Intelligent Medicine 5 (2025) 209–221

Contents lists available at ScienceDirect

Intelligent Medicine

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

Research Article

Source independent multiple-domain adaptation for knee osteoarthritis cartilage and meniscus segmentation in clinical magnetic resonance imaging

Sheheryar Khan[a,#], Siyue Li[b,#], Fan Xiao[b], Kevin Ho[c], Michael Ong[c], James Griffith[b], Weitian Chen[b,*]

[a] Division of Science, Engineering and Health Studies (SEHS), School of Professional Education and Executive Development, The Hong Kong Polytechnic University, Kowloon, Hong Kong 999077, China
[b] Department of Imaging and Interventional Radiology, Chinese University of Hong Kong, Lab of AI in radiology, Hong Kong 999077, China
[c] Department of Orthopedics and Traumatology, Chinese University of Hong Kong, Hong Kong 999077, China

ARTICLE INFO

Keywords:
Deep learning
Knee osteoarthritis
Generative network
Cartilage and meniscus segmentation

ABSTRACT

Background Generalized knee tissue segmentation, such as cartilage and meniscus in magnetic resonance imaging (MRI), plays a vital role in the clinical assessment of knee osteoarthritis (OA). However, domain variability between MRI datasets poses a significant challenge for the application of robust segmentation methods in real-world clinical settings. Existing unsupervised domain adaptation (UDA) approaches, which rely on one-to-one assumptions between the source and target domains, often fail to preserve knee tissues such as cartilage and meniscus, which are critical for OA diagnosis in diverse clinical settings.

Methods We propose a source-independent segmentation approach tailored for multi-domain knee MRI datasets. Our method emphasizes knee tissue regions to reduce domain gaps and label inconsistencies. By introducing a stepwise adaptation strategy, segmentation performance was refined progressively from intermediate domains to the final target domain. Pseudo-label attention mechanisms were integrated into the adaptation pipeline, enabling iterative fine-tuning of domain-specific segmentations while leveraging unidirectional generative adversarial networks to enhance tissue-specific adaptation. This iterative training process ensures the generation of reliable pseudo-labels, thereby improving segmentation accuracy in diverse clinical MRI datasets.

Results We demonstrated the effectiveness of our approach on the OA initiative dataset as the source domain and self-collected, T1-weighted fast field echo (T1FFE) as the intermediate domain and three-dimensional fast spin echo (3D FSE) as the final target domain. Our method achieved an average dice scores of 0.8701 and 0.7990 for source and target domains, respectively, surpassing the typical UDA methods explored in our experiments.

Conclusion The experiments conducted on clinical MRI data, spanning OA severity from healthy knees to KL Grades 1–4, validated the effectiveness of the proposed domain adaptation method in precise segmentation of the cartilage and meniscus.

1. Introduction

Knee osteoarthritis (OA) is a prevalent degenerative musculoskeletal disease that affects people worldwide and can eventually lead to walking disabilities [1]. Magnetic resonance imaging (MRI) offers superior soft tissue contrast in comparison to alternative imaging methods and can provide morphology and composition measurement for the assessment of OA. It is important to segment the knee tissues accurately to carry out these analyses.

Magnetic resonance images acquired using different hardware, software, and protocols can have numerous signal distributions, which presents a challenge known as ’domain shift’ in medical image analysis [2]. Despite the numerous deep learning (DL) methods proposed in the literature, they have primarily been trained and tested on images sourced from a single clinical center, which typically employs similar imaging protocols and hardware. Consequently, these models tend to show a notable decrease in their performance when tested on new and unfamiliar data owing to the domain shift problem. The problem also

---

[*] Corresponding author: Weitian Chen, Department of Imaging and Interventional Radiology, Chinese University of Hong Kong, Lab of AI in radiology, Hong Kong 999077, China.

E-mail address: wtchen@cuhk.edu.hk (Weitian Chen).

[#] These authors contributed equally to this work.

https://doi.org/10.1016/j.imed.2024.12.002
Received 2 November 2024; Received in revised form 12 December 2024; Accepted 17 December 2024

often arises due to pathological instances that are not part of the original training data. These data samples, including cases with OA progression or severity levels ranging from mild to high, are under-represented in the training data during deployment, which affects the ability of the model to be applied for general use.

Considering the diversity in MRI sequences, clinical studies have reported improved diagnostic values in assessing knee pathologies using three-dimensional fast spin echo (3D FSE) compared with 2D FSE [3,4]. Most MRI vendors now offer 3D FSE MRI to users, making it highly desirable to develop automatic segmentation of knee tissues on 3D FSE in clinical settings [5,6].

Generalization methods on different MRI sequences incorporate domain adaptation, which strive to learn domain adaptation characteristics to meet novel testing predictions [2,7–9].

DL networks trained in a supervised manner can considerably improve the segmentation performance of knee joint tissues [10–13]; however, this requires a large amount of manually annotated data.

A few studies have reported on the use of unsupervised domain adaptation (UDA) for knee joint segmentation. The segmentation network was incorporated into an adversarial network to achieve knee bone and cartilage segmentation on unannotated 2D FSE MR images but did not report on meniscus segmentation [6]. UDA was implemented via cross-domain alignment to improve cartilage segmentation of independent test datasets acquired from different scanners; however, the same pulse sequence in cross domains was investigated [14]. The study emphasized using the 3D DESS pulse sequence for training and presented the UDA results for cartilage segmentation using UDA, where the training and test sequences belong to the same pulse sequence. A source-free UDA method was proposed for knee joint segmentation when the source dataset was unavailable owing to concerns such as patient privacy [15]. UDA was used for the segmentation of knee tissues and applications in automatic OA phenotype grading [16]. These previous approaches to knee MRI segmentation presented UDA for tissues that contain primary cartilage and mainly focused on labeled data (3D DESS) to directly translate to the target domain. These approaches impose one-to-one assumptions while performing the domain adaptation tasks, which ignore the complementary information in the context of distinct target domains when applied to clinical scenarios.

UDA segmentation methods often assume availability to source and target domains and focus on a single source and target domain, aligning feature distributions or preserving anatomical structure in translations. These approaches implicitly assume that label distributions between the domains are similar, which may not always hold, particularly when dealing with complex multi-domain adaptations. An example is in knee MRI, where the segmentation of the cartilages and meniscus are difficult to capture in patients with OA owing to their thin nature, meniscus loss/tear, and low boundary contrast; therefore, it is challenging to set up a direct source to target translation and obtain an adequate tissue alignment.

In contrast to the aforementioned UDA methods, we incorporated source-independent multiple target domains in a joint iterative training procedure to effectively capture the domain-invariant representation. Our approach comprises multiple domains, including the source domain (3D DESS), an intermediate target domain (3D T1-weighted fast field echo (T1FFE)) and a final target domain (3D FSE). Our method does not rely on a direct translation from the source domain to the target domain. Instead, we set the source domain for initial training and a preliminary segmentation step for other domains, producing weak pseudo labels for the intermediate target domains, without requiring any labels from the intermediate and final target domains in training. In the subsequent stage, we introduced pseudo-label attention along with iterative adversarial learning between the intermediate and final target domains. We then used the weak labels to retrain the network and iterate the process to improve the prediction of weak labels and final adaptation in the final target domain.

2. Methods

We introduced multi-domain adaptation to efficiently and simultaneously learn the segmentation model. Given a source domain D_S with annotations L_S, our goal was to train $SegD_S$ in the presence of multiple domains D_X and D_Y by introducing supplementary guidance, without incorporating the ground truths of D_X and D_Y, where D_X is termed the intermediate domain and D_Y is termed the target domain. Notably, the source and target domains are acquired from different pulse sequences and vendors.

Figure 1 illustrates the pipeline of our proposed method for cartilage and meniscus segmentation, which consists of 3 main stages: (a) initial segmentation task; (b) multiple-domain learning; and (c) iterative training.

Initially, we trained a segmentation network $SegD_S$ based on the modified U-Net using an annotated dataset L_S in the source domain. The segmentation network could produce relatively accurate segmentation of knee tissues in the source domain D_S owing to supervised learning tasks. However, its performance was suboptimal when applied to the final target domain D_Y, where the imaging sequence differs. In our study, we used T1-weighted fast field echo (T1FFE), also known as spoiled gradient recalled echo (SPGR), as the intermediate target domain D_X and 3D FSE (VISTA, Philips Healthcare) as the final target domain D_Y. In the second stage, we built an adversarial network based on CycleGAN [17] and pseudo-label attention guidance to train a T1FFE-to-3D FSE transformation model D_{XY}, aiming to obtain a T1FFE-to-3D FSE transformation such that the local attention focuses on the targeted tissue regions instead of transferring the style of the whole image. In the third stage, the transformation model D_{XY} was used to retrain the segmentation model such that the segmentation prediction of T1FFE (L_X) can be treated as pseudo labels for 3D FSE images D_Y. The new composite pairings of actual predictions from T1FEE and pseudo-3D FSE images (L_X D_{XY}) are fed back iteratively to fine-tune the segmentation network. The iterative process involves using the T1FFE-to-3D FSE model and training the segmentation model in the loop. This process enhances the self-supervision of the network without the need for label information in either of the domains. Further details of each stage including the training methodology are provided in the later sections.

2.1. Segmentation network

In our proposed method, the segmentation network performs two key tasks: guiding domain transformation in adversarial learning through pseudo-label attention based on the available source domain and providing final segmentation through iterative learning. We have developed our semantic segmentation framework based on a modified U-Net architecture, illustrated in Figure 2. The network comprises a symmetrical encoder and decoder architecture that uses compound scaling from efficient net to capture contextual information [18]. Depth-wise connections were used to aggregate context information inside the network, revealing complex hierarchical information from each of the MR images. The later section of the network comprises a decoder and is responsible for extracting information maps and reconstructing the annotated image following the coarse-to-fine process. Figure 2 depicts the block configuration of our network, which introduces compound scaling to capture rich information maps from MR images. The network comprises mainly of MobileNet convolutional (MBConv) blocks [18], which effectively incorporate the squeeze-and-excitation (SE) optimization [19]. SE blocks collectively enhance the semantic segmentation of the network by establishing a direct connection between the initial and final stages of convolutional blocks. In the beginning, the depth of the features was improved by expanding activation maps whereas depth-wise convolution was used to reduce the channels of the last feature maps. This increase in depth and stacked convolutional blocks improves the outcome by computing discriminant feature maps. The sub-block configuration was also presented in Figure 2, where the number of blocks and filter sizes were

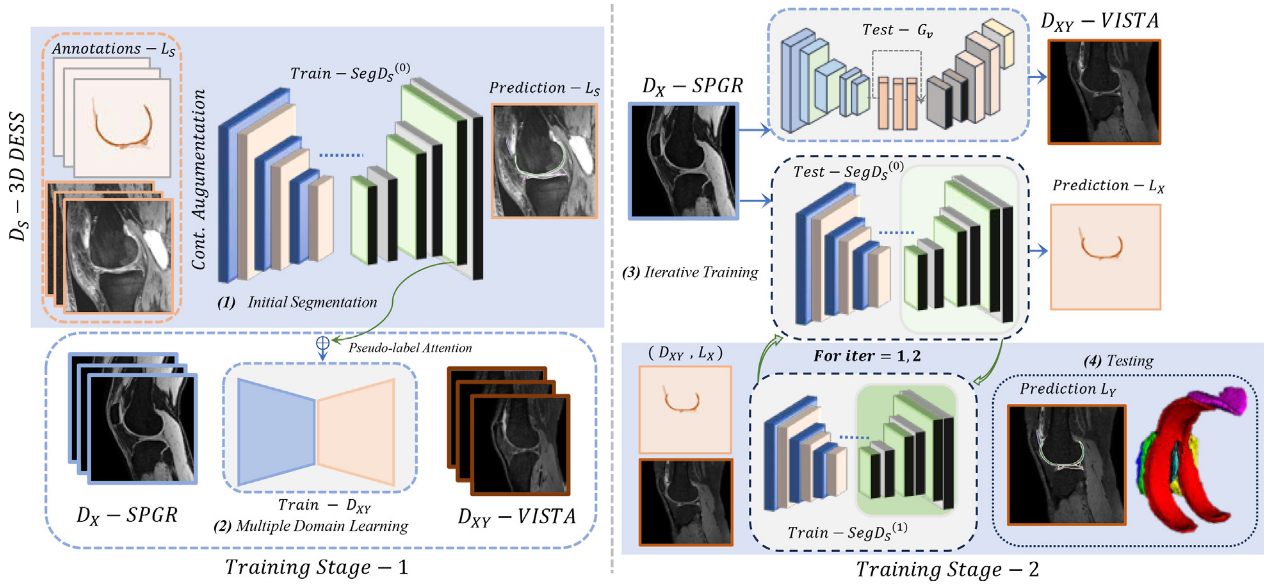

Figure 1. Block diagram of the initial segmentation module with domain adaptation and self-learning (1) Segmentation network with public dataset training. (2) Iterative adversarial learning for domain adaptation while considering new dataset without labels. (3) Iterative training with self-supervised labels on new domain. (4) Testing segmentation.

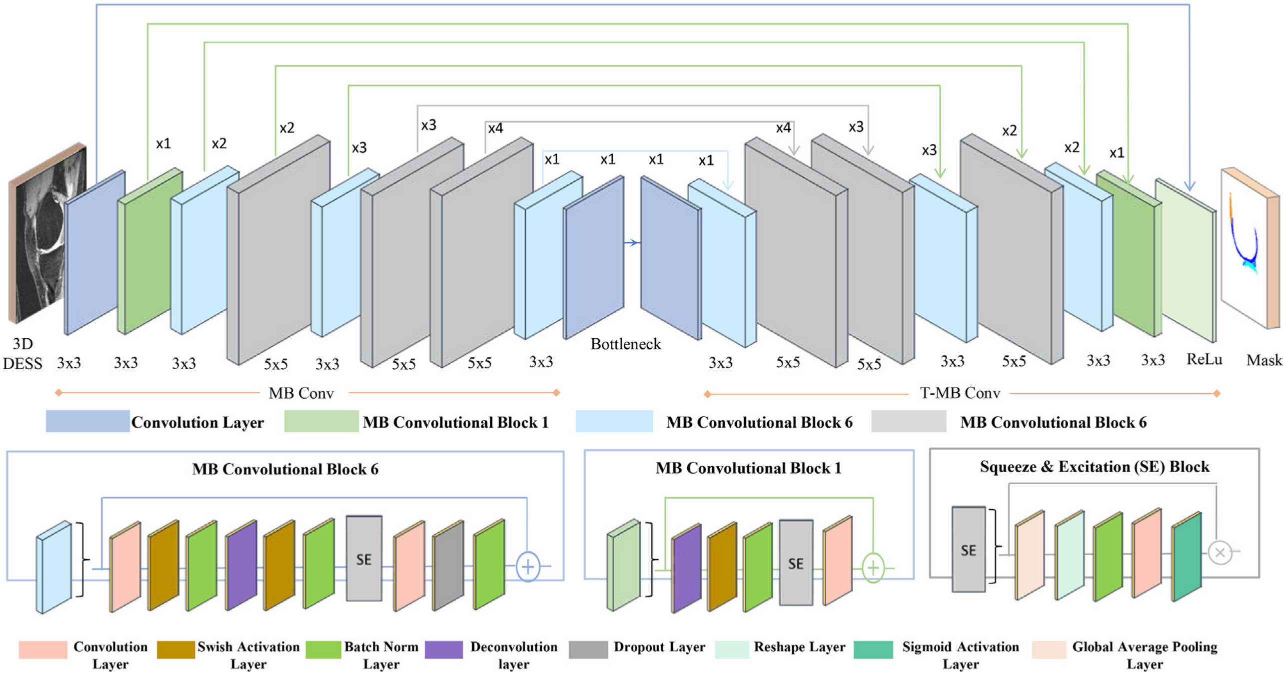

Figure 2. Block diagram of the segmentation task network for generating pseudo-label attention and final segmentation, implemented within the proposed framework. At each phase of the initial training of the encoder and decoder, the 3D DESS MRI input undergoes progressive convolution and is subsequently scaled down by a factor of 2 as indicated. The color scheme represents different blocks. The sub-block arrangements of each module along with the MobileNet (MB) segmentation blocks MBConv. Blocks 6, 1, and SE are also presented.

specified at the top of each block. MBConv and SE blocks are separately presented to show their respective convolutional layers. Finally, *softmax* activation was applied to the final layer to produce the probability map corresponding to each tissue class in knee MR images. Integrating SE blocks into a U-net-based architecture provides a means to leverage the strengths of both techniques, improved representation learning, and enhanced context awareness for precise tissue segmentation. These modifications are crucial for achieving accurate initial segmentation and preparing the composite image and pseudo-labels required for our domain adaptation method. Considering the class distribution discrepancy between the cartilage and meniscus volumes and rest of the MR volumes, our approach incorporates dice loss and weighted cross-entropy loss, given as follows

$$WCE(t,p) = \frac{1}{N(X)} \sum_{x \in X} \left[(t(x) \log(p(x)) w(t(x)) \right] \tag{1}$$

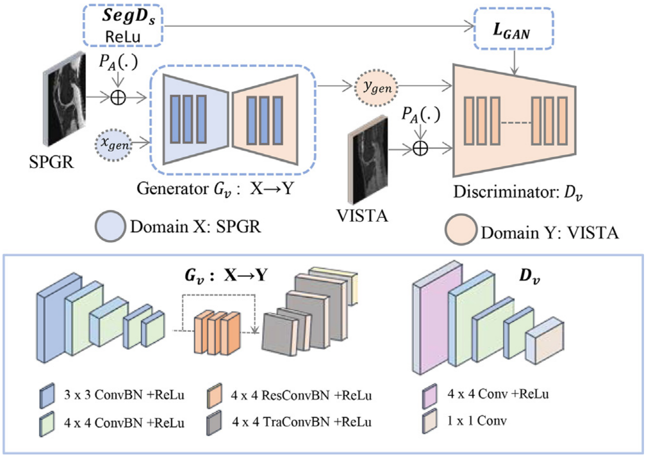

Figure 3. The iterative domain mapping with pseudo-label attention along with T1FFE (SPGR) and VISTA (3D FSE) inputs is shown for training and obtaining the T1FFE to 3DFSE pseudo representation. the block configuration of the generator and discriminator are shown below

$$DL(t,p) = 1 - \frac{2t(x)p(x)+1}{t(x)+p(x)+1} \tag{2}$$

where t(x) and p(x) represent the true value and predicted outcome values after the *softmax* activation function, and w(t(x)) shows the predefined weight from the class representation.

2.2. Self-supervised domain adaptation (SSDA)

Our self-supervised training framework benefits from a domain transformation from T1FFE to 3D FSE to generate pseudo-3D FSE images while introducing pseudo-label attention and iterative learning. The proposed pseudo-label attention consists of two steps, the first step involves the training of $SegD_S$ on source domain D_S, and in the next step, pseudo-label attention was used from the $SegD_S$ layer to guide the generative model and perform the transformation between intermediate and target domains. The aim was to train the (G_v) in the presence of pseudo-label attention and acquire a mapping among tissue regions rather than the whole MR image including the surrounding areas. After the training (G_v), the intermediate domain mapping in target domain images along with their previously acquired weak labels are passed to the segmentation network for fine-tuning.

2.2.1. Pseudo-label attention

Using an iterative training process, it is possible to improve the weak T1FFE labels and achieve superior segmentation results in later stages. To achieve this, we aimed to explore the unidirectional GAN-enabled framework coupled with pseudo-label attention, to produce the initial T1FFE to 3D FSE representation. The generator and discriminator arrangements for T1FFE and 3D FSE sequences are illustrated in Figure 3. A detailed description of the data can be found in later sections. We assumed that successful translation from T1FFE contrast to the target 3D FSE contrast, under tissue region guidance through the pseudo-label maps, will allow the precise prediction of T1FFE images to train the segmentation network on 3D FSE images in a self-supervised learning manner.

Using our transferring model, we aimed to learn the mapping function between the T1FFE and 3D FSE domains given the two training datasets, 3D FSE, and T1FFE. The mapping function is based on the generator network derived from the CycleGan [17]. Mathematically, this can be expressed as $G_{TV}: X \to Y$ and $G_{VT}: Y \to X$. This architecture comprises four neural networks: two encoder-decoder generators integrated with two discriminators. The first generator (G_v) is a transformation network that learns to map T1FFE images to their corresponding 3D FSE images, supervised by an adversarial loss. The first discriminator (D_v) distinguishes between generated and real 3D FSE images, enhancing its ability to judge. Conversely, the second generator (G_t) learns to map 3D FSE images to their corresponding T1FFE images. The trained $SegD_S$ network is used to generate the pseudo-label-based semantic information from the intermediate and target domains. The relu layer corresponding to the segmentation task module extracts the semantic pseudo-label attention and, is denoted as $P_A(.)$, such that for a given intermediate domain MRI D_X, the resulting pseudo-label attention map is $D_{XA} = P_A(D_X)$. The attention map is further separated into foreground and background maps using element-wise multiplication to achieve the mapped target domain image as:

$$X_A = D_{XA} \odot D_{XY} + (1 - D_{XA}) \odot D_X \tag{3}$$

2.2.2. Pseudo-label attention optimization objective

A cycle consistency loss is introduced to ensure that G_v learns the representative features of T1FFE images and produces images that are consistent with the information in 3D FSE and verse versa. This loss measures the difference between the real and synthesized images from the generator network. The second discriminator (D_t) uses a T1FFE image as an input and distinguishes whether the input is real or generated, using the same approach as D_v. We formally defined the loss functions used to train our networks, which receive data from two distributions: T1FFE images (X) and 3D FSE images (Y). The adversarial loss is used to encourage generators to produce more realistic images based on the performance of discriminators, and it supervises both GANs in the same manner. The generator aimed to minimize the loss function against the discriminator that aimed to maximize it:

$$\min_{G_v} \max_{D_v} L_{GAN}\left(G_v, D_v, X_A, Y_A\right) \tag{4}$$

The pseudo-label attention adversarial loss function guiding the discriminators and mapping T1FFE images to 3D FSE images is defined as follows:

$$\begin{aligned} L_{GAN}\left(G_v, D_v, X, Y\right) = {} & E_{y \to P(y)}\left[log D_v\left(D_{XA} \odot y\right)\right] \\ & + E_{x \longrightarrow P(x)}\left[\log\left(1 - D_v\left(G_v\left(D_{XA} \odot x\right)\right)\right)\right] \end{aligned} \tag{5}$$

Similarly, the adversarial loss of the reverse mapping can be formatted as follows:

$$\begin{aligned} L_{GAN}\left(G_t, D_t, Y, X\right) = {} & E_{x \to P(x)}\left[log D_t\left(D_{YA} \odot x\right)\right] \\ & + E_{y \longrightarrow P(y)}\left[\log\left(1 - D_t\left(G_t\left(D_{YA} \odot y\right)\right)\right)\right] \end{aligned} \tag{6}$$

The cycle consistency loss is used to ensure that the inverse transformation returns the same image as the input of the forwarding transformation. It is formulated as follows:

$$\begin{aligned} L_{cycle}\left(G_v, G_t\right) = {} & E_{x \to P(x)}\left[\left\|G_t\left(G_v(x)\right) - x\right\|_N\right] \\ & + E_{y \to P(y)}\left[\left\|G_v\left(G_t(y)\right) - y\right\|_N\right] \end{aligned} \tag{7}$$

where $x \to P(x)$ denotes the distribution of image x in domain X, and$\|.\|$ denotes any norm (e.g., l_1 or l_2). The total objective function combines all the aforementioned losses:

$$\begin{aligned} L_{total}\left(G_v, D_v, G_t, D_t\right) = {} & \lambda_1\ L_{GAN}\left(G_v, D_v, X, Y\right) \\ & + \lambda_2\ L_{GAN}\left(G_t, D_t, Y, X\right) \\ & + \lambda_{cycle} L_{cycle}\left(G_v, G_t\right) \end{aligned} \tag{8}$$

where λ_1 , λ_2 and λ_{cycle} are weight factors for the adversarial loss and cycle consistency loss, respectively, to balance the quality and cycle consistency of the synthesized image.

2.3. Iterative learning

The overall training consists of two stages. During the first stage of training, a 3D DESS-trained model and the domain mapping model trained between the intermediate and target domains were prepared. By incorporating multiple domains within domain adaptation frameworks, it is feasible to simultaneously train the segmentation network on the

Table 1 Training strategy

```
Require: Image Domains: $D_S$, $D_X$, and $D_Y$ and annotation labels $L_S$
Training Stage-1:
while not convergent do
    Augmentation ()
    Train $SegD_S$ on $L_S$ with Eq. (1)
    Train $G_v, D_v, G_t, D_t$ using $D_X$, and $D_Y$ with Eq. (6)
end while
Training Stage-2:
for iter = 1,2 do
    use trained $G_v$ to obtain $D_{XY}$
    use $D_X$ on trained $SegD_S$ to obtain $L_X$
    Retrain $SegD_S$ with $D_{XY}$ and $L_X$
end for
Output: $SegD_S$ $(D_Y)$, $L_Y$
```

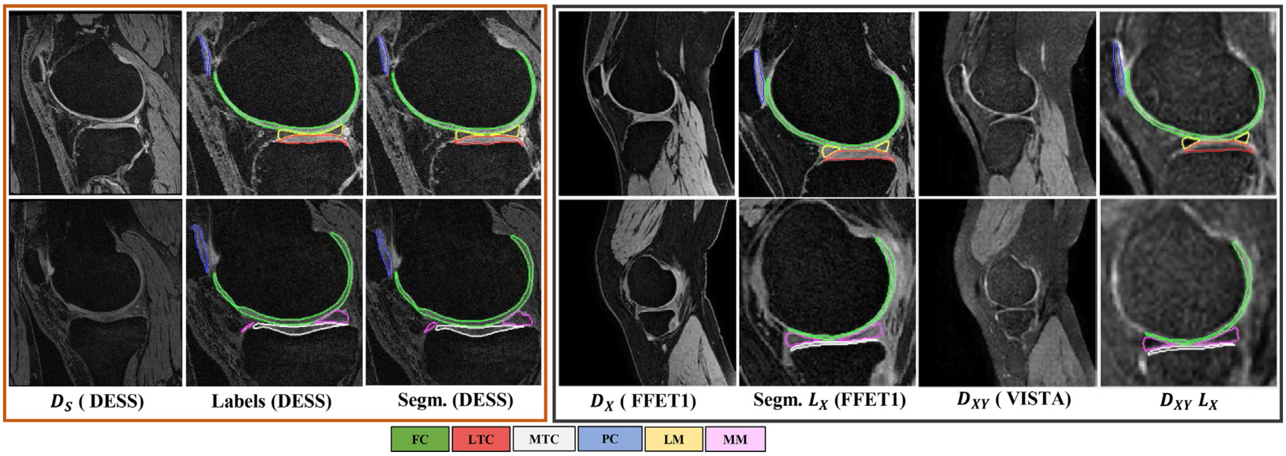

Figure 4. (Column 1-3) Example images from Dataset A and their corresponding segmentation (Segm.) results with ground truths. The tissue regions are color-coded and cropped for better visualization. (column 4-5) Example images from Dataset B showing the T1FFE images and their corresponding predicted segmentation masks, respectively. Column 6 shows the corresponding generated 3D FSE contrast. In the last coulmn, the predicted labels from T1FFE are overlayed on generated 3D FSE. Again, tissue regions are cropped to favorable regions for better visualization.

source domain (D_S) and adversarial network on multiple additional domains (D_S and D_Y). This results in the trained segmentation model, $SegD_S$, and the domain mapping generator, G_v. To train G_v with multiple domains, a limited number of subjects were used. In our study, we trained three subjects from D_X and D_Y to obtain G_v, which effectively generates satisfactory domain mappings and reduces the overall training time required. The second stage includes acquiring a pseudo target domain D_{XY} from the mapping model G_v and pseudo-labels L_X from $SegD_S$ (D_X). The composite pair $(D_{XY},\ L_X)$ resulting from both networks are fed back to the $SegD_S$. Specifically, the targeted composite pair is augmented by the contrast-selective schemes during the fine-tuning stage to obtain probability $p(x)$. Thus, the self-supervised loss L_{SS} is given by:

$$L_{SS} = WCE(L_X, p(x)) + DL(L_X, p(x)) \tag{9}$$

To improve the generalization of the segmentation network and adapt to new changes, the second stage was divided into iterations. In each iteration, the composite pairs $(D_{XY},\ L_X)$ are used for fine-tuning, allowing the segmentation network to benefit from the predictions of the previous stage. At each iteration, the composite pairing $(D_{XY},\ L_X)$, especially the pseudo-label L_X was improved, therefore, enhancing the $SegD_S$. This iterative approach enhances the overall performance and adaptability of the segmentation network in the presence of domain shifts and variations. The iterative training strategy is summarized in Table 1.

Figure 4 is included to illustrate the process at each stage. The network's predictions on 3D DESS MR slices and corresponding segmentation masks obtained from initial learning are shown in Figure 4 (first 3 columns). The network exhibits promising segmentation results in delineating tissue boundaries despite the inter-subject variability. The separately collected intermediate domain, T1FFE images (D_X FFET1), and initial segmentation outcomes of T1FEE (Segm. L_X FFET1) are shown in Figure 4. Pseudo images of the target domain are shown with D_{XY} VISTA, whereas the pseudo retraining labels are displayed with D_{XY}, L_X (segmentation overlayed). It can be observed that the weak labels of the meniscus and cartilage showed an encouraging alignment with the generated 3D VISTA samples. The process is further iterated to improve the weak labels and enhance the final segmentation of 3D VISTA.

2.4. Experiments

This section discusses the experiments conducted on the proposed framework. For cross-domain segmentation, we first established the modified U-Net (EF-Unet)-based semantic segmentation task module within a supervised learning framework. The suitability and performance of the EF-Unet were tested using separate experiments. In addition, we evaluated the performance of various CNN architectures, including the MPUnet, which uses a multiplanar sampling protocol [20], 3D U-Net [21], and PSP-net [22] to establish the preliminary segmentation network for generating pseudo-label attention and carrying out the domain adaptation.

To demonstrate the effectiveness of the proposed multi-domain adaptation approach, we first conducted experiments involving the direct mapping of single source domains to the target domain, known

Table 2 Evaluating initial segmentation results (DSC) on training and testing source domain 3D DESS images without domain adaptation on different tissue compartments

Method	Train/Test (3D)	Augmentation schemes	Loss	Cartilage				Meniscus		Average
				FC	LTC	MTC	PC	LM	MM	
EF-Unet (ours)	DESS	Mix	Dice	0.88 (±0.04)	0.87 (±0.04)	0.86 (±0.04)	0.85 (±0.04)	0.86 (±0.04)	0.84 (±0.04)	0.86
	DESS	Cont.S	Dice	0.88 (±0.04)	0.87 (±0.04)	0.85 (±0.04)	0.85 (±0.04)	0.85 (±0.04)	0.85 (±0.04)	0.86
	DESS	Cont.S	Dice+CE	0.89 (±0.02)	0.87 (±0.03)	0.87 (±0.01)	0.86 (±0.07)	0.86 (±0.04)	0.86 (±0.03)	0.87
PSP-net	DESS	Cont.S	Dice+CE	0.85 (±0.03)	0.84 (±0.04)	0.84 (±0.07)	0.81 (±0.19)	0.80(±0.03)	0.81 (±0.05)	0.83
3D U-net	DESS	Cont.S	Dice+CE	0.86 (±0.04)	0.87 (±0.04)	0.87 (±0.05)	0.81 (±0.12)	0.83 (±0.02)	0.82 (±0.06)	0.84
MPUnet	DESS	Cont.S	Dice+CE	0.87 (±0.02)	0.87 (±0.03)	0.85 (±0.03)	0.85 (±0.10)	0.85 (±0.03)	0.86 (±0.04)	0.86

FC: femoral cartilage; MM: medial meniscus; MTC: medial tibial cartilage; LM: lateral meniscus; LTC: lateral tibial cartilage; PC: patellar cartilage; Cont.S: contrast-selective.

as USDA (unsupervised single sources to target domain adaptation). Additionally, to highlight the adaptation capabilities in segmentation through iterative learning, we present the results and analysis of 2 iterations of the proposed methods: Self-supervised domain adaptation-I (SSDA-I) and the subsequent iteration, SSDA-II. These iterations aimed to refine and improve the segmentation performance by leveraging the knowledge gained from previous iterations. By showcasing the outcomes of USDA and SSDA-I/SSDA-II, we provided a comprehensive understanding of the impact and effectiveness of the proposed multi-domain adaptation approach in the context of segmentation tasks.

Moreover, we conducted a comprehensive evaluation of the proposed method by conducting comparative analyses with other state-of-the-art domain adaptation methods in knee MRI and medical image segmentation, including hybrid dual mean-teacher (HDMT) network [23], barely supervised calibrating label distribution knee MRI segmentation method (calibrating label distribution, CLD) [24], multi-modality unsupervised domain adaptation (MMUDA) [25], and mixup and adversarial UDA in knee MRI segmentation (mixupUDA) [14].

2.4.1. *Data acquisition and processing*

Our study used three distinct sagittal 3D knee MR image datasets. These datasets were acquired using different MRI pulse sequences. Seven classes of segmentation labels were created, including background, femoral cartilage (FC), medial meniscus (MM), medial tibial cartilage (MTC), lateral meniscus (LM), lateral tibial cartilage (LTC), and patellar cartilage (PC).

Dataset A: This dataset included 88 patients and can be downloaded from the public Osteoarthritis Initiative database, that can be accessed at http://www.oai.ucsf.edu. The images were acquired using a 3T Siemens MAGNETOM Trio scanner with a sagittal DESS water excitation imaging sequence. The imaging parameters used were repetition time (TR): 16.3 ms, echo time (TE): 4.7 ms, flip angle: 25°, voxel size: 0.37 ×0.37 × 0.7 mm, matrix size: 384 × 384, field of view (FOV): 140 mm, and 160 image slices. The manual segmentation was performed by a trained musculoskeletal radiologist following the Imorphics cartilage segmentation protocol, ensuring an intra-observer coefficient of variation below 3% on paired test images, and was reviewed by an expert segmenter [26]

Dataset B: This dataset was acquired in house using a 3T Philips Achieva scanner with an 8-channel knee coil and a sagittal T1 fast field echo (T1FFE) pulse sequence. The dataset consists of 3D volumes of knee MR images from 22 patients with the following parameters: pixel size: 0.5 mm × 0.5 mm, matrix size: 320 × 320, FOV: 100 mm, TR: 20 ms, TE: 5 ms, flip angle: 25 ° and slice thickness: 1 mm. No annotations are available for the tissues.

Dataset C: This dataset was acquired in house using the 3D FSE (VISTA) pulse sequence on a 3.0T Philips Achieva scanner. An 8-channel knee coil was used as the receiver. This dataset consists of 3D volumes of knee MR images from 22 patients. The imaging parameters were as follows: TR: 900 ms, TE: 35.5 ms, matrix size: 224 × 224, slice number: 150, resolution: 0.71 mm × 0.71 mm and slice thickness: 0.8 mm. Manual tissue segmentation was performed by a musculoskeletal radiologist using ITK-SNAP, and the labels were used solely for validation.

IRB approval: This study was conducted with IRB approval from the joint CUHK-NTEC Clinical Research Ethics Committee (reference number: CREC 2020.115 (15 Apr 2020)).

2.5. *Training*

We trained and evaluated our proposed SSDA segmentation methods using the method outlined in Table 1. The EF-Unet was trained with the Adam optimizer, batch normalization, a learning rate of 10^{-4}, and a batch size of 8 for 40 epochs with a fixed learning rate. The segmentation loss was calculated using a combination of the dice loss and categorical cross-entropy loss. For network prediction, we used a post-processing threshold of 0.5. The 3D DESS includes 176 scans resulting in 1,0560 images (176*160), whereas for the 3D VISTA scans 22 subjects resulting in 3,300 (22*150) manually labeled slices across 6 tissue classes, provides a valuable foundation for evaluating the performance of our method in knee MRI segmentation.

To avoid overfitting and improve the generation of initial weak labels for T1FFE, we designed a contrast-selective (Cont.S) augmentation. Specifically, flipping on the horizontal axis was applied to the images, with the probability of applying the transform (P =0.5). Additionally, we applied shift-scale-rotate (scale limits = 0.1, 0.5, 0.1) transformations to the input images and label information. In addition, we used hue saturation (P =1), perspective transformations (P =0.5), sharpen (P =1), and addition of Gaussian noise (P =0.2).

We used subject IDs provided by the dataset to separate the samples. We split Dataset A by subject ID into two halves, balancing them based on the severity of the OA status of the subjects, which was provided as their KL grade in the OAI dataset. To compare our methods, we used a 2-fold cross-validation scheme.

2.6. *Evaluation*

To evaluate the volume segmentation accuracy in knee tissues, the dice score coefficient (DSC) was used as an indicator of the reproducibility of the manual annotation and defined as follows:

$$DSC = \frac{2|S \cap G|}{|S| + |G|} = \frac{2TP}{2TP + FP + FN} \tag{10}$$

where S represents the network segmentation and G represents the manual segmentation ground truth. The DSC ranges from 0 to 1, with 1 indicating complete overlap and 0 indicating no intersection between S and G. Recall, also known as sensitivity, was used as another performance metric and was found to be sensitive to over- and under-segmentation. Specifically, Recall is defined as follows:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Moreover, relative absolute volume difference (RAVD) is used to evaluate segmentation accuracy and quantify the size difference between

Table 3 Evaluating the initial segmentation results (DSC) on training and testing source domain 3D DESS images without domain adaptation on different tissue compartments

Train 3D	Test 3D	Augmentation schemes	Loss	Cartilage				Meniscus		Average
				FC	LTC	MTC	PC	LM	MM	
DESS	FSE	Mix	Dice	0.51 (±0.19)	0.50 (±0.16)	0.43 (±0.15)	0.42 (±0.11)	0.46 (±0.13)	0.48 (±0.19)	0.47
DESS	FSE	Mix	Dice+CE	0.53 (±0.19)	0.49 (±0.16)	0.43 (±0.17)	0.41 (±0.11)	0.47 (±0.13)	0.49(±0.19)	0.47
DESS	FSE	Cont.S	Dice	0.62 (±0.09)	0.59 (±0.14)	0.57 (±0.12)	0.41 (±0.11)	0.57 (±0.11)	0.65 (±0.13)	0.57
DESS	FSE	Cont.S	Dice+CE	0.63 (±0.09)	0.61 (±0.13)	0.59 (±0.11)	0.42 (±0.10)	0.60 (±0.11)	0.68 (±0.14)	0.59

FC: femoral cartilage; MM: medial meniscus; MTC: medial tibial cartilage; LM: lateral meniscus; LTC: lateral tibial cartilage; PC: patellar cartilage; Cont.S: contrast-selective.

Table 4 Evaluation of segmentation metrics (DSC) for different tissue compartments in the 3D FSE sequence and their comparisons with different approaches along with the best baseline method

Metric	Method	Cartilage				Meniscus	
		FC	LTC	MTC	PC	LM	MM
DSC	Baseline	0.63 (±0.09)	0.61(±0.13)	0.59 (±0.11)	0.42 (±0.10)	0.60 (±0.11)	0.68 (±0.14)
	Bas. + USDA	0.72 (±0.06)	0.67 (±0.07)	0.68 (±0.10)	0.56 (±0.05)	0.72 (±0.08)	0.76 (±0.14)
	SSDA-I	0.81 (±0.02)	0.76 (±0.03)	0.77 (±0.04)	0.67(±0.02)	0.80 (±0.04)	0.83 (±0.14)
	SSDA-II	0.83 (±0.02)	0.77 (±0.03)	0.77 (±0.04)	0.76 (±0.02)	0.81(±0.04)	0.84 (±0.09)
Recall	Baseline	0.49 (±0.10)	0.48(±0.16)	0.45 (±0.12)	0.31 (±0.14)	0.49 (±0.11)	0.60 (±0.11)
	Bas. + USDA	0.59 (±0.08)	0.56 (±0.11)	0.56 (±0.12)	0.44 (±0.11)	0.65 (±0.09)	0.73 (±0.11)
	SSDA-I	0.85 (±0.04)	0.72 (±0.05)	0.75 (±0.06)	0.57 (±0.04)	0.77 (±0.07)	0.82 (±0.16)
	SSDA-II	0.84(±0.05)	0.77 (±0.06)	0.76 (±0.05)	0.76 (±0.04)	0.77 (±0.05)	0.83 (±0.10)
ASSD	Baseline	2.67 (±1.47)	8.60 (±0.66)	3.27 (±10.95)	3.70 (±0.61)	1.94 (±5.69)	1.73 (±3.22)
	Bas. + USDA	2.16 (±1.23)	8.36 (±0.50)	2.74 (±10.98)	2.96 (±0.39)	1.39 (±5.37)	1.20 (±3.23)
	SSDA-I	1.26 (±0.54)	5.67 (±0.79)	1.96 (±07.96)	2.05 (±0.34)	1.06 (±4.82)	0.84 (±2.24)
	SSDA-II	0.98 (±0.36)	5.65 (±0.42)	1.87 (±07.99)	1.60 (±0.11)	0.94 (±4.64)	0.71 (±2.01)
RAVD	Baseline	47.41 (±11.19)	43.11 (±19.91)	48.90 (±14.33)	72.58 (±18.57)	41.82 (±11.98)	26.80 (±54.45)
	Bas.+ USDA	36.92 (±09.53)	34.68 (±15.68)	37.01 (±14.98)	57.06 (±14.46)	22.21 (±11.00)	17.42 (±55.44)
	SSDA-I	12.04 (±10.91)	13.61 (±06.92)	10.15 (±07.33)	30.76 (±05.62)	11.47 (±06.29)	08.13 (±19.34)
	SSDA-II	08.97 (±07.18)	09.34 (±08.71)	09.39 (±06.59)	14.87 (±07.07)	11.51 (±06.32)	06.86 (±22.94)

FC: femoral cartilage; MM: medial meniscus; MTC: medial tibial cartilage; LM: lateral meniscus; LTC: lateral tibial cartilage; PC: patellar cartilage.

network segmentation and manual annotation tissues. It is formulated as follows:

$$RAVD = \frac{||S| - |G||}{|G|} \times 100 \tag{12}$$

To evaluate the average distance between the segmentation and ground truth, the average symmetric surface distance (ASSD) was computed. For each boundary pixel in segmentation, there is only one closest boundary pixel in the ground label and the distance between the 2 pixels can be calculated. The calculation rule is also appropriate to pixels in the ground truth. ASSD is defined as follows:

$$ASSD = \frac{1}{|B_{seg}|+|B_{GT}|}\Big(\sum_{x\in B_{seg}} \min_{y\in B_{GT}} \|x-y\| + \sum_{x\in B_{GT}} \min_{y\in B_{seg}} \|y-x\|\Big) \tag{13}$$

3. Results

In this section, we present the evaluation of our proposed method compared with those of other methods, and an analysis of the results.

3.1. Baseline segmentation

We conducted experiments on the 3D DESS dataset to establish a robust semantic segmentation on a pipeline that can serve as a strong baseline for cross-domain segmentation. We validated the performance based on the available ground truths. The results are shown in Table 2. In Table 2 the numbers represent the means of volumetric DSCs. The training was conducted only on the 3D DESS dataset, using publicly available ground truths. The testing results are presented for 3D DESS and 3D FSE, comparing mix and Cont.S augmentation schemes. These results show only supervised learning using 3D DESS labels; therefore, no domain adaptation was involved in deriving those results.

We investigated the performance of EF-Unet on 3D DESS using different augmentation schemes and found no significant difference between the DSCs of tissues. However, using the combined loss slightly improved the segmentation accuracy, as evident in Table 2. We compared EF-Unet with other methods based on the tissues and average DSCs. The highest average DSC of 0.8701 was achieved, whereas MPUnet, which uses multiplanar plain Unets, can achieve a DSC of 0.8616. The 3D U-net and PSP-net performed lower in DSCs, achieving 0.8439 and 0.8268, respectively. When comparing EF-Unet with MPUnet, the main difference was in the compound scaling and squeeze excitation of EF-Unet, which helped extract detailed information. In contrast, MPUnet uses only multiplane orientation to learn extra information.

Another crucial aspect of the experiment was the network's learning abilities for novel segmentation samples in 3D FSE. The results are presented in Table 3. We tested the 3D DESS-trained network on 3D FSE and found that the learned information is insufficient to cope with cross-domain diversity without augmentation. We tested the mix and Cont.S schemes on 3D FSE and found that a careful selection of the augmentation scheme can lead to significant improvement in DSCs. The mixed scheme produced a DSC of 0.4697, but the targeted tissues were not depicted, and the network failed to detect tissue regions accurately. Using a suitable combination of the augmentation scheme and loss, the maximum DSC achieved through the baseline can reach up to 0.5899, which is still less than satisfactory for any application. For the rest of our experiments, we will consider this outcome as our baseline to express the differences in segmentation.

3.2. Domain adaptation segmentation

We compared our baseline method and the proposed domain adaptation approaches in knee cartilage and menisci segmentation (Table 4). For the conventional domain adaptation approaches, where the labels from a single domain can be used to learn the target domain, the av-

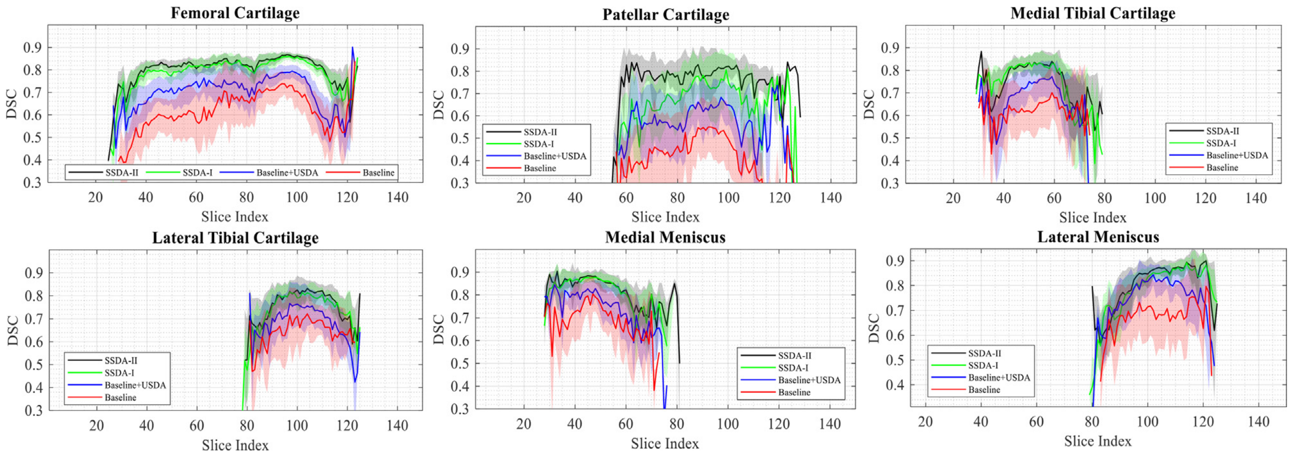

Figure 5. Slice-wise DSC comparison of the proposed method corresponding to each tissue in 3D FSE knee MRI. Solid lines in different colors show the mean value of the DSC spanned over the slice dimension of MRI, whereas the color band represents the 95% confidence levels of each method used in the comparison.

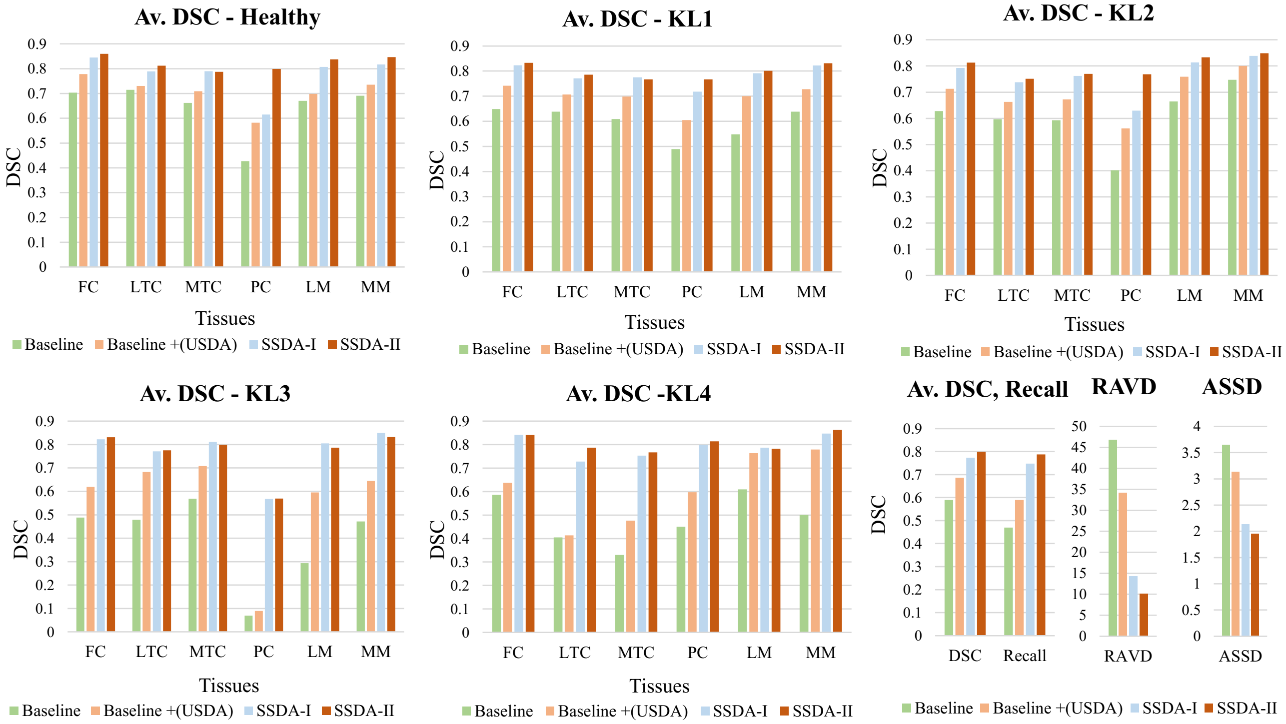

Figure 6. Segmentation performance comparison of subjects with different disease severity levels. The average DSC of each tissue is reported for different approaches. Overall average DSC, Recall, AVD, and ASSD are also shown to indicate the adaptation improvement.

erage DSC of the baseline can achieve 0.6865, whereas using the self-supervised approach at a single learning stage produces 0.7732. The second learning stage further improves segmentation accuracy by 0.7990, which is encouraging especially when no labels are available for the new domain.

Looking at the individual tissue performances, our proposed method, SSDA-II, had the most accurate performance in all knee tissues. On 3D FSE, the baseline method reached 0.634 for FC, 0.609 for LTC, 0.592 for MTC, 0.423 for PC, 0.604 for LM, and 0.679 for MM. The lack of model generalization toward new domains could have resulted in this weak performance. However, applying UDA led to a significant improvement in Dataset C. In this new setting, our model showed a significantly improved performance for FC (from 0.634 to 0.722), LTC (from 0.609 to 0.675), MTC (from 0.592 to 0.678), LM (from 0.604 to 0.725), and MM (from 0.679 to 0.760). Even in situations without 3D FSE annotation, UDA could learn specific features of the 3D FSE sequence using pseudo-3D FSE labels. The method with SSDA-II yielded the best DSCs (0.827 for FC, 0.772 for LTC, 0.772 for MTC, 0.764 for PC, 0.818 for LM, and 0.842 for MM). In addition to DSC, our approach consistently yielded better scores for Recall, RAVD, and ASSD.

3.3. Slice-wise segmentation

To understand the continuity and progression of each knee compartment independently, Figure 5 presents the distribution of the cartilage and meniscus tissues in depth-wise regions of interest against slice num-

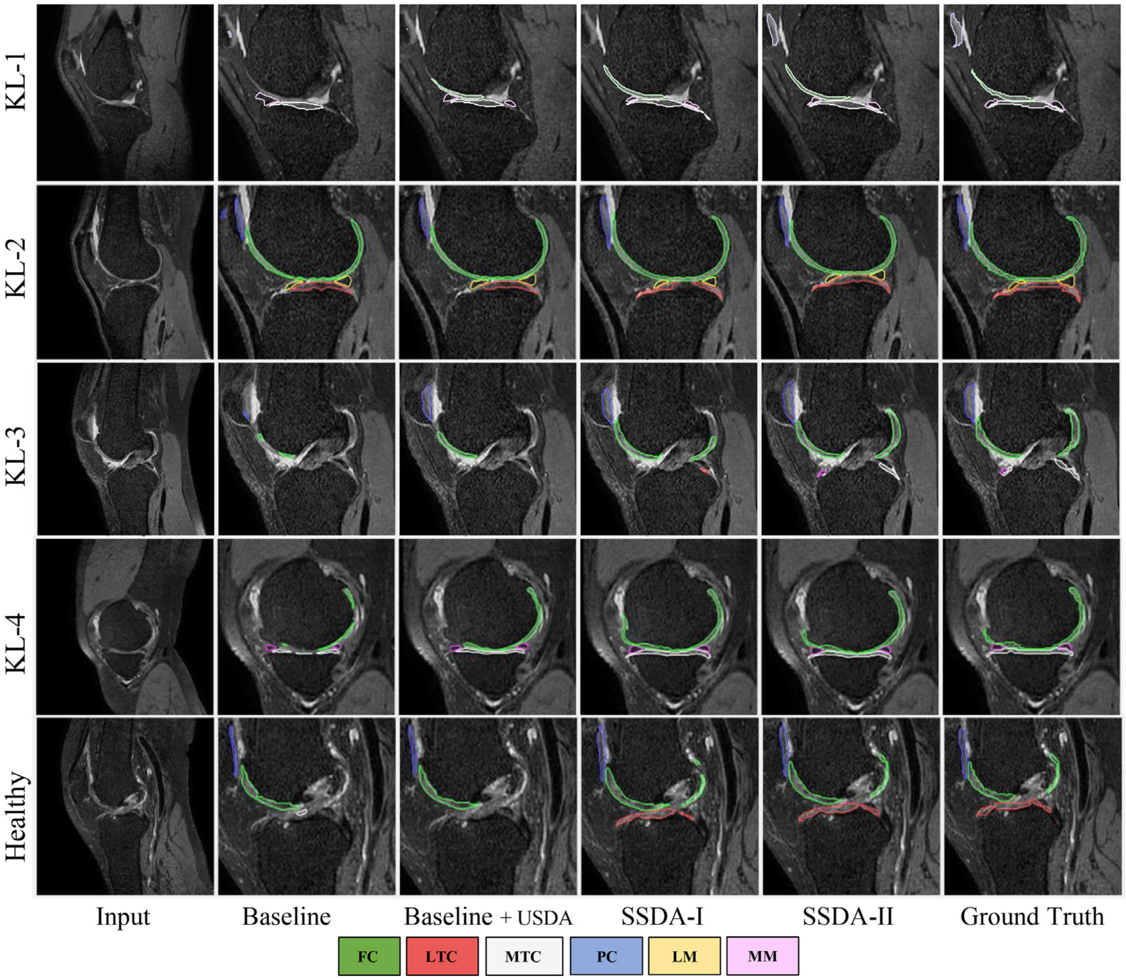

Figure 7. Segmentation results on 3D FSE sequence with representative slices from different KL grades. The comparison of segmented tissues against each method is presented, which shows the adaptation progress in each step.

bers. The average DSC comparison of different approaches is presented in the color graphs. FC spans over a larger number of slices, whereas 'Meniscus' and PC have shorter occurrences in the distribution. This makes the segmentation task even more challenging, especially when dealing with the meniscus. Our proposed methods dominate the baseline and UDA methods, especially in slice indices of 40–100 (FC), 60–100 (PC), 40–60 (MTC and MM), and 90–120 (LTC and LM). These regions belong to the medial and lateral femoral condyles and are responsible for the weight-bearing regions in the knee.

4. Discussion

The ability of source-independent trained CNNs to generalize to new and unseen distributions is crucial for its effective use in clinical MRI segmentation applications. One common approach to improve this capability is to train the CNN on a diverse dataset that covers a wide range of distributions. Manually annotating large amounts of data for each new and unknown distribution poses practical challenges. Although data augmentation within the source domain offers some benefits, it is not a complete solution for addressing this challenge. Our work introduces a generic framework for tackling this issue and proposes a source-independent multi-domain adaptation network, which involves an intermediate domain generator and segmentation task module specifically designed for knee MRI segmentation and are seamlessly integrated to produce knee tissue segmentation in a self-supervised iterative manner.

The segmentation performance improves after the introduction of the UDA. However, the maximum DSC of MTC, PC, and MM are 0.6781, 0.5596, and 0.7602, respectively, which are low values to be considered for clinical use. Using the proposed method (SSDA-1 in the first and SSDA-II in the second stage of retraining), the average DSC improves to 0.7990, and DSC of MTC, PC, and MM can reach 0.7715, 0.7640, and 0.8420, respectively.

4.1. Disease severity and segmentation

The Kellgren-Lawrence (KL) grading system is a widely used method for assessing the severity of OA based on radiographic findings, which categorize OA into 5 grades. To evaluate the segmentation performance of our proposed method against disease severity in terms of KL grades, Figure 6 shows DSC, Recall, RAVD, and ASSD in comparison with the Healthy, KL-1, KL-2, KL-3, and KL-4 grades. Unlike other approaches, where segmentation accuracy is affected by disease severity, i.e., Healthy *vs.* KL-4, our approaches are unaffected by the variation in disease severity. Figure 7 presents the visual comparison among the key MRI slices, chosen from healthy individuals to those with varying disease severity. Baseline segmentation and UDA showed inferior performance in patients with KL-4 grade OA. In contrast, the proposed method, compared with other approaches, achieved consistent DSC, Recall, RAVD, and ASSD measures due to the supporting intermediate domain

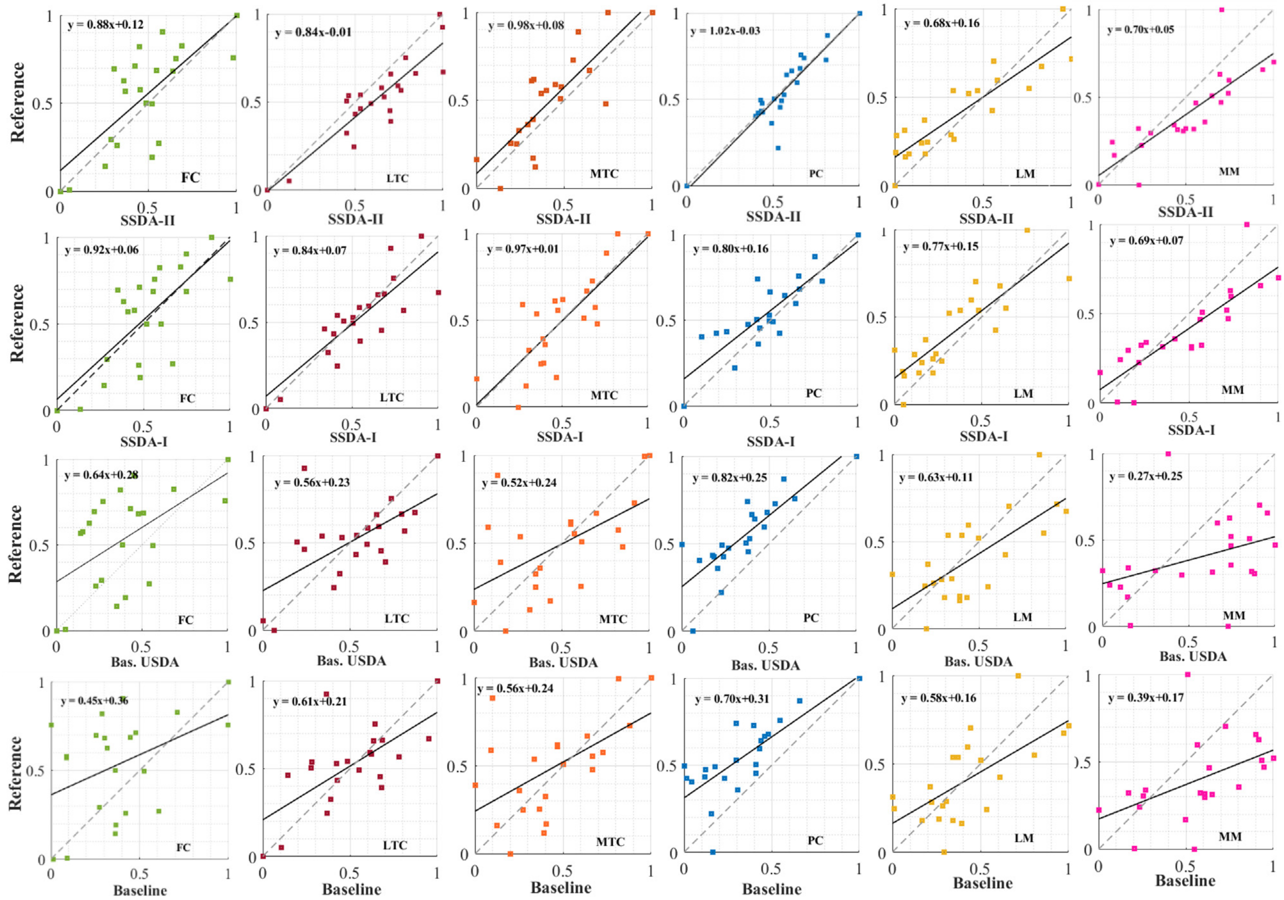

Figure 8. Correlation plots for DSCs corresponding to 6 tissues (cartilages and menisci) from different approaches. The slope equation is also shown for reference values and segmented values.

4.2. Individual patients with OA

We tested the individual segmentation performance on the 3D FSE of patients with OA and compared the results with the proposed methods. Figures 8 and 9 show the comparison of the correlation plots and bar charts corresponding to the 3D FSE of patients with OA, respectively. The correlation plots in Figure 8 illustrate the segmentation of DSCs compared with the baseline in different tissues. As shown in Figure 9, the results of the proposed methods (SSDA-I and SSDA-II) showed higher DSCs and less standard deviation compared with the baseline and UDA approaches. Based on the scatter plots and fitting values, the proposed methods were more consistent with the reference values, especially in the FC and MM tissue regions. Compared with the proposed methods, the baseline method and UDA approaches were less consistent owing to the under-segmentation of tissue regions.

4.3. Comparisons with other methods

To further validate our proposed multi-domain adaptation method, we compared the segmentation performance with the HDMT Network [23], which combines 2D and 3D mean-teacher networks to produce segmentation labels, MMUDA [25], knee MRI segmentation methods using domain adaptation mitigating label information, CLD [24], and closely-related mixupUDA [14]. Table 5 shows the DSCs against knee cartilage and meniscus compartments, whereas Table 6 shows the comparison of average DSCs, Recall, RAVD, and ASSD. Among all the methods, our proposed domain adaptation with a segmentation network shows better qualitative performance by a large margin. mixupUDA [14] with adversarial learning, which was specially designed for knee MRI, showed comparatively better DSCs owing to better alignment; however, its performance in FC, LTC, and MM was still below because of the sparse nature of the tissues and presence of OA severity in the clinical domain for testing. Different from the previous experimental settings, two experiments were conducted on the 3D FSE dataset, consisting of a random selection of 5% and 10% of the training set to form the labeled set. Notably, both partitions were fixed throughout each experiment to ensure consistency in the training data. The results are presented in Table 7. The DSCs improved from 0.799 to 0.852 and reached 0.869 with 5% and 10% labeled data. The HDMT [23] showed better improvement in DSCs with the availability of more labeled data and achieved 0.797, which is still lower than that of the proposed method, demonstrating the effectiveness of sparse tissue alignment in multi-domain settings. We also present the visual comparison of different methods in representative OA slices, as shown in Figure 10. The main reason behind the lack of performance is the under-segmentation of OA-degraded small regions, as pointed out in subfigures. MRI contrast adaptation corresponding to the key anatomical regions are crucial to conduct satisfactory domain adaptation in clinical MRI segmentation.

An important aspect of SSDA is its performance against the severity of OA in patients. It has been observed that the segmentation outcomes are consistent across different degrees of OA severity. Our proposed methods have little or no deviation against varying disease severity. However, other methods failed to achieve satisfactory performance, especially in MTC, PC, and LM, where the tissues are thin and difficult to segment due to disease. This phenomenon is more evident when the baseline misses tissue regions and can only detect a few. The UDA can only capture FC and miss the thin regions because of the inadaptability caused by imperfect domain transformation. Our proposed method

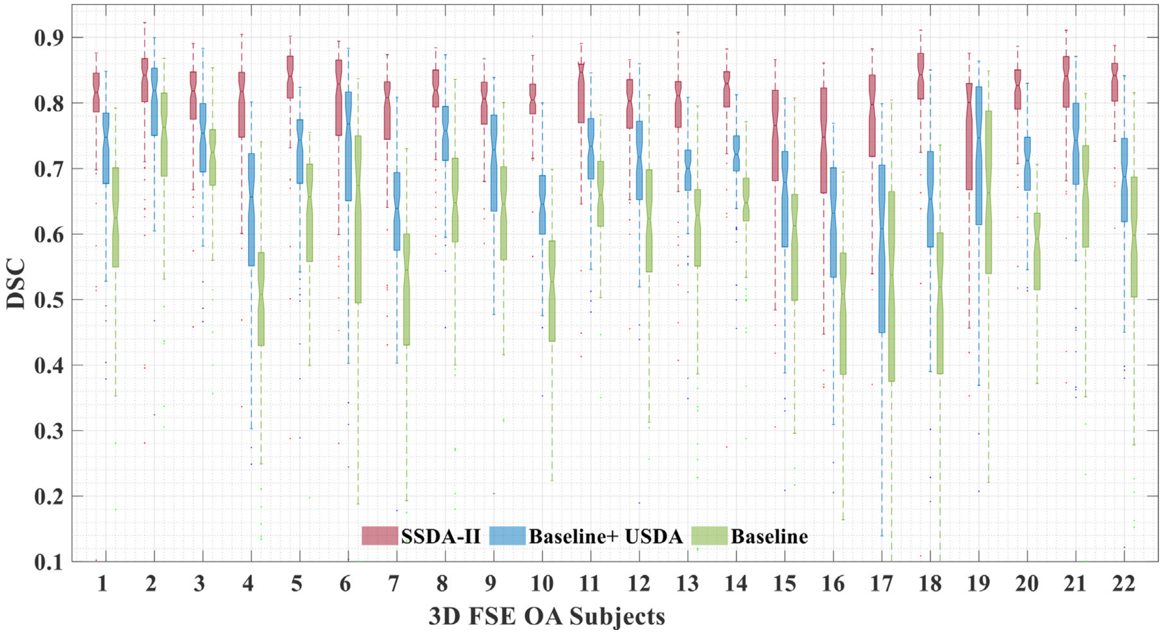

Figure 9. Segmentation comparison of different approaches on 3D FSE of patients with OA

Table 5 Dice scores comparison of cartilages and meniscus compartment segmentation with related methods

Methods	Cartilage				Meniscus		Average
	FC	LTC	MTC	PC	LM	MM	
Ours	0.83 (±0.02)	0.77 (±0.03)	0.77 (±0.04)	0.76 (±0.02)	0.82 (±0.04)	0.84 (±0.09)	0.80 (±0.04)
HDMT [23]	0.76 (±0.08)	0.65 (±0.04)	0.68 (±0.06)	0.63 (±0.09)	0.67 (±0.06)	0.79 (±0.12)	0.70 (±0.07)
CLD [24]	0.75 (±0.06)	0.64 (±0.04)	0.63 (±0.06)	0.74 (±0.04)	0.70 (±0.06)	0.75 (±0.13)	0.70 (±0.06)
mixupUDA [14]	0.73 (±0.07)	0.68 (±0.05)	0.74 (±0.04)	0.73 (±0.04)	0.78 (±0.04)	0.75 (±0.11)	0.73 (±0.05)
MMUDA [25]	0.65 (±0.11)	0.66 (±0.09)	0.71 (±0.05)	0.65 (±0.14)	0.71 (±0.05)	0.73 (±0.11)	0.68 (±0.09)

FC: femoral cartilage; MM: medial meniscus; MTC: medial tibial cartilage; LM: lateral meniscus; LTC: lateral tibial cartilage; PC: patellar cartilage.

Table 6 Segmentation performance comparison with other methods using average values of evaluation metrics

Methods	Dice ↑	Recall↑	RAVD ↓	ASSD ↓
Ours	0.80 (±0.04)	0.79 (±0.05)	10.15 (±9.80)	1.96 (±2.58)
HDMT [23]	0.70 (±0.07)	0.68 (±0.06)	24.10 (±17.45)	3.10 (±2.72)
CLD [24]	0.70 (±0.06)	0.69 (±0.06)	20.53 (±11.00)	2.70 (±2.95)
mixupUDA [14]	0.73 (±0.05)	0.73 (±0.05)	17.85 (±12.42)	2.40 (±2.87)
MMUDA [25]	0.68 (±0.09)	0.66 (±0.11)	26.63 (±13.20)	3.73 (±3.06)

shows better segmentation and resemblance to manual ground truths. The difference between SSDA-I and SSDA-II is more evident, where the difference occurs at tissue boundaries.

Although our study has shown promising results in knee MRI segmentation, we acknowledge the need for a more comprehensive validation to fully assess the clinical utility of our proposed method. In our future research, we plan to address this limitation by expanding our dataset and incorporating data from various MRI sequences and scanners. This will enable us to conduct more extensive investigations and provide a more robust evaluation of the effectiveness of our method in diverse clinical scenarios.

Our proposed solution, the self-supervised multi-domain adaptation, effectively addresses the multi-domain generalization of knee MR image segmentation in clinical MRIs under source-independent scenarios. This proposed approach involves a multi-domain generator guided by the segmentation task pseudo-label attentions across knee tissue regions and self-supervised segmentation that overcomes the challenges mentioned earlier and delivers accurate knee tissue segmentation without requiring any external label information of the target domain. We particularly focused on self-collected 3D FSE (VISTA) and T1FFE of patients with OA. The leverage of multi-domain 3D FSE (VISTA) and T1FFE data significantly improves the segmentation predictions of 3D FSE MR images, particularly in thin tissue regions, such as the PC and meniscus. The related methods based on UDA strategies showed inferior performance

Table 7 Dice scores performance comparison with other methods using different proportions of labeled data from 3D FSE

Methods	Labeled/Unlabeled (5%)	Dice scores ↑	Labeled/Unlabeled (10%)	Dice scores ↑
Ours	5/95	0.85 (±0.03)	10/90	0.87 (±0.03)
HDMT [23]	5/95	0.72 (±0.07)	10/90	0.80 (±0.05)
CLD [24]	5/95	0.78 (±0.05)	10/90	0.82 (±0.05)
mixupUDA [14]	5/95	0.76 (±0.05)	10/90	0.83 (±0.04)
MMUDA [25]	5/95	0.72 (±0.06)	10/90	0.76 (±0.06)

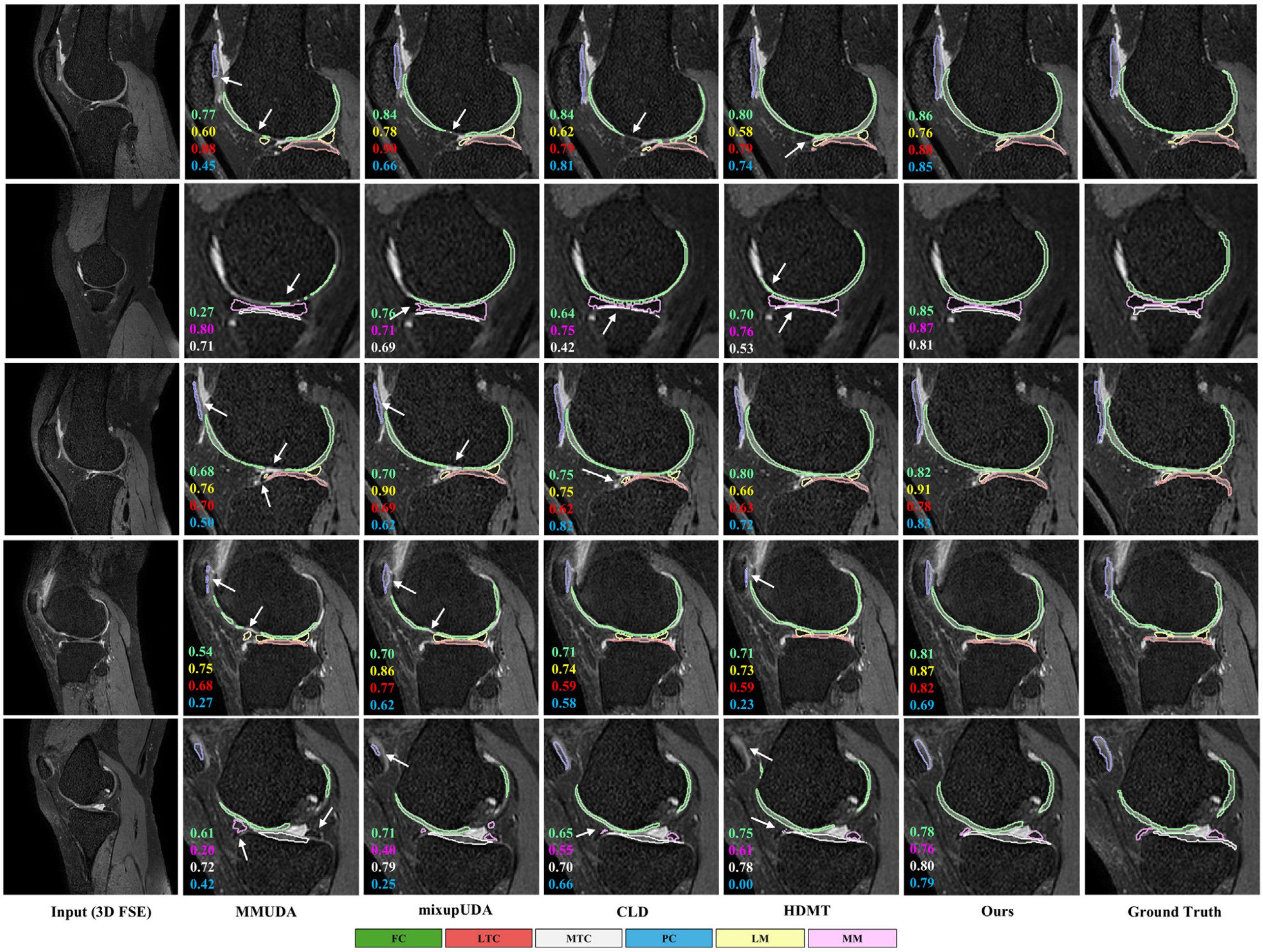

Figure 10. Visual comparison of segmentation results using benchmark domain adaptation methods with the proposed method. Each row displays sample slices from the 3D FSE MRI dataset alongside their corresponding reference ground truth. The colors of the Dice Similarity Coefficients (DSCs) match the tissue types. Arrows are positioned in different locations to highlight the segmentation errors.

on the segmentation of 3D FSE images, likely due to a more significant domain shift problem between the 3D DESS and 3D FSE. Using the proposed SSDA, a significant improvement was achieved in segmenting 3D FSE images without requiring annotation. We presented the results of the cartilage and meniscus segmentation in 3D FSE images and validated the results with the self-prepared manual ground truths.

Conflicts of interest statement

The authors declare no competing interests.

Fundings

This work is supported by a grant from the Innovation and Technology Commission of the Hong Kong SAR (Project MRP/001/18X), and a grant from the Research Grants Council of the Hong Kong SAR (UGC/FDS24/E18/22).

Author contributions

Sheheryar Khan: Conceptualization, Methodology, Formal analysis, Software, Writing – review & editing. **Siyue Li:** Writing – review & editing. **Fan Xiao:** Resources, Supervision, Writing – review & editing. **Kevin Ho:** Resources, Supervision, Writing – review & editing. **Michael Ong:** Resources, Supervision, Writing – review & editing. **James Griffith:** Resources, Supervision, Writing – review & editing. **Weitian Chen:** Supervision, Resources, Writing – review & editing.

References

[1] Galderisi U, Peluso G, Di Bernardo G. Clinical trials based on mesenchymal stromal cells are exponentially increasing: where are we in recent years? Stem cell rev and repo 2022;18(1):23–36. doi:10.1007/s12015-021-10231-w.

[2] Guan H, Liu M. Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng 2021;69(3):1173–85. doi:10.1109/TBME.2021.3053212.

[3] Kijowski R, Davis KW, Woods MA, et al. Knee joint: comprehensive assessment with 3d isotropic resolution fast spin-echo mr imaging–diagnostic performance compared with that of conventional mr imaging at 3.0 t. Radiology 2009;252(2):486–95. doi:10.1148/radiol. 2522081604.

[4] Lim D, Han Lee Y, Kim S, et al. Clinical value of fat-suppressed 3d volume isotropic spin-echo (vista) sequence compared to 2d sequence in evaluating internal structures of the knee. Acta Radiol 2016;57(1):66–73. doi:10.1177/0284185115583618.

[5] Almajalid R, Zhang M, Shan J. Fully automatic knee bone detection and segmentation on three-dimensional mri. Diagnostics 2022;12(1):123. doi:10.3390/diagnostics12010123.

[6] Liu F. Susan: segment unannotated image structure using adversarial network. Magn Reson Med 2019;81(5):3330–45. doi:10.1002/mrm.27609.

[7] Xu W, Li C, Bian Y, et al. Cross-modal consistency for single-modal mr image segmentation. IEEE Trans Biomed Eng 2024. doi:10.1109/TBME.2024.3380058.

[8] Schmidt AM, Desai AD, Watkins LE, et al. Generalizability of deep learning segmentation algorithms for automated assessment of cartilage morphology and mri relaxometry. J Magn Reson Imaging 2023;57(4):1029–39. doi:10.1002/jmri.28152.

[9] Felfeliyan B, Hareendranathan A, Kuntze G, et al. Mri knee domain translation for unsupervised segmentation by cyclegan (data from osteoarthritis initiative (oai)). In: Proc. IEEE Eng. Med. Biol. Soc. (EMBC). IEEE; 2021. doi:10.1109/EMBC46164.2021.9630324.

[10] Santomartino SM, Kung J, Yi PH. Systematic review of artificial intelligence development and evaluation for mri diagnosis of knee ligament or meniscus tears. Skeletal Radiol 2024;53(3):445–54. doi:10.1007/s00256-023-04416-2.

[11] Kessler DA, MacKay JW, Crowe VA, et al. The optimisation of deep neural networks for segmenting multiple knee joint tissues from mris. Comput Med Imaging Graph 2020;86:101793. doi:10.1016/.2020.101793.

[12] Jeon U, Kim H, Hong H, et al. Automatic meniscus segmentation using cascaded deep convolutional neural networks with 2d conditional random fields in knee mr images. In: Proc. SPIE Int. Workshop Adv. Imaging Technol. (IWAIT), SPIE; 2020. doi:10.1117/12.2566450.

[13] Khan S, Azam B, Yao Y, et al. Deep collaborative network with alpha matte for precise knee tissue segmentation from mri. Comput Methods Programs Biomed 2022;222:106963. doi:10.1016/j.cmpb. 2022.106963.

[14] Panfilov E, Tiulpin A, Klein S, et al. Improving robustness of deep learning based knee mri segmentation: Mixup and adversarial domain adaptation. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops; 2019. doi:10.1109/ICCVW.2019.00057.

[15] Li S, Zhao S, Zhang Y, et al. Source-free unsupervised adaptive segmentation for knee joint mri. Biomed Signal Process Control 2024;92:106028. doi:10.1016/j.bspc.2023.106028.

[16] Zhong J, Yao Y, Cahill DG, et al. Unsupervised domain adaptation for automated knee osteoarthritis phenotype classification. Quant Imaging Med Surg 2023;13(11):7444. doi:10.21037/qims-23-335.

[17] Zhu JY, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. IEEE Int. Conf. Comput.; 2017. doi:10.1109/ICCV.2017.244.

[18] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: Proc. Int. Conf. Mach. Learn.; 2019.

[19] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit.; 2018. doi:10.1109/CVPR.2018.00745.

[20] Sengar SS, Meulengracht C, Boesen MP, et al. Multi-planar 3d knee mri segmentation via unet inspired architectures. International Journal of Imaging Systems and Technology 2023;33(3):985–98. doi:10.1002/ima.22836.

[21] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D u-net: learning dense volumetric segmentation from sparse annotation. In: Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI). Springer; 2016. doi:10.1007/978-3-319-46723-8_49.

[22] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit.; 2017. doi:10.1109/CVPR.2017.660.

[23] Zhu J, Bolsterlee B, Chow BV, et al. Hybrid dual mean-teacher network with double-uncertainty guidance for semi-supervised segmentation of mri scans. Comp Med Imag and Grap 2024:115. doi:10.1016/j.compmedimag.2024.102383.

[24] Lin Y, Yao H, Li Z, et al. Calibrating label distribution for class-imbalanced barely-supervised knee segmentation. In: Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI). Springer; 2022. doi:10.1007/978-3-031-16452-1_11.

[25] Wu F, Li L, Zhuang X. Multi-modality cardiac segmentation via mixing domains for unsupervised adaptation. In: Proc. Int. Workshop Stat. Atlases Comput. Models Heart. Springer; 2021. doi:10.1007/978-3-030-72084-1_21.

[26] Paproki A, Engstrom C, Chandra SS, et al. Automated segmentation and analysis of normal and osteoarthritic knee menisci from magnetic resonance images–data from the osteoarthritis initiative. Osteoarthr Cartil 2014;22(9):1259–70. doi:10.1016/j.joca.2014.06.005.