This WACV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Learning Semi-Supervised Medical Image Segmentation from Spatial Registration

Qianying Liu
University of Glasgow
2665227L@student.gla.ac.uk

Paul Henderson†
University of Glasgow
paul.henderson@glasgow.ac.uk

Xiao Gu
University of Oxford
xiao.gu@eng.ox.ac.uk

Hang Dai
University of Glasgow
hang.dai@glasgow.ac.uk

Fani Deligianni†
University of Glasgow
fani.deligianni@glasgow.ac.uk

Abstract

Semi-supervised medical image segmentation has shown promise in training models with limited labeled data and abundant unlabeled data. However, state-of-the-art methods ignore a potentially valuable source of unsupervised semantic information—spatial registration transforms between image volumes. To address this, we propose CCT-R, a contrastive cross-teaching framework incorporating registration information. To leverage the semantic information available in registrations between volume pairs, CCT-R incorporates two proposed modules: Registration Supervision Loss (RSL) and Registration-Enhanced Positive Sampling (REPS). The RSL leverages segmentation knowledge derived from transforms between labeled and unlabeled volume pairs, providing an additional source of pseudo-labels. REPS enhances contrastive learning by identifying anatomically-corresponding positives across volumes using registration transforms. Experimental results on two challenging medical segmentation benchmarks demonstrate the effectiveness and superiority of CCT-R across various semi-supervised settings, with as few as one labeled case. Our code is available at https://github.com/kathyliu579/ContrastiveCross-teachingWithRegistration.

1. Introduction

Semantic segmentation is a foundational task in medical image analysis. However, supervised methods require meticulously annotated images, which are expensive and time-consuming to obtain. Alternatively, *Semi-Supervised Semantic Segmentation* (S4) minimizes the need for manual annotation by leveraging a large pool of unlabeled images alongside a limited set of labeled images [30].

Existing S4 methods try to extract useful information from unlabeled data in various ways. One line of work [2, 16] first performs self-supervised pretraining on unlabeled data to learn robust features, then fine-tunes with limited labeled data. Other works learn from unlabeled images via pseudo-labeling [27, 61, 70] or consistency regularization strategies [24, 63, 64], both of which retrain the model using its own predictions on unlabeled images as pseudo-supervision. Cross-teaching frameworks, like the teacher-student [70] and student-student paradigms [18, 56], learn from unlabeled data by encouraging consistency of predictions between different network branches. Supervised contrastive learning endows the S4 model with a stronger feature-extraction ability [12, 35, 79, 85], encouraging features of pixels with the same class (positives) to be similar, and features of different classes (negatives) to be dissimilar. State-of-the-art (SOTA) cross-teaching methods [48] also incorporate pixel-wise contrastive learning on multi-scale feature maps. However, learning a robust representation from numerous unlabeled images remains challenging due to potential noise in pseudo-labels.

Spatial registration is a related task that aims to find dense spatial correspondences between pairs of 3D image volumes [5,57]. Many methods, both classical and learning-based, do not require manual supervision, but are based on comparing pixel intensities or features. Still, spatial registration yields a wealth of semantic information, as points matched by the registration transformation should, in principle, have the same semantic labels. Indeed, registration techniques are commonly used in brain image analysis to directly propagate a segmentation map from a template image to another [52]. Despite the wide use of spatial registration in medical image analysis, the potential of harnessing registration for S4 remains under-explored.

In this work, we investigate how to improve S4 by leveraging the rich semantic information inherently available

†Equal advising

This work was supported in part by the China Scholarship Council and EPSRC (EP/W01212X/1).

through off-the-shelf spatial registration methods. By integrating this information into contrastive cross-teaching frameworks [48, 56] which currently represent the SOTA in S4 for medical images, we propose a novel method *CCT-R*, incorporating two techniques that give substantial improvements in S4 performance for medical images.

Firstly, we use registration-derived semantic information to generate additional pseudo-labels for unlabeled data, and introduce a new loss allowing these to guide the segmentation process. This is beneficial since the accuracy of existing cross-teaching methods is limited by the quality of pseudo-labels predicted by each network and used to supervise the other; these pseudo-labels are typically very noisy during the early stages of training. In contrast, registrations can be computed offline, prior to training, with relatively high accuracy. We can then use registration transforms to transfer annotations from labeled to unlabeled volumes. To mitigate poor-quality registrations, we develop a simple yet effective 'best registration selection' (BRS) strategy that uses cycle-consistency to identify the most useful registrations for generating high-quality labels, without requiring extra supervision. In this way, more reliable pseudo-labels are available early in the training process, which helps avoid confirmation bias from cross-teaching, accelerates learning, and improves final segmentation performance.

Secondly, we use registration to optimise the sampling of pairs during pixel-wise contrastive learning. The SOTA contrastive cross-teaching S4 approach, MCSC [48], selects positive pairs based on (potentially noisy) pseudo-labels, and only within the current minibatch. By employing registration transformations, we can go further, identifying spatially-corresponding pixels for each anchor point across different volumes. This allows us to sample spatially positive pairs across volumes for contrast, even when their current pseudo-labels are incorrect, e.g. early in training. Furthermore, to increase the diversity of registration guided positives, and avoid the constraints imposed by batch size, we construct a memory-bank of feature maps from across multiple volumes.

In summary, our main contributions are as follows:

- We propose CCT-R, the first registration-guided method for semi-supervised medical image segmentation, by integrating registration with a contrastive cross-teaching framework.
- We introduce a novel registration supervision loss that enhances cross-teaching, by providing additional and informative registered pseudo-labels early in training, automatically selecting the best registered volumes.
- We show how registration can be used to mitigate the noisiness of pseudo labels in supervised contrastive learning, by adding anatomically-corresponding positive pairs regardless the currently predicted class.

Our evaluation demonstrates that each of these strategies enhances accuracy when combined with several recent S4 algorithms including UAMT [84], CPS [18], CTS [56], and contrastive variants. Implementing both strategies simultaneously proves even more effective. Our proposed CCT-R (based on CTS) achieves SOTA performance across all settings with particularly impressive gains under minimal supervision conditions. With just a single labeled case, CCT-R improves Dice coefficient (DSC) by 33.6% and reduces Hausdorff Distance (HD) by 32.8 mm on ACDC cardiac MRI segmentation [9], while on Synapse abdominal CT [42] it improves DSC by 21.3% and HD by 58.1 mm.

2. Related Work

Consistency regularization in semi-supervised medical image segmentation. Semi-supervised learning is a very effective approach to address the challenge of limited annotations in medical image segmentation [10, 13, 43, 56, 64]. Researchers have proposed various consistency regularization approaches that enforce consistency between multiple branches, either through data augmentations [10, 64], network architectures [56], or task configurations [76]. For instance, Bortsova *et al*. [10] encouraged consistency between the predicted masks and the input images under spatial transformations. Peng *et al*. [64] used adversarial learning to encourage diverse predictions among a set of models, while Luo *et al*. [56] leveraged Transformer-CNN consistency. However, most of these methods focus on prediction consistency for each single slice, overlooking feature relationships between different slices [48]. Additionally, relying on models to generate pseudo-labels often results in inaccurate organ boundaries [49]. Addressing these limitations remains an open challenge. Our CCT-R encourages both output and feature consistency between two branches [48, 56], while uniquely using registration to provide richer information beyond cross-teaching alone.

Medical image registration. Spatial registration is the process of aligning images from various sources, times, or patients to a common coordinate system [57], enabling tasks like automatic segmentation [31, 72], mathematical modeling [60], and functional imaging [82]. Classical methods, such as those based on mutual information (MI) [75], and feature-based techniques like Demons registration [71], align images by optimizing a cost function to minimize misalignment. These approaches rely heavily on pixel intensities and anatomical features. Recent advances in deep learning have introduced learnt methods [5, 21], which automate feature extraction and optimization. These methods can be supervised (trained with labeled reference deformations) [23, 68] or unsupervised (optimize similarity metrics without ground truth) [5, 34, 36]. Both classical and learnt methods typically take a pair of images (fixed and moving) as input, and produce a transformation matrix or a dense deformation field that aligns them.

Combining segmentation and registration. Segmenta-

tion and registration are closely related tasks that can complement each other, as both require extracting similar information from images. Several methods achieve segmentation purely by propagating the labels from an atlas image to another after registration, such as for gray/white matter [26] or V1/V2/IT [8] regions of brain, cardiac MR images [54] and liver CT [67]. Conversely, segmentation can provide additional supervision (beyond image intensities) for registration [3], as well as serve as a mean to evaluate registration results [41]. Consequently, many studies have explored joint training of deep networks for registration and segmentation across various supervision levels: unsupervised [1,51], fully supervised [7,20,22,37], few shot [44,78] and semi-supervised [81]. The most relevant to our CCT-R, DeepAtlas [81], jointly learns registration and S4 using 3D networks. However, they leverage neither established registration techniques nor modern S4 strategies like co-training and contrastive learning, limiting their approach to simpler anatomies (knee and brain). Unlike these works, our approach does not aim to solve registration itself. Instead, it leverages an existing (imperfect) registration algorithms to boost the performance of S4.

Contrastive learning for segmentation and registration. Contrastive learning has been pivotal in self-supervised representation learning [15, 28, 32, 73]. Early contrastive learning approaches focused on image-level (global) representations [17,29,33,62], increasing similarity between positive pairs while differentiating negative pairs. To adapt contrastive learning to the segmentation task, which requires dense predictions, recent research has introduced pixel-level (local) self-supervised contrastive learning [77, 80]. Some methods [11] incorporate both local and global contrastive losses in segmentation. These self-supervised methods are prone to false negative predictions [40]; to mitigate this, existing works [12, 35, 48] have explored supervised local contrastive learning. In the field of natural images, the integration of semi-supervised learning and contrastive learning has become a popular trend. This has lead to the development of one-stage, end-to-end models that eliminate the need for self-supervised pre-training [25, 38, 45, 53, 83, 86]. This approach has also been successfully applied to medical image segmentation [6, 12, 35, 79, 85]. Lastly, some works use self-supervised contrastive learning for registration, aiming to achieve high mutual information between fixed and moving images at the level of whole images [47] or patches [19, 69]. Unlike the above works, our CCT-R is the first to use registration information to guide contrastive sampling for S4.

3. Method

We first describe our problem setup and overall learning framework (Section 3.1), which closely follows SOTA cross-teaching methods [18, 48, 56]. Next, we introduce the main technical contributions for our CCT-R: incorporating registration into the S4 framework (Section 3.2), followed by a detailed description of how this is accomplished through a Registration Supervision Loss (RSL) (Section 3.3) and by improving the quality of contrastive pairs with the Registration-Enhanced Positives Sampling (REPS) module (Section 3.4).

3.1. Preliminaries

S4 aims to obtain good segmentation performance by leveraging data comprising of few labeled 2D slices $D_l = \{(x_i^l, y_i^l)\}_{i=1}^K$ and many unlabeled slices $D_u = \{x_j^u\}_{j=1}^M$ (*i.e.* $M \gg K$). Let $V = \{v_n\}_{n=1}^N$ represents the set of all 3D volumes, from which the set $D = D_l \cup D_u$ is extracted.

Our overall learning framework is similar to cross pseudo supervision [18,56] (Fig. 1), and the input is a mini-batch $X = X^l \cup X^u$ including labeled images and unlabeled images. It uses two student models that are trained via a standard supervised loss $\mathcal{L}_{sup}$ on X^l, and via a cross pseudo supervision loss $\mathcal{L}_{cps}$ on X^u where each network learns from the predictions of the other.

The supervised loss combines Dice and cross-entropy terms, similar to [4, 48]:

$$\mathcal{L}_{sup} = \mathcal{L}_{dice}(P_*^l, Y^l) + \mathcal{L}_{ce}(P_*^l, Y^l). \quad (1)$$

Here P_*^l is the predicted class probability map of the labeled image batch X^l, calculated according to $P_*^l = C_*(E_*(X^l))$ where $E_*(\cdot)$ is a feature extractor, $C_*(\cdot)$ is a segmentation head yielding class probabilities for each pixel, Y^l is the ground-truth label maps and $*$ denotes the model A or B.

The cross pseudo supervision loss $\mathcal{L}_{cps}$ [18] enables model A and model B teach each other on the unlabeled X^u, encouraging their respective predictions to be consistent. Specifically, we define

$$\mathcal{L}_{cps(A)} = \mathcal{L}_{dice}(P_A^u, Y_B^u), \quad \mathcal{L}_{cps(B)} = \mathcal{L}_{dice}(P_B^u, Y_A^u). \quad (2)$$

Here the Dice loss $\mathcal{L}_{dice}$ for model A uses pseudo-labels Y_B^u predicted by model B as its target, instead of ground-truth labels as in $\mathcal{L}_{sup}$. Note that there is no gradient back-propagation between P_A^u and Y_B^u during training, nor between P_B^u and Y_A^u. In Section 3.3, we will show how using spatial registration information can improve accuracy by providing additional pseudo-labels that are often less noisy than the cross teaching predictions.

Supervised contrastive learning. In addition, we optionally incorporate a supervised contrastive learning loss $\mathcal{L}_{cl}$, to better capture high-level semantic relationships between distant regions of different cases across the entire dataset. Our contrastive loss follows [39], but with the key difference that it contrasts pixel features instead of whole-image features. We project each pixel to a shared embedding space then regularize in a supervised manner, encouraging features of anchor pixels to be similar to those of pixels having the same class (positives), and to be dissimilar to those of different classes (negatives).

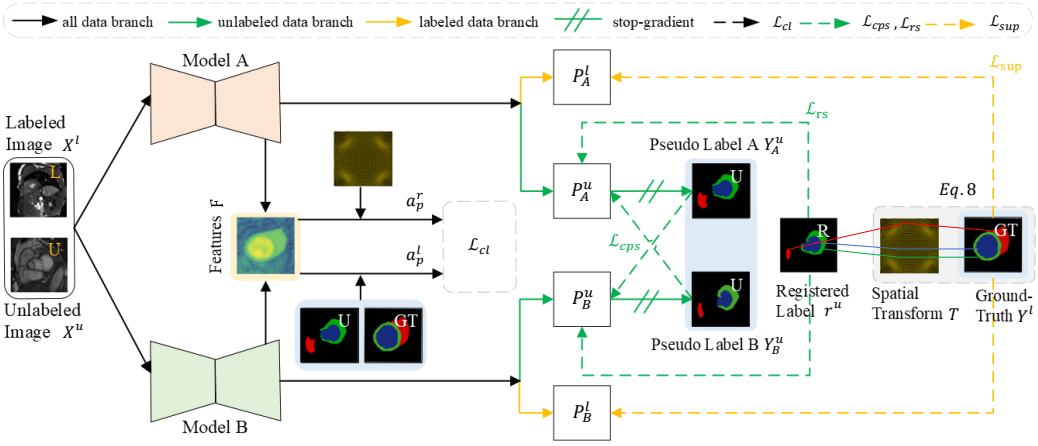


Figure 1. The overall architecture of our framework for semi-supervised medical image segmentation.

Specifically, as shown in Fig. 1, we extract a feature batch $F = F_A \cup F_B$, where $F_* = H_*(E_*(X))$ and $H_*(\cdot)$ is the projector. The choice of anchors, which serve as the comparison target of each class, has a great impact on learning; we therefore try to reduce the number of anchors with incorrect class labels. For every class in the current mini-batch, we sample pixels with high top-1 probability value as anchors A_c for class c, setting

$$A_c = \{f_i \,|\, (y_i = c) \wedge (p_i > h)\}, \tag{3}$$

where f_i is the i^{th} pixel feature in F, and the threshold h for top-1 probability value is set to 0.5 to only exclude hard samples.

The supervised contrastive loss $\mathcal{L}_{cl}$ is then computed as:

$$\mathcal{L}_{cl} = -\frac{1}{|C|}\sum_{c\in C}\frac{1}{|\text{an}_c|}\sum_{a_i\in \text{an}_c}\log\left\{\frac{\exp(a_i\cdot a_p/\tau)}{\exp(a_i\cdot a_p/\tau)+Z}\right\},$$

$$Z = \sum_{\substack{j\in C\\ j\neq c}}\sum_{a_k\in n_c^j}\exp(a_i\cdot a_k/\tau). \tag{4}$$

Here C is the number of classes, $\text{an}_c \subseteq A_c$ is the current anchor subset, *i.e.* N randomly sampled queries from the anchor set A_c, a_i represents the i^{th} anchor of class c, $n_c \subseteq N_c$ is the current negative set, *i.e.* O randomly sampled keys from N_c (the negative set of class c), $n_c^j \in n_c$ is the subset of negative keys with class j, $j \neq c$, and τ is a temperature constant. To prevent the background class from dominating the learning process, we limit the number of negative samples for each category. It ensures balanced contributions across classes and reduces memory usage, unlike [35] which simply discards background features. Note that in our experiments, $N = 1000$ and $O = 500$. The positive key $a_p = a_p^l$ is given by calculating the average of all other pixels of the same class, *i.e.* in the anchor set A_c:

$$a_p^l = \frac{1}{|A_c|}\sum_{a_i\in A_c} a_i. \tag{5}$$

Contrasting only an average positive instead of all positives is computationally cheaper, yet still allows reducing the average distance between the anchor and other samples of class c [49]. In Section 3.4 we will show how using spatial registration information can provide additional positives for contrastive learning.

Training and inference. The two models are trained simultaneously with separate losses. The total training loss $\mathcal{L}_A$ for model A is:

$$\mathcal{L}_A = \mathcal{L}_{sup(A)} + w_{cps}\mathcal{L}_{cps(A)} + w_{cl}\mathcal{L}_{cl}. \tag{6}$$

and similarly for model B. Here w_* are weighting factors used to balance each loss term. Overall, this setup yields comparable performance to the SOTA contrastive cross-teaching method, MCSC [48], while being significantly simpler, and easier to adapt to use registration information. For inference, we make predictions by averaging the logits from the two models.

3.2. Learning from spatial registration

We now describe how our CCT-R incorporates registration information into the learning framework described in Section 3.1. In CCT-R, spatial correspondences from registration serve as additional supervision, since points mapped together by an accurate registration transform share the same anatomical label across volumes.

A registration transform aligns a source to a target volume, by smoothly mapping source coordinates to anatomically-matching target coordinates. We assume pairwise 3D transforms, affine or deformable, are available between all volumes in V; these can be calculated using any standard off-the-shelf method. Affine transforms apply a 3×4 transform matrix to source coordinates (causing a global rotation, translation, scale and shear), while deformable transforms use a spatially-varying deformation field to achieve precise local alignment. Although the segmentation model remains 2D, operating on individual slices, each slice is now considered within the 3D space of its original volume. We define the set of registration transforms as $T = \{T_{ij}\}_{i=1,j=1}^{N}$, where T_{ij} maps points from volume v_i to v_j, and N is the total number of volumes.

Our CCT-R uses T in two ways. First, we go beyond

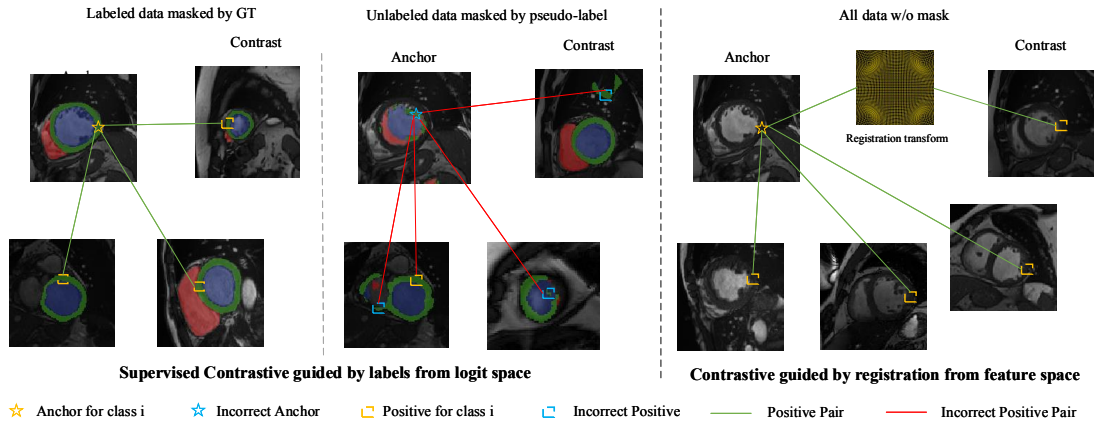


Figure 2. **Supervised contrastive learning guided by labels *vs.* registration**: In the semi-supervised setting, for unlabeled data, the supervised contrastive loss uses pseudo-label information to select pairs. However, pseudo-labels are unreliable, especially early in training. For example, in the middle panel, the anchor is wrongly labeled as Myo (green), which leads to an incorrect learning signal, due to contrasting with positives correctly labeled as Myo. In contrast, registration finds the anatomically-closest point to the anchor in each 3D volume, without relying on label predictions from models, enabling the contrastive loss to perform correct comparisons between cases.

cross-teaching, introducing a new loss that uses registration to transfer labels from labeled to unlabeled data (Sec. 3.3). Furthermore, traditional supervised contrastive learning typically relies on predicted logits, which can introduce errors. Our CCT-R mitigates this by using T to identify anatomically corresponding features across volumes, providing a complementary set of positives (Sec. 3.4).

3.3. Registration supervision loss

We use spatial transforms obtained by registration as an additional source of pseudo-labels to supervise the two models. Specifically, by transforming a point from an unlabeled volume to the corresponding point in a labeled volume, we can assume that these two points correspond to the same anatomical location. Thus, the label from the labeled volume can be used as supervision for the unlabeled slice. This provides much more accurate pseudo-labels early in training, and also helps to reduce the confirmation bias that can arise from cross-teaching.

Formally, we define a new loss $\mathcal{L}_{rs}$, that encourages each pixel to match the label of its corresponding location in the paired labeled volume:

$$\mathcal{L}_{rs} = -\frac{1}{M}\sum_{i=1}^{M}\left(\mathcal{L}_{dice}(p_i^u, r_i^u) + \mathcal{L}_{ce}(p_i^u, r_i^u)\right), \quad (7)$$

where p_i^u is the class probability map of the i^{th} unlabeled image x_i^u, and r_i^u is a new registered label found by registration. $\mathcal{L}_{rs}$ is then added to the overall loss function (Eq. 6).

Assuming that the slice x_i^u belongs to the unlabeled volume v_j^u, we define the registered label r_i^u by mapping the ground truth y_i^l from the labeled volume v_q^l:

$$r_i^u = T_{qj}(y_i^l), \quad (8)$$

where T_{qj} is the transform from v_q^u to v_j^l. This transform aligns the label y_i^l with the corresponding coordinates in the slice x_i^u, resulting in the r_i^u. This greatly improves the model's learning performance (see Sec. 4.4), especially in cases with minimal supervision (*e.g.* one labeled volume).

Best registration selection strategy. In practice, registrations are often imperfect, particularly for complex anatomical regions such as the abdomen. Moreover, the loss in Eq. 7 does not require every image to be paired with all others. We therefore design a strategy to choose which registered pairs should be used. Importantly, this strategy cannot rely on ground-truth labels, due to our semi-supervised setting. Specifically, we measure the cycle-consistency of the transforms from T (Sec. 3.2) between two volumes, say v_j^u and v_q^l. We apply the forward transform T_{jq} (j-to-q) and the reverse transform T_{qj} (q-to-j) on volume v_j^u:

$$\tilde{v}_j^u = T_{qj}(T_{jq}(v_j^u)). \quad (9)$$

Ideally, $\tilde{v}_j^u$ should be equal to the original volume v_j^u, meaning the composition of forward and reverse transformations approximates the identity function. We calculate the global similarity between v_j^u and $\tilde{v}_j^u$ using both mutual information (MI) [58] and root mean square error (RMSE), and use these to derive a composite score

$$S = w_{\text{rmse}} \cdot \text{RMSE} + w_{\text{mi}} \cdot \text{MI}, \quad (10)$$

where w_{rmse} and w_{mi} weight the importance of RMSE and MI, respectively. We then select the v_q^l that minimizes this composite score to generate the best additional pseudo-label r_i^u for the unlabeled slice x_i^u in v_j^u.

3.4. Registration-enhanced positive sampling

We next show how to use registration to improve the supervised contrastive learning loss in Eq. 4. Fig. 2 shows the shortcomings of standard positive sampling in comparison to our novel approach integrating registration. Positives a_p derived from (pseudo-)labels are sampled from any location within the same organ or class as shown in Eq. 4. In

contrast, registration-based positives correspond to the exact same anatomical location within the organ, albeit in different volumes or patients. Any noise in registration-based positives stems from registration inaccuracies and is independent of pseudo-label errors. Therefore, we augment the set of positive samples by incorporating registration-based examples. This approach reduces the confirmation bias that can arise when learning only from pseudo-labels.

Assume the xyz coordinate of anchor a_i in an image from volume v_q is denoted by p. We use a registration transform to get the corresponding positive coordinates p_j in v_j:

$$p_j = T_{qj}(p), \tag{11}$$

where $j \in \{1, 2, \ldots, N\}$ and $j \neq q$, i.e. we consider all other training volumes in V. Given the p_j, we extract the positive feature a^r_{pj} from the corresponding feature maps.

Since our minibatch comprise 2D slices rather than full 3D volumes, there is only a small probability that the feature map containing a given registered point p_j will in fact be available in the current minibatch. We therefore build a memory bank B to serve as a source of feature maps, which provides more diverse registered positive samples across different 3D volumes. The memory bank B stores feature maps of 2D slices. For every slice in each mini-batch, new feature maps are added to B. If a slice is not yet in B, it is added; otherwise, the existing slice is updated with the new features. Once B reaches its maximum capacity K, the oldest slices are removed in a first-in, first-out (FIFO) order. This provides the model with a more diverse set of features from various 3D volumes.

The positive features a^r_{pj} are averaged over the available j indices that exist in the memory bank:

$$a^r_p = \frac{1}{|J|} \sum_{j \in J} a_{pj}, \tag{12}$$

where J represents the set of volume indices for which the feature point exists in the memory bank. Note that J is a subset of the total volume indices $\{1, 2, \ldots, N\}$.

Finally, we combine with the pseudo-label-supervised positive key a^l_p from Eq. 5 to give a single combined positive key a_p for a_i:

$$a_p = w_1 a^l_p + w_2 a^r_p. \tag{13}$$

We use these positives in the contrastive loss Eq. 4, but otherwise keep it unchanged.

4. Experiments

Datasets. We evaluate CCT-R using two challenging benchmark datasets. **ACDC [9]** comprises of 200 short-axis cardiac MR volumes from 100 cases, with segmentation masks provided for the left ventricle (LV), myocardium (Myo), and right ventricle (RV). We allocate 70 cases (1930 slices) for training, 10 for validation, and 20 for testing as in [56], and match their choice of labeled cases. **Synapse [42]** consists of abdominal CT volumes from 30 cases, with eight labeled organs: aorta, gallbladder, spleen, left kidney, right kidney, liver, pancreas, and stomach. As in [14], we use 18 cases (2212 slices) for training and 12 for testing.

Metrics. For quantitative evaluation, we use two widely-recognized metrics for 2D segmentation: Dice coefficient (DSC) and 95% Hausdorff Distance (HD).

Baselines. We first compare with a registration baseline that is not learning-based—we use the transforms to propagate labels from the labeled training cases to the test images, similar to [8, 26, 54], selecting labeled cases with our BRS. We compare a joint registration and segmentation model, DeepAtlas [81]; this learns registration from scratch simultaneously with segmentation. To stay consistent with our CCT-R, we reimplemented it using a 2D U-Net segmentation model. We evaluate several recent S4 methods with the U-Net [66] backbone: MT [70], DCT [65], UAMT [84], ICT [74], CCT [63], CPS [18], and CTS [56]. In addition, we include the SOTA S4 method with contrastive learning, MCSC [48]. As a reference we also train the U-Net backbone from the S4 methods on only the labeled subset of cases (LS) and full supervision (FS), and the SOTA fully-supervised methods BATFormer [46] (on ACDC) and nnFormer [87] (on Synapse). We retrain all baselines using their recommended hyperparameters, and report the best results from [56] or our replication.

Implementation details. Further implementation details for CCT-R and the baselines are given in the supplementary material. We precomputed a composite pairwise registration (affine for ACDC and affine + B-spline deformable for Synapse) for all training data using ITK [55, 59]. The compute time required for each affine registration pair is approximately 2 minutes, while each deformable pair takes around 3 hours. This cost is only incurred once per dataset; it does not need to be repeated for each training run.

4.1. Comparison with Existing Methods

ACDC. Table 1 presents quantitative results from our CCT-R and baselines, under three different levels of supervision (7, 3, and 1 labeled cases). When trained on 7 labeled cases (10%), significantly outperforms the baseline CTS, with more than a 4% improvement in DSC and a reduction of 7 mm in HD. With just 5% of labeled data (3 cases), our CCT-R surpasses CTS and SOTA MCSC by an impressive margin of 20% and 12% in DSC and reduction of 14 mm and 8.5 mm in HD, respectively. When the supervision is reduced to one labeled case, our approach outperforms the SOTA by an even larger margin (DSC of 80.4 vs. 58.6 for MCSC), highlighting its robustness in scenarios with extremely limited labeled data. DeepAtlas, a joint registration and segmentation method, underperforms. This may be due to its lack of advanced S4 techniques, and its online learning of registration, which means registrations are inaccurate early in training and provide poor guidance for segmentation. Qualitative results in Fig. 3 (left) further illustrate the superiority of CCT-R, showing more accurate segmentation with fewer under-segmented regions for the RV (bottom)

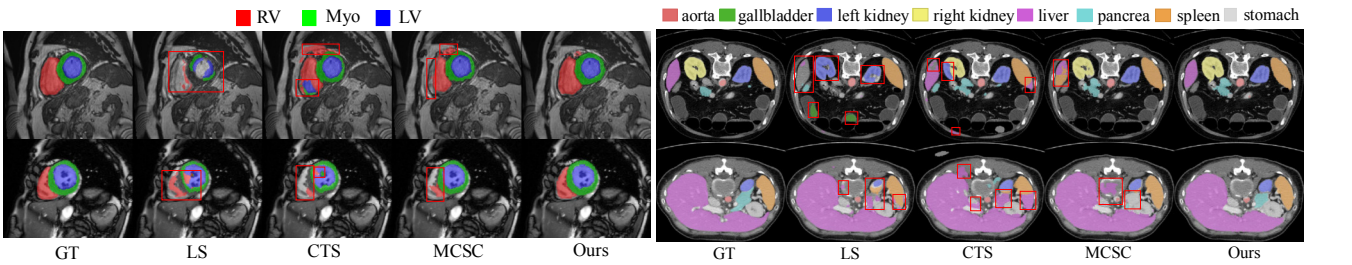


Figure 3. Qualitative results from our CCT-R and baselines. **Left:** ACDC, trained on 3 labeled cases; **right:** Synapse, 2 labeled cases

Table 1. Segmentation results on ACDC for our method and baselines, according to DSC (%) and HD (mm).

Labeled	Methods	Mean		Myo		LV		RV	
		DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓
70 (100%)	UNet-FS	91.7	4.0	89.0	5.0	94.6	5.9	91.4	1.2
	BATFormer [46]	92.8	8.0	90.26	6.8	96.3	5.9	91.97	11.3
7 (10%)	Reg. only (Aff)	30.7	16.4	19.7	13.9	42.0	14.4	30.5	20.8
	DeepAtlas [81]	79.4	8.0	79.0	11.7	81.9	3.2	77.3	9.0
	UNet-LS	75.9	10.8	78.2	8.6	85.5	13.0	63.9	10.7
	MT [70]	80.9	11.5	79.1	7.7	86.1	13.4	77.6	13.3
	DCT [65]	80.4	13.8	79.3	10.7	87.0	15.5	75.0	15.3
	UAMT [84]	81.1	11.2	80.1	13.7	87.1	18.1	77.6	14.7
	ICT [74]	82.4	7.2	81.5	7.8	87.6	10.6	78.2	3.2
	CCT [63]	84.0	6.6	82.3	5.4	88.6	9.4	81.0	5.1
	CPS [18]	85.0	6.6	82.9	6.6	88.0	10.8	84.2	2.3
	CTS [56]	86.4	8.6	84.4	6.9	90.1	11.2	84.8	7.8
	MCSC [48]	89.4	2.3	**87.6**	**1.1**	**93.6**	3.5	87.1	2.1
	Ours (Affine)	**90.3**	**1.6**	87.4	1.4	92.7	**2.2**	**90.9**	**1.3**
3 (5%)	Reg. only (Aff)	32.0	17.8	18.0	15.7	43.9	16.0	34.0	21.7
	DeepAtlas [81]	59.0	8.6	62.8	5.4	67.8	7.7	46.4	12.6
	UNet-LS	51.2	31.2	54.8	24.4	61.8	24.3	37.0	44.4
	MT [70]	56.6	34.5	58.6	23.1	70.9	26.3	40.3	53.9
	DCT [65]	58.2	26.4	61.7	20.3	71.7	27.3	41.3	31.7
	UAMT [84]	61.0	25.8	61.5	19.3	70.7	22.6	50.8	35.4
	ICT [74]	58.1	22.8	62.0	20.4	67.3	24.1	44.8	23.8
	CCT [63]	58.6	27.9	64.7	22.4	70.4	27.1	40.8	34.2
	CPS [18]	60.3	25.5	65.2	18.3	72.0	22.2	43.8	35.8
	CTS [56]	65.6	16.2	62.8	11.5	76.3	15.7	57.7	21.4
	MCSC [48]	73.6	10.5	70.0	8.8	79.2	14.9	71.7	7.8
	Ours (Affine)	**85.7**	**2.0**	**83.8**	**1.4**	**89.9**	**2.4**	**83.5**	**2.1**
1 (1.4%)	Reg. only (Aff)	23.4	19.7	13.6	18.7	31.6	19.0	25.1	21.4
	DeepAtlas [81]	40.4	18.5	42.2	11.7	34.7	29.2	44.4	14.6
	UNet-LS	26.4	60.1	26.3	51.2	28.3	52.0	24.6	77.0
	CTS [56]	46.8	36.3	55.1	5.5	64.8	**4.1**	20.5	99.4
	MCSC [48]	58.6	31.2	64.2	13.3	78.1	12.2	33.5	68.1
	Ours (Affine)	**80.4**	**3.5**	**78.3**	**3.2**	**83.6**	4.3	**79.3**	**2.9**

Best is bold, Second Best is underlined.

and fewer false positives (top) compared to CTS.

Synapse. We evaluate performance on the Synapse using 4, 2, and 1 labeled cases. Although Synapse is more challenging than ACDC due to greater class imbalance and anatomical variability, CCT-R demonstrates even larger improvements (Table 2). With 4 labeled cases, DSC increases from 64.0% to 71.4%, outperforming CTS by 7.4% and MCSC by 2.9%. Even with just one case, CCT-R still excels at segmenting challenging small organs like the aorta, kidney, and pancreas, where others struggle. It significantly outperforms MCSC, improving DSC by 13.6% and reducing HD by 15.4 mm. This robustness to extreme class imbalance and limited supervision emphasizes the value of registration information. Furthermore, our approach is robust across varying registration qualities. Even with simpler affine registrations, inaccurate for complex abdominal anatomy, it significantly improves segmentation (*Ours (Affine)* rows) over not using registration, though results are better still with deformable transforms (*Ours (Affine+Deform)*). Fig. 3 (right) shows CCT-R accurately segments small structures like the gallbladder and pancreas, often missed or over-

Table 2. Segmentation results on Synapse for ours method and baselines, according to DSC (%) and HD (mm).

Labeled	Methods	DSC↑	HD↓	Aorta	Gallb	Kid_L	Kid_R	Liver	Pancr	Spleen	Stom
18(100%)	UNet-FS	75.6	42.3	88.8	56.1	78.9	72.6	91.9	55.8	85.8	74.7
	nnFormer	86.6	10.6	92.0	70.2	86.6	86.3	96.8	83.4	90.5	86.8
4(20%)	Reg. only (Affine)	27.0	39.6	16.0	7.5	36.4	33.0	56.8	13.1	28.5	25.1
	Reg. only (Aff+Def)	32.5	36.5	29.7	4.8	36.5	29.4	65.5	14.2	48.0	31.7
	DeepAtlas [81]	56.1	85.3	69.2	43.3	50.8	55.2	88.8	30.5	62.7	48.0
	UNet-LS	47.2	122.3	67.6	29.7	47.2	50.7	79.1	25.2	56.8	21.5
	UAMT [84]	51.9	69.3	75.3	33.4	55.3	40.8	82.6	27.5	55.9	44.7
	CPS [18]	57.9	62.6	75.6	41.4	60.1	53.0	88.2	26.2	69.6	48.9
	CTS [56]	64.0	56.4	79.9	38.9	66.3	63.5	86.1	41.9	75.3	60.4
	MCSC [48]	68.5	24.8	76.3	**44.4**	**73.4**	**72.3**	91.8	46.9	79.9	62.9
	Ours (Affine)	70.0	23.2	79.8	34.5	71.0	70.7	92.8	**49.6**	87.4	**74.4**
	Ours (Affine+Deform)	**71.4**	**21.1**	**80.4**	42.3	73.0	70.0	**93.7**	49.4	**87.9**	74.2
2(10%)	Reg. only (Affine)	25.4	36.8	17.5	3.5	32.7	27.5	53.4	12.6	33.4	22.5
	Reg. only (Aff+Def)	29.1	44.0	27.2	11.3	28.6	26.5	66.4	12.7	29.7	30.3
	DeepAtlas [81]	44.0	67.1	68.0	24.9	37.9	46.0	82.7	18.4	44.2	30.6
	UNet-LS	45.2	55.6	66.4	27.2	46.0	48.0	82.6	18.2	39.9	33.4
	UAMT [84]	49.5	62.6	71.3	21.1	62.6	51.4	79.3	22.8	58.2	29.0
	CPS [18]	48.8	65.6	70.9	21.3	58.0	45.1	80.7	23.5	58.0	32.7
	CTS [56]	55.2	45.4	71.5	25.6	62.6	67.5	78.2	26.3	75.9	34.3
	MCSC [48]	61.1	32.6	73.9	26.4	69.9	72.7	90.0	33.2	79.4	43.0
	Ours (Affine)	65.1	22.5	75.7	28.4	74.5	**75.0**	91.8	38.0	**82.3**	55.1
	Ours (Affine+Deform)	**66.5**	**19.7**	**77.6**	**34.4**	**75.1**	74.2	**92.6**	**39.5**	82.1	**56.1**
1(5%)	Reg. only (Affine)	26.4	45.0	16.3	6.6	35.8	32.8	53.5	14.4	28.7	22.7
	Reg. only (Aff+Def)	27.4	52.2	26.4	11.3	30.5	27.1	61.6	12.8	26.3	23.6
	DeepAtlas [81]	16.1	72.3	18.4	14.9	1.2	10.1	57.1	0.6	14.4	12.2
	UNet-LS	13.7	116.5	11.6	**17.8**	0.8	1.8	56.9	0.1	8.7	11.6
	UAMT [84]	10.7	90.2	8.0	9.3	0.3	8.1	31.7	1.1	13.1	14.3
	CPS [18]	15.0	123.5	19.6	9.6	5.6	6.9	59.4	2.3	9.4	7.2
	CTS [56]	26.3	96.5	44.6	4.0	11.2	5.5	60.3	9.6	54.1	21.2
	MCSC [48]	34.0	53.8	50.9	13.0	17.6	54.6	64.3	5.5	43.1	23.5
	Ours (Affine)	43.4	40.8	62.5	13.3	17.9	**71.0**	**77.0**	**11.4**	65.4	**28.7**
	Ours (Affine+Deform)	**47.6**	**38.4**	**65.5**	9.3	**50.6**	70.2	72.7	11.1	**73.9**	27.8

Best is bold, Second Best is underlined.

segmented by LS and CTS. Our approach also correctly identifies the spleen and distinguishes it from the liver, a common error in other methods. It also provides more precise segmentation of the liver and stomach, significantly outperforming MCSC. This figure shows the robustness in handling challenging, imbalanced datasets.

Segmentation via registration only. We also test whether simply propagating labels based on either affine or deformable registration achieves adequate segmentation performance (*Reg.only* rows in Tables 1 & 2). We see this performs substantially worse than the learning-based methods.

4.2. Benefit of Our Registration-Based Modules Applied on Different Baselines

Our main experiments build on CTS; however to show the wide applicability of our approach, we measure performance when it is integrated with alternative SSL baselines (Table 3). We include UAMT [84], a classic teacher-student framework with two U-Nets, CPS [18], a student-student framework with two cross-teaching U-Nets, and CTS [56],

Table 3. Benefit of our modules combined with different baselines, on Synapse with 10% labeled data.

	UAMT [84]		CPS [18]		CTS [56]	
	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓
Baselines	49.5	62.6	48.8	65.6	55.2	45.4
+ RSL	52.3	60.3	57.3	42.4	65.4	28.5
+ RSL + SCL + REPS	54.6	55.6	59.1	37.5	66.5	19.7

Table 4. Comparisons with SOTA contrastive learning methods combined with CTS, on ACDC and Synapse.

Contrastive learning method		ACDC 3 (5 %)		/ 1 (1.4 %)		Synapse 4 (20 %)		/ 2 (10%)	
		DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓
Patch-level	GLCL [35] (MICCAI'21)	71.7	3.8	47.4	35.8	67.7	42.6	59.7	34.6
	MCSC [48] (BMVC'23)	73.6	10.5	58.6	31.2	68.5	24.8	61.1	32.6
Slice-level	ReCo [50] (ICLR'22)	70.2	6.1	48.3	33.5	68.3	25.9	60.4	20.7
	Ours	**85.4**	**2.6**	**80.0**	**4.2**	**71.4**	**21.1**	**66.5**	**19.7**
None (Vanilla CTS [56])		65.6	16.2	46.8	36.3	64.0	56.4	57.2	45.7

Best is bold.

Table 5. Ablation study for the primary components of our CCT-R. SCL: typical supervised local contrastive loss. RSL: registration supervision loss. BRS: best registration selection strategy for registered labels r^u. REPS: registration-enhanced positive sampling module (using positives from registration in SCL).

SCL	RSL	BRS	REPS	1 (5%)		2 (10%)	
				DSC↑	HD↓	DSC↑	HD↓
				26.3	96.5	55.2	45.4
	✓			29.0	46.9	64.2	33.9
	✓	✓		–	–	65.4	28.5
✓				27.5	59.8	63.1	29.1
✓	✓	✓		28.1	53.9	64.8	20.6
✓			✓	31.4	55.2	63.9	29.7
✓	✓	✓	✓	47.6	38.4	66.5	19.7

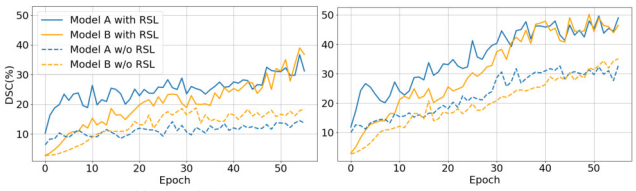
which improves CPS by replacing one of the U-Nets with Swin-UNet. With each baseline, we measure the benefit of adding RSL only, and RSL in conjunction with contrastive learning and registration-based positive selection (*SCL + REPS* row). Our registration-derived modules boost all baselines. Enhanced UAMT approaches CTS performance, while improved CPS surpasses CTS by 4% on DSC. CTS with our modules remains the top performer.

4.3. Comparison with Alternative Supervised Contrastive Learning Losses

In Table 4, we compare our proposed approach with SOTA contrastive S4 method MCSC [48], and with incorporating other recent patch-level and slice-level contrastive learning techniques (GLCL [35] and ReCo [50]) into CTS. While all the contrastive losses improve on vanilla CTS, our CCT-R achieves higher segmentation accuracy on nearly all datasets and labelling rates.

4.4. Ablation Studies and Analysis

We conduct an ablation study on Synapse, measuring the importance of various aspects of our proposed CCT-R (Table 5). CTS, as our baseline, achieves Dice of 26.3% and 55.2% for one and two labeled cases respectively (top row). Our registration supervision loss (RSL) improves the baseline by +2.7% and 9.0%. The best registration selection strategy (BRS), which is only applicable for two



(a) 1 case (5%) (b) 2 cases (10%)

Figure 4. DSC of pseudo-labels from two models on unlabeled data during the early training stages, for Synapse (a) 1 labeled case, and (b) 2 labeled cases.

or more labeled cases, further boosts performance by an additional +1.2% in DSC and reduces HD by -5.4 mm. Adding a standard supervised local contrastive learning (SCL) improves the baseline by +1.2% and 7.9% respectively even without registration; also incorporating RSL gives further improvements of 0.6% and 1.7%, indicating that contrastive learning and RSL are complementary strategies. The registration-enhanced positive sampling (REPS), which mitigates bias towards single pseudo-label supervision in SCL, yields significant improvements: a +3.9% DSC and -4.6 mm HD for one labeled case and +0.8% for two labeled cases versus just SCL. Lastly, when combining all components, our full method achieves substantial Dice score improvement compared to the CTS baseline of 21.3% for 1 labeled case (from 26.3% to 47.6%) and 11.3% for 2 labeled cases (from 55.2% to 66.5%).

Analysing the quality of pseudo-labels. We measured the DSC of pseudo-labels predicted for unlabeled training data and used for cross-teaching, illustrating the noisiness of pseudo-labels and demonstrating how the proposed RSL mitigates this issue. Fig. 4 shows that early in training, cross-teaching models without RSL (dashed lines) yield suboptimal results due to the insufficient training. This limitation persists even in later training stages, as the model struggles to generalize and often converges to local optima, especially in the 5% labeled setting. In contrast, the supervision provided by registrations, RSL, offers consistent and reliable guidance throughout the training process (solid lines), significantly mitigating these issues and enabling more effective learning from limited data.

5. Conclusion

We have introduced CCT-R, a registration-guided method for semi-supervised medical image segmentation. This builds on cross-teaching methods, and improves segmentation via two novel modules: the Registration Supervision Loss and Registration-Enhanced Positive Sampling module. The RSL uses segmentation knowledge derived from transforms between labeled and unlabeled volume pairs, providing an additional source of supervision for the models. With the REPS, supervised contrastive learning can sample anatomically-corresponding positives across volumes. Without introducing extra training parameters, CCT-R achieves the new SOTA on popular S4 benchmarks.

References

[1] Julia Andresen, Timo Kepp, Jan Ehrhardt, Claus von der Burchard, Johann Roider, and Heinz Handels. Deep learning-based simultaneous registration and unsupervised non-correspondence segmentation of medical images with pathologies. *International Journal of Computer Assisted Radiology and Surgery*, 17(4):699–710, 2022. 3

[2] Rajath C Aralikatti, SJ Pawan, and Jeny Rajan. A dual-stage semi-supervised pre-training approach for medical image segmentation. *IEEE Transactions on Artificial Intelligence*, 5(2):556–565, 2023. 1

[3] Brian B Avants, Nick Tustison, Gang Song, et al. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009. 3

[4] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11514–11524, 2023. 3

[5] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019. 1, 2

[6] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19786–19797, June 2023. 3

[7] Laurens Beljaards, Mohamed S Elmahdy, Fons Verbeek, and Marius Staring. A cross-stitch architecture for joint registration and segmentation in adaptive radiotherapy. In *Medical Imaging with Deep Learning*, pages 62–74. PMLR, 2020. 3

[8] Noah C Benson, Omar H Butt, David H Brainard, and Geoffrey K Aguirre. Correction of distortion in flattened representations of the cortical surface allows prediction of v1-v3 functional organization from anatomy. *PLoS computational biology*, 10(3):e1003538, 2014. 3, 6

[9] Olivier Bernard et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE T Med Imaging*, 37(11):2514–2525, 2018. 2, 6

[10] Gerda Bortsova et al. Semi-supervised medical image segmentation via learning consistency under transformations. In *MICCAI*, pages 810–818. Springer, 2019. 2

[11] Krishna Chaitanya et al. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Adv Neur In*, 33:12546–12558, 2020. 3

[12] Krishna Chaitanya et al. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Med Image Anal*, 87:102792, 2023. 1, 3

[13] Chen Chen et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*, 7:25, 2020. 2

[14] Jieneng Chen et al. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 6

[15] Ting Chen et al. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020. 3

[16] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020. 1

[17] Xinlei Chen et al. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 3

[18] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1, 2, 3, 6, 7, 8

[19] Neel Dey, Jo Schlemper, Seyed Sadegh Mohseni Salehi, Bo Zhou, Guido Gerig, and Michal Sofka. Contrareg: Contrastive learning of multi-modality unsupervised deformable image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 66–77. Springer, 2022. 3

[20] Wangbin Ding, Lei Li, Junyi Qiu, Sihan Wang, Liqin Huang, Yinyin Chen, Shan Yang, and Xiahai Zhuang. Aligning multi-sequence cmr towards fully automated myocardial pathology segmentation. *IEEE Transactions on Medical Imaging*, 42(12):3474–3486, 2023. 3

[21] Yuzhen Ding, Hongying Feng, Yunze Yang, Jason Holmes, Zhengliang Liu, David Liu, William W Wong, Nathan Y Yu, Terence T Sio, Steven E Schild, et al. Deep-learning based fast and accurate 3d ct deformable image registration in lung cancer. *Medical physics*, 50(11):6864–6880, 2023. 2

[22] Mohamed S. Elmahdy, Laurens Beljaards, Sahar Yousefi, Hessam Sokooti, Fons Verbeek, Uulke A. Van Der Heide, and Marius Staring. Joint registration and segmentation via multi-task learning for adaptive radiotherapy of prostate cancer. *IEEE Access*, 9:95551–95568, 2021. 3

[23] Koen AJ Eppenhof and Josien PW Pluim. Pulmonary ct registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*, 38(5):1097–1105, 2018. 2

[24] Jiashuo Fan, Bin Gao, Huan Jin, and Lihui Jiang. Ucc: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9947–9956, 2022. 1

[25] Bei Fang, Xian Li, Guangxin Han, and Juhou He. Rethinking pseudo-labeling for semi-supervised facial expression recognition with contrastive self-supervised learning. *IEEE Access*, 11:45547–45558, 2023. 3

[26] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002. 3, 6

[27] Geoff French, Timo Aila, Samuli Laine, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations. In *Proceedings of the IEEE/CVF International Conference on Learning Representations*, 2019.

[28] Jean-Bastien Grill et al. Bootstrap your own latent-a new approach to self-supervised learning. *NIPS*, 33:21271–21284, 2020. 3

[29] Jean-Bastien Grill et al. Bootstrap your own latent-a new approach to self-supervised learning. *NIPS*, 33:21271–21284, 2020. 3

[30] Xiao Gu, Fani Deligianni, Jinpei Han, Xiangyu Liu, Wei Chen, Guang-Zhong Yang, and Benny Lo. Beyond supervised learning for pervasive healthcare. *IEEE Reviews in Biomedical Engineering*, 17:42–62, 2024. 1

[31] Gilion Hautvast, Steven Lobregt, Marcel Breeuwer, and Frans Gerritsen. Automatic contour propagation in cine cardiac magnetic resonance images. *IEEE transactions on medical imaging*, 25(11):1472–1482, 2006. 2

[32] Kaiming He et al. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 3

[33] Kaiming He et al. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 3

[34] Thao Thi Ho, Woo Jin Kim, Chang Hyun Lee, Gong Yong Jin, Kum Ju Chae, and Sanghun Choi. An unsupervised image registration method employing chest computed tomography images and deep neural networks. *Computers in Biology and Medicine*, 154:106612, 2023. 2

[35] Xinrong Hu et al. Semi-supervised contrastive learning for label-efficient medical image segmentation. In *MICCAI*, pages 481–490. Springer, 2021. 1, 3, 4, 8

[36] Yipeng Hu, Marc Modat, Eli Gibson, Wenqi Li, Nooshin Ghavami, Ester Bonmati, Guotai Wang, Steven Bandula, Caroline M Moore, Mark Emberton, et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*, 49:1–13, 2018. 2

[37] Bin Huang, Yufeng Ye, Ziyue Xu, Zongyou Cai, Yan He, Zhangnan Zhong, Lingxiang Liu, Xin Chen, Hanwei Chen, and Bingsheng Huang. 3d lightweight network for simultaneous registration and segmentation of organs-at-risk in ct images of head and neck cancer. *IEEE Transactions on Medical Imaging*, 41(4):951–964, 2022. 3

[38] Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18145–18155, 2023. 3

[39] Prannay Khosla et al. Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *NIPS*, volume 33, pages 18661–18673. Curran Associates, Inc., 2020. 3

[40] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 3

[41] Arno Klein and Joy Hirsch. Mindboggle: a scatterbrained approach to automate brain labeling. *NeuroImage*, 24(2):261–280, 2005. 3

[42] Bennett Landman et al. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *MICCAI*, volume 5, page 12, 2015. 2, 6

[43] Tao Lei et al. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. 2022. 2

[44] Yiwen Li, Yunguan Fu, Iani JMB Gayo, Qianye Yang, Zhe Min, Shaheer U Saeed, Wen Yan, Yipei Wang, J Alison Noble, Mark Emberton, et al. Prototypical few-shot segmentation for cross-institution male pelvic structures with spatial registration. *Medical Image Analysis*, 90:102935, 2023. 3

[45] Huibin Lin, Chun-Yang Zhang, Shiping Wang, and Wenzhong Guo. A probabilistic contrastive framework for semi-supervised learning. *IEEE Transactions on Multimedia*, 25:8767–8779, 2023. 3

[46] Xian Lin et al. Batformer: Towards boundary-aware lightweight transformer for efficient medical image segmentation. *IEEE JBHI*, 2023. 6, 7

[47] Lihao Liu, Angelica I Aviles-Rivero, and Carola-Bibiane Schönlieb. Contrastive registration for unsupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 3

[48] Qianying Liu, Xiao Gu, Paul Henderson, and Fani Deligianni. Multi-scale cross contrastive learning for semi-supervised medical image segmentation. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 1, 2, 3, 4, 6, 7, 8

[49] Shikun Liu et al. Bootstrapping semantic segmentation with regional contrast. In *ICRL*, 2022. 2, 4

[50] Shikun Liu, Shuaifeng Zhi, Edward Johns, and Andrew Davison. Bootstrapping semantic segmentation with regional contrast. In *International Conference on Learning Representations (ICLR)*, 2022. 8

[51] Yang Liu and Shi Gu. Co-learning semantic-aware unsupervised segmentation for pathological image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 537–547. Springer, 2023. 3

[52] Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008. 1

[53] Zijun Long, George Killick, Lipeng Zhuang, Richard McCreadie, Gerardo Aragon Camarasa, and Paul Henderson. Elucidating and overcoming the challenges of label noise in supervised contrastive learning. In *European Conference on Computer Vision*, 2024. 3

[54] Maria Lorenzo-Valdés, Gerardo I Sanchez-Ortiz, Raad Mohiaddin, and Daniel Rueckert. Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I 5*, pages 642–650. Springer, 2002. 3, 6

[55] Bradley C Lowekamp, David T Chen, Luis Ibáñez, and Daniel Blezek. The design of simpleitk. *Frontiers in neuroinformatics*, 7:45, 2013. 6

[56] Xiangde Luo, Minhao Hu, Tao Song, Guotai Wang, and Shaoting Zhang. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In *International Conference on Medical Imaging with Deep Learning*, pages 820–833. PMLR, 2022. 1, 2, 3, 6, 7, 8

[57] J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998. 1, 2

[58] David Mattes, David R Haynor, Hubert Vesselle, Thomas K Lewellen, and William Eubank. Pet-ct image registration in the chest using free-form deformations. *IEEE transactions on medical imaging*, 22(1):120–128, 2003. 5

[59] Matthew McCormick, Xiaoxiao Liu, Julien Jomier, Charles Marion, and Luis Ibanez. Itk: enabling reproducible research and open science. *Frontiers in neuroinformatics*, 8:13, 2014. 6

[60] Seungjong Oh, David Jaffray, and Young-Bin Cho. A novel method to quantify and compare anatomical shape: application in cervix cancer radiotherapy. *Physics in Medicine & Biology*, 59(11):2687, 2014. 2

[61] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, and Lennart Svensson. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1369–1378, 2021. 1

[62] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 3

[63] Yassine Ouali et al. Semi-supervised semantic segmentation with cross-consistency training. In *CVPR*, pages 12674–12684, 2020. 1, 6, 7

[64] Jizong Peng et al. Deep co-training for semi-supervised image segmentation. *Lect Notes Comput Sc*, 107:107269, 2020. 1, 2

[65] Siyuan Qiao et al. Deep co-training for semi-supervised image recognition. In *ECCV*, pages 135–152, 2018. 6, 7

[66] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015. 6

[67] László Ruskó, György Bekes, and Márta Fidrich. Automatic segmentation of the liver from multi-and single-phase contrast-enhanced ct images. *Medical Image Analysis*, 13(6):871–882, 2009. 3

[68] Hessam Sokooti, Bob De Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 232–239. Springer, 2017. 2

[69] Xinrui Song, Hanqing Chao, Xuanang Xu, Hengtao Guo, Sheng Xu, Baris Turkbey, Bradford J Wood, Thomas Sanford, Ge Wang, and Pingkun Yan. Cross-modal attention for multi-modal image registration. *Medical Image Analysis*, 82:102612, 2022. 3

[70] Antti Tarvainen et al. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *NIPS*, 30, 2017. 1, 6, 7

[71] J.P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998. 2

[72] Maria Thor, Jørgen BB Petersen, Lise Bentzen, Morten Høyer, and Ludvig Paul Muren. Deformable image registration for contour propagation from ct to cone-beam ct scans in radiotherapy of prostate cancer. *Acta Oncologica*, 50(6):918–925, 2011. 2

[73] Yuandong Tian et al. Understanding self-supervised learning dynamics without contrastive pairs. In *ICML*, pages 10268–10278. PMLR, 2021. 3

[74] Vikas Verma et al. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 145:90–106, 2022. 6, 7

[75] Paul Viola and William M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997. 2

[76] Kaiping Wang et al. Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. *Med Image Anal*, 79:102447, 2022. 2

[77] Xinlong Wang et al. Dense contrastive learning for self-supervised visual pre-training. In *CVPR*, pages 3024–3033, 2021. 3

[78] Zhiwei Wang, Xiaoyu Zeng, Chongwei Wu, Xu Zhang, Wei Fang, Qiang Li, et al. Styleseg v2: Towards robust one-shot segmentation of brain tissue via optimization-free registration error perception. *arXiv preprint arXiv:2405.03197*, 2024. 3

[79] Huisi Wu et al. Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In *CVPR*, pages 11666–11675, 2022. 1, 3

[80] Zhenda Xie et al. Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In *CVPR*, pages 16684–16693, 2021. 3

[81] Zhenlin Xu and Marc 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*, pages 420–429. Springer, 2019. 3, 6, 7

[82] Tokihiro Yamamoto, Sven Kabus, Jens Von Berg, Cristian Lorenz, and Paul J Keall. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 79(1):279–288, 2011. 2

[83] Fan Yang et al. Class-aware contrastive semi-supervised learning. In *CVPR*, pages 14421–14430, 2022. 3

[84] Lequan Yu et al. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *MICCAI*, pages 605–613. Springer, 2019. 2, 6, 7, 8

[85] Xiangyu Zhao et al. Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *arXiv preprint arXiv:2301.05500*, 2023. 1, 3

[86] Yuanyi Zhong et al. Pixel contrastive-consistent semi-supervised semantic segmentation. In *CVPR*, pages 7273–7282, 2021. 3

[87] Hong-Yu Zhou et al. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*, 2021. 6