Semi-automatic muscle segmentation in MR images using deep registration-based label propagation

Nathan Decaux, Pierre-Henri Conze, Juliette Ropars, Xinyan He, Frances T Sheehan, Christelle Pons, Ben Salem, Sylvain Brochard, François Rousseau

Semi-automatic muscle segmentation in MR images using deep registration-based label propagation

Nathan Decaux[a,b,*], Pierre-Henri Conze[a,b], Juliette Ropars[a,c], Xinyan He[b], Frances T. Sheehan[d], Christelle Pons[a,c,e], Douraied Ben Salem[a,c], Sylvain Brochard[a,c], François Rousseau[a,b]

[a] *LaTIM UMR 1101, Inserm, Brest, France*
[b] *IMT Atlantique, Brest, France*
[c] *University Hospital of Brest, Brest, France*
[d] *Rehabilitation Medicine, NIH, Bethesda, USA*
[e] *Fondation ILDYS, Brest, France*

Abstract

Fully automated approaches based on convolutional neural networks have shown promising performances on muscle segmentation from magnetic resonance (MR) images, but still rely on an extensive amount of training data to achieve valuable results. Muscle segmentation for pediatric and rare diseases cohorts is therefore still often done manually. Producing dense delineations over 3D volumes remains a time-consuming and tedious task, with significant redundancy between successive slices. In this work, we propose a segmentation method relying on registration-based label propagation, which provides 3D muscle delineations from a limited number of annotated 2D slices. Based on an unsupervised deep registration scheme, our approach ensures the preservation of anatomical structures by penalizing deformation compositions that do not produce consistent segmentation from one annotated slice to another. Evaluation is performed on MR data from lower leg and shoulder joints. Results demonstrate that the proposed semi-automatic multi-label segmentation model outperforms state-of-the-art techniques.

Keywords: semi-automatic segmentation, musculoskeletal system, label propagation, deep registration

*Corresponding Author.
Email address: nathan.decaux@imt-atlantique.fr (Nathan Decaux)

1. Introduction

Pediatric musculoskeletal segmentation is a key pre-processing stage for clinical decision-making, as the generated 3D muscle and bone models enable the extraction of key biomarkers related to disease progression, which, in turn, optimizes identifying therapeutic interventions. In this context, manual segmentation remains the gold standard to achieve valuable delineations [1]. Indeed, the variability and scarcity of musculoskeletal data, especially in pediatric cohorts, is an obstacle to the generalization of automatic segmentation models. Thus, 3D atlas-based label propagation methods, whether based on registration [2, 3] or local patch similarity [4], are not always well suited. To reduce the processing-time while maintaining a quality close to the inter-expert variability, semi-automatic techniques requiring few annotated images appeared to be the best trade-off to date [5].

Fully supervised methods based on deep learning (DL) have been investigated with sparse annotated 2D slices [6, 7]. Despite superior performance to atlas-based methods [8], these DL-based approaches require a significant number of training examples [9]. Learning from few examples remains an open issue. Interactive approaches have also been explored in the field of DL for medical image segmentation purposes. Point or scribble-based methods allow for the least possible interaction, but suffer from a lack of reproducibility and thus require complex optimization procedures [10, 11, 12]. Chanti et al. explored the use of long short-term memory (LSTM) units to propagate an annotated sub-volume through 3D ultrasound images, but still requires hundreds of annotated slices for training [13].

As an alternative, intra-volume semi-automatic segmentation enables to deal with the limited size of annotated data [5]. In recent medical imaging literature, semi-automatic segmentation by label propagation, which involves spreading annotated slices throughout a volume or sequence, has not been widely studied. This technique is more commonly seen in the field of video segmentation, where the first annotated frame is propagated throughout the sequence [14, 15]. In the context of medical images, morphological-based interpolation of distant annotated slices provides a fast strategy to propagate labels over a 3D volume, but may also require an important number of interactions and often fails on multi-structure objects [16, 17]. In addition, this method does not capture local structure variations, as it relies solely on segmentation contours. To address this issue, Ogier et al. investigated a 2D registration framework for propagating annotated slices towards 3D

segmentation maps [18]. In the same spirit as [19, 20] for static or dynamic images, the approach proposed in [18] allowed to obtain contour delineations for a sequence of slices delimited by two annotated slices. The anatomical coherence is reinforced by exploiting both distant deformations between annotations and successive deformations between images, weighting each contribution according to their distance to the closest annotated slices.

Our approach differs from these previous works by using an unsupervised deep learning framework for deformation generation. Numerous studies have been conducted in the topic of deep learning registration, with promising outcomes, especially for its generalizability and its lower processing time compared to direct optimization methods [21, 22, 23]. Among these methods, Voxelmorph reaches similar performances of popular 3D free-form registration algorithms between subjects such as ANTs SyN [24, 25]. In this study, we propose to use a similar approach to perform 2D registration. One key distinction between this study and the work of Voxelmorph is the focus on 2D registration within a single volume, rather than on 3D registration between multiple volumes.

In this work, our objective is to develop an intra-subject 3D segmentation method for pediatric muscles from magnetic resonance (MR) images based on very few manually annotated 2D slices. To this end, we investigate the use of a learning-based registration framework to propagate labels through the full volume. More precisely, the 3D segmentation problem is modeled as a 2D label propagation problem based on unsupervised DL-based registration. The registration approach relies on intensity similarity between successive slices and on muscle shapes from annotated slices. A regularization term is introduced via the definition of a dedicated loss from combined deformation fields. Furthermore, propagated masks from different manual segmentations are merged through a novel weighting method based on image similarity measures. The proposed approach is evaluated on two clinical datasets and compares favorably to state-of-the-art methods. The PyTorch implementation of this work is publicly available[1].

2. Methods

Let us consider a volumetric MR image as a stack of 2D slices. The use-case considered in this work consists of a clinical expert who minimally

[1] `https://github.com/nathandecaux/labelprop`

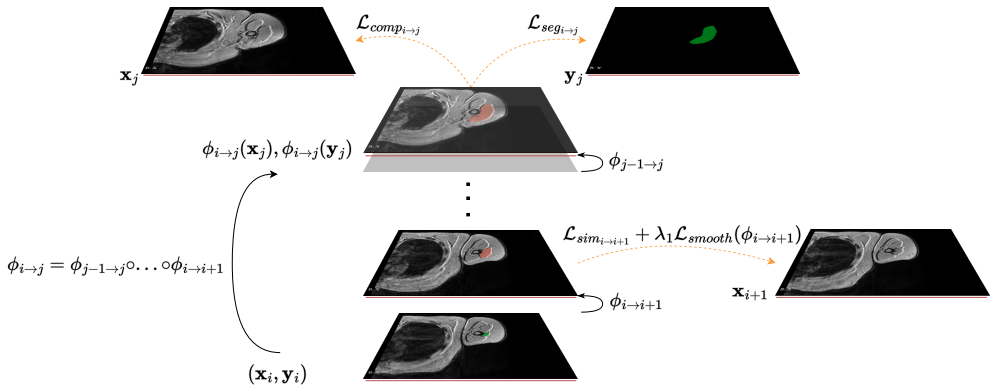


Figure 1: Deep registration-based label propagation from an annotated slice $\{\mathbf{x}_i, \mathbf{y}_i\}$ to the next one $\{\mathbf{x}_j, \mathbf{y}_j\}$ towards dense muscle segmentation in MR images. Please refer to the text for complete description of notations.

delineates muscles of interest in few slices only. The objective is to provide a full 3D segmentation map of the muscles by propagating the 2D annotations. We model this problem into an intra-patient registration-based propagation framework. We denote a full volume $\mathcal{X}$ as a set of stacked 2D images $\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ and the corresponding segmentation map $\mathcal{Y}$ as a set of stacked sparsely annotated 2D images $\mathcal{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)$, i.e. the majority of the slices in the set are not annotated. In such a registration-based propagation framework, the key assumption is that muscle shapes in two consecutive 2D images are highly similar. As recently highlighted in [5], methods based on intra-subject non-linear registration currently provide state-of-the-art results. The purpose of the proposed approach is to propagate a small amount of annotated 2D slices through the volume $\mathcal{X}$ to estimate a full 3D muscle segmentation map. To this end, we propose a DL-based approach that simultaneously registers consecutive 2D MR slices and propagates annotated muscle contours.

2.1. Unsupervised 2D intensity-based registration

The proposed label propagation framework relies on non-linear registration of consecutive 2D slices. Let $\phi_{k\to k+1}$ be a non-linear mapping that maps coordinates of $\mathbf{x}_k$ to coordinates of $\mathbf{x}_{k+1}$. Such a mapping $\phi_{k\to k+1}$ is usually estimated by solving the following optimization problem:

$$\hat{\phi}_{k\to k+1} = \arg\min_{\phi} \mathcal{L}_{sim_{k\to k+1}}(\mathcal{X}, \phi) + \lambda_1 \mathcal{L}_{smooth}(\phi), \tag{1}$$

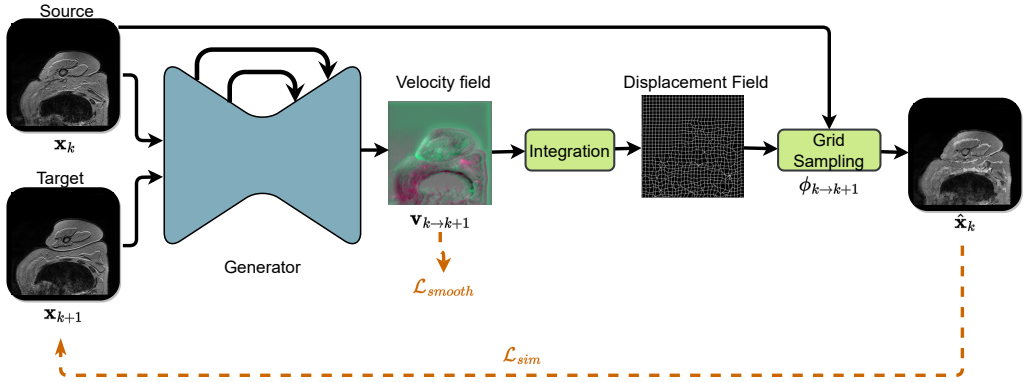


Figure 2: Unsupervised 2D intensity-based registration model between two successive slices $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$. Please refer to the text for complete description of notations.

where $\mathcal{L}_{sim_{k\to k+1}}(\mathcal{X}, \phi)$ is an image dissimilarity measure between the consecutive slices $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$, $\mathcal{L}_{smooth}$ a regularization term and λ_1 a trade-off weight between both terms. In this work, the non-linear mapping ϕ is modeled as $\phi = Id + \mathbf{u}$ where Id is the identity transform and $\mathbf{u}$ is the displacement field. We consider a diffeomorphic mapping ϕ through the integral of a stationary velocity vector field $\mathbf{v}$ such that ϕ preserves the topology and is invertible [26]. Moreover, the inverse mapping $\phi^{-1}_{k\to k+1}$ is obtained using the negative velocity field -$\mathbf{v}$, computed with a similar process as in forward integration.

$\mathcal{L}_{sim_{k\to k+1}}(\mathcal{X}, \phi)$ penalizes the dissimilarity between the slice $\mathbf{x}_k$ deformed using $\phi_{k\to k+1}$ and the slice $\mathbf{x}_{k+1}$, and reciprocally using $\phi^{-1}_{k\to k+1}$. To be robust to any intensity variations inside the full volume $\mathcal{X}$, the dissimilarity loss used in this work is the normalized local cross-correlation (NCC):

$$\mathcal{L}_{sim_{k\to k+1}} = -\texttt{NCC}(\phi_{k\to k+1}(\mathbf{x}_k), \mathbf{x}_{k+1}) - \texttt{NCC}(\phi^{-1}_{k\to k+1}(\mathbf{x}_{k+1}), \mathbf{x}_k). \quad (2)$$

To ensure smooth deformation fields, the regularization term $\mathcal{L}_{smooth}$ penalizes for each pixel $\mathbf{p}$ the spatial derivatives of the velocity field $\mathbf{v}$ [24]: $\mathcal{L}_{smooth}(\phi_{k\to k+1}) = \sum_{\mathbf{p}} \|\nabla(\mathbf{v}_{k\to k+1}(\mathbf{p}))\|^2$. Similarly to VoxelMorph [24], the deformation field is computed in an unsupervised way, where the velocity field is modeled using a neural network.

2.2. Propagation constraints

Our objective is to segment a full MR volume by propagating 2D annotations using estimated deformation fields between consecutive slices. We propose to consider a loss related to segmentation to make the registration-based label propagation more accurate and anatomically plausible. Let consider two annotated slices $\mathbf{x}_i$ and $\mathbf{x}_j$ and their corresponding annotations $\mathbf{y}_i$ and $\mathbf{y}_j$. The slice $\mathbf{x}_i$ can be mapped to $\mathbf{x}_j$ using $\phi_{i\to j}$, the combination of non-linear mappings $\phi_{i\to j} = \phi_{j-1\to j} \circ ... \circ \phi_{i\to i+1}$. To improve the anatomical correspondences, we consider a overlap-based segmentation loss $\mathcal{L}_{seg_{i\to j}}$ between the deformed annotated slice $\phi_{i\to j}(\mathbf{y}_i)$ and $\mathbf{y}_j$: $\mathcal{L}_{seg_{i\to j}} = -\texttt{DSC}(\phi_{i\to j}(\mathbf{y}_i), \mathbf{y}_j)$ where DSC is the Dice similarity coefficient quantifying the surface overlap for a given muscle. The combination of non-linear mappings can lead to an accumulation of registration errors between annotated sections. Therefore, to regularize the estimated displacement fields through the propagation, we propose to investigate the use of the following intensity-based criterion: $\mathcal{L}_{comp_{i\to j}} = -\texttt{NCC}(\phi_{i\to j}(\mathbf{x}_i), \mathbf{x}_j)$.

2.3. Label propagation framework

Considering two slices $(\mathbf{x}_i, \mathbf{x}_j)$ and their annotations $(\mathbf{y}_i, \mathbf{y}_j)$, the corresponding loss function $\mathcal{L}_{i\to j}$ is defined as follows:

$$\mathcal{L}_{i\to j} = \frac{1}{j-i}\sum_{k=i}^{j-1}\left(\mathcal{L}_{sim_{k\to k+1}} + \lambda_1 \mathcal{L}_{smooth}(\phi_{k\to k+1})\right) + \lambda_2 \mathcal{L}_{seg_{i\to j}} + \lambda_3 \mathcal{L}_{comp_{i\to j}}, \tag{3}$$

with weighting factors λ_2 and λ_3.
For more robustness, label propagation can be performed in both directions by considering a bidirectional loss $\mathcal{L} = \mathcal{L}_{i\to j} + \mathcal{L}_{j\to i}$. The generated dense annotations are then merged by giving more weight to the pixels that are most likely to be correctly registered:

$$\hat{\mathbf{y}}_k = \beta_k(\mathcal{X}, \phi).(\phi_{i\to k}(\mathbf{y}_i)) + (1 - \beta_k(\mathcal{X}, \phi)).(\phi_{j\to k}(\mathbf{y}_j)), \tag{4}$$

with β_k the weighting function. As an approximation of the DSC evolution along the propagation axis, [18] suggested using $\beta_k = 1 - \frac{\texttt{arctan}(C(k-(j-i)/2))}{\pi}$ as a weighting function, where C is a smoothing factor. The approach therefore assumes a symmetry in the propagation quality in both directions, which may not be the case. This function is denoted as *distance-based weighting* (DW) in Sect. 3.7.

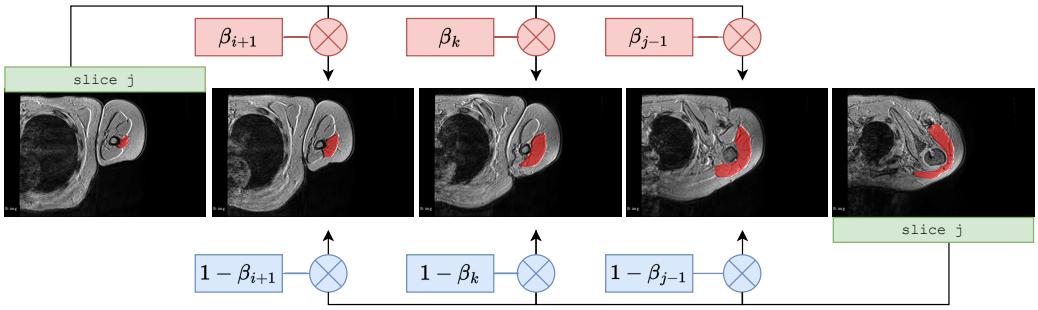


Figure 3: Label propagation framework scheme. Labels are propagated in both direction, resulting in two prediction for intermediate slices, that are fused using a weighting function β. Please refer to the text for complete description of notations.

Instead of distance, we hypothesize a correlation between image similarity and segmentation quality by considering now $\beta_k(\mathcal{X}, \phi)$ as a function of the propagated slices $\phi_{i\to k}(\mathbf{x}_i)$ and $\phi_{j\to k}(\mathbf{x}_j)$, such that :

$$\beta_k(\mathcal{X}, \phi) = \frac{\texttt{sim}(\phi_{i\to k}(\mathbf{x}_i), \mathbf{x}_k)}{\texttt{sim}(\phi_{i\to k}(\mathbf{x}_i), \mathbf{x}_k) + \texttt{sim}(\phi_{j\to k}(\mathbf{x}_j), \mathbf{x}_k)}. \tag{5}$$

Three $\texttt{sim}(\phi_{i\to k}(\mathbf{x}_i), \mathbf{x}_k)$ functions as composition of the pixel-level function $\texttt{NCC}(.,.)$ have been investigated: (1) the average cross-correlation map $\mathbf{p}$: $\frac{1}{N}\sum_p^N \texttt{NCC}(\phi_{i\to k}(\mathbf{x}_i), \mathbf{x}_k)(p)$, denoted as *correlation-based weighting* `(CW)`; (2) the average of the masked map $\frac{\sum_p^N \phi_{i\to k}(\mathbf{y}_i)(p)\texttt{NCC}(\phi_{i\to k}(\mathbf{x}_i),\mathbf{x}_k)(p)}{\sum_p^N \phi_{i\to k}(\mathbf{y}_i)(p)}$, denoted as *masked correlation-based weighting* `(MCW)`; (3) the pixel-level map $\texttt{NCC}(\phi_{i\to k}(\mathbf{x}_i))$ where each pixel is weighted independently, denoted as *pixel-wise correlation-based weighting* `(PCW)` in Sect. 3.7.

3. Experiments and results

3.1. MR muscle datasets

We evaluate the proposed approach on two muscle MR datasets, acquired from two distinct anatomies: shoulder and thigh.

The first one (referred to as `MR Shoulders`) is a shoulder dataset collected from a previous IRB approved study [27] originally established to investigate

the muscle volume-strength relationship in 12 children with unilateral obstetrical brachial plexus palsy (averaged age of 12.1±3.3 years). Informed consents from a legal guardian and assents from the participants were obtained for all subjects. Data from children with atrophic muscles exhibit a large variability in muscle structure and texture between subjects, which represents a challenge for learning-based segmentation algorithms. For each subject, 3D axial-plane T1-weighted gradient-echo MR images were acquired for the affected shoulder. For each MR volume, a sparse set of 2D axial slices were selected to delineate four rotator cuff muscles: deltoid, infraspinatus, supraspinatus and subscapularis (Fig.4, *top*). Manual annotations were performed by an expert in physical medicine and rehabilitation. Size for axial slices is constant for each subject (416×312 pixels). Resolution varies from 0.55×0.55 to 0.63×0.63mm. The number of axial slices fluctuates from 192 to 224 while slice thickness is fixed to 1.2mm. The average axial slice number of interest is 106±28 for deltoid, 63±16 for infraspinatus, 24±8 for subscapularis and 71±16 for supraspinatus.

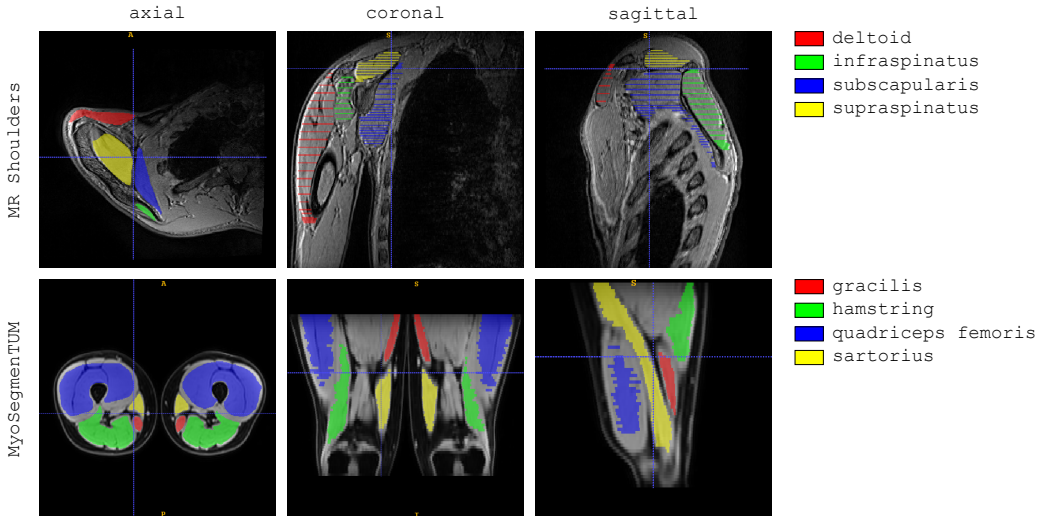


Figure 4: Sample images from the two evaluated shoulder MR dataset in each plane, with corresponding labels. (*Top*) 3D MR scan of a pediatric shoulder of a child with obstetrical brachial plexus palsy from the `MR Shoulders` dataset. Four muscles of the rotator cuff are represented in color: deltoid, infraspinatus, subscapularis and supraspinatus muscles. (*Bottom*) 3D MR scan of a healthy adult thigh extracted from the `MyoSegmenTUM` dataset. Four muscle groups are represented in color: gracilis, hamstring, quadriceps femoris and sartorius. Each muscle is differentiated by leg (left-right).

The second dataset employed to evaluate our method is the publicly available `MyoSegmenTUM` dataset [28], consiststing of Dixon water MR scans of thighs, with dense segmentation of four muscles: quadriceps femoris, sartorius, gracilis, hamstring, differentially marked for left and right thigh (Tab.4, *bottom*). The dataset is composed of a population of 15 healthy adults (80% of males) with age, weight, and height of resp. 29±8 years old, 86±13kg, and 180±9cm. Image size varies from 400×400 (63 axial slices) to 560×560 (65 axial slices), while resolution is 0.93×0.93 with slice thickness of 4mm. The average number of annotated axial slices is 48±10, all muscles included.

3.2. Implementation details and evaluation

A U-Net architecture is used for the velocity field generator, designed as a contraction/expansion path with skip-connections. It takes as input two successive images of the volume and estimates as output two maps representing the velocity fields in both directions of the plane. The displacement field is obtained by repeatedly integrating the velocity field over small-time steps, following the method described in the DARTEL [26] paper. The velocity field is first scaled by a factor of $\frac{1}{2^N}$ where N is the number of squaring steps. Then, the scaled field is integrated by self-application N times, which is considered an approximation of the computation of the exponential of the velocity field. The number of squaring steps is set to 7, as suggested in DARTEL method [26]. The implementation used is a publicly available U-Net network from the MONAI library[2]. The number of features is set to $(16, 32, 32, 32)$ and the number of residual units to 2, other parameters being kept by default. Final convolution layer, velocity field integration, deformation field mapping function and original losses are based on the official Voxelmorph Pytorch implementation[3]. As in [24] when using `NCC`, the weighting parameter λ_1 (Eq. 1) is set to 1. λ_2 and λ_3 is empirically set to 1. NCC sliding-window size is set to 9×9 in $\mathcal{L}_{sim}$ and $\mathcal{L}_{comp}$ while set to 31×31 in Eq. 5. The total number of parameters is similar to the one involved in Voxelmorph ($114k$ versus $110k$).

For all experiments, the model is trained using the Adam optimizer with a learning rate fixed at 10^{-3} and a weight decay of 10^{-8}. The best trained model over 200 epochs is selected by keeping the one that provides the highest

[2] `https://docs.monai.io/en/stable/networks.html#unet`

[3] `https://github.com/voxelmorph/voxelmorph`

Dice score between propagated labels and ground-truth annotated training slices. The prediction evaluated is the segmentation map resulting from the fusion of the propagation in both directions, according to Eq. 4, using `PCW` function.

To provide a comprehensive assessment, the metrics employed for evaluation purposes are the Dice coefficient (`DSC`), the average surface distance (`ASD`) as well as the Hausdorff distance (`HD`), computed in 2D according to the axial plane.

3.3. Existing methods

The proposed approach is compared with two segmentation methods: a morphology-based interpolation method [17] as well as a supervised U-Net deep segmentation model [29]. The morphology-based approach is an iterative method that can explicitly handle inter-slice topology changes by decomposing a many-to-many correspondence. This method is very suitable in the scenario considered in this work, namely a semi-automatic segmentation strategy which propagates annotated slices within the same volume. This can thus be considered as a reference method. A publicly available code is provided by Kitware in ITK[4]. This approach is denoted as Morph in Tab. 1 and Tab. 2.

The second method is a supervised DL technique, relying on the U-Net architecture, similar to previous studies [7, 30, 31]. The network is trained using a cross-entropy loss and data augmentation with random affine transformations (rotations between -20 and 20 degrees, scaling from 0.8 to 1.2 and shearing from -20 to 20). In the context of semi-automatic segmentation, the U-Net model is trained during 200 epochs and tested in a supervised learning fashion for each subject independently. The implementation is based on a publicly available Pytorch implementation[5]. It follows the original architecture and number of filters, with the addition of a batch normalization operation after each convolution. The number of parameters is approximately 14 million. This approach is denoted as U-Net in Tab. 1 and Tab. 2.

The final method presented in the paper is a semi-automatic deep learning approach for video label propagation tasks. It uses space-time graphs to represent video sequences, where patches from each frame are the nodes and

[4] `https://github.com/KitwareMedical/ITKMorphologicalContourInterpolation`
[5] `https://github.com/milesial/Pytorch-UNet`

transitions between patches in adjacent frames are the edges. The model learns a representation that defines the transition probability of a random walk on the graph. The representation is optimized to place high probability along paths of similar nodes, and is trained using cycle-consistency without supervision. At test time, the first frame annotation is propagated to the entire sequence using the most likely transitions for each patch. To remain fair to our method, we apply the same label propagation framework during inference, described in Sec. 2.3, propagating for each subsequence (i.e. between two annotated slices) in both directions, and merging the predicted segmentations as a function of distance from the nearest annotated slice (see DW function in Sec. 2.3). A model is trained for each dataset separately, using every sequence without annotation. Models are initialized using pretrained weights provided in the public implementation[6] of the method, and trained during 200 epochs. During training, sequences are divided into clips of 8 frames, dropout is set to 0.1, softmax temperature to 0.05, and learning-rate and batch size are respectively set to 10^{-4} and 32. For data augmentation, we use the random flip function provided in the code, which provides the best performance in our case. Other hyperparameters are set to default. The number of parameters is approximately 2.7 million. This approach is denoted as VideoWalk in Tab. 1 and Tab. 2.

3.4. Intra-Subject Segmentation with Minimal Supervision

In this section, we consider a difficult intra-subject segmentation scenario: only three slices are annotated for each MR volume to segment. The middle slice allows a fine delineation of the muscle while the two other slices provide its spatial extent (in z-direction). In this case, the typical distance between two consecutive annotations is 32 ± 17 for shoulders and 24 ± 5 for thighs. These three slices are the only annotation seen for each methods, as we perform the experiment for each subject separately.

Tab. 1 reports the `DSC`, `HD` and `ASD` scores for the shoulder dataset averaged over each slice and muscle. With `DSC` scores of 76.6, 81.0, 80.7, and 72.7% for deltoid, infraspinatus, subscapularis, and supraspinatus muscles, respectively, our method outperforms interpolation-based techniques [17], U-Net [29] and VideoWalk [14] for all muscles. Significant improvements in `ASD` stability (3.0 ± 4.6 versus 5.3 ± 16.8 mm with [29] for supraspinatus) and `HD`

[6] `https://github.com/ajabri/videowalk`

(18.8±16.1 versus 32.8±26.4 mm with [17] for infraspinatus) are achieved. Overall, VideoWalk shows the lowest performance on this shoulder dataset, presumably because the approach is not suitable for the propagation of fine structures on low contrast images such as these T1w images. . The DSC scores for supraspinatus are overall lower than for the other muscles due to its thin and elongated shape. For a more complete assessment, we also include in Tab. 1 the results from [7], based on a U-Net model trained in a leave-one-out (LOO) manner, i.e., using every subject except the one tested. Compared to [7], our method shows better results for all muscles despite a much smaller amount of annotated data. A strong improvement in DSC is moreover visible for infraspinatus (from 71.4 to 81.0% in DSC) although only 3 slices are used for label propagation.

Similar conclusions can be noted for MyoSegmenTUM in Tab. 2, where the proposed approach consistently outperformed other methods in DSC, ASD, and HD metrics. Specifically, we find that small muscles such as gracilis and sartorius challenge both U-Net [29] and interpolation-based technique [17] in this semi-automatic intra-subject segmentation case (10.2 for [29] and 30.6% for [17] in average DSC for gracilis and sartorius muscles and left/right legs versus 81.4% for our approach). Thanks to the good contrast of the water sequences in this dataset, the VideoWalk [14] approach achieves better performance on these small muscles in DSC, but remains significantly less accurate compared to our method (47.0% for gracilis and sartorius muscles and left/right legs versus 81.4% for our approach).

The proposed method shows similar ASD scores with [17] on larger muscles, but shows significant improvements in HD (12.9±12.0 versus 30.7±27.2 mm for [17] on right quadriceps femoris). The VideoWalk method [14] shows similar performance for the global quadriceps femoris in terms of DSC, with a score of 82.9% for their method compared to 86.7% for our method. However, their method did not perform as well in terms of ASD and HD, with scores of 5.3 in ASD and 26.7 in HD compared to our scores of 3.1 in ASD and 13.0 in HD.

Moreover, Fig.5 shows the evolution in the z-direction of the averaged DSC of both datasets. It appears that the proposed label propagation is less sensitive to distance from the nearest annotated slice, compared to interpolation-based approaches [17], U-Net [29] and VideoWalk [14]. Results also demonstrate a higher overall robustness with our method, as evoked by the standard deviation (colored area around the mean curves).

		no learning	intra-subject learning			LOO learning
metric	muscle	Morph [17]	U-Net [29]	VideoWalk [14]	Proposed	[7]
DSC ↑	deltoid	45.4±29.2	56.8±28.7	49.7±32.6	**76.3±15.5**	68.9±29.9
	infraspinatus	53.4±27.6	67.0±27.7	46.1±31.6	**81.0±11.5**	71.4±24.7
	subscapularis	55.6±30.5	66.3±25.5	56.7±25.9	**80.7±12.1**	78.1±18.1
	supraspinatus	38.6±25.1	50.2±29.0	39.1±29.1	**72.7±18.3**	64.9±28.0
ASD ↓	deltoid	3.4±5.7	3.2±9.0	4.8±4.9	**2.9±3.2**	-
	infraspinatus	3.6±7.5	**1.8±4.3**	7.3±8.8	1.9±1.3	-
	subscapularis	**2.1±4.2**	2.3±7.8	7.8±5.9	2.4±1.2	-
	supraspinatus	4.0±5.5	5.3±16.8	7.4±6.6	**3.0±4.6**	-
HD ↓	deltoid	39.6±32.3	42.7±44.2	37.8±35.2	**27.6±27.4**	-
	infraspinatus	32.8±26.4	25.7±29.6	36.1±28.8	**18.8±16.1**	-
	subscapularis	18.6±12.9	20.6±21.4	31.2±20.6	**12.8±9.3**	-
	supraspinatus	33.9±19.5	35.9±33.3	33.7±21.5	**22.5±17.9**	-

Table 1: Quantitative assessment over the MR Shoulder dataset in the minimal supervision setting. Best scores are highlighted in bold.

3.5. Ablation study

We now evaluate the contribution of $\mathcal{L}_{seg}$ and $\mathcal{L}_{comp}$ losses and the propagation schemes: forward (F) model, backward (B) model (i.e. propagated in the opposite direction as F) and the fusion of these two predictions as described in Eq. 4). To this end, 7 annotated slices per subject are used for training. Results are reported in Tab. 3 for the two datasets. First, it can be seen that, as expected, adding a segmentation-based loss improves the label propagation quality, especially while propagating in a single direction (82.3±14 versus 71.1±21.2 in DSC for MR Shoulders in backward direction and 90.5±7.1 versus 83.8±15.4 for MyoSegmenTUM in forward direction). Second, adding a composition loss to regularize the estimation of the deformation fields does not lead to a significant improvement in the segmentation results. Finally, the bidirectional fusion strategy allows in a large majority of cases an improvement of the segmentation results, and more particularly when $\mathcal{L}_{seg}$ is not used (DSC scores of 82.2±13.9 versus 73.7±19.2 and 71.1±21.2 in single directions for MR Shoulders, 90.1±7.4 versus 83.8±15.4 and 86.5±10.3 for MyoSegmenTUM).

3.6. Pre-training

The training stage of the proposed method can be performed in a subject-specific label propagation context for each volume independently, similarly to [18]. However, it is possible to train the neural network from the available unannotated data set as well. We focus in this section on the influence

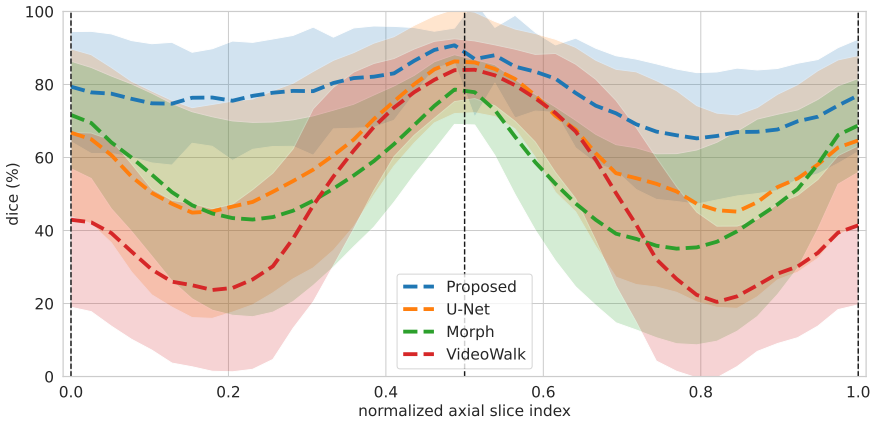
			no learning	intra-subject learning		
metric	muscle	leg	Morph [17]	U-Net [29]	VideoWalk [14]	Proposed
DSC ↑	gracilis	R	27.2±29.6	8.7±21.5	40.8±30.4	**83.6±12.2**
		L	27.8±30.9	2.5±7.5	38.0±30.3	**80.6±16.6**
	hamstring	R	78.6±14.3	36.6±31.3	76.9±14.5	**88.0±10.2**
		L	78.0±15.8	34.7±29.1	77.6±13.2	**86.7±11.0**
	quadriceps femoris	R	76.8±18.2	38.5±31.8	83.1±9.7	**87.4±8.8**
		L	75.6±20.0	37.0±35.2	82.7±10.8	**86.0±10.1**
	sartorius	R	36.1±31.0	12.1±20.6	56.9±25.5	**82.9±16.8**
		L	31.5±27.4	17.7±27.0	52.5±27.4	**78.5±21.4**
ASD ↓	gracilis	R	7.8±3.7	152.4±70.5	2.0±1.9	**0.9±0.4**
		L	7.9±3.9	174.7±60.3	2.1±2.0	**1.1±0.9**
	hamstring	R	3.1±1.3	90.1±71.1	4.8±4.4	**2.1±0.9**
		L	3.1±1.7	94.9±64.5	5.1±4.2	**2.2±1.0**
	quadriceps femoris	R	3.3±2.7	113.3±73.3	5.2±4.6	**2.9±2.5**
		L	**3.2±1.6**	112.5±80.6	5.4±4.7	3.3±3.0
	sartorius	R	9.3±6.2	137.2±76.0	2.8±2.9	**1.5±2.7**
		L	9.5±5.8	116.4±83.7	2.7±2.1	**1.7±1.7**
HD ↓	gracilis	R	15.2±5.3	263.6±104.7	7.8±7.4	**3.1±1.9**
		L	15.6±5.9	295.3±69.7	8.1±7.8	**3.8±3.8**
	hamstring	R	16.5±12.0	210.8±131.6	20.3±15.1	**10.3±8.4**
		L	16.9±12.8	229.2±114.2	21.3±14.9	**11.3±10.1**
	quadriceps femoris	R	30.7±27.2	251.6±114.7	24.9±22.4	**12.9±12.0**
		L	28.4±26.8	249.8±141.7	28.6±28.3	**13.1±12.8**
	sartorius	R	24.7±14.9	259.7±122.3	9.8±8.5	**4.8±5.7**
		L	25.4±13.1	218.9±141.1	10.5±8.7	**5.1±4.3**

Table 2: Quantitative assessment over the MyoSegmenTUM dataset in the minimal supervision setting. Best scores are highlighted in bold.

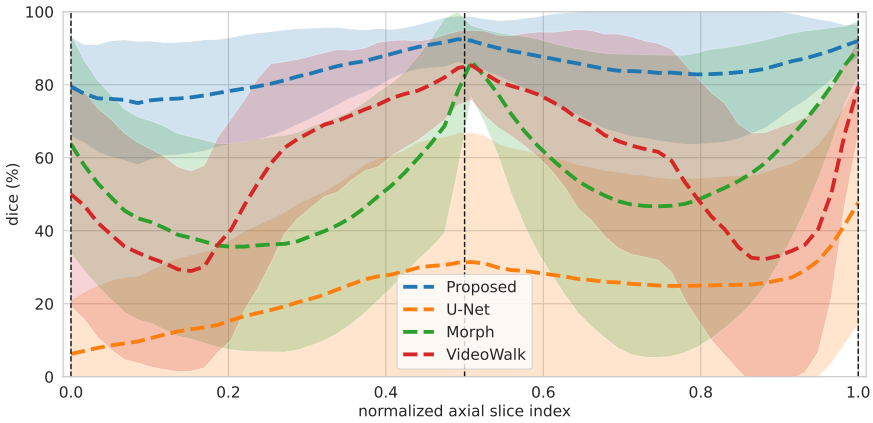
of such a pre-training step. More precisely, we evaluate performance and convergence time with or without pre-training. For a given dataset and a given propagation direction, the pre-training stage corresponds to a joint learning of every subject successive deformations, without any annotation. This corresponds then to minimizing Eq. 1 for each slice of each subject, in both directions.

We consider the minimal supervision learning scenario described as in Sect. 3.4. Results for both datasets are reported in Tab. 4. The experiment shows the interest of pre-training in a difficult context of learning from only 3 slices, with improvements in mean DSC score from 2.7 to 7.2% for MR Shoulders, and from 4.0 to 22.6% for MyoSegmenTUM, as well as for other metrics.

DSC convergence over epochs on the MR shoulder dataset is studied in



(a) `MR Shoulders`



(b) `MyoSegmenTUM`

Figure 5: Averaged Dice (`DSC`) for each annotated slice over the `MR Shoulders` (a) and `MyoSegmenTUM` (b) datasets, in a minimal supervision setting. DSC is displayed with respect to the normalized axial slice number obtained by linearly scaling slice number from $[z_{min}, z_{max}]$ to $[0, 1]$ where z_{min} (z_{max}) is the minimal (maximal) slice index displaying a muscle. Vertical dotted black lines represents the location of annotated slices used for training. Colored areas deal with standard deviation.

Fig. 6. It shows that the pre-training stage provides a good initialization (58.1 versus 41.1 at epoch 0). Moreover, the pre-trained models reach on

			MR Shoulders			MyoSegmenTUM		
direction	$\mathcal{L}_{seg}$	$\mathcal{L}_{comp}$	DSC ↑	HD ↓	ASD ↓	DSC ↑	HD ↓	ASD ↓
F			73.7±19.2	22.4±23.0	5.1±8.2	83.8±15.4	8.0±10.7	1.6±1.0
B	✗	✗	71.1±21.2	23.3±25.1	3.4±4.2	86.5±10.3	7.6±10.6	2.0±3.1
Fused			**82.2±13.9**	**14.0±15.8**	**2.3±4.3**	**90.1±7.4**	**5.4±8.1**	**1.3±1.8**
F			83.8±11.1	15.5±19.0	2.7±4.8	90.5±7.1	5.9±9.6	**1.2±0.7***
B	✓	✗	82.3±14.0	16.5±22.1	**1.9±2.9**	90.6±6.7	5.8±9.3	1.3±1.5
Fused			**84.4±11.6**	**13.1±16.9**	2.0±3.7	**91.1±6.4***	**5.0±7.9**	1.2±1.2
F			76.4±18.4	21.8±23.0	4.8±7.9	84.5±15.6	8.1±10.7	1.7±1.1
B	✗	✓	74.2±20.0	22.6±24.6	3.3±4.6	87.4±9.7	7.5±10.6	1.9±2.8
Fused			**83.1±13.3**	**14.0±15.8**	**2.3±4.2**	**89.9±8.3**	**5.6±8.5**	**1.4±1.8**
F			84.6±10.8	15.3±18.3	2.7±4.7	90.3±7.2	6.0±9.6	**1.2±0.7***
B	✓	✓	83.2±13.8	16.5±22.4	1.9±4.0	90.6±6.4	5.8±9.1	1.3±1.6
Fused			**85.3±11.2***	**12.3±15.4***	**1.8±3.4***	**91.0±6.4**	**4.9±7.5***	1.2±1.2

Table 3: Ablation study. Influence of $\mathcal{L}_{seg}$ and $\mathcal{L}_{comp}$ losses for different propagation schemes. Best results for a given loss combination is in bold. The asterisk (*) indicates the best results over all loss combinations.

average per epoch performance of non pre-trained models from 14 epochs (70.5 at epoch 14 versus 70.0 at epoch 175).

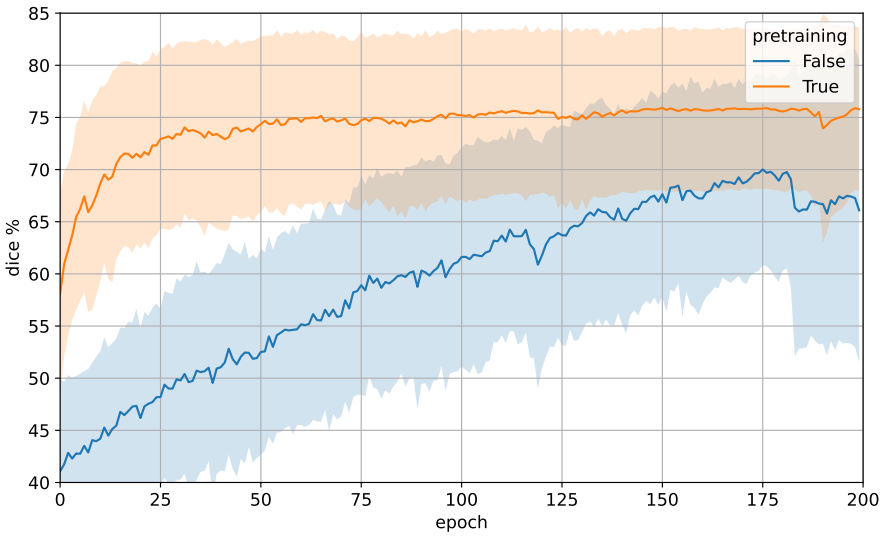


Figure 6: Performance evolution over training epochs on MR shoulders dataset, with or without pretraining. Shown metric is the 2D DSC averaged over all subjects and muscles. Colored areas deal with standard deviation.

	DSC ↑		ASD ↓		HD ↓	
pre-training	No	Yes	No	Yes	No	Yes
muscle						
deltoid	69.5±16.4	**76.3±15.5**	4.4±4.4	**2.9±3.2**	34.2±29.7	**27.6±27.4**
infraspinatus	78.3±12.4	**81.0±11.5**	2.6±2.2	**1.9±1.3**	21.9±18.9	**18.8±16.1**
subscapularis	77.8±13.3	**80.7±12.1**	3.2±1.8	**2.4±1.2**	15.2±9.7	**12.8±9.3**
supraspinatus	65.5±21.9	**72.7±18.3**	4.7±6.1	**3.0±4.6**	25.4±19.5	**22.5±17.9**

(a)

		DSC ↑		ASD ↓		HD ↓	
	pre-training	No	Yes	No	Yes	No	Yes
muscle	leg						
gracilis	L	64.8±24.3	**80.6±16.6**	2.4±2.5	**1.1±0.9**	7.8±7.3	**3.8±3.8**
	R	61.0±30.7	**83.6±12.2**	2.1±2.2	**0.9±0.4**	7.0±5.6	**3.1±1.9**
hamstring	L	82.5±12.6	**86.7±11.0**	3.0±1.4	**2.2±1.0**	13.0±7.8	**11.3±10.1**
	R	84.0±11.6	**88.0±10.2**	2.9±1.5	**2.1±0.9**	12.7±7.2	**10.3±8.4**
quadriceps femoris	L	81.6±13.6	**86.0±10.1**	4.9±4.3	**3.3±3.0**	18.4±14.0	**13.1±12.8**
	R	83.4±10.8	**87.4±8.8**	4.8±4.3	**2.9±2.5**	19.8±14.2	**12.9±12.0**
sartorius	L	61.7±28.2	**78.5±21.4**	4.4±6.3	**1.7±1.7**	12.9±13.1	**5.1±4.3**
	R	61.9±29.8	**82.9±16.8**	4.2±5.7	**1.5±2.7**	12.4±12.9	**4.8±5.7**

(b)

Table 4: Evaluation of pre-training step on `MR Shoulders` (*a*) and `MyoSegmenTUM` (*b*) datasets. Best results are highlighted in bold.

3.7. Study of weighting strategies

In this experiment, we compare performances of each weighting strategies described in above Sect. 2.3. We study their influence on both datasets, and with a different number of annotated slices per subject : 3, 5 and 7.

Results for `DSC`, `ASD` and `HD` scores are reported in Tab. 5. Pixel-wise fusion (`PCW`) shows the best average improvements compared to the baseline distance fusion, with `DSC` gains from 0.1 to 1.2% and up to -1.3 in `HD`. In `MR Shoulders`, the local similarity strategy (`MCW`) reaches the lowest `ASD` even though it gives the lowest `DSC` as well.

Visual comparison between weighting strategies based on distance and pixel-wise similarity is provided in Fig. 9. The `PCW` approach shows more robustness to structure splitting of the quadriceps femoris (blue label) and hamstring.

		MR Shoulders			MyoSegmenTUM		
annotated slices	strategy	DSC ↑	HD ↓	ASD ↓	DSC ↑	HD ↓	ASD ↓
3	DW	75.9±16.7	22.8±24.7	2.6±3.1	83.8±14.6	9.3±12.5	1.9±1.3
	MCW	75.3±19.9	22.3±24.0	**2.3±2.8**	83.1±17.8	8.8±11.3	2.1±2.9
	CW	76.6±16.3	22.4±24.2	2.5±3.0	**84.5±14.3**	8.7±11.6	**1.8±1.2**
	PCW	**77.1±15.5**	**21.5±20.4**	2.6±3.2	84.3±14.3	**8.2±9.3**	2.0±2.1
5	DW	82.5±13.1	16.1±20.8	1.9±2.8	90.7±7.6	5.1±7.6	**1.1±1.0**
	MCW	82.5±15.0	15.2±19.6	**1.8±2.8**	90.5±9.8	**4.8±6.9**	1.1±1.2
	CW	83.0±13.1	15.7±20.7	**1.8±2.8**	90.9±7.4	4.9±7.2	1.2±1.3
	PCW	**83.3±11.9**	**15.2±17.5**	2.2±3.7	**90.9±7.3**	4.9±7.4	1.2±1.4
7	DW	84.8±11.6	13.4±18.2	1.7±3.0	92.0±6.8	**4.3±6.8**	**1.0±1.1**
	MCW	84.6±13.9	12.8±17.5	**1.6±2.8**	92.0±6.9	4.3±7.1	1.0±1.2
	CW	**85.3±11.2**	13.1±17.9	1.7±2.9	92.0±6.9	**4.3±6.8**	**1.0±1.1**
	PCW	**85.3±11.2**	**12.3±15.4**	1.8±3.4	**92.1±6.7**	4.4±7.4	1.0±1.3

Table 5: Performance table summarizing average metrics per dataset, number of annotated slices, and by weighting strategy. Please refer to the Sect. 2.3 for details about weighting strategies

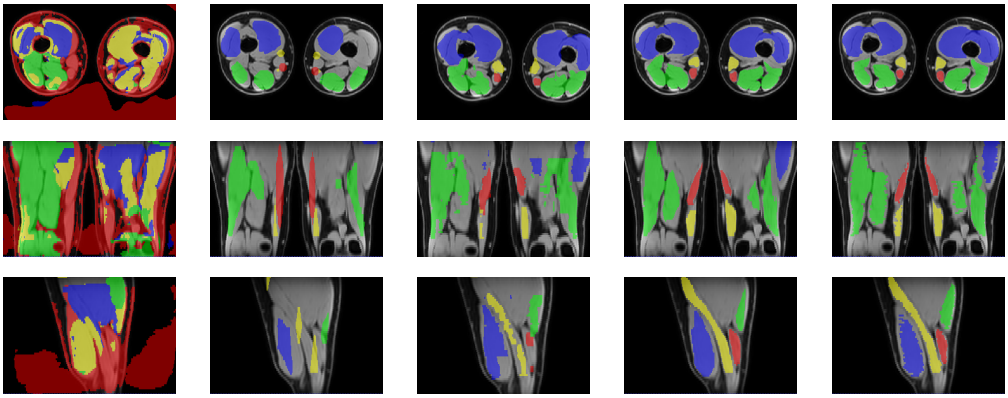


Figure 7: Visual comparison (first row : axial, second : sagittal, last : coronal) on MyoSegmenTUM dataset of existing methods and our approach in the minimal supervision setting (Sect. 3.3).

3.8. Influence of annotated slice spacing

In this work, annotated slices are uniformly distributed across the volume, with the first and last annotations defining the region of interest to be segmented. Intermediate slices are then chosen to divide the MR volumes into approximately equal sub-volumes. In Sec. 3.4 and 3.6, only 3 annotated

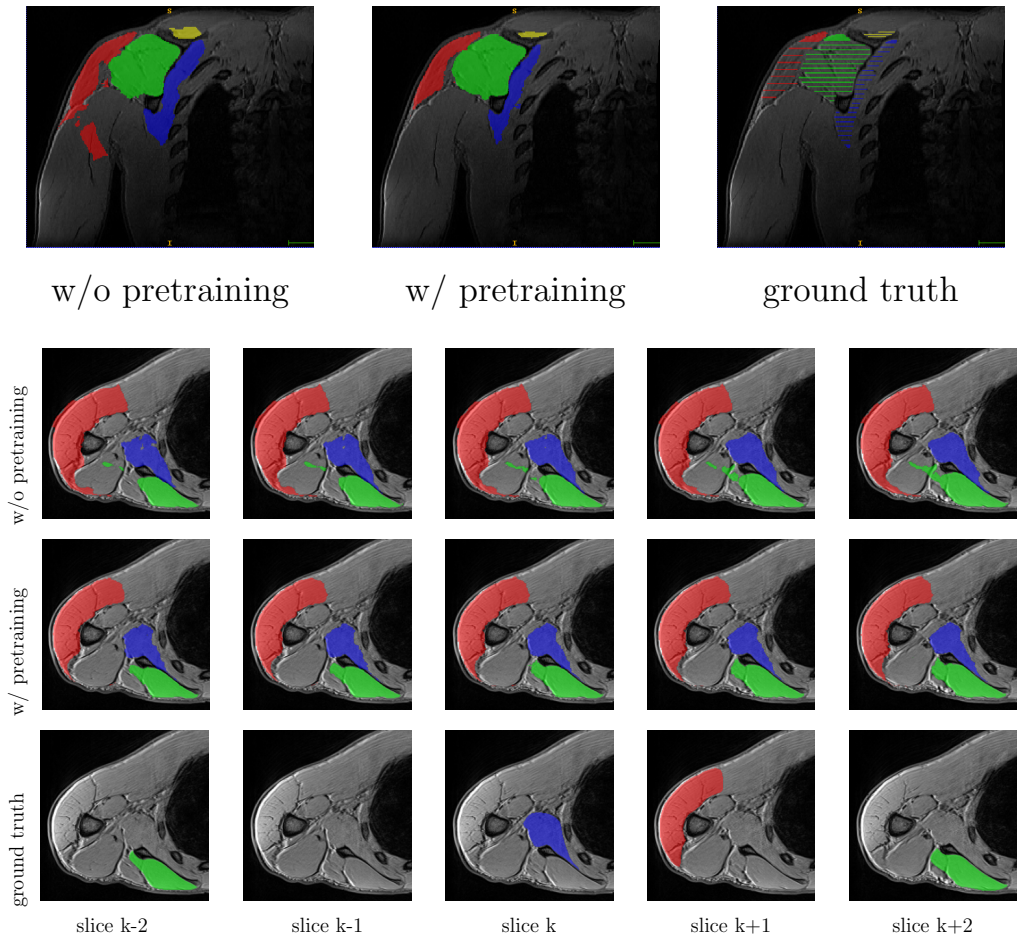


Figure 8: Visual comparison (first row : coronal plane, last rows : axial) on MR Shoulder dataset of pretraining stage influence in the minimal supervision setting.

slices are used, while the number of annotated slices is set to 7 in Sec. 3.5. In Sec. 3.7, we have studied the influence of the number of annotated slices per volume: 3, 5, and 7 slices. In the previous sections, the experiments were performed with a fixed number of annotated slices, rather than an interval between annotations, to demonstrate the application of our approach for a fixed interaction time, invariant to the subject morphology. The advantage of this approach is that the user knows how long it takes to segment a dataset using the proposed propagation method. However, the same anatomical region of interest may occupy a different number of slices depending on the

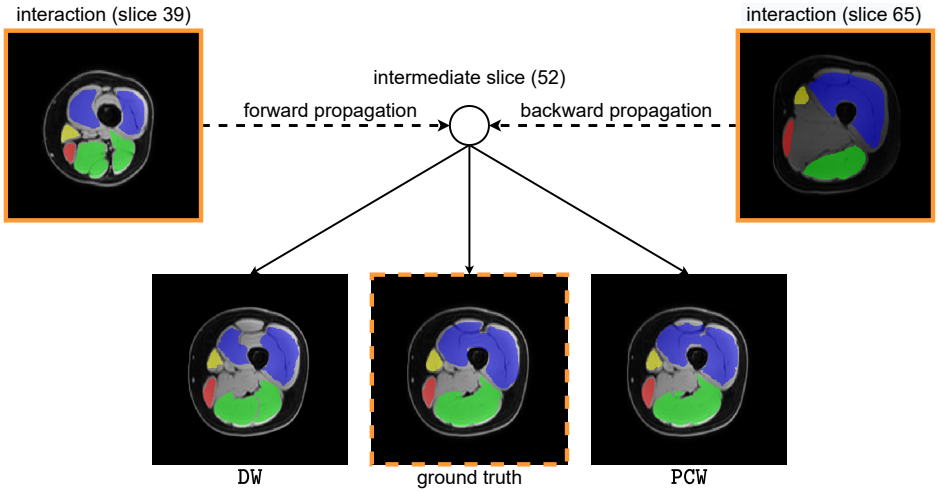
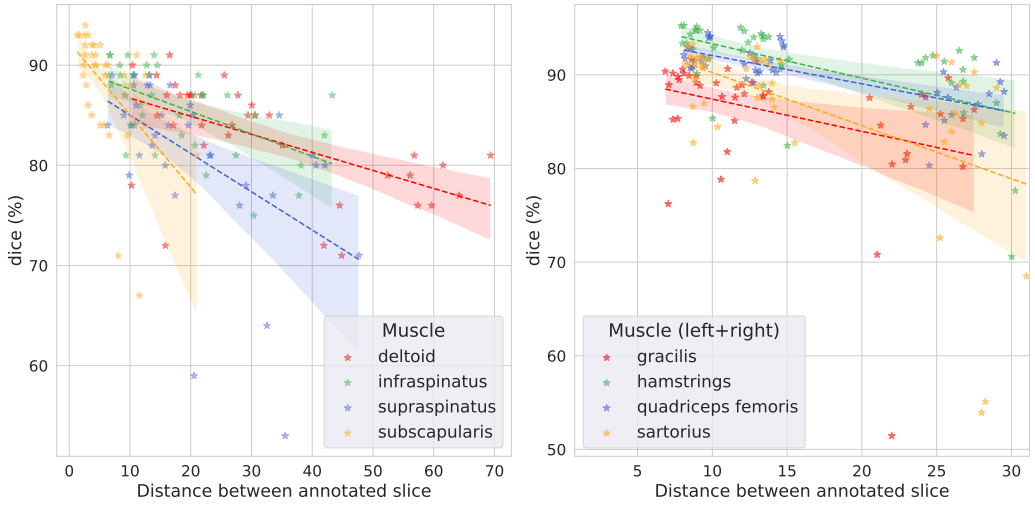


Figure 9: Visual comparison of axial slice predictions for different weighting strategies on MyoSegmenTUM dataset, in the three annotated slices setting. Images surrounded by solid lines yellow lines are annotated slices used for training and propagation.

subject's height, weight, or age. Longer muscles, for a given slice thickness, therefore imply larger propagation distances for a given number of annotated slices. In this section, we examine the performance of the proposed approach as a function of the distance between annotated slices, with the hypothesis that a greater propagation induces an accumulation of registration errors and thus a lower segmentation accuracy. Figure 10 shows the results of the experiment described in Sec. 3.7 with the pixel-wise correlation-based weighting `PCW` strategy for the three configurations of the number of annotated slices (3, 5 and 7) according to the propagation distance. Each point corresponds to the average DICE for a given subject and muscle as a function of the distance between two annotated slices used for propagation. As expected, the smaller the distance between the annotated slices, the better the segmentation is propagated. For example, it can be noted that for the deltoid, the DICE remains above 75% of `DSC` despite large distances between annotated slices (up to 69 slices). Figure 10 shows the variability of performance depending on the muscle to be segmented. The choice of the optimal distance between the annotated slices depends on the anatomy of the subject, the size of the structures to be segmented but also the contrast in the images. These results require further investigation. Indeed, these results should be interpreted with caution, as the `DSC` coefficient is sensitive to the size of the structure to be

segmented. The determination of the optimal distance for each muscle and each subject remains an avenue to be explored for future work, for example with a view to using the approach in a routine clinical case.



(a) `MR Shoulders` (b) `MyoSegmenTUM`

Figure 10: Average Dice (`DSC`) per subject as a function of the propagation distance, over the `MR Shoulders` (a) and `MyoSegmenTUM` (b) datasets. Dotted lines show linear regressions per label, with 95% confidence interval in colored area.

4. Conclusion

In this paper, we describe a deep registration-based label propagation method for intra-subject muscle segmentation from MR images. The propagation process is guided by image intensity, muscle shape and registration consistency. This operation is performed in both directions, resulting in two segmentation predictions that are merged with image registration quality estimation as weighting guidance. To improve performance and speed up convergence, an unsupervised pretraining stage is suggested. To obtain smooth and accurate fusions of segmentations from bidirectional propagation, several weighting functions have been studied. Experiments on clinical data show that 3D full scan segmentation can be achieved from very limited manual annotations using the proposed approach. One limitation of using diffeomorphic registration for anatomical structure segmentation is that it does not

adequately handle topological discontinuities. In such cases, the annotated slices must accurately describe these changes in topology. A potential solution to this problem is to incorporate an active learning mechanism that guides the annotator in selecting the most informative slices for manual segmentation. On the other hand, our approach relies on context information in the form of annotated extreme slices. This requirement can be challenging for annotators because extreme slices are often difficult to segment manually and provide little information about the label structure. One possible way to address this limitation is to investigate the use of weak annotations, such as clicks or scribbles, for these slices. In our study, we focused on single-axis propagation, which is more applicable to MR images where only one plane is likely to be isotropic. However, a multi-view propagation approach could be considered for isotropic volume segmentation, such as in CT or 3D ultrasound imaging. In the context of musculoskeletal medical images analysis, for which there is very little annotated data, the described approach can serve as a basis for the development of interactive user-friendly segmentation tools. It is also very likely adaptable to other modalities that benefit from high spatial resolution in all planes (CT, 3D US) or organs.

5. Acknowledgment

This work was partially funded by ANR (AI4Child project, grant ANR-19-CHIA-0015) and Innoveo from CHRU de Brest.

6. Compliance with ethical standards

Data acquisition was performed in line with the principles of the Declaration of Helsinki. Informed consents from the participants were obtained for all subjects. Authors declare that they do not have any conflicts of interest.

References

[1] S. I. Muzic, M. Paoletti, F. Solazzo, E. Belatti, R. Vitale, N. Bergsland, S. Bastianello, A. Pichiecchio, Reproducibility of manual segmentation in muscle imaging, Acta Myologica 40 (3) (2021) 116.

[2] M. I. Miller, G. E. Christensen, Y. Amit, U. Grenander, Mathematical textbook of deformable neuroanatomies, Proceedings of the National Academy of Sciences 90 (24) (1993) 11944–11948.

[3] B. Sambaturu, A. Gupta, C. Jawahar, C. Arora, Scribblenet: Efficient interactive annotation of urban city scenes for semantic segmentation, Pattern Recognition 133 (2023) 109011.

[4] F. Rousseau, P. A. Habas, C. Studholme, A supervised patch-based approach for human brain labeling, IEEE Transactions on Medical Imaging 30 (10) (2011) 1852–1862.

[5] A. C. Ogier, M.-A. Hostin, M.-E. Bellemare, D. Bendahan, Overview of MR image segmentation strategies in neuromuscular disorders, Frontiers in Neurology 12 (2021) 255.

[6] P.-H. Conze, C. Pons, V. Burdin, F. T. Sheehan, S. Brochard, Deep convolutional encoder-decoders for deltoid segmentation using healthy versus pathological learning transferability, in: International Symposium on Biomedical Imaging, 2019, pp. 36–39.

[7] P.-H. Conze, S. Brochard, V. Burdin, F. T. Sheehan, C. Pons, Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders, Computerized Medical Imaging and Graphics 83 (2020) 101733.

[8] W. Chen, Y. Li, B. A. Dyer, X. Feng, S. Rao, S. H. Benedict, Q. Chen, Y. Rong, Deep learning vs. atlas-based models for fast auto-segmentation of the masticatory muscles on head and neck CT images, Radiation Oncology 15 (1) (2020) 1–10.

[9] R. Cheng, M. Crouzier, F. Hug, K. Tucker, P. Juneau, E. McCreedy, W. Gandler, M. J. McAuliffe, F. T. Sheehan, Automatic quadriceps and patellae segmentation of MRI with cascaded U2-Net and SASSNet deep learning model, Medical Physics 49 (1) (2022) 443–460.

[10] R. Feng, X. Zheng, T. Gao, J. Chen, W. Wang, D. Z. Chen, J. Wu, Interactive few-shot learning: Limited supervision, better medical image segmentation, IEEE Transactions on Medical Imaging 40 (10) (2021) 2575–2588.

[11] J. Zhang, Y. Shi, J. Sun, L. Wang, L. Zhou, Y. Gao, D. Shen, Interactive medical image segmentation via a point-based interaction, Artificial Intelligence in Medicine 111 (2021) 101998.

[12] T. Sakinis, F. Milletari, H. Roth, P. Korfiatis, P. Kostandy, K. Philbrick, Z. Akkus, Z. Xu, D. Xu, B. J. Erickson, Interactive segmentation of medical images through fully convolutional neural networks, arXiv preprint arXiv:1903.08205 (2019).

[13] D. Al Chanti, V. G. Duque, M. Crouzier, A. Nordez, L. Lacourpaille, D. Mateus, IFSS-Net: Interactive few-shot siamese network for faster muscle segmentation and propagation in volumetric ultrasound, IEEE Transactions on Medical Imaging 40 (10) (2021) 2615–2628.

[14] A. Jabri, A. Owens, A. A. Efros, Space-time correspondence as a contrastive random walk, in: Advances in Neural Information Processing Systems, 2020.

[15] J. Sun, Y. Mao, Y. Dai, Y. Zhong, J. Wang, Munet: Motion uncertainty-aware semi-supervised video object segmentation, Pattern Recognition (2023) 109399.

[16] M. Azimbagirad, G. Dardenne, D. Ben Salem, O. Rémy-Néris, V. Burdin, Towards the definition of a patient-specific rehabilitation program for TKA: A new MRI-based approach for the easy volumetric analysis of thigh muscles, in: International Conference of the IEEE Engineering in Medicine & Biology Society, 2021, pp. 3141–3144.

[17] A. B. Albu, T. Beugeling, D. Laurendeau, A morphology-based approach for interslice interpolation of anatomical slices from volumetric images, IEEE Transactions on Biomedical Engineering 55 (8) (2008) 2022–2038.

[18] A. Ogier, M. Sdika, A. Foure, A. Le Troter, D. Bendahan, Individual muscle segmentation in MR images: A 3D propagation through 2D non-linear registration approaches, in: International Conference of the IEEE Engineering in Medicine and Biology Society, 2017, pp. 317–320.

[19] W. Feng, H. Nagaraj, H. Gupta, S. G. Lloyd, I. Aban, G. J. Perry, D. A. Calhoun, L. J. Dell'Italia, T. S. Denney, A dual propagation contours technique for semi-automated assessment of systolic and diastolic cardiac function by cmr, Journal of Cardiovascular Magnetic Resonance 11 (1) (2009) 1–13.

[20] F. Khalvati, A. Salmanpour, S. Rahnamayan, G. Rodrigues, H. R. Tizhoosh, Inter-slice bidirectional registration-based segmentation of the

prostate gland in MR and CT image sequences, Medical Physics 40 (12) (2013) 123503.

[21] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, X. Yang, Deep learning in medical image registration: a review, Physics in Medicine & Biology 65 (20) (2020).

[22] N. J. Tustison, B. B. Avants, J. C. Gee, Learning image-based spatial transformations via convolutional neural networks: A review, Magnetic Resonance Imaging 64 (2019) 142–153.

[23] D. Wei, L. Zhang, Z. Wu, X. Cao, G. Li, D. Shen, Q. Wang, Deep morphological simplification network (ms-net) for guided registration of brain magnetic resonance images, Pattern Recognition 100 (2020) 107171.

[24] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, A. V. Dalca, Voxelmorph: a learning framework for deformable medical image registration, IEEE Transactions on Medical Imaging 38 (8) (2019) 1788–1800.

[25] B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, Medical Image Analysis 12 (1) (2008) 26–41.

[26] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (1) (2007) 95–113.

[27] C. Pons, F. T. Sheehan, H. S. Im, S. Brochard, K. E. Alter, Shoulder muscle atrophy and its relation to strength loss in obstetrical brachial plexus palsy, Clinical Biomechanics 48 (2017) 80–87.

[28] S. Schlaeger, F. Freitag, E. Klupp, M. Dieckmeyer, D. Weidlich, S. Inhuber, M. Deschauer, B. Schoser, S. Bublitz, F. Montagnese, et al., Thigh muscle segmentation of chemical shift encoding-based water-fat magnetic resonance images: the reference database myosegmentum, PLoS One 13 (6) (2018) e0198200.

[29] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234–241.

[30] E. Ahmad, M. Goyal, J. S. McPhee, H. Degens, M. H. Yap, Semantic segmentation of human thigh quadriceps muscle in magnetic resonance images, arXiv preprint arXiv:1801.00415 (2018).

[31] J. Ding, P. Cao, H.-C. Chang, Y. Gao, S. H. S. Chan, V. Vardhanabhuti, Deep learning-based thigh muscle segmentation for reproducible fat fraction quantification using fat–water decomposition MRI, Insights Into Imaging 11 (1) (2020) 1–11.