Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders

Pierre-Henri Conze, Sylvain Brochard, Valérie Burdin, Frances T Sheehan-Gavelli, Christelle Pons

Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders

Pierre-Henri Conze[a,b,*], Sylvain Brochard[b,c,d], Valérie Burdin[a,b], Frances T. Sheehan[e], Christelle Pons[b,c,d]

[a] *IMT Atlantique, LaTIM UMR 1101, UBL, Technopôle Brest-Iroise, 29238 Brest, France*
[b] *Inserm, LaTIM UMR 1101, IBRBS, 22 rue Camille Desmoulins, 29238 Brest, France*
[c] *Rehabilitation Medicine, University Hospital of Brest, 2 avenue Foch, 29200 Brest, France*
[d] *SSR pediatric, Fondation ILDYS, Ty Yann, rue Alain Colas, 29218 Brest, France*
[e] *Rehabilitation Medicine, NIH, 10 Center Drive, MD 20892, Bethesda, USA*

Abstract

Fully-automated segmentation of pathological shoulder muscles in patients with musculo-skeletal diseases is a challenging task due to the huge variability in muscle shape, size, location, texture and injury. A reliable automatic segmentation method from magnetic resonance images could greatly help clinicians to diagnose pathologies, plan therapeutic interventions and predict interventional outcomes while eliminating time consuming manual segmentation. The purpose of this work is three-fold. First, we investigate the feasibility of automatic pathological shoulder muscle segmentation using deep learning techniques, given a very limited amount of available annotated pediatric data. Second, we address the learning transferability from healthy to pathological data by comparing different learning schemes in terms of model generalizability. Third, extended versions of deep convolutional encoder-decoder architectures using encoders pre-trained on non-medical data are proposed to improve the segmentation accuracy. Methodological aspects are evaluated in a leave-one-out fashion on a dataset of 24 shoulder examinations from patients with unilateral obstetrical brachial plexus palsy and focus on 4 rotator cuff muscles (deltoid, infraspinatus, supraspinatus and subscapularis). The most accurate segmentation model is partially pre-trained on the large-scale ImageNet dataset and jointly

*corresponding author: pierre-henri.conze@imt-atlantique.fr

exploits inter-patient healthy and pathological annotated data. Its performance reaches Dice scores of 82.4%, 82.0%, 71.0% and 82.8% for deltoid, infraspinatus, supraspinatus and subscapularis muscles. Absolute surface estimation errors are all below 83mm^2 except for supraspinatus with 134.6mm^2. The contributions of our work offer new avenues for inferring force from muscle volume in the context of musculo-skeletal disorder management.

Keywords: shoulder muscle segmentation, musculo-skeletal disorders, deep convolutional encoder-decoders, healthy versus pathological transferability, obstetrical brachial plexus palsy

1. Introduction

The rapid development of non-invasive imaging technologies over the last decades has opened new horizons in studying both healthy and pathological anatomy. As part of this, pixel-wise segmentation has become a crucial task in medical image analysis with numerous applications such as computer-assisted diagnosis, surgery planning, visual augmentation, image-guided interventions and extraction of quantitative indices from images. However, the analysis of complex magnetic resonance (MR) imaging datasets is cumbersome and time-consuming for radiologists, clinicians and researchers. Thus, computerized assistance methods, including robust automatic image segmentation techniques, are needed to guide and improve image interpretation and clinical decision making.

Although great strides have been made in automatically delineating cartilages and bones [1, 2], there is a great need for accurate muscle delineations in managing musculo-skeletal disorders. The task of segmenting muscles from MR images becomes more difficult when the pathology alters the size, shape, texture and global MR appearance of muscles [3] (Fig.1). Further, the large variability across patients, arising from age-related development and injury, impacts the ability to delineate muscles. To circumvent these difficulties, muscle segmentation is traditionally performed manually, in a slice-by-slice fashion [4]. However, manual segmentation is a time-consuming task and is often impre-

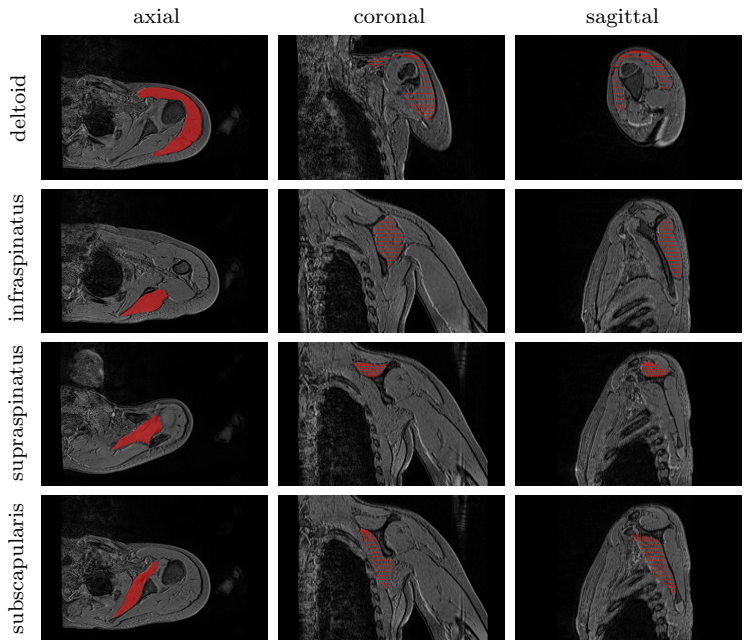

Figure 1: Groundtruth segmentation of pathological shoulder muscles including deltoid as well as infraspinatus, supraspinatus and subscapularis from the rotator cuff. Axial, coronal and sagittal slices are extracted from a 3D MR examination acquired for a child with obstetrical brachial plexus palsy.

cise due to intra- and inter-expert variability. Therefore, most musculo-skeletal diagnoses are based on 2D analyses of single images, despite the utility of 3D volume exploration. Recently, there has been a growing interest in developing automatic techniques for 3D muscle segmentation, particularly in the area of deploying deep learning methodologies using convolutional encoder-decoders [5].

Obstetrical brachial plexus palsy (OBPP), among the most common birth injuries [6], is one such pathology in which accurate 3D automatic muscle segmentation could help to quantify a patient's level of impairment, guide interventional planning or track treatment progress. OBPP occurs most often during the delivery phase when lateral traction is applied to the head to permit shoulder clearance [7]. It is characterized by the disruption of the peripheral nervous system conducting signals from the spinal cord to shoulders, arms and hands, with an incidence of around 1.4 every 1000 live births [8]. This nerve injury

leads to variable muscle denervation, resulting in muscle atrophy with fatty infiltration, growth disruption, muscle atrophy and force imbalances around the shoulder [9]. Treatment and prevention of shoulder muscle strength imbalances are main therapeutic goals for children with OBPP who do not fully recover [10]. Patient-specific information related to the degree of muscle atrophy across the shoulder is therefore needed to plan interventions and predict interventional outcomes. Recent work, reporting a clear relationship between muscle atrophy and strength loss for children with OBPP [6], demonstrates that an ability to accurately quantify 3D muscle morphology directly translates into an understanding of the force capacity of shoulder muscles. In this direction, shoulder muscle segmentation on MR images is needed to both quantify individual muscle involvement and analyze shoulder strength balance in children with OBPP.

Therefore, the purpose of our study is to develop and validate a robust and fully-automated muscle segmentation pipeline, which will support new insights into the evaluation, diagnosis and management of musculo-skeletal diseases. The specific aims are three-fold. First, we aim at studying the feasibility of automatically segmenting pathological shoulder muscle using deep convolutional encoder-decoder networks, based on an available, but small, annotated dataset in children with OBPP [6]. Second, our work addresses the learning transferability from healthy to pathological data, focusing particularly on how available data from both healthy and pathological shoulder muscles can be jointly exploited for pathological shoulder muscle delineation. Third, extended versions of deep convolutional encoder-decoder architectures, using encoders pre-trained on non-medical data, are investigated to improve the segmentation accuracy. Experiments extend our preliminary results [11] to four shoulder muscles including deltoid, infraspinatus, supraspinatus and subscapularis.

2. Related works

To extract quantitative muscle volume measures, from which forces can be derived [6], muscle segmentation is traditionally performed manually in a slice-

by-slice manner [4] from MR images. This task is extremely time-consuming and requires tens of minutes to get accurate delineations for one single muscle. Thus, it is not applicable for large volumes of data typically produced in research studies or clinical imaging. In addition, manual segmentation is prone to intra- and inter-expert variability, resulting from the irregularity of muscle shapes and the lack of clearly visible boundaries between muscles and surrounding anatomy [12]. To facilitate the process, a semi-automatic processing, based on transversal propagations of manually-drawn masks, can be applied [13]. It consists of several ascending and descending non-linear registrations applied to manual masks to finally achieve volumetric results. Although semi-automatic methods achieve volume segmentation in less time then manual segmentation, they are still time-consuming.

A model-based muscle segmentation incorporating a prior statistical shape model can be employed to delineate muscles boundaries from MR images. A patient-specific 3D geometry is reached based on the deformation of a parametric ellipse fitted to muscle contours, starting from a reduced set of initial slices [14, 15]. Segmentation models can be further improved by exploiting *a-priori* knowledge of shape information, relying on internal shape fitting and auto-correction to guide muscle delineation [16]. Baudin et al. [17] combined a statistical shape atlas with a random walks graph-based algorithm to automatically segment individual muscles through iterative linear optimization. Andrews et al. [18] used a probabilistic shape representation called generalized log-ratio representation that included adjacency information along with a rotationally invariant boundary detector to segment thigh muscles.

Conversely, aligning and merging manually segmented images into specific atlas coordinate spaces can be a reliable alternative to statistical shape models. In this context, various single and multi-atlas methods have been proposed for quadriceps muscle segmentation [19, 20] relying on non-linear registration. Engstrom et al. [21] used a statistical shape model constrained with probabilistic MR atlases to automatically segment quadratus lumborum. Segmentation of muscle versus fatty tissues has been also performed through possibilistic clus-

tering [22], histogram-based thresholding followed by region growing [23] and active contours [24] techniques.

However, all the previously described methods are not perfectly suited for high inter-subject shape variability, significant differences of tissue appearance due to injury and delineations of weak boundaries. Moreover, many of the previously described methods are semi-automatic and hence require prior knowledge, usually associated with high computational costs and large dataset requirements. Therefore, developing a robust fully-automatic muscle segmentation method remains an open and challenging issue, especially when dealing with pathological pediatric data.

Huge progress has been recently made for automatic image segmentation using deep Convolutional Neural Networks (CNN). Deep CNNs are entirely data-driven supervised learning models formed by multi-layer neural networks [25]. In contrast to conventional machine learning which requires hand-crafted features and hence specialized knowledge, deep CNNs automatically learn complex hierarchical features directly from data. CNNs obtained outstanding performance for many medical image segmentation tasks [5, 26], which suggests that robust automated delineation of shoulder muscles from MR images may be achieved using CNN-based segmentation. To our knowledge, no other study has been conducted on shoulder muscle segmentation using deep learning methods.

The simplest way to perform segmentation using deep CNNs consists in classifying each pixel individually by working on patches extracted around them [27]. Since input patches from neighboring pixels have large overlaps, the same convolutions are computed many time. By replacing fully connected layers with convolutional layers, a Fully Convolutional Network (FCN) can take entire images as inputs and produce likelihood maps instead of single pixel outputs. It removes the need to select representative patches and eliminates redundant calculations due to patch overlaps. In order to avoid outputs with far lower resolution than input shapes, FCNs can be applied to shifted versions of the input images [28]. Multiple resulting outputs are thus stitched together to get results at full resolution.

Further improvements can be reached with architectures comprising a regular FCN to extract features and capture context, followed by an up-sampling part that enables to recover the input resolution using up-convolutions [5]. Compared to patch-based or *shift-and-stitch* methods, it allows a precise localization in a single pass while taking into account the full image context. Such architecture made of paired networks is called Convolutional Encoder-Decoder (CED).

U-Net [29] is the most well-known CED in the medical image analysis community. It has a symmetrical architecture with equal amount of down-sampling and up-sampling layers between contracting and expanding paths (Fig.3*a*). The encoder gradually reduces the spatial dimension with pooling layers whereas the decoder gradually recovers object details and spatial dimension. One key aspect of U-Net is the use of shortcuts (so-called *skip connections*) which concatenate features from the encoder to the decoder to help in recovering object details while improving localization accuracy. By allowing information to directly flow from low-level to high-level feature maps, faster convergence is achieved. This architecture can be exploited for 3D volume segmentation [30] by replacing all 2D operations with their 3D counterparts but at the cost of computational speed and memory consumption. Processing 2D slices independently before reconstructing 3D medical volumes remains a simpler alternative. Instead of cross-entropy used as loss function, the extension of U-Net proposed in [31] directly minimizes a segmentation error to handle class imbalance between foreground and background.

3. Material and methods

In this work, we develop and validate a fully-automatic methodology for pathological shoulder muscle segmentation through deep CEDs (Sect.2), using a pediatric OPBB dataset (Sect.3.1). Healthy versus pathological learning transferability is addressed in Sect.3.2. Extended deep CED architectures with pre-trained encoders are proposed in Sect.3.3. Assessment is performed using dedicated evaluation metrics (Sect.3.4).

3.1. Imaging dataset

Data collected from a previous study [6] investigating the muscle volume-strength relationship in 12 children with unilateral OPBB (averaged age of 12.1± 3.3 years) formed the basis of the current study. In this IRB approved study, informed consents from a legal guardian and assents from the participants were obtained for all subjects. If a participant was over 18 years of age, only informed consent was obtained from that participant. For each patient, two 3D axial-plane T1-weighted gradient-echo MR images were acquired: one for the affected shoulder and another for the unaffected one. For each image set, equally spaced 2D axial slices were selected for four different rotator cuff muscles: deltoid, infraspinatus, supraspinatus and subscapularis. These slices were annotated by an expert in pediatric physical medicine and rehabilitation to reach pixel-wise groundtruth delineations. Image size for axial slices are constant for each subject (416×312 pixels). Image resolution varies from 0.55×0.55 to 0.63×0.63mm, allowing a finer resolution for smaller subjects. The number of axial slices fluctuates from 192 to 224, whereas slice thickness remains unchanged (1.2mm). Overall, we had 374 (resp. 395) annotated axial slices for deltoid, 306 (367) for infraspinatus, 238 (208) for supraspinatus and 388 (401) for subscapularis across 2400 (2448) axial slices arising from 12 affected (unaffected) shoulders. Among these 24 MR image sets, pairings between affected and unaffected shoulders are known. Due to sparse annotations (Fig.1), deep CEDs exploit as inputs 2D axial slices and produce 2D segmentation masks which can be then stacked to recover a 3D volume for clinical purposes. Among the images from the affected side, 8 are from right shoulders (`R-P-{0134,0684,0382,0447,0660,0737,0667,0277}`) whereas 4 correspond to left shoulders (`L-P-{0103,0351,0922,0773}`). Training images displaying a right (left) shoulder are flipped when a left (right) shoulder is considered for test.

3.2. Healthy versus pathological learning transferability

In the context of OBPP, the limited availability of both healthy and pathological data for image segmentation brings new queries related to the learning

transferability from healthy to pathological structures. This aspect is particularly suitable to musculo-skeletal pathologies for two reasons. First, despite different shapes and sizes due to growth and atrophy, healthy and pathological muscles may share common characteristics such as anatomic locations and overall aspects. Second, combining healthy and pathological data for deep learning-based segmentation can act as a smart data augmentation strategy when faced with limited annoted data. In exploring the combined use of healthy and pathological data for pathological muscle segmentation, determining the optimal learning scheme is crucial. Thus, three different learning schemes (Fig.2) employed with deep CEDs are considered:

- **pathological only** (`P`): the most common configuration consists in exploiting groundtruth annotations made on impaired shoulder muscles only, making the hypothesis that CED features extracted from healthy examinations are not suited enough for pathological anatomies.

- **healthy transfer to pathological** (`HP`): another strategy deals with transfer learning and fine tuning from healthy to pathological muscles. In this context, a first CED is trained using groundtruth segmentations from unaffected shoulders only. The weights of the resulting model are then used as initialization for a second CED network which is trained using pathological inputs only.

- **simultaneous healthy and pathological** (`A`[1]): the last configuration consists in training a CED with a groundtruth dataset comprising annotations made on both healthy and pathological shoulder muscles, which allows to benefit from a more consequent dataset.

By comparing these different training strategies, we evaluate the benefits brought by combining healthy and pathological data together in terms of model generalizability. The balance between data augmentation and healthy versus

[1] `A` stands for 'all'

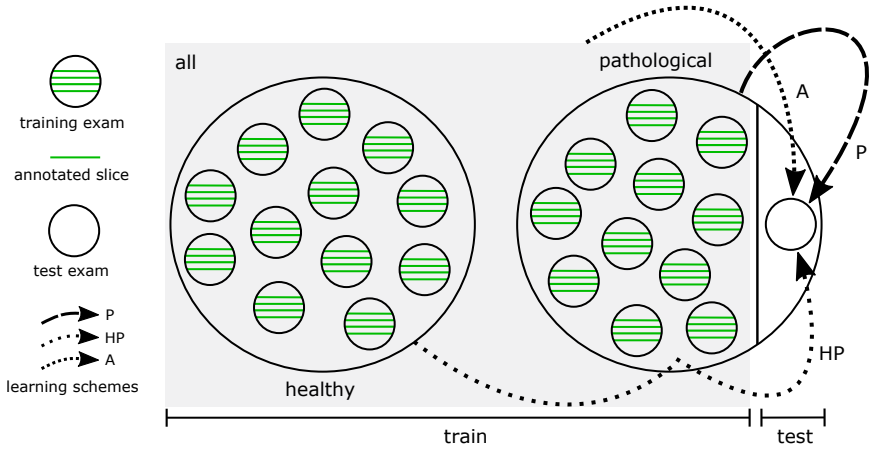

Figure 2: Three different learning schemes (`P`, `HP`, `A`) involved in a leave-one-out setting for deep learning-based pathological shoulder muscle segmentation.

pathological muscle variability is a crucial question which has never been investigated for muscle segmentation. These three different schemes, referred as `P` (pathological only), `HP` (healthy transfer to pathological) and `A` (simultaneous healthy and pathological) are compared in a leave-one-out fashion (Fig.2). The overall dataset is divided into healthy and pathological MR examinations. Iteratively, one pathological examination is extracted from the pathological dataset and considered as test examination for muscle segmentation. To avoid any bias for `HP` and `A`, annotated data from the healthy shoulder of the patient whose pathological shoulder is considered for test is not used during training.

For all schemes, deep CED networks are trained using data augmentation since the amount of available training data is limited. Training 2D axial slices undergo random scaling, rotation, shearing and shifting on both directions to teach the network the desired invariance and robustness properties [29]. In practice, 100 augmented images are produced for one single training axial slice. Comparisons between `P`, `HP` and `A` schemes are performed using standard U-Net [29] with 10 epochs, a batch size of 10 images, an *Adam* optimizer with 10^{-4} as learning rate for stochastic optimization, a fuzzy Dice score as loss function and randomly initialized weights for convolutional filters. Models were implemented using Keras and trained with a single Nvidia GeForce GTX 1080 Ti GPU with

11Gb/s. Once training is performed, predictions for one single axial slice take 28ms only which is suitable for routine clinical practice.

3.3. Extended architectures with pre-trained encoders

Contrary to deep classification networks which are usually pre-trained on a very large image dataset, CED architectures used for segmentation are typically trained from scratch, relying on randomly initialized weights. Reaching a generic model without over-fitting is therefore challenging, especially when only a small amount of images is available. As suggested in [32], the encoder part of a deep CED network can be replaced by a well-known classification network whose weights are pre-trained on an initial classification task. It allows to exploit transfer learning from large datasets such as ImageNet [33] for deep learning-based segmentation. In the literature, the encoder part of a deep CED has been already replaced by pre-trained VGG-11 [32] and ABN WideResnet-38 [34] with improvements compared to their randomly weighted counterparts.

Following this idea, we propose to extend the standard U-Net architecture (Sect.2) by exploiting another simple network from the VGG family [35] as encoder, namely the VGG-16 architecture. To improve performance, this encoder branch is pre-trained on ImageNet [33]. This database has been designed for object recognition purposes and contains more than 1 million natural images from 1000 classes. Pre-training our deep CED dedicated to muscle image segmentation using non-medical data is an efficient way to reduce the data scarcity issue while improving model generalizability [36]. Pre-trained models can not only improve predictive performance but also require less training time to reach convergence for the target task. In particular, low-level features captured by first convolutional layers are usually shared between different image types which explains the success of transfer learning between tasks.

The VGG-16 encoder (Fig.3*b*) consists of sequential layers including 3×3 convolutional layers followed by Rectified Linear Unit (ReLU) activation functions. Reducing the spatial size of the representation is handled by 2×2 max pooling layers. Compared to standard U-Net (Fig.3*a*), the first convolutional

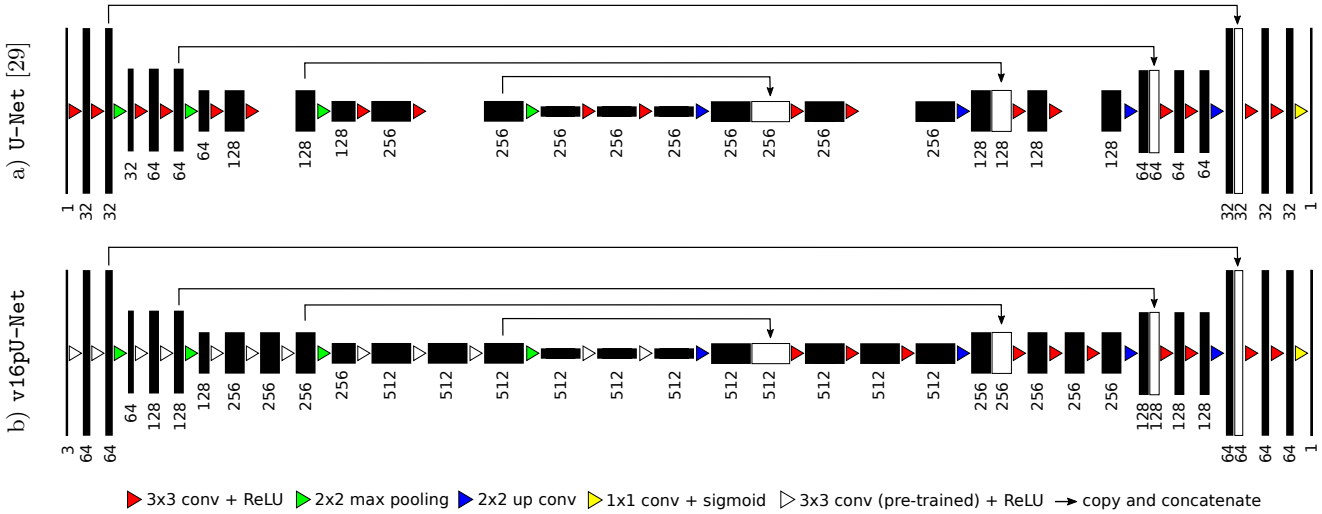

Figure 3: Extension of U-Net [29] by exploiting as encoder a slightly modified VGG-16 [35] with weights pre-trained on ImageNet [33], following [32, 34]. The decoder is modified to get an exactly symmetrical construction while keeping *skip connections*.

layer generates 64 channels instead of 32. As the network deepens, the number of channels doubles after each max pooling until it reaches 512 (256 for classical U-Net). After the second max pooling operation, the number of convolutional layers differ from U-Net with patterns of 3 consecutive convolutional layers instead of 2, following the original VGG-16 architecture. In addition, input images are extended from one single greyscale channel to 3 channels by repeating the same content in order to respect the dimensions of the RGB ImageNet images used for encoder pre-training. The only differences with VGG-16 rely in the fact that the last convolutional layer as well as top layers including fully-connected layers and softmax have been omitted. The two last convolutional layers taken from VGG-16 serve as central part of the CED and separate both contracting and expanding paths.

The extension of the U-Net encoder is transferred to the decoder branch by adding 2 convolutional layers as well as more feature channels to get an exactly symmetrical construction while keeping *skip connections*. Contrary to encoder weights which are initialized using pre-training performed on ImageNet, decoder

weights are set randomly. As for U-Net, a final 1×1 convolutional layer followed by a sigmoid activation function achieves pixel-wise segmentation masks whose resolution is the same as input slices.

Pathological shoulder muscle segmentation using the standard U-Net architecture [29] as well as the proposed extension without (`v16U-Net`) and with (`v16pU-Net`) weights pre-trained on ImageNet is performed through leave-one-out experiments. In this context, we rely on training scheme `A` combining both healthy and pathological data (Sect.3.2). As previously, networks are trained with data augmentation, 10 epochs, a batch size of 10 images, an *Adam* optimizer and a fuzzy Dice score used as loss function. Learning rates change from U-Net and `v16pU-Net` (10^{-4}) to `v16U-Net` (5×10^{-5}) to avoid divergence for deep networks trained with randomly selected weights.

3.4. Segmentation assessment

To assess both healthy versus pathological learning transferability (Sect.3.2) and extended pre-trained deep convolutional architectures (Sect.3.3), the accuracy of automatic pathological shoulder muscle segmentation is quantified based on Dice ($\frac{2TP}{2TP+FP+FN}$), sensitivity ($\frac{TP}{TP+FN}$), specificity ($\frac{TN}{TN+FP}$) and Jaccard ($\frac{TP}{TP+FP+FN}$) scores (in %) where TP, FP, TN and FN are the number of true or false positive and negative pixels. Evaluations also rely on the Cohen's kappa coefficient ($\frac{p_o-p_e}{1-p_e}$) in % where p_o and p_e are the relative observed agreement and the hypothetical probability of chance agreement. In practice, $p_o = \frac{TP+TN}{TP+FN+FP+TN}$ which corresponds to the accuracy and $p_e = \frac{(TP+FN)\times(TP+FP)}{TP+FN+FP+TN} + \frac{(FP+TN)\times(FN+TN)}{TP+FN+FP+TN}$. Finally, we exploit an absolute surface estimation error (`ASE`) which compares groundtruth and estimated muscle surfaces defined in mm^2 from segmentation masks. These scores tend to provide a complete assessment of the ability of CED models to provide contours identical to those manually performed. Reported results are averaged among all annotated slices arising from the 12 pathological shoulder examinations. Network parameters are those reaching the best fuzzy Dice test scores during training.

4. Results and discussion

4.1. Healthy versus pathological learning transferability

The highest performance is achieved when both healthy and pathological data are simultaneously used for training (`A`), with Dice scores of 78.32% for deltoid, 81.58% for infraspinatus and 81.41% for subscapularis (Tab.1). Scheme `A` outperforms transfer learning and fine tuning (`HP`) from 4 to 7% in terms of Dice. However, this conclusion does not apply to supraspinatus for which `A` and `HP` schemes achieve the same performance in Dice ($\approx 65.7\%$) and Cohen's kappa ($\approx 65.6\%$). In particular, `A` increases the sensitivity (65.55% instead of 63.16%) but provides a slightly smaller specificity, compared to `HP`. In this specific case, medians are nevertheless rather in favour of `A` compared to means (Fig.4). Comparing `ASE` from `HP` to `A` reveals improvements for all shoulder muscles, including deltoid whose surface estimation error decreases from 268 to 105.5mm^2. The same finding arises when studying Jaccard scores whose gains are 7.8% and 6.5% for deltoid and subscapularis. The Cohen's kappa coefficient jumps from 70.73% (76.85%) to 78.15% (81.45%) for deltoid (infraspinatus). Therefore, directly combining healthy and pathological data appears a better strategy than dividing training into two parts, focusing on first healthy and then pathological data via transfer learning. Further, exploiting annotations for the pathological shoulder muscles only (`P`) is the worst training strategy (Tab.1, Fig.4), especially for deltoid (Dice loss of 10% from `A` to `P`). However, results for subscapularis deviate from this result, with higher similarity scores (except for kappa) compared to `HP` combined with the best `ASE` (94.56mm^2). In general, the CED features extracted from healthy examinations are suited enough for pathological anatomies while acting as an efficient data augmentation strategy.

Accuracy scores for supraspinatus are globally worse than for other muscles (Fig.4) since its thin and elongated shape can strongly vary across patients [16]. Moreover, we notice the presence of a single severely atrophied supraspinatus (`L-P-0922`) among the set of pathological examinations. Dice results for this single muscle is 42.99% for `P` against 38.59% and 32.33% for `HP` and `A` respec-

metric	scheme	P	HP	A		
	network	U-Net [29]			v16U-Net	v16pU-Net
dice ↑	deltoid	68.94±29.9	71.05±29.5	*78.32*±24.4	80.05±23.1	**82.42**±20.4
	infraspinatus	71.38±24.7	77.00±22.5	*81.58*±18.3	81.91±19.0	**81.98**±18.6
	supraspinatus	64.94±28.0	*65.69*±29.6	65.68±30.7	67.30±29.4	**70.98**±28.7
	subscapularis	78.10±18.1	74.55±25.2	*81.41*±15.0	81.58±15.2	**82.80**±14.4
sens ↑	deltoid	70.85±30.5	70.74±29.5	*78.92*±25.4	81.45±23.7	**83.80**±21.3
	infraspinatus	72.12±26.4	79.45±23.1	***84.61***±18.2	83.74±18.6	83.48±19.0
	supraspinatus	64.02±31.8	63.16±33.2	*65.55*±34.5	67.21±33.0	**68.60**±32.3
	subscapularis	78.89±19.7	74.75±27.3	*82.53*±18.1	81.75±18.8	**84.36**±16.5
spec ↑	deltoid	99.61±0.80	99.56±1.07	***99.85***±0.19	99.82±0.22	99.84±0.22
	infraspinatus	99.82±0.23	99.82±0.22	*99.84*±0.18	**99.86**±0.17	**99.86**±0.18
	supraspinatus	99.86±0.18	*99.90*±0.13	99.88±0.15	99.86±0.17	**99.91**±0.12
	subscapularis	99.86±0.13	99.83±0.28	*99.87*±0.13	**99.88**±0.12	99.86±0.15
jacc ↑	deltoid	59.27±29.7	61.68±29.3	*69.48*±26.0	71.46±24.9	**74.00**±22.8
	infraspinatus	60.32±25.6	66.91±24.0	*72.00*±20.4	72.63±20.6	**72.71**±21.0
	supraspinatus	53.61±27.1	55.27±29.3	*55.70*±30.1	56.98±28.7	**61.31**±28.7
	subscapularis	66.93±19.6	64.31±24.7	*70.83*±17.6	71.13±17.7	**72.72**±17.16
kappa ↑	deltoid	68.63±30.0	70.73±29.7	*78.15*±24.4	79.89±23.2	**82.28**±20.5
	infraspinatus	71.19±24.7	76.85±22.5	*81.45*±18.3	81.79±19.0	**81.86**±18.7
	supraspinatus	64.76±28.0	*65.56*±29.6	65.55±30.7	67.16±29.4	**70.87**±28.7
	subscapularis	77.95±18.1	79.23±17.1	*81.27*±15.0	81.45±15.2	**82.67**±14.4
ASE ↓	deltoid	252.0±421.6	268.0±507.8	*105.5*±178.9	94.23±139.2	**80.38**±127.5
	infraspinatus	156.8±228.7	92.37±105.9	***74.47***±92.8	80.11±96.2	79.17±96.9
	supraspinatus	174.8±164.0	159.9±153.5	*153.9*±146.0	147.5±129.4	**134.6**±135.5
	subscapularis	*94.56*±95.5	102.0±110.7	95.19±109.0	94.06±111.3	**82.95**±86.88

Table 1: Quantitative assessment of convolutional encoder-decoders (U-Net [29], `v16U-Net`, `v16pU-Net`) embedded with learning schemes `P`, `HP` and `A` over the pathological dataset in Dice, sensitivity, specificity, Jaccard, Cohen's kappa (%) as well as absolute surface error (mm^2). Best results are in bold. Italic underlined scores highlight best results among learning schemes employed with U-Net.

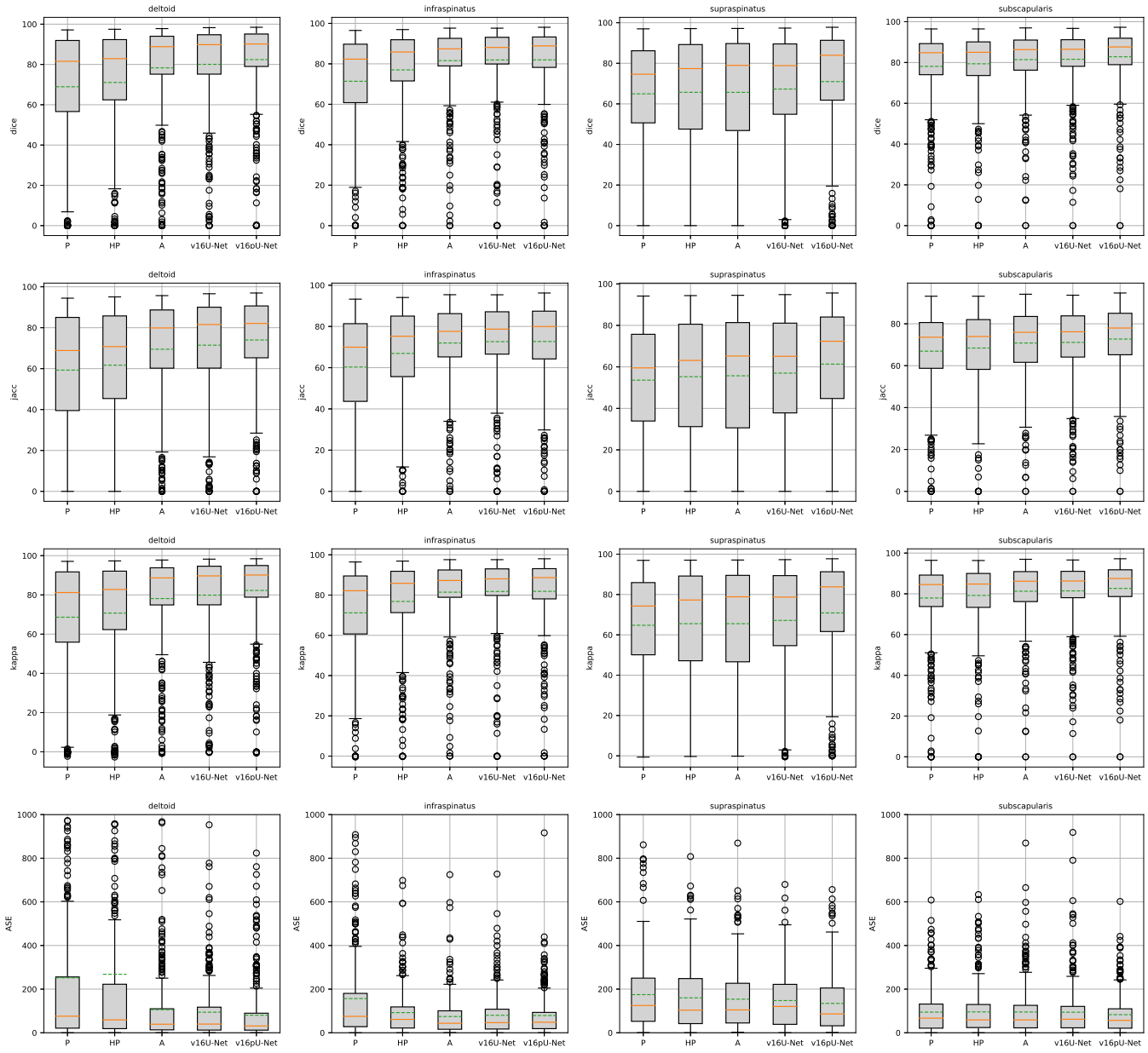

Figure 4: Box plots on Dice, Jaccard, Cohen's kappa and absolute surface error (`ASE`) slice-wise scores over the pathological dataset using convolutional encoder-decoders (U-Net [29], `v16U-Net`, `v16pU-Net`) embedded with learning schemes `P`, `HP` and `A`. Dashed green and solid orange lines respectively represent means and medians.

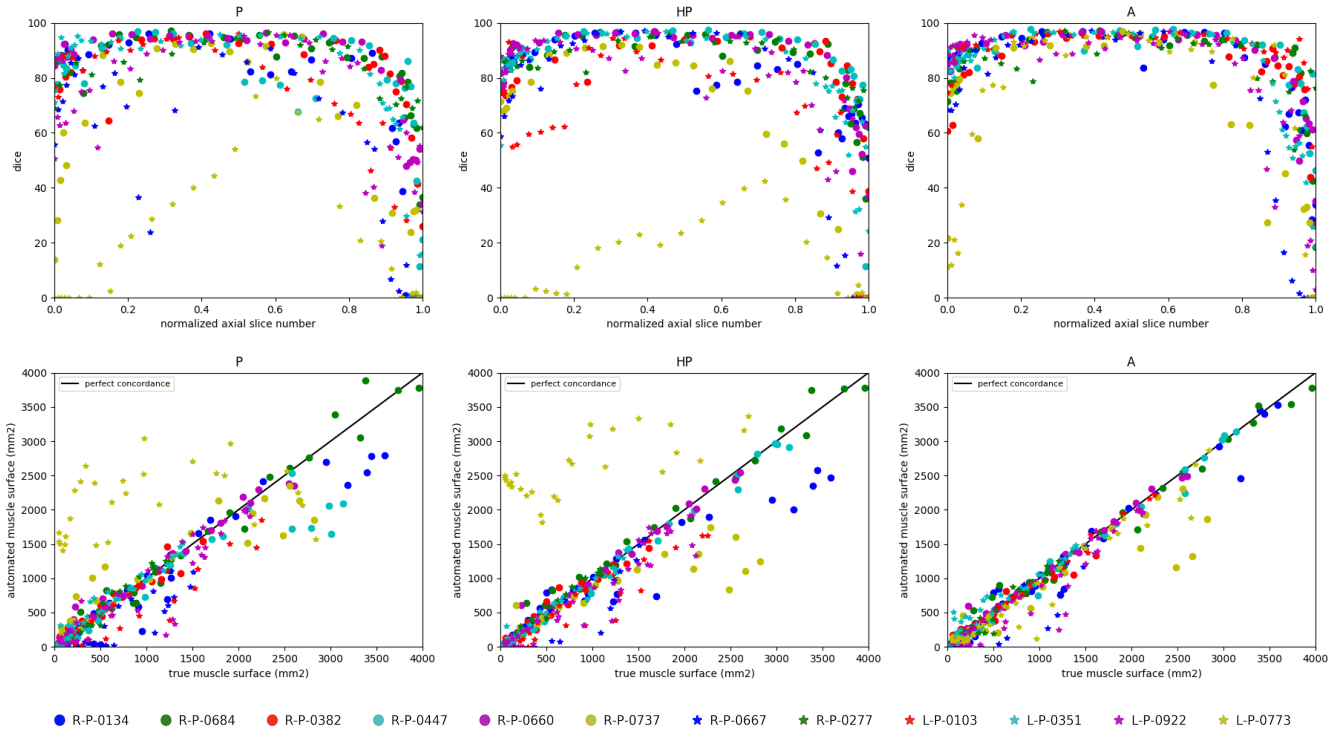

Figure 5: Deltoid segmentation accuracy using U-Net [29] with learning schemes `P`, `HP` and `A` for each annotated slice of the whole pathological dataset. Top raw shows Dice scores (%) with respect to the normalized axial slice number obtained by linearly scaling slice number from $[z_{min}, z_{max}]$ to $[0, 1]$ where $\{z_{min}, z_{max}\}$ are the minimal and maximal axial slice indices displaying the deltoid. Bottom row displays concordance between groundtruth and predicted deltoid muscle surfaces in mm^2. Black line indicates perfect concordance.

tively. It suggests that muscles undergoing very strong degrees of injury must be processed separately, relying either on pathological data only or manual delineations. Nevetheless, learning scheme `A` appears globally better suited from weak to moderately severe muscle impairments.

Overall, the segmentation results for all three learning schemes are more accurate for mid-muscle regions than for both base and apex, where muscles appear smaller with strong appearance similarities with surrounding tissues (Fig.5, top row). Above conclusions (`A` > `HP` > `P`) are confirmed with much more individual Dice scores grouped on the interval $[75, 95\%]$ for `A`. The concordance between predicted and groundtruth deltoid surfaces (Fig.5, bottom row), demonstrates a

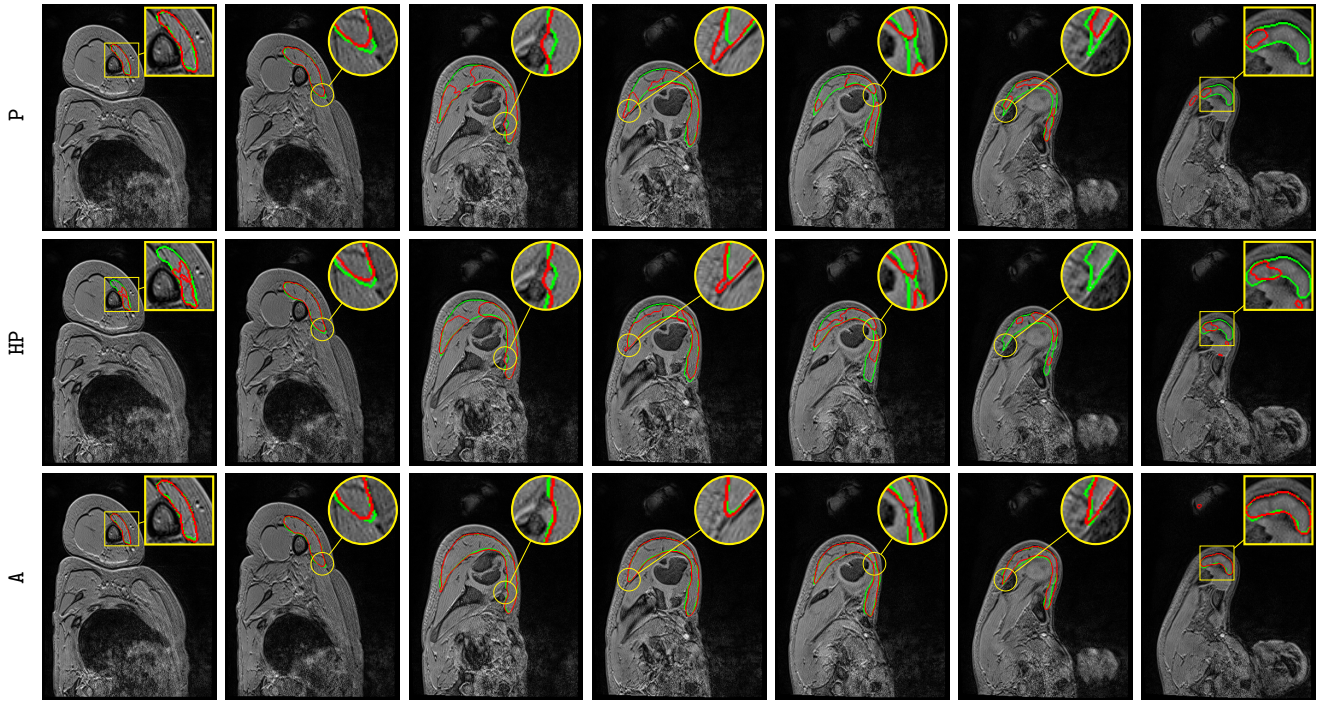

Figure 6: Automatic pathological deltoid segmentation using U-Net [29] embedded with learning schemes P, HP and A. Groundtruth and estimated delineations are in green and red respectively. Displayed results cover the whole muscle spatial extent for L-P-0103 examination.

stronger correlation for A than for P and HP with individual estimations closer to the line of perfect concordance (L-P-0773 is the most telling example), in agreement with similarity scores reported for each learning scheme (Tab.1, Fig.4).

Visually comparing both manual and automatic segmentation for deltoid (P, HP and A, Fig.6) and other rotator cuff muscles (A only, Fig.7) further supports the validity of automatic segmentation. A very accurate deltoid delineation is achieved for A whereas P and HP tend to under-segment the muscle area (Fig.6). Complex muscle shapes and subtle contours (Fig.7) are relatively well captured. In addition, we can notice outstanding performance near muscle insertion regions (Fig.7) whose contours are usually very hard to extract, even visually. These results confirm that using simultaneously healthy and pathological data for training helps in providing good model generalizability despite the data scarcity issue combined with a large appearance variability.

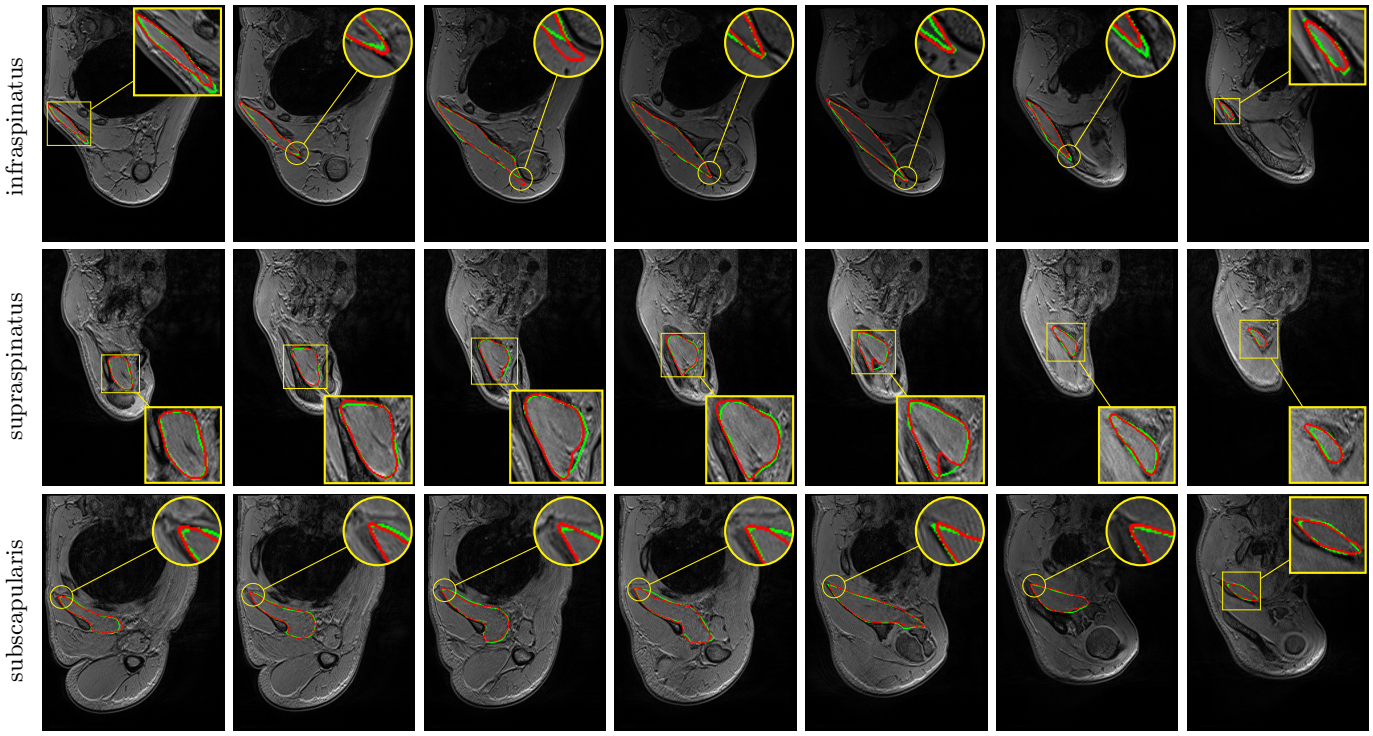

Figure 7: Automatic pathological segmentation of infraspinatus, supraspinatus and subscapularis using U-Net [29] with training on both healthy and pathological data simultaneously (`A`). Groundtruth and estimated delineations are in green and red respectively. Displayed results cover the whole muscle spatial extents for `R-P-0447` (top), `R-P-0660` (middle) and `R-P-0134` (bottom) examinations.

4.2. Extended architectures with pre-trained encoders

The `v16pU-Net` architecture globally outperforms both U-Net and `v16U-Net` networks (Fig.4) with Dice scores of 82.42% for deltoid, 81.98% for infraspinatus, 70.98% for supraspinatus and 82.80% for subscapularis (Tab.1). On the contrary, `v16U-Net` (U-Net) obtains 80.05% (78.32%) for deltoid, 81.91% (81.58%) for infraspinatus, 67.30% (65.68%) for supraspinatus and 81.58% (81.41%) for subscapularis. In one hand, despite slightly worse scores compared with U-Net for infraspinatus in terms of sensitivity (83.74 against 84.61%) and `ASE` (80.11 against 74.47mm^2), `v16U-Net` is most likely to provide good predictive performance and model generalizability thanks to its deeper architecture. On the other hand, comparisons between `v16U-Net` and `v16pU-Net` reveal that pre-training the encoder using ImageNet brings non-negligible improvements (Fig.4). For instance, `v16pU-Net` provides significant gains (Tab.1) for deltoid (supraspinatus) whose Jaccard score goes from 71.46 (56.98) to 74% (61.31%). The Cohen's kappa coefficient enhancement is around 2.4% (3.7%). Surface estimation errors are among the lowest obtained with only 80.38mm^2 for deltoid and 82.95mm^2 for subscapularis. Medians and first quartiles (Fig.4) globally highlight significant segmentation gains, especially for supraspinatus. Despite their non-medical nature, the large amount of ImageNet images used for pre-training makes the network converge towards a better solution. `v16pU-Net` is therefore the most able to efficiently discriminate individual muscles from surrounding anatomical structures, compared to U-Net and `v16U-Net`. In average among the four shoulder muscles, gains for Dice, sensitivity, Jaccard and kappa reach 2.8, 2.7, 3.2 and 2.8% from U-Net to `v16pU-Net`.

Above conclusions (`v16pU-Net` > `v16U-Net` > U-Net) are further supported by statistical analysis (Tab.2). Except for infraspinatus, Student's paired t-tests between `v16pU-Net` and `v16U-Net` or U-Net globally indicate that extended architectures with pre-trained encoders really bring non-negligible improvements (p-values < 0.05 for similarity metrics and `ASE`). This finding is all the more verified between `v16pU-Net` embedded with learning scheme `A` and U-Net [29] with `P`, `HP` or `A` for all muscles including infraspinatus.

metric	scheme	P	HP	A	
	network	U-Net [29]			v16U-Net
dice	deltoid	$\mathbf{3.7}\times10^{-29}$	$\mathbf{8.8}\times10^{-21}$	$\mathbf{9.7}\times10^{-11}$	$\mathbf{7.5}\times10^{-6}$
	infraspinatus	$\mathbf{2.3}\times10^{-19}$	$\mathbf{5.4}\times10^{-7}$	0.491	0.935
	supraspinatus	$\mathbf{4.3}\times10^{-6}$	$\mathbf{4.8}\times10^{-5}$	$\mathbf{3.2}\times10^{-6}$	$\mathbf{7.0}\times10^{-5}$
	subscapularis	$\mathbf{1.1}\times10^{-13}$	$\mathbf{1.5}\times10^{-7}$	**0.001**	**0.006**
sens	deltoid	$\mathbf{1.2}\times10^{-23}$	$\mathbf{3.7}\times10^{-23}$	$\mathbf{2.2}\times10^{-9}$	$\mathbf{4.4}\times10^{-4}$
	infraspinatus	$\mathbf{6.1}\times10^{-16}$	$\mathbf{3.9}\times10^{-4}$	0.117	0.788
	supraspinatus	$\mathbf{8.6}\times10^{-4}$	$\mathbf{8.7}\times10^{-5}$	**0.016**	0.135
	subscapularis	$\mathbf{2.7}\times10^{-12}$	$\mathbf{2.8}\times10^{-10}$	**0.002**	$\mathbf{1.5}\times10^{-6}$
spec	deltoid	$\mathbf{7.9}\times10^{-9}$	$\mathbf{2.6}\times10^{-7}$	0.457	0.069
	infraspinatus	$\mathbf{1.3}\times10^{-5}$	$\mathbf{1.2}\times10^{-6}$	**0.010**	0.387
	supraspinatus	$\mathbf{8.2}\times10^{-6}$	0.118	$\mathbf{1.0}\times10^{-5}$	$\mathbf{7.8}\times10^{-9}$
	subscapularis	0.924	**0.005**	0.078	$\mathbf{2.5}\times10^{-4}$
jacc	deltoid	$\mathbf{5.1}\times10^{-35}$	$\mathbf{9.9}\times10^{-25}$	$\mathbf{7.0}\times10^{-11}$	$\mathbf{1.1}\times10^{-5}$
	infraspinatus	$\mathbf{1.6}\times10^{-23}$	$\mathbf{1.6}\times10^{-9}$	0.251	0.917
	supraspinatus	$\mathbf{4.9}\times10^{-10}$	$\mathbf{8.3}\times10^{-7}$	$\mathbf{1.8}\times10^{-7}$	$\mathbf{1.7}\times10^{-6}$
	subscapularis	$\mathbf{1.2}\times10^{-17}$	$\mathbf{1.0}\times10^{-9}$	$\mathbf{2.6}\times10^{-4}$	**0.002**
kappa	deltoid	$\mathbf{2.6}\times10^{-29}$	$\mathbf{7.9}\times10^{-21}$	$\mathbf{8.9}\times10^{-11}$	$\mathbf{6.8}\times10^{-6}$
	infraspinatus	$\mathbf{1.7}\times10^{-19}$	$\mathbf{4.6}\times10^{-7}$	0.476	0.928
	supraspinatus	$\mathbf{3.4}\times10^{-6}$	$\mathbf{4.3}\times10^{-5}$	$\mathbf{2.9}\times10^{-6}$	$\mathbf{5.8}\times10^{-5}$
	subscapularis	$\mathbf{9.5}\times10^{-14}$	$\mathbf{1.4}\times10^{-7}$	**0.001**	**0.006**
ASE	deltoid	$\mathbf{2.2}\times10^{-15}$	$\mathbf{7.9}\times10^{-13}$	**0.001**	**0.021**
	infraspinatus	$\mathbf{8.4}\times10^{-10}$	**0.021**	0.308	0.813
	supraspinatus	$\mathbf{7.3}\times10^{-5}$	$\mathbf{8.2}\times10^{-4}$	**0.004**	**0.036**
	subscapularis	**0.009**	**0.009**	**0.005**	**0.011**

Table 2: Statistical analysis between v16pU-Net embedded with learning scheme A and all other configurations (U-Net [29] with P, HP and A as well as v16U-Net with A) through Student's paired t-tests using Dice, sensitivity, specificity, Jaccard, Cohen's kappa scores as well as absolute surface error over the pathological dataset. Bold p-values (< 0.05) highlight statistically significant results.

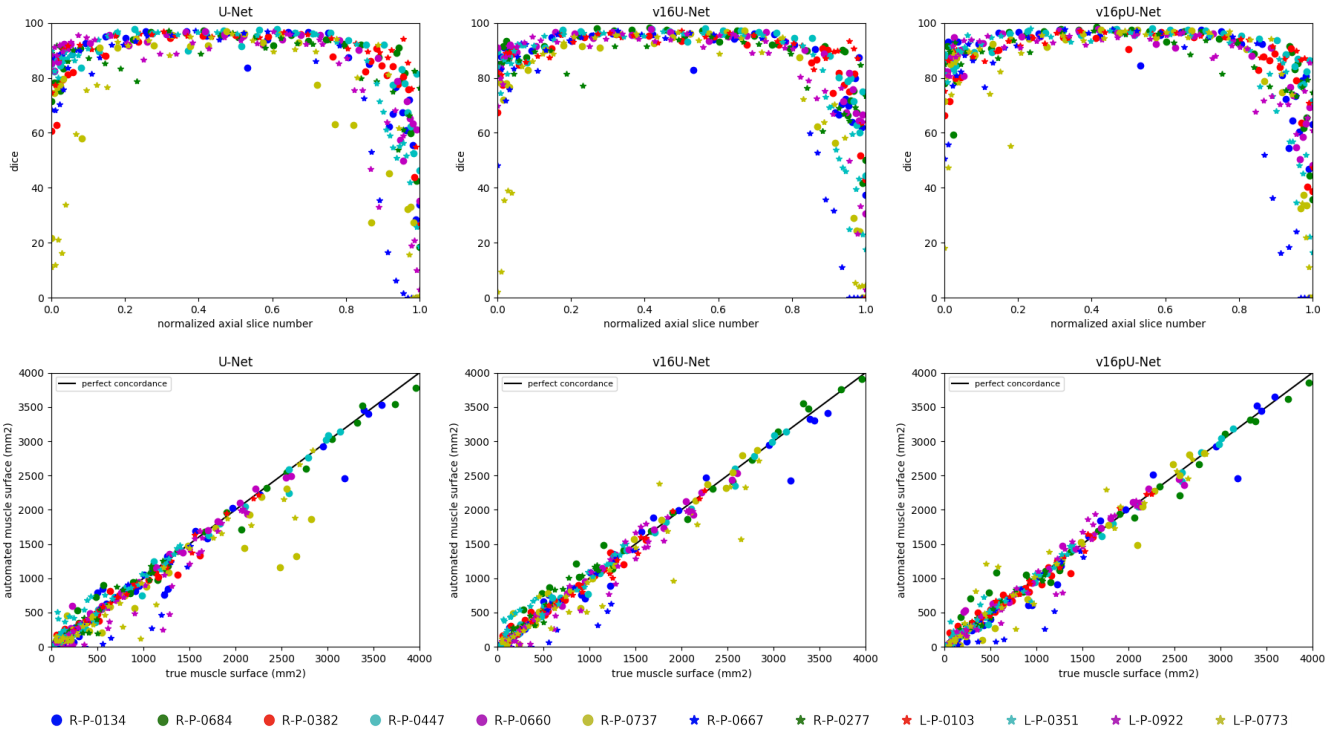

Figure 8: Deltoid segmentation accuracy using U-Net [29], `v16U-Net` and `v16pU-Net` with learning scheme `A` for each annotated slice of the whole pathological dataset. Top raw shows Dice scores with respect to normalized axial slice number. Bottom row displays concordance between groundtruth and predicted deltoid surfaces. Black line indicates perfect concordance.

From U-Net to `v16pU-Net`, individual Dice scores (Fig.8, top row) are slightly pushed towards the upper limit (100%) with less variability and an increased overall consistency along the axial axis, as for `R-P-0737` and `L-P-0773`. Extreme axial slices are much better handled in the case `v16pU-Net`, especially when normalized slice numbers approach zero. In addition, a slightly stronger correlation between predicted and groundtruth deltoid surface can be seen for `v16pU-Net` with respect to U-Net and `v16U-Net` (Fig.8, bottom row). In particular, great improvements for `R-P-0737` and `L-P-0773` can be highlighted.

Globally, compared to U-Net and `v16U-Net`, better contour adherence and shape consistency are reached by `v16pU-Net` whose ability to mimmic expert annotations is notable (Fig.9). The great diversity in terms of textures (smooth

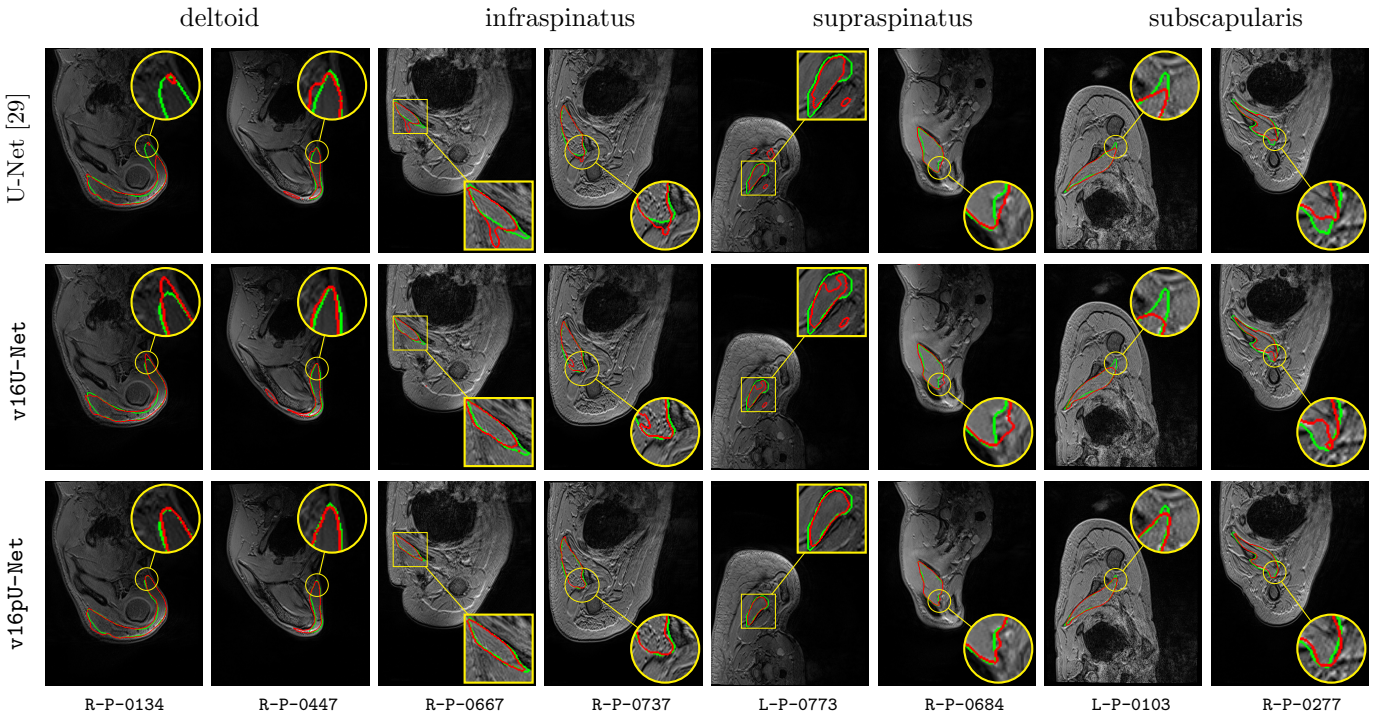

Figure 9: Automatic pathological segmentation of deltoid, infraspinatus, supraspinatus and subscapularis using U-Net [29], `v16U-Net` and `v16pU-Net` with training on both healthy and pathological data simultaneously (`A`). Groundtruth and estimated delineations are in green and red respectively. 8 pathological examinations among the 12 available are involved to provide valuable insight into the overall performance.

in `R-P-0684` versus granular in `R-P-0737`) is accurately captured despite high similar visual properties with surrounding structures. Visual results also reveal that `v16pU-Net` has a good behavior for complex muscle insertion regions (`R-P-0447`). Despite a satisfactory overall quality, U-Net and `v16U-Net` are frequently prone to under- (`R-P-0134`, `R-P-0277`) or over-segmentation (`R-P-0684`). Some examples report inconsistent shapes (`R-P-0667`, `R-P-0737`), sometimes combined with false positive areas which can be located far away from the groundtruth muscle location (`R-P-0447`, `L-P-0773`). Using a pre-trained and complex architecture such as `v16pU-Net` to simultaneously process healthy and pathological data provides accurate automated delineations of pathological shoulder muscles for patients with OPBB.

4.3. Benefits for clinical practice

The key contribution of this work deals with the possibility of automatically providing robust MR delineations for shoulder pathological muscles, despite the strong diversity in shape, size, location, texture and injury (Fig.9). First, it has the advantages of reducing the burden of manual segmentation and avoiding the subjectivity of experts. Second, it paves the way for the automated inference of individual morphological parameters [6] which are not accessible with simple clinical examinations. This can therefore be useful to guide the rehabilitative and surgical management of children with OBPP. The benefit of the proposed technology in real clinical use can be also involved for other very frequent shoulder muscular disorders such as rotator cuff tears in order to provide objective predictors of successful surgical repair [37].

Despite specific segmentation difficulties in shoulder muscles related to complex shapes and reduced sizes, our contributions show good performance with, in particular, excellent specificity (Tab.1). In shoulder muscles, better segmentation results are highlighted for mid muscle regions (Fig.8) where muscles appear bigger and well differentiated from surrounding tissues. Thus, we can assume that our approach could have very good performance for larger muscles with stable shapes like most of arm, forearm, thigh and leg muscles. Additionally, it provides interesting perspectives for other muscular disorders, for which objective and non-invasive biomarkers are required to effectively monitor both disease progression and treatment response.

At a research level, it could document effects of innovative treatments like genetic therapies for neuromuscular disorders [38] or improve the understanding of particular symptoms or diseases [39]. It could also be integrated into biomechanical models [40, 41] to help clinicians for intervention planning.

5. Conclusion

In this work, we successfully addressed automatic pathological shoulder muscle MRI segmentation for patients with obstetrical brachial plexus palsy by

means of deep convolutional encoder-decoders. In particular, we studied healthy to pathological learning transferability by comparing different learning schemes in terms of model generalizability against large muscle shape, size, location, texture and injury variability. Moreover, convolutional encoder-decoder networks were expanded using VGG-16 encoders pre-trained on ImageNet to improve the accuracy reached by standard U-Net architectures. Our contributions were evaluated on four different shoulder muscles: deltoid, infraspinatus, supraspinatus and subscapularis. First, results clearly show that features extracted from unimpaired limbs are suited enough for pathological anatomies while acting as an efficient data augmentation strategy. Compared to transfer learning, combining healthy and pathological data for training provides the best segmentation accuracy together with outstanding delineation performance for muscle boundaries including insertion areas. Second, experiments reveal that convolutional encoder-decoders involving a pre-trained VGG-16 encoder strongly outperforms U-Net. Despite the non-medical nature of pre-training data, such deeper networks are able to efficiently discriminate individual muscles from surrounding anatomical structures. These conclusions offer new perspectives for the management of musculo-skeletal disorders, even if a small and heterogeneous dataset is available. The proposed approach can be easily extended to other muscle types and imaging modalities to provide decision support in various applications including neuro-muscular diseases, sports related injuries or any other muscle disorders. Methodological perspectives on domain adaptation should deserve further investigation to take advantage of multi-centric data. Clinically, our method can be useful to distinguish between pathologies, evaluate the effect of treatments and facilitate surveillance of neuro-muscular disease course. It could be exploited together with bio-mechanical models to improve the understanding of complex pathologies and help clinicians to plan surgical interventions.

Conflicts of interest

None of the authors of this manuscript have any financial or personal relationships with other people or organizations that could inappropriately influence and bias this work.

References

[1] F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao, R. Kijowski, Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging, Magnetic Resonance in Medicine 79 (4) (2018) 2379–2391.

[2] A. Boutillon, B. Borotikar, V. Burdin, P.-H. Conze, Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images, in: IEEE International Symposium on Biomedical Imaging, 2020.

[3] Y. Barnouin, G. Butler-Browne, T. Voit, D. Reversat, N. Azzabou, G. Leroux, A. Behin, J. S. McPhee, P. G. Carlier, J.-Y. Hogrel, Manual segmentation of individual muscles of the quadriceps femoris using MRI: a reappraisal, Journal of Magnetic Resonance Imaging 40 (1) (2014) 239–247.

[4] M. J. Tingart, M. Apreleva, J. T. Lehtinen, B. Capell, W. E. Palmer, J. J. Warner, Magnetic resonance imaging in quantitative analysis of rotator cuff muscle volume, Clinical Orthopaedics and Related Research 415 (2003) 104–110.

[5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. Van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.

[6] 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.

[7] P. O'Berry, M. Brown, L. Phillips, S. H. Evans, Obstetrical brachial plexus palsy, Current Problems in Pediatric and Adolescent Health Care 47 (7) (2017) 151–155.

[8] S. P. Chauhan, S. B. Blackwell, C. V. Ananth, Neonatal brachial plexus palsy: incidence, prevalence, and temporal trends, in: Seminars in Perinatology, Vol. 38, 2014, pp. 210–218.

[9] S. Brochard, K. Alter, D. Damiano, Shoulder strength profiles in children with and without brachial plexus palsy, Muscle & Nerve 50 (1) (2014) 60–66.

[10] P. M. Waters, J. T. Monica, B. E. Earp, D. Zurakowski, D. S. Bae, Correlation of radiographic muscle cross-sectional area with glenohumeral deformity in children with brachial plexus birth palsy, The Journal of Bone and Joint Surgery 91 (10) (2009) 2367.

[11] 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: IEEE International Symposium on Biomedical Imaging, 2019.

[12] C. Pons, B. Borotikar, M. Garetier, V. Burdin, D. B. Salem, M. Lempereur, S. Brochard, Quantifying skeletal muscle volume and shape in humans using MRI: a systematic review of validity and reliability, PloS One 13 (11) (2018) 1–26.

[13] 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: IEEE International Engineering in Medicine and Biology Conference, 2017, pp. 317–320.

[14] I. Südhoff, J. A. de Guise, A. Nordez, E. Jolivet, D. Bonneau, V. Khoury, W. Skalli, 3D-patient-specific geometry of the muscles involved in knee motion from selected MRI images, Medical & Biological Engineering & Computing 47 (6) (2009) 579–587.

[15] E. Jolivet, E. Dion, P. Rouch, G. Dubois, R. Charrier, C. Payan, W. Skalli, Skeletal muscle segmentation from MRI dataset using a model-based approach, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2 (3) (2014) 138–145.

[16] S. Kim, D. Lee, S. Park, K.-S. Oh, S. W. Chung, Y. Kim, Automatic segmentation of supraspinatus from MRI by internal shape fitting and autocorrection, Computer Methods and Programs in Biomedicine 140 (2017) 165–174.

[17] P.-Y. Baudin, N. Azzabou, P. G. Carlier, N. Paragios, Prior knowledge, random walks and human skeletal muscle segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2012, pp. 569–576.

[18] S. Andrews, G. Hamarneh, The generalized log-ratio transformation: learning shape and adjacency priors for simultaneous thigh muscle segmentation, IEEE Transactions on Medical Imaging 34 (9) (2015) 1773–1787.

[19] E. Ahmad, M. H. Yap, H. Degens, J. S. McPhee, Atlas-registration based image segmentation of MRI human thigh muscles in 3D space, in: Medical Imaging: Image Perception, Observer Performance, and Technology Assessment, 2014.

[20] A. Le Troter, A. Fouré, M. Guye, S. Confort-Gouny, J.-P. Mattei, J. Gondin, E. Salort-Campana, D. Bendahan, Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches, Magnetic Resonance Materials in Physics, Biology and Medicine 29 (2) (2016) 245–257.

[21] C. M. Engstrom, J. Fripp, V. Jurcak, D. G. Walker, O. Salvado, S. Crozier, Segmentation of the quadratus lumborum muscle using statistical shape modeling, Journal of Magnetic Resonance Imaging 33 (6) (2011) 1422–1429.

[22] V. Barra, J.-Y. Boire, Segmentation of fat and muscle from MR images of the thigh by a possibilistic clustering algorithm, Computer Methods and Programs in Biomedicine 68 (3) (2002) 185–193.

[23] S. Purushwalkam, B. Li, Q. Meng, J. McPhee, Automatic segmentation of adipose tissue from thigh magnetic resonance images, in: International Conference Image Analysis and Recognition, 2013, pp. 451–458.

[24] S. Orgiu, C. L. Lafortuna, F. Rastelli, M. Cadioli, A. Falini, G. Rizzo, Automatic muscle and fat segmentation in the thigh from T1-weighted MRI, Journal of Magnetic Resonance Imaging 43 (3) (2016) 601–610.

[25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[26] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, X. Ding, Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, Medical Image Analysis (2020).

[27] D. Ciresan, A. Giusti, L. M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.

[28] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[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] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 424–432.

[31] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: International Conference on 3D Vision, 2016, pp. 565–571.

[32] V. Iglovikov, A. Shvets, TernausNet: U-Net with VGG11 encoder pre-trained on imagenet for image segmentation, arXiv preprint arXiv:1801.05746 (2018).

[33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., ImageNet large scale visual recognition challenge, International Journal of Computer Vision 115 (3) (2015) 211–252.

[34] V. Iglovikov, S. Seferbekov, A. Buslaev, A. Shvets, TernausNetV2: Fully convolutional network for instance segmentation, arXiv preprint arXiv:1806.00844 (2018).

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[36] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.

[37] D. Laron, S. P. Samagh, X. Liu, H. T. Kim, B. T. Feeley, Muscle degeneration in rotator cuff tears, Journal of Shoulder and Elbow Surgery 21 (2) (2012) 164–174.

[38] J. Ropars, F. Gravot, D. B. Salem, F. Rousseau, S. Brochard, C. Pons, Muscle MRI: A biomarker of disease severity in Duchenne muscular dystrophy? A systematic review, Neurology 94 (3) (2020) 117–133.

[39] C. M. Engstrom, D. G. Walker, V. Kippers, A. J. Mehnert, Quadratus lumborum asymmetry and l4 pars injury in fast bowlers: a prospective mr study, Medicine & Science in Sports & Exercise 39 (6) (2007) 910–917.

[40] K. R. Holzbaur, W. M. Murray, S. L. Delp, A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control, Annals of Biomedical Engineering 33 (6) (2005) 829–840.

[41] S. S. Blemker, S. L. Delp, Three-dimensional representation of complex muscle architectures and geometries, Annals of Biomedical Engineering 33 (5) (2005) 661–673.