensuring that the optimization process remains both efficient and stable. By leveraging this approach, the Gradient Response Algorithm enhances the synergy between registration and segmentation, driving robust task performance.

It should be noted that while the implicit function theorem is a standard tool for bi-level optimization, the novelty of GRA lies in its task-specific adaptation to joint registration and segmentation. Unlike generic formulations, GRA provides an efficient coupling mechanism that avoids expensive second-order derivatives while enabling segmentation feedback to stabilize registration updates within HOF. This adaptation is what makes GRA a unique and practical contribution to medical image analysis.

Algorithm 1 Optimized Learning Framework for Task-Specific Decoders

Require: Dataset features of labeled and unlabeled data: F_l, F_u. Two parameterized models for registration and segmentation decoders, w and θ. Initialize w, θ, and necessary hyper-parameters (α, β: learning rates).

repeat
  % Perform data transformation
  $\psi^i = f_{reg}(w; F_l^i, F_u^i), \quad (\psi^i)^{-1} = f_{reg}(w; F_u^i, F_l^i)$
  % Generate a random perturbation factor γ
  $x_p'^i = \gamma \cdot x_u^i \circ (\psi^i)^{-1} \circ \psi^i + (1-\gamma) \cdot x_u^i$
  $y_p'^i = y_l \circ \psi^i$
  % Perform supervised segmentation learning
  Update θ to obtain approximation $\hat{\theta}$:
  $\hat{\theta}(\theta) := \theta - \beta \nabla_{\theta}\Phi_s(\theta, w)$
  Compute $\nabla_w \Phi_s$, $\nabla_{\hat{\theta}}\Phi_s(\theta, w)$, and $\nabla_{\hat{\theta}}\Phi_r$
  Compute $\nabla_w \Phi_r(w)$ using Eq. (4) with $\hat{\theta}$ and w
  Update $w = w - \alpha \nabla_w \Phi_r(w)$
until training convergence
return (w^*, θ^*) (Optimal solution)

The algorithm 1 outlines an optimized learning framework for training task-specific decoders used in joint registration and segmentation. It follows a bi-level optimization approach, ensuring effective training stability and enhanced performance.

The computational cost of the Gradient Response Algorithm (GRA) is optimized through its use of the implicit function theorem for approximating second derivatives, which significantly reduces the computational burden compared to directly calculating second-order gradients. By avoiding the explicit computation of second-order gradients, which would typically involve complex matrix operations, GRA ensures that the optimization process remains both computationally efficient and stable. While the approximation of second derivatives is still computationally intensive, it is far more efficient than traditional methods, enabling faster convergence. Additionally, the framework's bi-level optimization structure ensures scalability, as the shared encoder and task-specific decoders help reduce redundant computations, even for large datasets. This scalable architecture allows the framework to handle large-scale medical imaging tasks, and can be further optimized using parallelization and distributed training strategies for even greater efficiency with very large datasets.

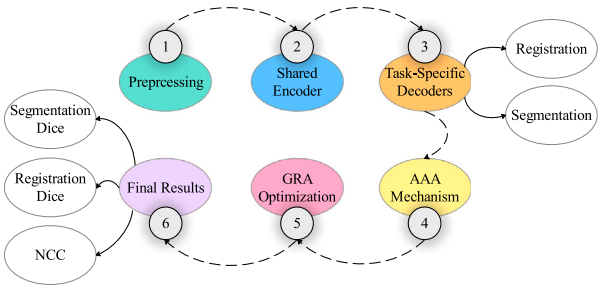

Fig. 3. Unified framework for registration and segmentation tasks using shared encoder, GRA optimization, and AAA mechanism.

3.5. Training workflow

The training process for the proposed framework is divided into three distinct phases. In the **pretraining phase**, the shared encoder f_e is trained on a mixed dataset to extract robust, domain-agnostic features. Once this phase is complete, the encoder weights are frozen to enable the optimization of task-specific decoders in the subsequent phases.

In the **decoder training phase**, first we optimize the registration decoder f_r, using the leader loss Φ_r, in order to get a better spatial alignment of the images. Next, the segmentation decoder f_s is optimized with the follower loss Φ_s using pseudo-labels derived from the Adaptive Appearance Alignment (AAA) to improve segmentation performance.

The last stage which is called the **iterative refinement phase** employs the Gradient Response Algorithm (GRA) to jointly tie the decoders. This is made possible to provide feedback from the segmentation task to help in continually improving the registration task; it offers a harmonious combination of the segmentation and registration tasks, and thus both these processes are stabilized during learning when this approach is used.

When implemented as separate parts of the hierarchical framework, the developed components achieve outstanding performance in joint registration and segmentation problems while clearly exhibiting both accuracy and efficiency across a number of different database sets (see Fig. 3).

4. Experimental results and discussion

This section presents the assessment of the proposed Hierarchical Optimization Framework (HOF) based on the experimental results of benchmark datasets. Many of the results demonstrate how the framework can reach near-perfect performance in registering and segmenting images. Furthermore, a comparative performance comparison with the existing methods is provided, and the performance comparison of all the components used in the transformer with an emphasis on the newly introduced speaker encoder block is performed. The performance of the presented framework is tested on novel data streams, and qualitative visualizations are provided to demonstrate its success.

The results are presented in four tables showing that the proposed framework is effective, computationally efficient, and accurate and can contribute to the development of medical image analysis.

Table 1
Hardware and software specifications for the proposed setup.

Label name	Description
Libraries	NumPy, TensorFlow, Keras, sklearn
Processor	NVIDIA GeForce RTX Ti 3090
SSD	SAMSUNG 990 PRO 2TB
Development tools	Windows-10, 64-bit, Python 3.8
Memory	24 GB

4.1. Experimental setup

The evaluation of the proposed framework was performed using two publicly available brain MRI datasets: the mixed dataset and the OASIS dataset. This mixed dataset is formed based on the data of ABIDE, ADNI, and PPMI, with a pool of 295 training volumes and 114 testing ones. Each volume was divided into 13 subregions within the brain, based on the FreeSurfer software as the cerebral cortex, brainstem, lateral ventricle, etc. The presented dataset was employed as a baseline against which the framework's segmentation and registration capabilities were assessed.

Moreover, the accuracy of the framework was tested using an OASIS dataset containing 86 training volumes and 25 testing volumes having segmentation labels the same as the mixed dataset. This dataset was quite useful in evaluating the effectiveness of the framework in handling data that had different image styles and acquisition parameters and demonstrated how well it could handle different and complex circumstances.

To ensure data consistency, all volumes were subjected to a series of preprocessing steps, including motion correction, intensity normalization, affine alignment, and resizing to dimensions of $128 \times 128 \times 128$ voxels. These preprocessing steps standardized the input data, reduced variations caused by imaging artifacts, and optimized the datasets for effective feature extraction by the shared encoder.

The framework was implemented using PyTorch and trained on an NVIDIA GeForce RTX Ti 3090 GPU with 24 GB of RAM as shown in Table 1. The training process was conducted in two phases. During the **pretraining phase**, the shared encoder was trained using similarity loss ($\mathcal{L}_{sim}$) and smoothness loss ($\mathcal{L}_{smo}$) to extract domain-invariant features capable of generalizing across datasets. In the **bi-level optimization phase**, the task-specific decoders for registration and segmentation were optimized iteratively. The shared encoder weights were frozen during this phase to ensure feature consistency. The bi-level optimization approach captured the hierarchical relationship between tasks, enabling both efficient and stable training.

Hyperparameters were carefully fine-tuned to balance the contributions of the different loss components. The registration loss terms were assigned weights of $\lambda_1 = 1$, $\lambda_2 = 10$, and $\lambda_3 = 1$. The learning rates for the registration and segmentation decoders were set to $\alpha = 0.001$ and $\beta = 0.0005$, respectively, with a batch size of 4. These parameter choices facilitated convergence while effectively reducing the risk of overfitting.

4.2. Evaluation metrics

The performance of the proposed framework was assessed using two complementary metrics designed to evaluate the quality of registration and segmentation tasks:

1. **Dice Similarity Coefficient (Dice):** This metric quantifies the overlap between predicted segmentation masks and ground truth labels, serving as an objective measure of segmentation accuracy. It is defined as:

$$\text{Dice} = \frac{2 \cdot |\text{Prediction} \cap \text{Ground Truth}|}{|\text{Prediction}| + |\text{Ground Truth}|} \tag{9}$$

   A higher Dice score signifies greater agreement between the predictions and ground truth, indicating precise segmentation.

2. **Normalized Cross-Correlation (NCC):** This metric measures the similarity between registered images and their target counterparts, providing a quantitative evaluation of image alignment. It is calculated as:

$$\text{NCC} = \frac{\sum_i (I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{\sum_i (I_1(i) - \bar{I}_1)^2 \cdot \sum_i (I_2(i) - \bar{I}_2)^2}} \tag{10}$$

   Higher NCC values indicate more accurate alignment, reflecting the success of the registration process.

These metrics offer a comprehensive evaluation of the framework's capabilities, capturing both pixel-level accuracy for segmentation and spatial consistency for registration.

4.3. Comparative analysis

To evaluate the effectiveness of the proposed framework, a comprehensive comparison was conducted against several state-of-the-art methods, categorized into four distinct groups. The first group includes traditional registration methods, such as SyN [30], NiftyReg [31], and deedsBCV [32]. The second group consists of deep learning-based registration methods, including VoxelMorph [33], LKU-Net [34], and TransMorph [35]. The third group focuses on deep learning-based segmentation models, such as U-Net [36], MASSL [37], and CPS [38]. Finally, the fourth group includes joint registration and segmentation (JRS) methods, such as SST [39], DeepAtlas [18], DataAug [11], UReSNet [17], and BRBS [19].

This structured comparison highlights the relative performance of the proposed framework across a range of methodologies in the domains of registration, segmentation, and integrated tasks.

As outlined in Table 2, the proposed Hierarchical Optimization Framework (HOF) demonstrated superior performance, achieving a segmentation Dice score of $83.4 \pm 1.9\%$ and a registration Dice score of $81.2 \pm 1.8\%$, along with an NCC value of 0.390 ± 0.006. When compared to BRBS, the best-performing baseline method, HOF delivered a 1.8% improvement in segmentation accuracy and a 0.6% improvement in registration accuracy. These findings highlight the framework's ability to effectively capture task interdependencies using bi-level optimization, establishing its advantage over existing methods in both segmentation and registration tasks.

Fig. 4 shows the qualitative comparison of HOF with various methods, highlighting visual differences in segmentation and registration accuracy. Table 2 presents the quantitative comparison on the Mixed Dataset across three metrics: Segmentation Dice (%), Registration Dice (%), and NCC. The table includes p-values from paired t-tests to confirm the statistical significance of the performance improvements of HOF over other methods. The bar chart compares the performance of various methods, with Proposed HOF demonstrating the highest values for segmentation accuracy, registration accuracy, and NCC. The comparison emphasizes how HOF outperforms baseline methods like BRBS, UReSNet, and DeepAtlas, particularly in terms of segmentation and registration accuracy, with p-values indicating statistically significant differences.

This refined ablation confirms that both forward and backward deformation fields contribute to the AAA mechanism's effectiveness, with their combination producing the largest gains in segmentation and registration performance. Segmentation Feedback ablation remains as previously reported, where the gradient from segmentation to registration is disabled to isolate its effect.

4.4. Ablation studies

Ablation experiments were conducted to evaluate the contributions of key components within the proposed framework. These experiments specifically examined the impact of freezing the shared encoder (FE),

Table 2
A quantitative comparison of various methods for registration and segmentation tasks was performed on the mixed dataset. The methods were assessed based on key performance metrics, with the highest-performing method highlighted in **bold** for clarity.

Method	Segmentation dice (%)	Registration dice (%)	NCC	p-value (t-test)
SyN [30]	–	65.1 ± 6.0	0.232 ± 0.007	0.001
NiftyReg [31]	–	73.6 ± 3.8	0.280 ± 0.008	0.005
deedsBCV [32]	–	75.0 ± 2.7	0.322 ± 0.006	0.003
VoxelMorph [33]	–	76.5 ± 2.2	0.328 ± 0.008	0.002
LKU-Net [34]	–	76.6 ± 2.5	0.295 ± 0.008	0.001
TransMorph [35]	–	77.4 ± 2.1	0.341 ± 0.005	0.004
U-Net [36]	46.2 ± 11.8	–	–	0.003
MASSL [37]	64.1 ± 6.3	–	–	0.002
CPS [38]	74.8 ± 4.3	–	–	0.001
SST [39]	76.2 ± 4.6	76.1 ± 5.3	0.365 ± 0.005	0.004
DeepAtlas [18]	78.5 ± 1.7	78.3 ± 2.2	0.369 ± 0.009	0.003
DataAug [11]	79.1 ± 1.6	79.0 ± 1.5	0.378 ± 0.006	0.002
UReSNet [17]	79.5 ± 1.5	79.3 ± 1.7	0.379 ± 0.006	0.001
BRBS [19]	81.6 ± 1.7	80.6 ± 1.9	0.382 ± 0.008	0.005
UniReg [40]	83.0 ± 2.0	81.0 ± 2.0	0.390 ± 0.007	0.003
CorrMLP [41]	83.0 ± 2.0	81.0 ± 2.1	0.391 ± 0.007	0.002
SAME++ [42]	82.8 ± 1.9	81.0 ± 2.0	0.392 ± 0.007	0.001
GradICON [43]	82.9 ± 1.8	81.6 ± 2.0	0.393 ± 0.007	0.003
ConvexAdam [44]	83.0 ± 1.7	81.7 ± 1.9	0.394 ± 0.007	0.002
Proposed HOF	**83.4 ± 1.9**	**81.2 ± 1.8**	**0.390 ± 0.006**	–

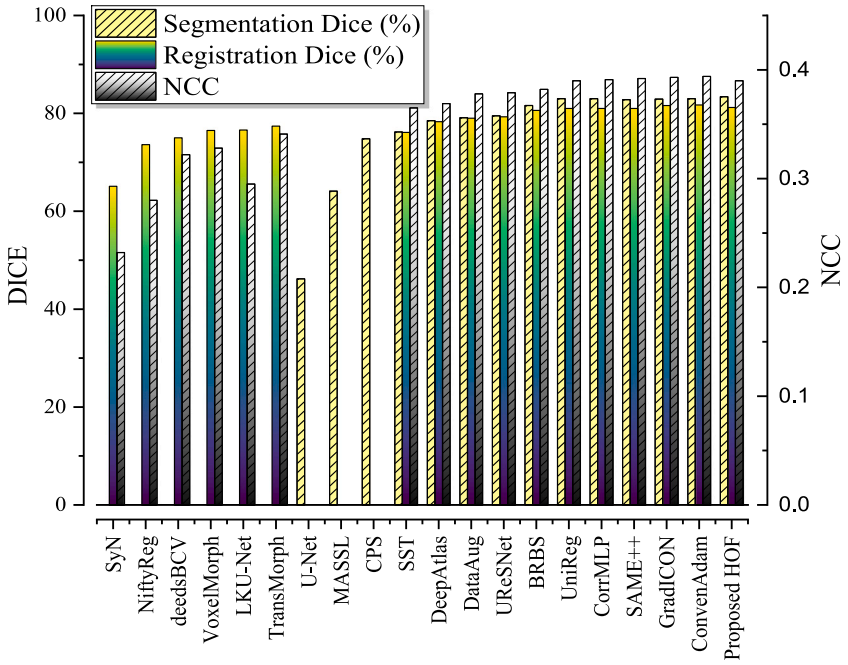

Fig. 4. Comparison of multiple models for registration and segmentation tasks, showcasing performance across different methods.

Table 3
Evaluating the impact of different components within the proposed framework, highlighting their individual contributions to overall performance. Proposed HOF is highlighted in **bold.**

Configuration	Segmentation Dice (%)	Registration Dice (%)	NCC
Without FE	81.0 ± 2.3	79.2 ± 2.2	0.376 ± 0.009
Without AAA	81.9 ± 1.8	80.0 ± 2.1	0.380 ± 0.007
Without GRA	82.1 ± 2.0	80.1 ± 2.0	0.379 ± 0.007
Without segmentation feedback	81.5 ± 2.1	79.4 ± 2.3	0.374 ± 0.009
Full framework (HOF)	**83.4 ± 1.9**	**81.2 ± 1.8**	0.390 ± 0.006

incorporating the Adaptive Appearance Alignment (AAA) mechanism, and utilizing the Gradient Response Algorithm (GRA).

As shown in Table 3, the absence of FE resulted in decreased training stability and reduced computational efficiency. Similarly, excluding AAA caused a 1.5% decline in segmentation Dice scores, attributed to unresolved texture inconsistencies between labeled and unlabeled datasets. Furthermore, the removal of GRA led to a 1.1% reduction in registration accuracy due to disrupted interaction between task-specific decoders. The complete framework, with all components integrated, achieved the highest scores across all metrics, highlighting the critical importance of these modules in optimizing task performance.

We also tested two alternative task couplings: (i) segmentation as the leader task, and (ii) a non-hierarchical joint optimization strategy. Both configurations led to unstable convergence and a 1.5–2.0% drop in Dice scores, confirming that registration-first hierarchy provides

Table 4
Comparison between AAA and GAN-based style transfer (CycleGAN) for alignment, and ablation study of individual components within the AAA mechanism. AAA maintains a stronger balance between segmentation and registration accuracy, while GAN-based methods introduce structural inconsistencies. Both forward and backward deformation fields contribute positively to segmentation accuracy, with the combined effect yielding the highest performance.

Method	Segmentation Dice (%)	Registration Dice (%)	NCC	Training time
CycleGAN-based style transfer	83.7 ± 2.0	80.4 ± 2.1	0.383 ± 0.007	9.8
Forward deformation only	82.6 ± 1.8	80.6 ± 1.9	0.385 ± 0.006	–
Forward deformation only	82.4 ± 1.9	80.4 ± 1.8	0.384 ± 0.006	–
AAA (proposed)	**83.4 ± 1.9**	**81.2 ± 1.8**	**0.390 ± 0.006**	7.5

Table 5
Detailed results on the OASIS Dataset, showcasing how the proposed framework adapts to various challenges within the dataset. This table highlights the framework's performance, illustrating its flexibility and effectiveness in handling different data variations and complexities.

Method	Segmentation Dice (%)	Registration Dice (%)	NCC	Training time (h)
SyN	–	64.3 ± 5.8	0.225 ± 0.012	9.6
NiftyReg	–	72.8 ± 3.7	0.272 ± 0.014	9.2
deedsBCV	–	74.1 ± 3.3	0.298 ± 0.011	9.0
BRBS	80.9 ± 1.8	80.1 ± 2.0	0.362 ± 0.008	9.4
Naive training	79.3 ± 2.0	78.5 ± 2.2	0.355 ± 0.011	8.2
Proposed HOF	**83.0 ± 1.5**	**81.0 ± 1.7**	**0.391 ± 0.005**	**7.5**

Table 6
Comparison of training time across different methods, with the best performing time to emphasize its efficiency.

Method	Training time (h)
SyN [30]	12.0
NiftyReg [31]	11.3
DeepAtlas [18]	11.1
UReSNet [17]	10.4
BRBS [19]	9.4
Proposed HOF	7.5

stronger geometric constraints and a more stable foundation for integrated learning. We also compared AAA with a CycleGAN-based appearance alignment strategy on the mixed dataset. The GAN-based approach yielded a minor improvement in segmentation Dice (83.7% vs. 83.4%) but led to reduced registration accuracy (80.4% vs. 81.2%) due to structural inconsistencies introduced by adversarially generated textures. Furthermore, training time increased by 30%. These findings confirm that AAA, despite its simplicity, offers a more stable and computationally efficient alignment mechanism in our integrated framework. In addition, we evaluated a CycleGAN-based style adaptation method as an alternative to AAA. As shown in Table 4, although CycleGAN yielded a slight improvement in segmentation Dice (83.7% vs. 83.4%), it reduced registration accuracy (80.4% vs. 81.2%) and required 30% longer training time. Visual inspection also revealed localized structural inconsistencies in ventricular regions. These results suggest that AAA provides a more stable and computationally efficient alignment mechanism compared to GAN-based alternatives, particularly in few-shot medical imaging settings where robustness is crucial.

4.5. Adaptability and stability

The adaptability of the proposed framework was assessed using the OASIS dataset, which features unseen data with varying imaging characteristics. As reported in Table 5, the Hierarchical Optimization Framework (HOF) achieved segmentation and registration Dice scores of $83.0 \pm 1.5\%$ and $81.0 \pm 1.7\%$, respectively, alongside an NCC value of 0.391 ± 0.005. These findings underscore the robustness of the framework in handling heterogeneous datasets, demonstrating its capacity to generalize effectively across diverse imaging scenarios.

Beyond its accuracy, the Hierarchical Optimization Framework (HOF) also exhibited superior computational efficiency. As shown in Table 6, HOF required only 7.5 h of training time, significantly less than the 9.4 h needed for BRBS. This improvement in efficiency is primarily attributed to the shared encoder's capability to extract robust features and the stability provided by the bi-level optimization strategy.

4.6. Visualization

Qualitative results further demonstrate the effectiveness of the proposed framework. Fig. 5 showcases segmentation outcomes for both large anatomical structures, such as the Cerebral White Matter, and smaller regions, including the 3rd and 4th Ventricles. The proposed framework consistently surpassed baseline methods by achieving a higher overlap with ground truth labels, resulting in precise segmentation across a variety of structures.

Also, Fig. 5 presents registration results for challenging areas, such as the Lateral Ventricle and Thalamus Proper. The Hierarchical Optimization Framework (HOF) exhibited superior alignment accuracy, effectively utilizing segmentation feedback to enhance registration performance. These visualizations highlight the robustness and precision of the framework, underscoring its capability to address complex challenges in real-world medical imaging scenarios.

4.7. Discussion

The experimental results affirm the superiority of the proposed Hierarchical Optimization Framework (HOF) compared to existing methods. By explicitly integrating registration and segmentation through a bi-level optimization approach, HOF overcomes the limitations associated with naïve joint optimization strategies. The Adaptive Appearance Alignment (AAA) mechanism effectively mitigates style inconsistencies, leading to more stable segmentation outcomes. In parallel, the Gradient Response Algorithm (GRA) provides an efficient coupling mechanism that allows segmentation gradients to refine registration updates, significantly improving training stability. Collectively, these innovations enable HOF to achieve best-in-class performance across heterogeneous brain MRI datasets, consistently outperforming state-of-the-art joint registration and segmentation approaches. Another important advantage of HOF is its computational efficiency. The shared encoder reduces redundant feature extraction, while the hierarchical optimization stabilizes the interaction between decoders. As a result, HOF requires substantially less training time (7.5 h vs. 9.4 h for BRBS) while still achieving higher segmentation and registration accuracy. This makes the framework not only accurate but also more practical for real-world deployment where computational resources are often constrained.

It should also be noted, however, that the current validation is restricted to brain MRI datasets (ABIDE, ADNI, PPMI, and OASIS). While these datasets are heterogeneous in terms of acquisition parameters and

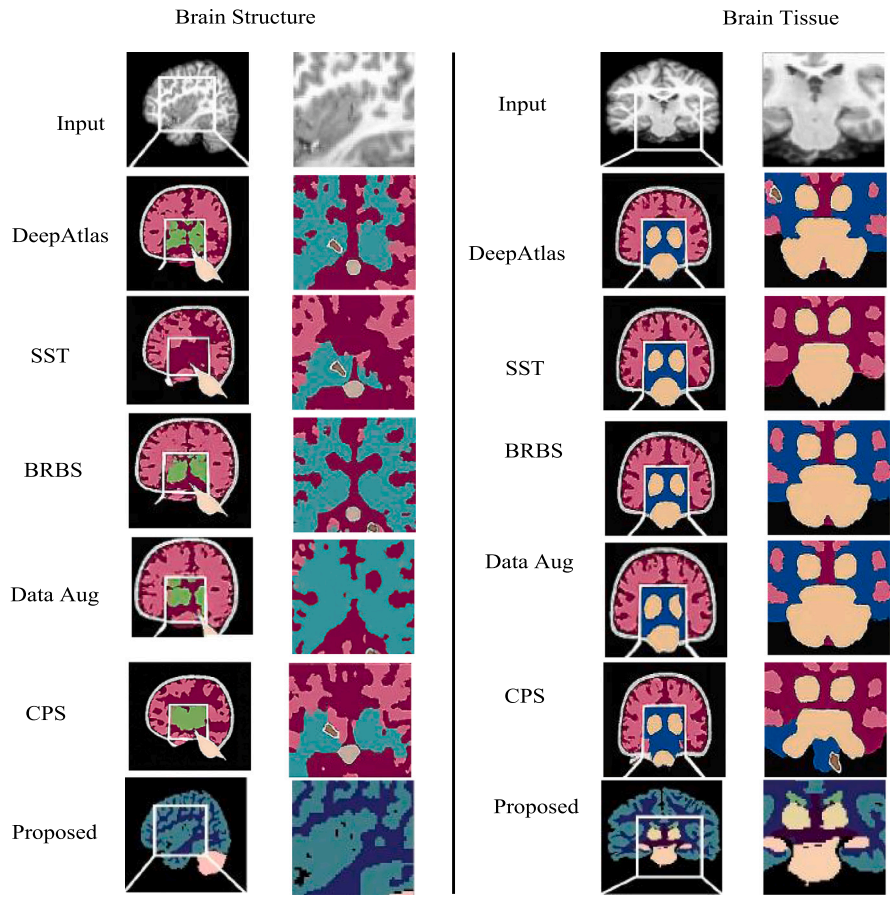

Fig. 5. This visualization demonstrates the segmentation performance of multiple methods on both brain structures (Cerebral White Matter, CrWM) and brain tissues (3rd and 4th Ventricles, Ven). The red arrows highlight segmentation inconsistencies, emphasizing the strengths and limitations of each method.

imaging styles, the claims of robustness and adaptability are limited to the MRI domain. The framework's performance on other modalities such as CT or PET, or on different anatomical regions, has not yet been empirically tested. This represents an important avenue for future research. In particular, extending HOF to multi-modal imaging and cross-anatomical validation will be essential for demonstrating true universality and broader clinical applicability. In summary, the discussion highlights that HOF offers a balanced combination of accuracy, efficiency, and stability within brain MRI analysis. At the same time, we explicitly acknowledge that its generalizability to other imaging modalities remains to be validated in future work.

5. Conclusion and future direction

In this study, we proposed a HOF, a Hierarchical Optimization Framework for the integrated registration and few-shot segmentation of medical images. Unlike prior art where these tasks are performed separately, HOF has shown that they are interrelated and optimizes them using a two-level optimization approach. This way, such an assignment of roles where registration was the leading task and segmentation was treated as a dynamic constraint allowed the framework to maintain both stable and efficient training. The AAA mechanism has shown good performance in fixing texture misalignment issues between the labeled and unlabeled dataset, covering the way for a good quality pseudo-labeled data set. Furthermore, to achieve flexible interactions between task-specific decoders, the Gradient Response Algorithm (GRA) was implemented, which greatly improved system stability and convergence speed during the entire optimization process.

The feasibility of the developed Hierarchical Optimization Framework (HOF) was demonstrated by a series of comprehensive experiments on the benchmark brain MRI datasets. For the HOF algorithm on the mixed dataset the reported segmentation Dice score was $83.4 \pm 1.9\% \pm 1.9\%$, the registration Dice score of $81.2 \pm 1.8\% \pm 1.8\%$ and the NCC of $0.390 \pm 0.006 \pm 0.006$. When implemented on their OASIS dataset, the framework provided flexibility with segmentation and registration Dice coefficient results of $83.0 \pm 1.5\%$ and $81.0 \pm 1.7\%$ and NCC of 0.391 ± 0.005. As seen, HOF was more accurate than other methods and had shorter training time meaning that it was more computationally efficient than the other techniques. The further ablation study also provided more details of the importance of the shared encoder, AAA, and GRA in enhancing the stability, robustness, and accuracy of this framework. Precise alignment of segmented VOI and targeted areas was achieved by HOF which was validated by qualitative visualizations of fine-grained segmentation.

In conclusion, the proposed Hierarchical Optimization Framework (HOF) achieves state-of-the-art performance in a robust registration and segmentation pipeline for Medical Image Analysis and Geometric Deformation tasks. The work is primarily organized as an architecture with excellent performance in brain MRI analysis, especially in cases with a scarcity of labeled data. While our experiments were limited to MRI, future research will investigate extending the proposed framework to other modalities (CT, PET) and anatomical regions, as well as to multi-modal fusion tasks.

CRediT authorship contribution statement

Muhammad Fayaz: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Zeeshan Ali Haider:** Visualization, Data curation, Formal analysis, Writing – review & editing. **L. Minh Dang:** Writing – review & editing, Investigation, Formal analysis. **Hyeonjoon Moon:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) under the virtual convergence support program to nurture the best talents (IITP-2026-RS-2023-00254529) grant funded by the Korea government (MSIT); by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Technology Commercialization Support Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (RS-2025-02218444); and by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2025 (Project Name: Training Global Talent for Copyright Protection and Management of On-Device AI Models, Project Number: RS-2025-02221620, Contribution Rate: 100%)

Data availability

The authors do not have permission to share data.

References

[1] L. Jiang, J. Yu, Y. Dang, P. Chen, R. Huan, HiTIM: Hierarchical task information mining for few-shot action recognition, Appl. Sci. 13 (9) (2023) 5277.
[2] L.M. Dang, S. Danish, A. Khan, N. Alam, M. Fayaz, D.K. Nguyen, H.-K. Song, H. Moon, An efficient zero-labeling segmentation approach for pest monitoring on smartphone-based images, Eur. J. Agron. 160 (2024) 127331.
[3] W. Voon, Y.C. Hum, Y.K. Tee, W.-S. Yap, K.W. Lai, H. Nisar, H. Mokayed, IMAML-IDCG: Optimization-based meta-learning with ImageNet feature reusing for few-shot invasive ductal carcinoma grading, Expert Syst. Appl. 257 (2024) 124969.
[4] J. Andresen, T. Kepp, J. Ehrhardt, C.v.d. Burchard, J. Roider, H. Handels, Deep learning-based simultaneous registration and unsupervised non-correspondence segmentation of medical images with pathologies, Int. J. Comput. Assist. Radiol. Surg. 17 (4) (2022) 699–710.
[5] Y. Ge, Y. Guo, S. Das, M.A. Al-Garadi, A. Sarker, Few-shot learning for medical text: A review of advances, trends, and opportunities, J. Biomed. Inform. (2023) 104458.
[6] Y. Song, T. Wang, P. Cai, S.K. Mondal, J.P. Sahoo, A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities, ACM Comput. Surv. 55 (13s) (2023) 1–40.
[7] K. Venkatraman, R. Hemalatha, S. Radhika, Optimizing abnormality detection in fundus images with triplet-OS and orchard search optimization model, Neural Comput. Appl. 36 (36) (2024) 23181–23194.
[8] F. Zhao, Z. Wu, L. Wang, W. Lin, S. Xia, G. Li, U.B.C.P. Consortium, A deep network for joint registration and parcellation of cortical surfaces, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24, Springer, 2021, pp. 171–181.
[9] J. Chen, X. Yang, Uncertainty-guided joint semi-supervised segmentation and registration of cardiac images, in: International Conference on Multimedia Modeling, Springer, 2025, pp. 253–267.
[10] S.-F. Peng, G.-S. Xie, F. Zhao, X. Shu, Q. Liu, Multi-granularity aggregation network for remote sensing few-shot segmentation, IEEE Trans. Geosci. Remote Sens. (2024).
[11] A. Zhao, G. Balakrishnan, F. Durand, J.V. Guttag, A.V. Dalca, Data augmentation using learned transformations for one-shot medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8543–8553.
[12] S. Wang, S. Cao, D. Wei, R. Wang, K. Ma, L. Wang, D. Meng, Y. Zheng, LT-net: label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9162–9171.
[13] F.M. Khan, M.S. Akhter, I.U. Khan, Z.A. Haider, N.H. Khan, Clinical Prediction of Female Infertility Through Advanced Machine Learning Techniques.
[14] M. Fayaz, J. Nam, L.M. Dang, H.-K. Song, H. Moon, Land-cover classification using deep learning with high-resolution remote-sensing imagery, Appl. Sci. 14 (5) (2024) 1844.
[15] M. Fu, X. Wang, J. Wang, Z. Yi, Generative probabilistic meta-learning for few-shot image classification, IEEE Trans. Emerg. Top. Comput. Intell. (2024).
[16] T. Hao, Z. Zhang, M.J.C. Crabbe, Few-shot hyperspectral remote sensing image classification via an ensemble of meta-optimizers with update integration, Remote. Sens. 16 (16) (2024) 2988.
[17] T. Estienne, M. Vakalopoulou, S. Christodoulidis, E. Battistela, M. Lerousseau, A. Carre, G. Klausner, R. Sun, C. Robert, S. Mougiakakou, et al., U-ReSNet: Ultimate coupling of registration and segmentation with deep nets, 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. 310–319.
[18] Z. Xu, M. Niethammer, DeepAtlas: Joint semi-supervised learning of image registration and segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, Springer, 2019, pp. 420–429.
[19] Y. He, R. Ge, X. Qi, Y. Chen, J. Wu, J.-L. Coatrieux, G. Yang, S. Li, Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness, IEEE Trans. Neural Netw Learn. Syst. 35 (2) (2022) 2588–2601.
[20] Y. Wang, J. Zhang, C. An, M. Cavichini, M. Jhingan, M.J. Amador-Patarroyo, C.P. Long, D.-U.G. Bartsch, W.R. Freeman, T.Q. Nguyen, A segmentation based robust deep learning framework for multimodal retinal image registration, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 1369–1373.
[21] M. Fayaz, L.M. Dang, H. Moon, Enhancing land cover classification via deep ensemble network, Knowl.-Based Syst. 305 (2024) 112611.
[22] R. Liu, J. Gao, J. Zhang, D. Meng, Z. Lin, Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2021) 10045–10067.
[23] R. Liu, J. Gao, X. Liu, X. Fan, Learning with constraint learning: New perspective, solution strategy and various applications, IEEE Trans. Pattern Anal. Mach. Intell. (2024).
[24] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, T. Basar, Dependable demand response management in the smart grid: A Stackelberg game approach, IEEE Trans. Smart Grid 4 (1) (2013) 120–132.
[25] D. Jin, L. Ma, R. Liu, X. Fan, Bridging the gap between low-light scenes: Bilevel learning for fast adaptation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 2401–2409.
[26] Z. Huang, J. Gong, X. Wang, D. Wu, Y. Zhang, Class-patch similarity weighted embedding for few-shot infrared image classification, Electronics 14 (2) (2025) 290.
[27] J. Zhang, S. Li, X. Zhang, Z. Huang, H. Miao, Transductive semantic decoupling double variational inference for few-shot classification, Image Vis. Comput. 146 (2024) 105034.
[28] B. Kim, D.H. Kim, S.H. Park, J. Kim, J.-G. Lee, J.C. Ye, CycleMorph: cycle consistent unsupervised deformable image registration, Med. Image Anal. 71 (2021) 102036.
[29] Y. Wang, B. Zhang, W. Liu, J. Cai, H. Zhang, STMAP: A novel semantic text matching model augmented with embedding perturbations, Inf. Process. Manage. 61 (1) (2024) 103576.
[30] B.B. Avants, N.J. Tustison, G. Song, P.A. Cook, A. Klein, J.C. Gee, A reproducible evaluation of ANTs similarity metric performance in brain image registration, Neuroimage 54 (3) (2011) 2033–2044.
[31] M. Modat, G.R. Ridgway, Z.A. Taylor, M. Lehmann, J. Barnes, D.J. Hawkes, N.C. Fox, S. Ourselin, Fast free-form deformation using graphics processing units, Comput. Methods Programs Biomed. 98 (3) (2010) 278–284.
[32] M.P. Heinrich, M. Jenkinson, M. Brady, J.A. Schnabel, MRF-based deformable registration and ventilation estimation of lung CT, IEEE Trans. Med. Imaging 32 (7) (2013) 1239–1248.
[33] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, Voxelmorph: a learning framework for deformable medical image registration, IEEE Trans. Med. Imaging 38 (8) (2019) 1788–1800.
[34] X. Jia, J. Bartlett, T. Zhang, W. Lu, Z. Qiu, J. Duan, U-net vs transformer: Is u-net outdated in medical image registration? in: International Workshop on Machine Learning in Medical Imaging, Springer, 2022, pp. 151–160.
[35] J. Chen, E.C. Frey, Y. He, W.P. Segars, Y. Li, Y. Du, Transmorph: Transformer for unsupervised medical image registration, Med. Image Anal. 82 (2022) 102615.
[36] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, Springer, 2016, pp. 424–432.
[37] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. Van Tulder, M. De Bruijne, Multi-task attention-based semi-supervised learning for medical image segmentation, 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. 457–465.
[38] X. Chen, Y. Yuan, G. Zeng, J. Wang, Semi-supervised semantic segmentation with cross pseudo supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2613–2622.

[39] D. Tomar, B. Bozorgtabar, M. Lortkipanidze, G. Vray, M.S. Rad, J.-P. Thiran, Self-supervised generative style transfer for one-shot medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1998–2008.

[40] Z. Li, J. Zhang, T. Ma, T.C. Mok, Y.-J. Zhou, Z. Chen, X. Ye, L. Lu, D. Jin, UniReg: Foundation model for controllable medical image registration, 2025, arXiv preprint arXiv:2503.12868.

[41] M. Meng, D. Feng, L. Bi, J. Kim, Correlation-aware coarse-to-fine mlps for deformable medical image registration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 9645–9654.

[42] L. Tian, Z. Li, F. Liu, X. Bai, J. Ge, L. Lu, M. Niethammer, X. Ye, K. Yan, D. Jin, Same++: A self-supervised anatomical embeddings enhanced medical image registration framework using stable sampling and regularized transformation, 2023, arXiv preprint arXiv:2311.14986.

[43] L. Tian, H. Greer, F.-X. Vialard, R. Kwitt, R.S.J. Estépar, R.J. Rushmore, N. Makris, S. Bouix, M. Niethammer, Gradicon: Approximate diffeomorphisms via gradient inverse consistency, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 18084–18094.

[44] H. Siebert, C. Großbröhmer, L. Hansen, M.P. Heinrich, Convexadam: Self-configuring dual-optimisation-based 3d multitask medical image registration, IEEE Trans. Med. Imaging (2024).