

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Cronin, Neil J.; Finni, Taija; Seynnes, Olivier

**Title:** Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images

**Year:** 2020

**Version:** Published version

**Copyright:** © 2020 the Author(s)

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Cronin, N. J., Finni, T., & Seynnes, O. (2020). Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images. *Computer methods and programs in biomedicine*, 196, Article 105583. <https://doi.org/10.1016/j.cmpb.2020.105583>



# Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images

Neil J. Cronin<sup>a,b,c,\*</sup>, Taija Finni<sup>a</sup>, Olivier Seynnes<sup>d</sup>

<sup>a</sup>Neuromuscular Research Centre, Faculty of Sport and Health Sciences, University of Jyväskylä, Finland

<sup>b</sup>Department for Health, Bath University, UK

<sup>c</sup>School of Sport & Exercise, University of Gloucestershire, Gloucestershire, UK

<sup>d</sup>Norwegian School of Sport Sciences, Oslo, Norway

## ARTICLE INFO

### Article history:

Received 3 February 2020

Accepted 29 May 2020

### Key words:

Ultrasound

Muscle

Deep learning

Medical imaging

Generative adversarial network

cycleGAN

Synthetic image

## ABSTRACT

**Background and objective:** Deep learning approaches are common in image processing, but often rely on supervised learning, which requires a large volume of training images, usually accompanied by hand-crafted labels. As labelled data are often not available, it would be desirable to develop methods that allow such data to be compiled automatically. In this study, we used a Generative Adversarial Network (GAN) to generate realistic B-mode musculoskeletal ultrasound images, and tested the suitability of two automated labelling approaches.

**Methods:** We used a model including two GANs each trained to transfer an image from one domain to another. The two inputs were a set of 100 longitudinal images of the gastrocnemius medialis muscle, and a set of 100 synthetic segmented masks that featured two aponeuroses and a random number of 'fascicles'. The model output a set of synthetic ultrasound images and an automated segmentation of each real input image. This automated segmentation process was one of the two approaches we assessed. The second approach involved synthesising ultrasound images and then feeding these images into an ImageJ/Fiji-based automated algorithm, to determine whether it could detect the aponeuroses and muscle fascicles.

**Results:** Histogram distributions were similar between real and synthetic images, but synthetic images displayed less variation between samples and a narrower range. Mean entropy values were statistically similar (real: 6.97, synthetic: 7.03;  $p = 0.218$ ), but the range was much narrower for synthetic images (6.91 – 7.11 versus 6.30 – 7.62). When comparing GAN-derived and manually labelled segmentations, intersection-over-union values- denoting the degree of overlap between aponeurosis labels- varied between 0.0280 – 0.612 (mean  $\pm$  SD: 0.312  $\pm$  0.159), and pennation angles were higher for the GAN-derived segmentations (25.1° vs. 19.3°;  $p < 0.001$ ). For the second segmentation approach, the algorithm generally performed equally well on synthetic and real images, yielding pennation angles within the physiological range (13.8–20°).

**Conclusions:** We used a GAN to generate realistic B-mode ultrasound images, and extracted muscle architectural parameters from these images automatically. This approach could enable generation of large labelled datasets for image segmentation tasks, and may also be useful for data sharing. Automatic generation and labelling of ultrasound images minimises user input and overcomes several limitations associated with manual analysis.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

In recent years, the use of machine- and deep learning approaches to analyse image data has rapidly accelerated. These

methods are now commonly used to identify features from various types of images, and in some cases, deep learning approaches are even used to support the medical decision-making process [1]. However, this kind of approach often relies on so-called supervised learning, whereby a large volume of training images is accompanied by hand-crafted labels for each image. The labels essentially identify which part(s) of the image are of interest, and a model

\* Corresponding author.

E-mail address: [neil.j.cronin@jyu.fi](mailto:neil.j.cronin@jyu.fi) (N.J. Cronin).

is then trained with the goal of learning a mapping between the original images and their corresponding labels.

Access to labelled data is often the biggest challenge in this domain, perhaps because it is difficult to acquire a large volume of suitable data, and/or because the labelling process requires expertise, and can thus be prohibitively expensive in terms of financial or time costs [2,3]. An additional challenge is the variability of the labelling process between different individuals, i.e. subjectivity. It would be desirable to develop methods that allow large volumes of annotated data to be compiled automatically, avoiding the need for excessive human effort and overcoming the effects of labelling variability. These data could then be used to train deep learning models.

Deep learning models often include some form of data augmentation to offset the effect of a small training dataset [4]. For example, training images can be rotated, rescaled or flipped, thereby creating additional images and expanding the size of the training set. However, these modifications only introduce minor additional diversity, and the augmented images are likely to be correlated with the original images. An alternative approach is to generate synthetic images using Generative Adversarial Networks (GANs) [5]. GANs have recently been used to generate realistic images based on a training set of real images in various fields, including medical imaging (e.g. [6]). Synthetic images have several possible applications. For example, GAN-based methods have been used to translate images between different domains, to correct imaging-related motion artefacts, and to denoise images [7]. This kind of approach can also be used to anonymise data to facilitate sharing [8], to improve image resolution [9], and to assist in various segmentation and classification tasks (for a review of medical imaging applications see [10]).

One area where synthetic images may be of value is the field of musculoskeletal ultrasound, where images are often taken of superficial muscles and tendons [11]. Although longitudinal ultrasound images often exhibit features such as muscle borders (aponeuroses) and muscle fascicles (bundles of muscle fibres), the ability to reliably identify these features may be limited by artefacts such as attenuation, scattering and refraction. Moreover, the data collection process is not well standardised between different labs and devices, so the same muscle scanned with different devices may result in quite different images. These issues make it difficult to develop robust analysis methods. As a result, the analysis of musculoskeletal ultrasound images still requires a lot of manual labour. Recently, fully automated analysis approaches for longitudinal ultrasound images have started to emerge [12,13]. However, it is not yet clear whether such approaches are sufficiently robust for broad, unsupervised applications. In this respect, a deep learning approach may be advantageous, but automating the analysis process using current deep learning techniques would likely require a large, diverse volume of labelled data for training. Although GANs have been used to generate ultrasound images, the focus of such work has been on fetal [14] and intravascular applications [15], and generally aimed at technical issues such as improving image resolution [16].

In this study, we investigated the ability to use a GAN framework to generate musculoskeletal ultrasound images that are realistic and statistically similar to real images. Since the efficacy of deep learning models relies on accurate labelling, a second aim was to test the suitability of two automated labelling approaches: 1) using an approach called Unsupervised Data to Content Transformation (UDCT [17]) to automatically segment images in an unsupervised manner (i.e. without any user input), and 2) using an existing automated analysis algorithm [12]. In both cases, the goal was to generate automatically-derived labels to accompany the synthetic images, thus yielding a labelled dataset without requiring any user input.

## 2. Methods

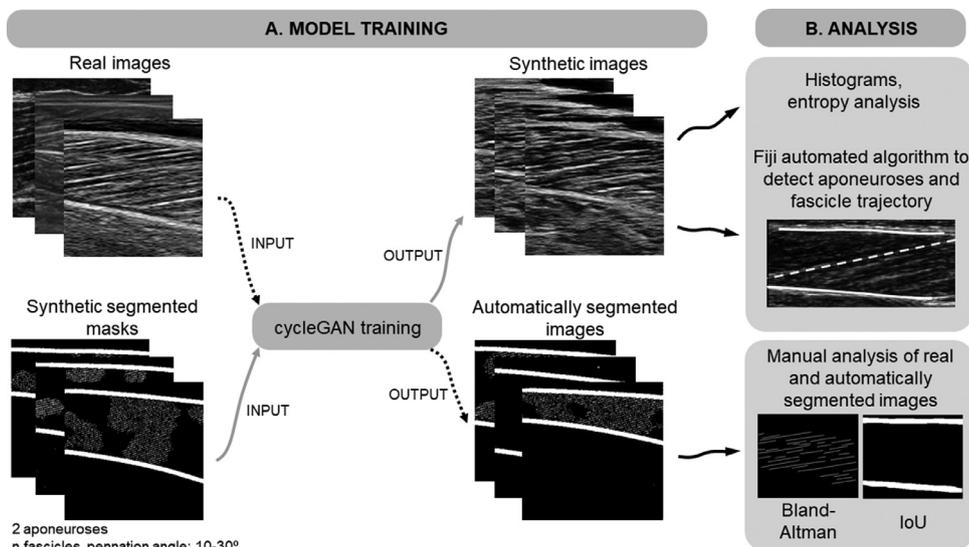
### 2.1. Data

To train a GAN model to produce synthetic images, we first manually curated a set of 100 longitudinal images of the gastrocnemius medialis (MG) muscle from the right legs of 52 different individuals (1–2 images per person). To increase data diversity, images were obtained using three different devices (Aloka alpha-10, Telemed Echoblaster 128, Phillips HD11; probe frequencies: 7–10 MHz; probe lengths: 50–60 mm). Images were acquired from healthy individuals aged between approximately 20–45, and were chosen on the basis that the superficial and deep aponeuroses, as well as muscle fascicles, were visible (Fig. 1). All images were from the authors' previous studies that had received institutional review board approval, and were collected by the authors, all of whom have >10 years of experience with this methodology. Anonymised images were manually cropped to  $256 \times 256$  pixels centred around the MG muscle and its superficial and deep aponeuroses using Matlab software (The Mathworks Inc., v2019b). This particular image size was used because larger images would have required excessive computational power and training time. A set of 100 synthetic segmented masks (also  $256 \times 256$  pixels) was also generated using some pre-coded features, namely that the generated image should include two aponeuroses and an unspecified number of 'fascicles' that extend at an angle that could realistically correspond to MG pennation angle (in this case between  $\sim 10$  and  $30^\circ$ , [18]). The fascicles in these masks randomly extended between the aponeuroses and so did not necessarily appear as complete fascicles, as is often the case in real ultrasound images. The real images and the synthetic masks served as the two inputs to the UDCT GAN model for training (described below), and the result was a set of fully synthetic ultrasound images and a set of automatically segmented images, i.e. segmented versions of the original (real) images that were input (Fig. 1A). This segmentation is achieved by using the information encoded in the synthetic masks, i.e. the synthetic masks help to transfer their encoded information to the domain of the original image to achieve the segmentation.

### 2.2. Approach

We used an approach called UDCT [17], which builds on a previous GAN architecture known as cycle-consistent generative adversarial networks (cycleGAN) [19]. The architecture consists of two GANs that are each trained to transfer an image from one domain to another. A GAN consists of a generator network and a discriminator network. The generator network is given a real input image and produces another one using a convolutional neural network architecture. The discriminator is a classifier which receives a generated image from the generator and predicts whether it is genuine or fake (synthetic). In cycleGAN, the generator takes images from its respective domain and creates images from the opposite domain (Fig. 1A). The generator is trained to produce images of such high quality that they are able to fool the discriminator into classifying them as real images. The discriminator is trained to distinguish synthetic images from real ones. In the optimal case, this adversarial procedure eventually results in the creation of realistic images [20].

In the UDCT architecture, the generators are deep convolutional neural networks, and the deepest encoding level is built with residual layers [21]. To speed up generator training, instance normalisation is applied layer-wise [22]. The discriminators use a PatchGAN architecture [23] with a  $70 \times 60$  receptive field. Because the UDCT approach requires two inputs (real images and synthetic masks), and this implementation of cycleGAN generates images from the opposite domain to each input, there are also



**Fig. 1.** Schematic of the approach. A. The model is trained using a batch of raw (real) images and a same-sized batch of synthetic segmented masks. The synthetic masks were designed to encode the main features of real images, namely that they should include two aponeuroses and an unspecified number of 'fascicles', with a pseudo-random pennation angle that should be within physiological limits. After cycleGAN training, the network is able to generate fully synthetic, realistic ultrasound images, as well as automatically-generated segmentations of input images. All image sizes are  $256 \times 256$  pixels. B: Details of the analysis procedures. IoU: intersection over union.

two outputs. The synthetic ultrasound images are one output, and the other is an automated segmentation of each real input image (Fig. 1A). We trained several models using different-sized datasets ranging from 20 to 100 real images. For smaller models, images were randomly selected from the larger pool of 100 images. For each model, there was always a matching number of synthetically generated segmentation masks. All model development was done in Python software (version 3) via Anaconda, using a modified version of the original UDCT implementation (<https://github.com/UDCTGAN/UDCT>) with a Tensorflow backend (additional code available on request).

### 2.3. Analyses

To assess the quality of the synthetic ultrasound images that were generated (other than by visual inspection), we computed histograms of greyscale values from the cropped real images as well as the synthetic images (all  $256 \times 256$  pixels) using Matlab, and compared the shape of their distributions and entropy values (a measure of image texture or information content).

As stated above, one study aim was to assess whether the segmented images produced by UDCT could be used to identify features of interest in ultrasound images. As well as visual interpretation, we compared these segmentations with manual labelling of the same images (100 images; labelling done by the first author). The manual process involved identifying the two aponeuroses, which was done by creating a binary mask for each image in Fiji software [24]. All visible fascicle parts within the image were also identified by creating a separate binary mask. Intersection over union (IoU, also known as the Jaccard index) was used to determine the overlap between the UDCT-derived and manually-derived aponeurosis locations. IoU was computed as the area of overlap divided by the area of union (i.e. the total area of both labels), so that a value of 0 denotes no overlap and a value of 1 denotes perfect overlap. Mean pennation angle was computed from the real and segmented images as the average of 5 measures from different sites within each image (10, 30, 50, 70 and 90% of the distance along the x-axis), and these data were used to produce a Bland-Altman plot [25]. We did not attempt to compute muscle fascicle length or muscle thickness because the UDCT approach generates synthetic images of a specific pixel size, but the images

do not inherently contain any absolute size information in mm. However, the effect of adding this information by scaling the image size is demonstrated for different scaling factors in Supplementary Fig. 2. Thus, a user wishing to deploy this method could select a suitable scaling factor in order to produce synthetic images with realistic dimensions.

The second segmentation approach that we examined involved first synthesising ultrasound images, and then feeding these images into an open source ImageJ/Fiji-based automated algorithm [12] (Fig. 1B), to determine whether it could automatically detect the aponeuroses and muscle fascicle trajectories, and whether the resulting pennation angles were also within the range of expected values for this muscle [18]. The output of this analysis included the locations of the detected aponeuroses and the mean trajectory of a single muscle fascicle.

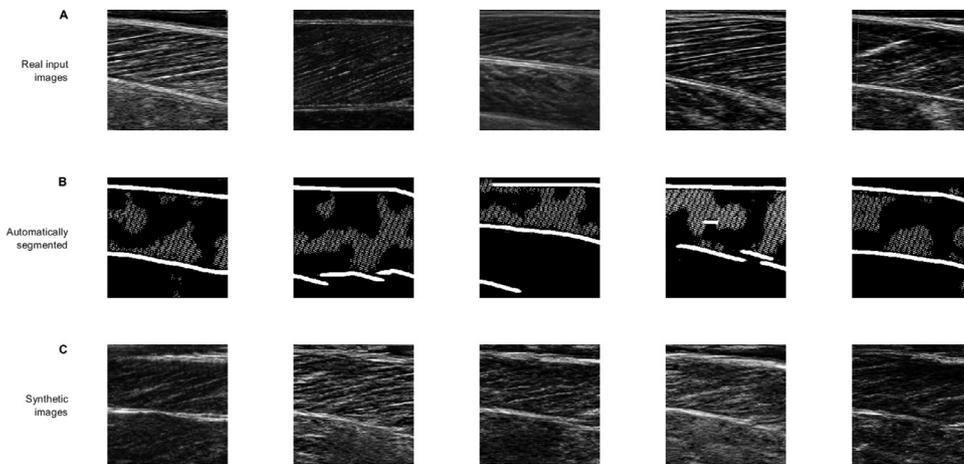
For comparisons between manually- and synthetically-derived parameters, we performed independent samples t-tests in Matlab using the `ttest2` function.

## 3. Results

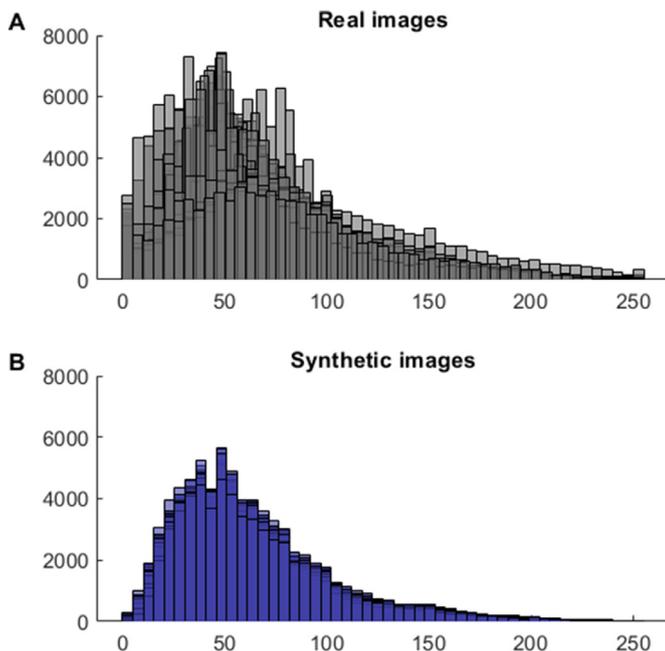
The training of GAN models was done with 200–300 iterations and a batch size of 1–2 (due to memory constraints). The latter stages of training were visualised in real-time to ensure that the model was producing realistic images, and training was interrupted when the model appeared to collapse, as evidenced by a clear, rapid decline in synthetic image quality or contrast. In these cases, the most recent checkpoint storing the model weights prior to the collapse was used to run inference. The duration of training using an Nvidia GeForce RTX GPU (8 GB ram) was about 2–4 h depending on the number of training images and iterations. Inference time, i.e. the time required to generate a new synthetic image, was less than 1 s using the same hardware.

### 3.1. Synthetic image generation

In general, as the training progressed, the subjective quality of the generated ultrasound images and the automatic segmentations improved (Supplementary Fig. 1). After training was complete, the model produced synthetic images that could often conceivably be mistaken for real images (Fig. 2C).



**Fig. 2.** Examples of real images (A), corresponding automatically segmented images (B), and a sample of fully synthetic images generated after training of the cycleGAN model (C). The images in C are independent of those in A and B. In some but not all cases, the segmented images in B appear to correspond well with the patterns in the real images (A). These examples are all taken from a model trained on 20 images.



**Fig. 3.** Histograms comparing the properties of real (A) and synthetic images (B). Histograms were computed from a random sample of 30 real images (taken from the larger set of 100) and 30 synthetic images, so that the patterns are easier to identify.

Histograms were computed from a batch of real and synthetic images (Fig. 3). The distribution of the values was statistically similar between the two sets of images, whereby both showed a right skewed distribution (mean skewness of real images: 1.03, synthetic images: 4.70;  $p = 0.568$ ). However, synthetic images displayed less variation between samples, as reflected by the different y-axis values in Fig. 3, as well as a narrower range of peaks on the x-axis. Mean entropy values for the images represented in Fig. 3 did not differ statistically (real: 6.97, synthetic: 7.03;  $p = 0.218$ ). However, the range of values was much narrower for the synthetic images (6.91 – 7.11) than the real images (6.30 – 7.62).

### 3.2. UDCT image segmentation

In general, the UDCT segmentations identified the correct number of aponeuroses from the real images, and in some cases, the

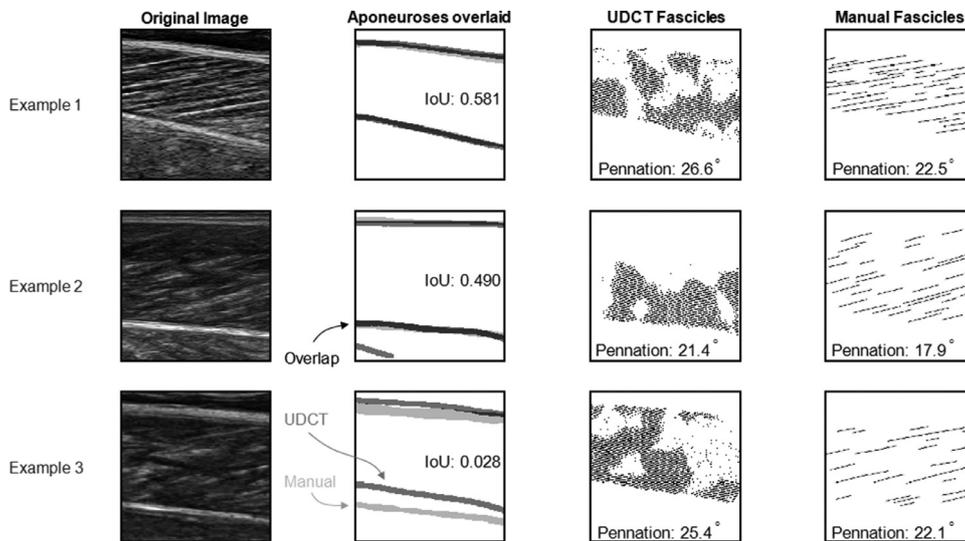
detected ‘fascicle’ structures also showed similar pennation angles to those in the real images. However, this was not a consistent finding (Fig. 2B; Figs. 4 and 5), and only very rarely did we observe an image where both fascicles and aponeuroses appeared to be correctly segmented. When comparing the UDCT-derived segmentations with manually labelled results, IoU values- denoting the degree of overlap between aponeurosis labels- varied between 0.0280 and 0.612 (mean  $\pm$  SD:  $0.312 \pm 0.159$ ). As shown in Fig. 4 and the Bland-Altman plot in Fig. 5B, pennation angle values (averaged from 5 sites) were consistently higher for the UDCT segmentations than the manually labelled values ( $25.1^\circ$  vs.  $19.3^\circ$ ;  $p < 0.001$ ).

### 3.3. Automated analysis of synthetic images using Fiji software

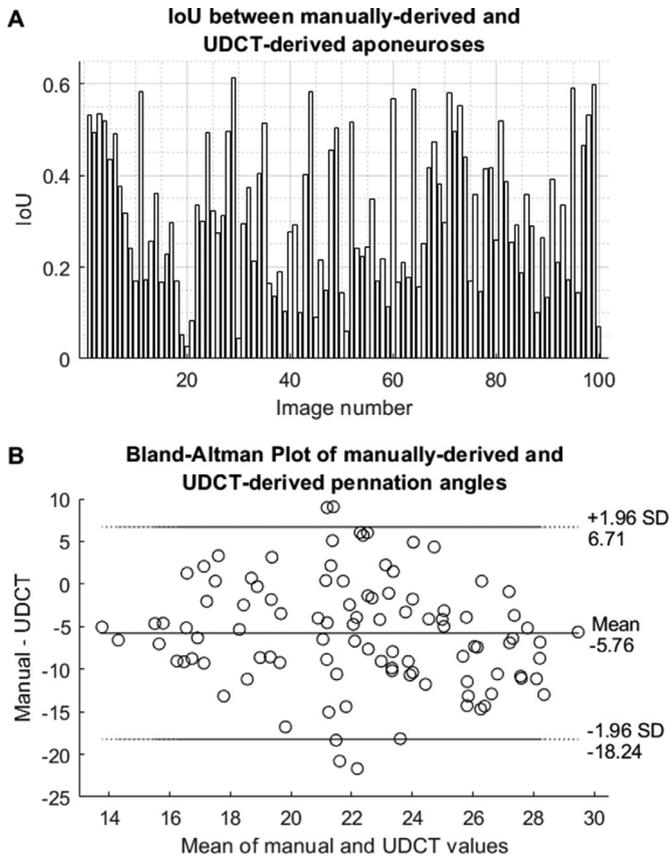
We fed a random set of 20 of the synthesised ultrasound images into an automated algorithm [12]. The output of this analysis included the locations of the detected aponeuroses and the mean trajectory of a single muscle fascicle. This analysis revealed that, in about 90% of cases, the algorithm was reliably able to identify the two aponeuroses and the dominant direction of the muscle fascicles in both synthetic and real images (Fig. 6). The resulting pennation angle values for synthetic images were also within the physiological range for this muscle ( $13.8\text{--}20^\circ$  for the images in Fig. 6). To demonstrate the effect of scaling the synthetic images to yield muscle thickness and fascicle length values in mm, the fascicle length and thickness values shown for the images in Fig. 6 (in pixels) were rescaled using a range of scale factors (Supplementary Figure 2).

## 4. Discussion

We present a method that generates fully synthetic ultrasound images in an unsupervised manner, requiring only a set of real images and a set of synthesised masks that encode the broad properties of the real images. The synthetic images that were generated were visually similar to real images, although there was less variability in the range and amplitude of the greyscale values between synthetic images. Moreover, although mean entropy values were similar between real and synthetic images, the variability in entropy between images was again smaller for synthetic images, implying less variation in image texture or information content. However, this issue is unlikely to limit the usefulness of the method in clinical settings, since the synthetic images were sufficiently realistic for our automated approach to be able to analyse them. Therefore, we believe that the approach presented here is a first step



**Fig. 4.** Comparison of image segmentation by the UDCT approach and by a human researcher. Each row shows a different example image, as well as the aponeurosis and fascicle labels derived by the two approaches. For each image, IoU values are shown, representing the overlap in aponeurosis locations between the UDCT and human labels. Manually-derived pennation angles averaged from 5 sites per image are also shown. For the aponeurosis overlays, light grey denotes manually-drawn aponeuroses, dark grey the UDCT-derived aponeuroses, and black regions show regions of overlap.



**Fig. 5.** A. Individual IoU values (for 100 images) denoting the degree of overlap between aponeurosis labels derived from the UDCT approach and manual labelling (0: no overlap, 1: perfect overlap). B. Bland-Altman plot of pennation angles derived from the two approaches. The solid horizontal line denotes the mean difference between datasets, and the dotted lines represent the 95% confidence intervals.

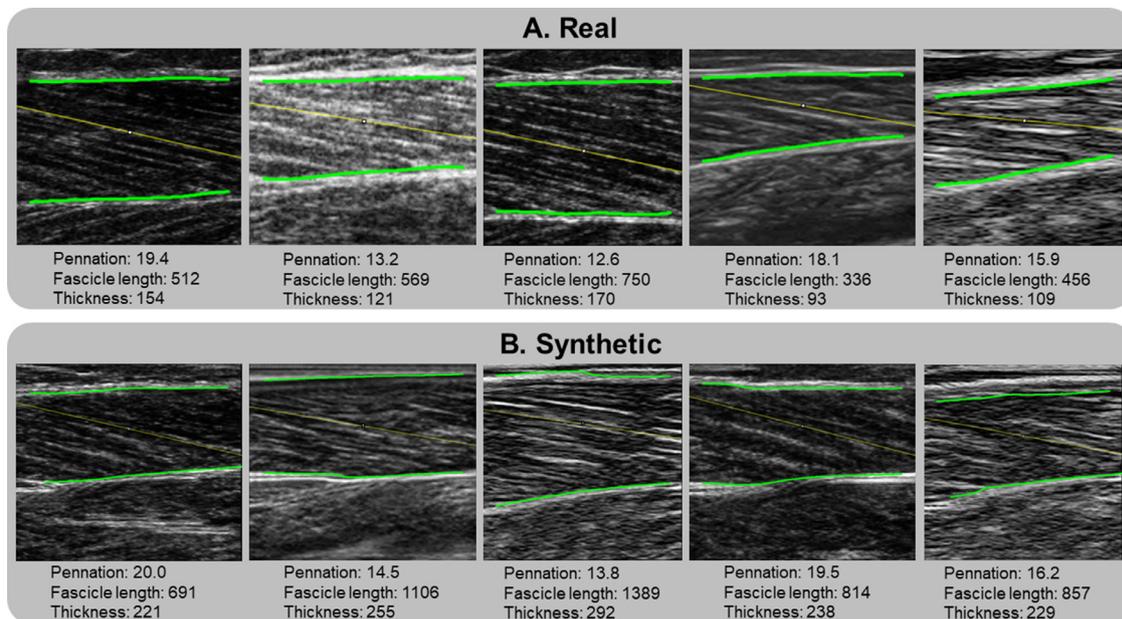
towards a fully automated pipeline for generating and annotating musculoskeletal ultrasound images. These images may in turn be useful for training larger deep learning models in cases where there is insufficient labelled training data available. To increase the

diversity of synthetic images, it may be possible to increase the diversity of the training set, e.g. by using images from several different muscles.

Although generating realistic images was the goal of this work, this approach may have an additional use case beyond training deep learning models. Laws concerning sharing of medical imaging data are very strict, especially with the recent introduction of GDPR in Europe. In some cases, it may not be necessary to share the actual medical data. Instead, synthetic data that are statistically representative of real data could suffice [26]. Synthetic images contain no metadata and by definition, cannot be traced back to an individual, thus avoiding one of the biggest obstacles of sharing this kind of data [8].

In addition to synthesising images, we tested two approaches for segmenting images automatically. For the first approach, we tested whether the segmentations produced by the UDCT method could be used to detect muscle architectural features. In some cases, the segmentation of the aponeuroses was reasonably good (IoU values close to 0.6), although this was the exception rather than the norm. It should be noted however that for muscle architectural analysis, only one border of each aponeurosis needs to be accurately identified (e.g. when calculating muscle thickness). In this respect, the IoU metric may be stricter than necessary for our purposes. The ‘pennation angle’ of muscle fascicles in the UDCT segmentations was generally higher than that obtained via manual analysis. In theory, an unsupervised segmentation approach like this could solve the issue of automated 2D ultrasound analysis, since a simple contour analysis of the segmented image would allow all relevant parameters to be extracted. It seems feasible that such an approach could be developed, perhaps using some variant of the methods used here, but the notorious difficulty of training stable GAN models [27] means that it may be challenging to produce a trained model that consistently yields accurate segmentation of all images.

The second segmentation approach we used relied on our recently developed ImageJ/Fiji-based software [12], which uses various thresholding techniques and measurements of gradient structure tensors to detect muscle aponeuroses and the mean orientation of muscle fascicles within regions of interest. The software failed in around 10% of cases, usually because the aponeuroses in the synthetic images were not sufficiently distinct from



**Fig. 6.** Output of the automated analysis of real and synthetic images using our previously developed approach [12]. A: A selection of real images. Pennation angle ( $^{\circ}$ ), fascicle length (pixels) and apparent muscle thickness (pixels) values are displayed on each image. B: The same analysis repeated for a selection of synthetic images. Note that we did not attempt to scale fascicle length or thickness values since the synthetic images do not inherently possess any scaling information.

surrounding tissues. However, in general, when the synthetic images appeared to be visually similar to real ones, the analysis yielded excellent segmentations, including pennation angle values that are consistent with published values for this muscle [18]. Thus, by combining the UDCT approach with our automated analysis method, it is possible to generate realistic synthetic images, and then analyse those images, all without the user needing to manually interpret or label any data. Musculoskeletal ultrasound analysis is often still reliant on manual (and thus subjective) methods. The analysis pipeline presented here could help to remove one of the major bottlenecks of working with large datasets in this field. Nonetheless, it should be noted that we did not actually train a deep learning model using synthetically-generated images to detect muscle architecture, so it remains to be determined whether such an approach would be sufficiently robust.

## 5. Conclusions

We used a cycleGAN approach to generate realistic ultrasound images, and we extracted muscle architectural parameters from these images automatically using open source software. This approach could be used to generate large labelled datasets for the purpose of training deep neural networks for image segmentation tasks, and may also be useful for sharing data anonymously. The ability to automate both the generation and labelling of ultrasound images means that the user input is minimal, thereby overcoming the subjectivity (and time costs) associated with data analysis. Code and training data from this project are available upon request.

## Declaration of Computing Interest

None of the authors have any conflicts to declare, financial or otherwise.

## Acknowledgements

The authors are grateful to Stephan Ihle for valuable discussions and assistance with the implementation of the UDCT method. No

specific ethical approval was required for this study. The authors have no competing interests to declare. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: [10.1016/j.cmpb.2020.105583](https://doi.org/10.1016/j.cmpb.2020.105583).

## References

- [1] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nat. Med.* 25 (2019) 44–56, doi:[10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7).
- [2] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, (2018). <https://doi.org/10.1016/j.neucom.2018.09.013>.
- [3] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, *IEEE Trans. Med. Imaging.* 35 (2016) 1153–1159, doi:[10.1109/TMI.2016.2553401](https://doi.org/10.1109/TMI.2016.2553401).
- [4] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: *Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105, doi:[10.1145/3065386](https://doi.org/10.1145/3065386).
- [5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, (2014). <http://arxiv.org/abs/1406.2661> (Accessed 14 November 2019).
- [6] P. Costa, A. Galdran, M.I. Meyer, M.D. Abràmoff, M. Niemeijer, A.M. Mendonça, A. Campilho, Towards Adversarial Retinal Image Synthesis, n.d.
- [7] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, B. Yang, MedGAN: medical image translation using GANs, *Comput. Med. Imaging Graph.* 79 (2020) 101684, doi:[10.1016/j.compmedimag.2019.101684](https://doi.org/10.1016/j.compmedimag.2019.101684).
- [8] H.C. Shin, N.A. Tenenholtz, J.K. Rogers, C.G. Schwarz, M.L. Senjem, J.L. Gunter, K.P. Andriole, M. Michalski, Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: *SASHIMI 2018 Simul. Synth. Med. Imaging*, Springer Verlag, 2018, pp. 1–11, doi:[10.1007/978-3-030-00536-8\\_1](https://doi.org/10.1007/978-3-030-00536-8_1).
- [9] C. You, W. Cong, M.W. Vannier, P.K. Saha, E.A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE), *IEEE Trans. Med. Imaging.* 39 (2020) 188–203, doi:[10.1109/TMI.2019.2922960](https://doi.org/10.1109/TMI.2019.2922960).
- [10] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: a review, *Med. Image Anal.* 58 (2019) 101552, doi:[10.1016/j.media.2019.101552](https://doi.org/10.1016/j.media.2019.101552).
- [11] N. Cronin, G. Lichtwark, The use of ultrasound to study muscle-tendon function in human posture and locomotion, *Gait Posture* 37 (2013) 305–312 <http://www.sciencedirect.com/science/article/pii/S0966636212002949>.

- [12] O.R. Seynnes, N.J. Cronin, Simple Muscle Architecture Analysis (SMA): an ImageJ macro tool to automate measurements in B-mode ultrasound scans, *PLoS ONE* 15 (2020) e0229034, doi:[10.1371/journal.pone.0229034](https://doi.org/10.1371/journal.pone.0229034).
- [13] C. Caresio, M. Salvi, F. Molinari, K.M. Meiburger, M.A. Minetto, Fully automated Muscle Ultrasound Analysis (MUSA): robust and accurate muscle thickness measurement, *Ultrasound Med. Biol.* 43 (2017) 195–205, doi:[10.1016/j.ultrasmedbio.2016.08.032](https://doi.org/10.1016/j.ultrasmedbio.2016.08.032).
- [14] Y. Hu, E. Gibson, L.L. Lee, W. Xie, D.C. Barratt, T. Vercauteren, J.A. Noble, Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks, in: RAMBO 2017, C. 2017, Switch 2017 Mol. Imaging, Reconstr. Anal. Mov. Body Organs, Stroke Imaging Treat, Springer Verlag, 2017, pp. 105–115, doi:[10.1007/978-3-319-67564-0\\_11](https://doi.org/10.1007/978-3-319-67564-0_11).
- [15] F. Tom, D. Sheet, Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning, in: Proc. - Int. Symp. Biomed. Imaging, IEEE Computer Society, 2018, pp. 1174–1177, doi:[10.1109/ISBI.2018.8363780](https://doi.org/10.1109/ISBI.2018.8363780).
- [16] R. Wang, Z. Fang, J. Gu, Y. Guo, S. Zhou, Y. Wang, C. Chang, J. Yu, High-resolution image reconstruction for portable ultrasound imaging devices, *EURASIP J. Adv. Signal Process.* 2019 (2019) 1–12, doi:[10.1186/s13634-019-0649-x](https://doi.org/10.1186/s13634-019-0649-x).
- [17] S.J. Ihle, A.M. Reichmuth, S. Girardin, H. Han, F. Stauffer, A. Bonnin, M. Stamparoni, K. Pattisapu, J. Vörös, C. Forró, Unsupervised data to content transformation with histogram-matching cycle-consistent generative adversarial networks, *Nat. Mach. Intell.* 1 (2019) 461–470, doi:[10.1038/s42256-019-0096-2](https://doi.org/10.1038/s42256-019-0096-2).
- [18] M.V. Narici, T. Binzoni, E. Hiltbrand, J. Fasel, F. Terrier, P. Cerretelli, In vivo human gastrocnemius architecture with changing joint angle at rest and during graded isometric contraction, *J. Physiol.* 496 (1996) 287–297.
- [19] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, (2017). <http://arxiv.org/abs/1703.10593> (Accessed 14 November 2019).
- [20] T. Karras, S. Laine, T. Aila, A Style-Based Generator Architecture for Generative Adversarial Networks, (2018). <http://arxiv.org/abs/1812.04948> (Accessed 14 November 2019).
- [21] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *ArXiv*. (2015). <https://arxiv.org/abs/1512.03385v1> (Accessed 15 November 2018).
- [22] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance Normalization: the Missing Ingredient for Fast Stylization, *ArXiv*. (2016). <http://arxiv.org/abs/1607.08022v3> (Accessed 16 April 2020).
- [23] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, *ArXiv*. (2016). <http://arxiv.org/abs/1611.07004v3> (Accessed 16 April 2020).
- [24] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.Y. Tinevez, D.J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis, *Nat. Methods* 9 (2012) 676–682, doi:[10.1038/nmeth.2019](https://doi.org/10.1038/nmeth.2019).
- [25] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet* 1 (1986) 307–310 <http://www.ncbi.nlm.nih.gov/pubmed/2868172>.
- [26] B.K. Beaulieu-Jones, Z.S. Wu, C. Williams, R. Lee, S.P. Bhavnani, J.B. Byrd, C.S. Greene, Privacy-preserving generative deep neural networks support clinical data sharing, *Circ. Cardiovasc. Qual. Outcomes* 12 (2019) e005122, doi:[10.1161/CIRCOUTCOMES.118.005122](https://doi.org/10.1161/CIRCOUTCOMES.118.005122).
- [27] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved Techniques for Training GANs, *ArXiv*. 1606.03498 (2016). <https://arxiv.org/abs/1606.03498> (Accessed 15 November 2019).