

# Spatio-temporal Consistency and Negative Label Transfer for 3D freehand US Segmentation

Vanessa Gonzalez Duque, Dawood Al Chanti, Marion Crouzier, Antoine Nordez, Lilian Lacourpaille, Diana Mateus

▶ **To cite this version:**

Vanessa Gonzalez Duque, Dawood Al Chanti, Marion Crouzier, Antoine Nordez, Lilian Lacourpaille, et al.. Spatio-temporal Consistency and Negative Label Transfer for 3D freehand US Segmentation. the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention„ Oct 2020, Lima, Peru. 10.1007/978-3-030-59710-8_69 . hal-02734902

**HAL Id: hal-02734902**
**https://hal.science/hal-02734902**

Submitted on 2 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatio-temporal Consistency and Negative Label Transfer for 3D freehand US Segmentatione*

Vanessa Gonzalez Duque[1], Dawood Al Chanti[1], Marion Crouzier[2], Antoine Nordez[2], Lilian Lacourpaille[2], and Diana Mateus[1]

[1] Ecole Centrale de Nantes, Laboratoire des Sciences du Numérique de Nantes LS2N, UMR CNRS 6004 Nantes, France.
[2] Université de Nantes, Laboratoire "Movement, Interactions, Performance", EA 4334 Nantes, France.

**Abstract.** The manual segmentation of multiple organs in 3D ultrasound (US) sequences and volumes towards their quantitative analysis is very expensive and time-consuming. Fully supervised segmentation methods still require the collection of large volumes of annotated data while unlabeled images are abundant. In this work, we propose a semi-automatic deep learning approach modeled as a weak-label learning problem: given a few 2-D incomplete annotations for selected slices, the goal is to propagate the masks to the entire sequence. To this end, we make use of both positive and negative constraints induced by incomplete labels to penalize the segmentation loss function. Our model is composed of one encoder and two decoders to model the segmentation and an auxiliary reconstruction task. Moreover, we consider the spatio-temporal information by deploying a Convolutional Long Short Term Memory module. Our findings suggest that the reconstruction decoder and the spatio-temporal information lead to a better geometrical estimation of the mask shape. We apply the model to the task of low-limb muscle segmentation in a dataset of 44 patients and 6160 images.

**Keywords:** 3-D ultrasound · weakly supervised learning · guided backpropagation · convolutional LSTM · fully convolutional neural networks.

## 1 Introduction

Duchenne Muscular Dystrophy (DMD) is a degenerative muscular disorder in which muscle fibers are replaced with fat. Treatment follow-up is commonly done through imaging of the lower limb muscles under Magnetic Resonance (MR) [18, 14]. Due to the early onset of the disease, patients are often children, for whom MRI is unpractical. 3-D US-imaging is rapidly evolving [13] offering an inexpensive and portable alternative, yet needs further clinical validation. An identified imaging bio-marker for DMD evolution is muscle volume [17]. Quantifying such

* This work has been supported in part by the European Regional Development. Fund, the Pays de la Loire region on the Connect Talent scheme (MILCOM Project) and Nantes Métropole (Convention 2017-10470),

information often requires the segmentation of the 3-D images. Towards the validation of 3-D ultrasound as a viable alternative for DMD follow-up, we propose an automatic segmentation algorithm for 3-D freehand ultrasound images.

Segmentation in US images comes with unique challenges including low imaging quality, high variability, attenuation, speckle and shadows [13]. Furthermore, the contrast between areas of interest is often low [3], among others due to and orientation dependence of the acquisition. These difficulties exist not only for automatic segmentation algorithms but also for the clinical experts who annotate the "ground-truth" [25]. Moreover, for 3-D US or ultrafast acquisitions, it is unpractical to label every imageFinally, in certain images, the structures can be more easy to segment than others, which may lead to incomplete masks.

It is essential to develop advanced methods that handle the above challenges to make assessment more objective and accurate. Herein, we propose a deep learning segmentation approach that relies on the spatial coherence of an image sequence (or contiguous slices of a 3D volume) to better exploit incompletely annotated data. We design a network architecture based on the encoder-decoder topology, built using depthwise separable convolutions. Moreover, we rely on spatio-temporal information to partially compensate the missing sequential annotations and recover better boundaries.

Similar to [16], we bring negative evidence from complementary masks from other organs, to constraint the area of prediction. However, we further propagate the information across the sequence by means of a Convolutional Long Short Term Memory (CLSTM)[23] placed in the bottleneck of an encoder-decoder architecture. The CLSTM captures the possible short and long range muscle deformations, while preventing to propagate incorrect or noisy information via the gated mechanism learning. To improve network convergence and to help preventing over-fitting, separable depth-wise convolutions [24] are favorable as they have less parametersFinally, to preserve the boundaries and the spatial structure, we enforce the encoding path to learn a compact representation that preserves the geometrical properties of the input sequence via an auxiliary reconstruction decoder trained in a fully unsupervised manner.

We evaluated our method over a total of 44 participants and 6160 images to evaluate our model performance for muscle segmentation. We performed an ablation study to evaluate the effectiveness and usefulness of each novelty. We show that the proposed method produces muscle segmentation results of high quality, scores an average dice similarity coefficient up to 94.5% with full annotation and up to 70.8% with 50% of the annotations.

## 2 Related Work

The well-known U-net architecture [19] has been successfully extended in [2] to fuse features between the encoding and the decoding path in a non-linear way using LSTMs. Such connections have the advantage of enhancing feature propagation and encourage feature reuse. Both [19] and [2] are limited by their inability to incorporate temporal information, that can facilitate the segmen-

tation task with sequential or volumetric data. To exploit the dynamics, we integrated a CLSTM within the U-net, similar to [4, 1, 21]. CLSTM can perceive the entire spatio-temporal context and provide more discriminative features. Our integration of CLSTM is done at the bottleneck of encoding path, wherein two CLSTMs for the spatio-temporal encoding path are deployed and followed by another two CLSTMs for the temporal decoding path. So far, CLSTM is used in fully supervised way but non of [4, 1, 21] exploit it for feature propagation when dealing with incomplete masks. In particular, we integrate prior information, e.g the masks of other easier to segment organs provided by clinical experts, to guide the network back-propagation. In this way, an interactive semi-automatic segmentation can be leveraged.

To reduce the cost of full pixel-wise image annotations, weak segmentation methods exist. [10] proposed a method that seed with weak localization cues, e.g. object position, and then to expand objects based on information about which classes can occur in an image. Then, the segmentation is constrained to coincide with object boundaries. The disadvantage is the dependency of the segmentation quality with respect to the combination of those terms in the loss function. [5] investigated bounding box annotations as an alternative source of supervision to train Fully Convolutional Network (FCN) and to recover segmentation masks. Their method deploys region proposal networks [12] to generate candidate segmentation masks. The FCN is then trained under the supervision of these approximate masks. Despite the competitive performance demonstrated for object recognition, this method requires a huge amount of bounding boxes, up to 123k. Medical data cannot afford such amount of annotation. Simple yet effective approaches [15, 16] address the weak supervision by compensating the missing annotations for a specific organ by incorporating background labels. As missing annotations are usually considered as background pixel classes, prior knowledge of other known classes could be leveraged to restrict the area of prediction for the missing class. Although interesting, these methods do not propagate the spatio-temporal coherence of sequential data. We consider both the spatio-temporal feature propagation and the use of prior regarding the other organs that are easily annotated, to generate a true negative mask related to the background which compensate the absence of the true positive mask. With the presence of the true positive annotated mask, e.g at time step $t - 8$, and using CLSTM for feature propagation, mask approximated is attained.

Multi-task learning [11, 8] shown to improve the performance of different tasks with auxiliary objective functions. We explore an unsupervised reconstruction task that seeks to reproduce the sequential US slices to aid the weak supervision of the segmentation task. We build a model that operates on the the same encoding path. Our finding shows that the reconstruction task helps to efficiently preserve the geometrical and appearance structure of the segmented mask, yielding better shape estimation. To handle multi-task learning, we propose a principled way of solving multiple loss functions to simultaneously learn multiple objectives instead of a naive weighted sum combination [6]. The novelty lies in its strength to update the network parameters twice at the same time for

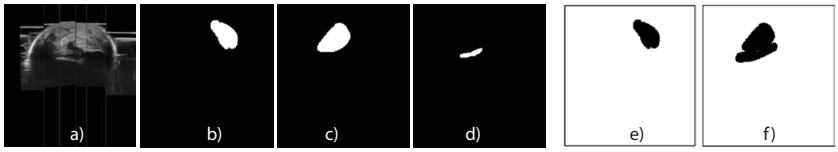



**Fig. 1.** (a) 2-D US image; its manual segmentation masks ((b):GM, (c):GL, (d):SOL); (e) The background of annotated GM. (f) Generated true negative background.

each iteration. Hence, if the segmentation network parameters are stuck in a saddle point, the optimizer of reconstruction task has the opportunity to re-update the gradient into a better position. We could think about it as fine-tuning a model trained for reconstruction task and then it is tuned for the segmentation task, with main difference, that it happened concurrently.

## 3 Method

With the long term objective of assisting the volume and other quantitative measurements during the follow-up of DMD patients, we address here the problem of segmenting muscles in series of 2-D images. Given a 3-D US image of the lower leg, the goal is to retrieve a segmentation mask for one of the three muscles: the Gastrocnemius Medialis (GM) and Lateralis (GL), and the Soleus (SOL). Given the difficulties of manually annotating such difficult sequences, we present a FCN model and a training strategy that rely on incomplete 2-D annotations, where only some of the slices are annotated and not necessarily with all the muscle masks (see Fig. 1). To train a deep learning model under these constraints, we devise a training strategy capable of handling and propagating partial annotations while exploiting all the available information, i.e. the location of other muscles is advantageous for constraining the extent of the foreground prediction.

We propose a spatio-temporal multi-task approach performing two important and complementary tasks: 1) segmentation and 2) image reconstruction. The segmentation relies on a spatio-temporal U-Net with a CLSTM in the bottleneck ensuring the propagation of information across slices. Furthermore, two competing masks are considered: the foreground mask containing the muscle of interest, and the background mask filled with negative evidence from other annotated organs. The purpose of the auxiliary reconstruction task is to compress and store the important spatio-temporal information into a compact representation.

**Model Architecture**. The core of the FCN model is a combination of two decoders sharing the same encoding path (see Fig. 2). The encoding path extracts compact low resolution features with convolutional blocks. The feature maps from the last encoder layer are fed to a CLSTM module to capture the spatio-temporal transition within inter-slices and help compensating for missing annotations. The output of CLSTM is then passed to two decoders, the first focusing on reconstructing the original image, while the second on the segmentation task. The last layer of the reconstruction decoder is mapped into one

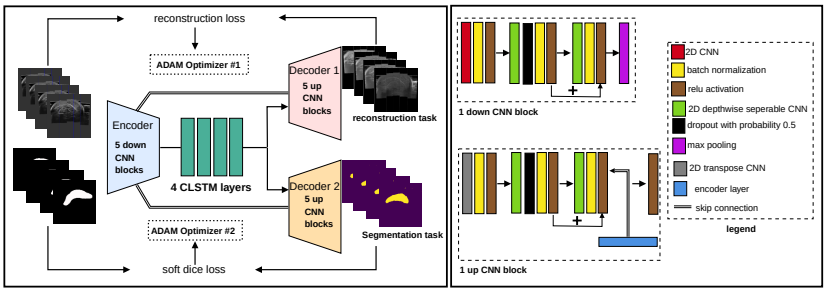



**Fig. 2.** Schematic representation of our network architecture.

channel while the segmentation decoder is mapped onto $C$ maps, where $C$ represent the number of classes. Afterwards, the output feature maps $C$ are passed to a pixel-wise softmax layer to generate the probabilities of the predicted masks.

*Architecture details:* Our model is composed of an encoder with structure of 5 residual depthwise-separable convolutional blocks and max-pooling operations, gradually projecting the gray-scale channel into 16, 32, 64, 128 and 256 feature maps at each layer respectively. The CLSTM at the bottleneck is composed of 4 stacked cells of 256 feature maps. The decoder mirrors this structure.

**Loss functions**. The reconstruction objective function is the average Mean Square Error (st-MSE) between an input sequence $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ of $T$ frames, and the corresponding output reconstructions $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_T\}$, which is achieved by the reconstruction network $f_\theta(.)$ parameterized by $\theta$. Then the reconstruction loss is:

$$\text{st-MSE}(\mathbf{X}, \hat{\mathbf{X}}, \theta) = \sum_{t=1}^{T} (\mathbf{x}_t - f_\theta(\mathbf{x}_t)) \tag{1}$$

where $\mathbf{x}_t$ is the $t$-th 2-D slice, and $f_\theta(\mathbf{x}_t)$ denotes the corresponding output of the reconstruction branch.

The segmentation loss is the Soft Dice Coefficient (SDC) adapted to suit sequential data, we refer to as (st-SDC). Consider a sequence of $T$ annotations $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$ corresponding to the input sequence $\mathbf{X}$, and the masks $\hat{\mathbf{Y}} = \{g_\omega(\mathbf{x}_1), \ldots, \hat{g}_\omega(\mathbf{x}_T)\}$ estimated by the segmentation network $g_\omega(.)$ parameterized by $\omega$. Then the segmentation loss is:

$$\text{dice}(\mathbf{y}_t, g_\omega(\mathbf{x}_t)) = \left( \frac{2 \sum_{\text{pixels}} \mathbf{y}_t \, g_\omega(\mathbf{x}_t)}{\sum \mathbf{y}_t^2 + \sum g_\omega(\mathbf{x}_t)^2} \right) \tag{2}$$

$$\text{st-SDC}(\mathbf{Y}, \hat{\mathbf{Y}}, \omega) = \frac{1}{T} \sum_{t=1}^{T} (1 - \text{dice}(\mathbf{y}_t, g_\omega(\mathbf{x}_t))) \tag{3}$$

where $\mathbf{y}_t$ is the available ground truth and $g_\omega(\mathbf{x}_t)$ is the output of the segmentation branch. The normalization ensures the loss to be between 0 and 1. The optimisation of the loss is conditioned upon on the type of available annotations

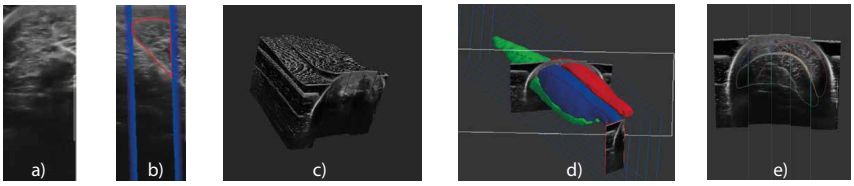



**Fig. 3.** a) B-mode image. b) manual annotation over B-mode. c) 3D volume. d) 3D segmented volumes. e) cross-section with GM,GL and SOL annotations

at time $t$. Consider a set of annotated frame $\mathbf{y}^a$ in which their true positive foreground is available. In this case, eq. (3) will be minimized. On the other hand, consider a set of un-annotated frame $\mathbf{y}^{un}$ in which their true negative background is generated, then eq. (3) will be maximized (or minimizing its negative).

**Multi-task learning.** Multi-task learning is typically treated as a weighted sum of two criteria. However, as different tasks may conflict finding suitable weighting hyperparameters is complex. Instead, we follow a multi-objective approach, similar to [20]. To this end, we optimise each objective function separately using two ADAM optimisers [9]. Since both networks $f_\theta(.)$ and $g_\omega(.)$ share the same encoder path, we alternatively update the networks parameters $\theta$ and $\omega$. With the proposed multi-task training, the encoder learns to extract a compact representation that not only serves the segmentation task, but also the reconstruction. Conditioning the encoder to preserve the spatio-temporal information required to reconstruct a sequence of images, favors a better geometrical representation in the bottleneck. The improved representation leads then to better segmentation masks, especially in the boundaries.

## 4 Experimental Validation

**Data acquisition.** A total of 59 acquisitions taken from 44 volunteers aged between 18 and 45 years old is recorded. Every acquisition consists of a sequence of 2-D B-mode US images (Fig3-a) acquired with a Supersonix Ultrasound machine and a 40mm linear VERMON probe. Images are recorded every 5 mm displacement in low speed mode. The probe is followed with an optical tracking system and then the 3-D volumes is reconstructed (Fig3-c) using stradwin software [7]. The B-mode images are of size $227 \times 544$ with a pixel spacing of 0.176mm/pixel. Volumes are recovered to fill a grid of $1372 \times 632 \times 2270$. The manual annotation of the GM, GL and SOL muscles is performed over $300 \pm 96$ muscles in B-mode US images (Fig3-b). For some cases, it was possible to segment GM and GL only. Hence, a second acquisition was acquired to segment in particular the SOL.

After volume reconstruction (Fig3-c), 3-D annotations (Fig3-d) are obtained through a surface fitting algorithm [22]. The comparison between the annotations from two experts lead only to a 3% volumetric difference validating the approach. The masks used for training were extracted as transverse-sections of the surface models (Fig3-e) computed from the more expert examiner annotations.

**Experimental setup**. For each participant, we extract low resolution cross-sections ($300 \times 400$ pixels) from the reconstructed volumes and select a sub-volume with 140 images. The data split consists in 29 train sequences ($29 \times 140 = 4060$ images) coming from participants in which the ground-truth mask of the GM, GL and SOL was provided over a single slice. For validation and test purposes, we used the data of other 5 and 10 (700 and 1400 images) participants with two sequences (coming from different acquisitions).

In our experimental setting, we consider different ratios of annotations to train our model. The 100% annotation setting corresponds to the 140 images available for a subvolume along with their ground truth for the muscle of interest (e.g. GM). However, having 30% of annotations means that only 42 masks (e.g. GM) out of 140 images are given to the network. Instead of ignoring the 98 images with un-annotated GM masks, we generated a true negative mask using the prior information about SOL and GL.

In our experimental validation, we perform first an ablation study to determine the performance contribution of each proposed component. We then present the results for different amounts of annotation ratios. Finally, we compare our model performance with the upper bound setting of full supervision with different % of annotation, but without prior information from other muscles. We use the validation set for hyperparameter tuning, and report the performance on the test set. We use Dice (DSC), mean Intersection over Union (mIoU), and Hausdorff distance error (HDE) to quantify the accuracy of the predicted segmentation map, and present also qualitative results.

### 4.1 Model Ablation study

To validate the contribution of the novel model components with respect to a plain U-net, we perform a comparative study with full supervision (100% of the SOL masks for training are used). SOL was chosen being the hardest muscle to segment. First, we replace the fully convolutional operators with separable depth-wise convolutions, we refer to this model as *Unet-S*. Then, we integrate the second decoder for reconstruction in the *Unet-S-R* model. Finally, we evaluate our full model including *CLSM* module on the top of *Unet-S-R* and we refer to as *Unet-S-R-CLSM*. Table 1 demonstrates the effectiveness of using separable depthwise convolutions with Unet architecture when deployed over a moderate

| Ablation studies | DSC (%) | mIoU (%) |
|---|---|---|
| Unet | 78 | 76 |
| Unet-S | 82 | 79 |
| Unet-S-R | 87 | 85 |
| **Unet-S-R-CLSM (ours)** | **91** | **89** |

**Table 1.** Comparison of the baseline (Unet) for segmenting SOL muscle vs the different variants proposed.

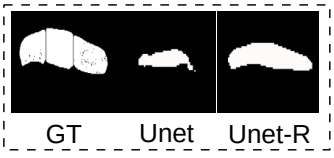



**Fig. 4.** Predicted mask from Unet and Unet with reconstruction.

| label % | DSC | mIoU | HDE |
|---|---|---|---|
| 100 | 94.5 | 91 | 2.42 |
| 90 | 90.1 | 88 | 2.85 |
| 70 | 83.4 | 76 | 4.20 |
| 50 | 70.8 | 61 | 4.87 |
| 30 | 50.6 | 42 | 6.50 |

a) under different annotation ratios.

| label % | Our model | Supervised setting |
|---|---|---|
| 90 | 90.1 | **91.4** |
| 70 | 83.4 | **88.8** |
| 50 | **70.8** | 66.2 |
| 30 | **50.6** | 43.1 |

b) with and without negative priors.

**Table 2.** Performance of the proposed model under different annotation setting.

size dataset. Moreover, our finding suggests that deploying the additional reconstruction decoder improves both the mIoU and the DSC measures.Deploying CLSTM dramatically improves the performance as it exploits the full spatio-temporal structure of the input data. Finally, Figure 4 shows the visual differences of the predicted mask between the Unet and Unet with reconstruction decoder. The geometrical structure is better preserved with the latter.

### 4.2 Decreasing amounts of annotations and negative priors.

To cope with the difficulty of the manual annotations, we evaluate the model's capacity to learn from fewer annotations and to exploit available true negative knowledge from other muscle masks. The results for the GM muscle are presented in Table 2. The model obtains reasonable DSC scores with fewer annotation.

To analyse the contribution of negative priors we compare the results above to a version of the model run on the ground truth masks of the relevant muscle only, without the prior knowledge from other muscles. The results are presented in Table 2(b). The first column is the model performance with true negative evidence. The second column reports the performance when it use only the available % of the relevant muscle annotations. The contribution of prior negative knowledge is relevant when the % of un-annotated data is less than 50%.

A second way to transfer knowledge from one muscle to another is fine-tuning. We train our model over 100% annotations of GL muscle. We then fine-tune the model with 50% of the GM annotations. We obtain 68.8% of DSC, which is comparable but less interesting than the 70.8% DSC obtained with our model in the experiments (Table 2(a)) with 50% annotations.

## 5 Discussion and Conclusions

In this paper, we proposed a deep learning approach to segment muscles in 3D freehand ultrasound data. Our model benefits from the spatio-temporal structure of the data at the feature level as well from an auxiliary reconstruction task. For the latter we also present a multi-objective training strategy that avoids the need of finding loss weights. We also explore different means to transfer prior knowledge from complementary masks and study the behavior of the different components under less annotations. Experimental results show that with good

amounts of supervision, the spatio-temporal consistency enforced through the CLSTM, as well as the addition of a parallel reconstruction decoder are effective tools to improve the segmentation results. The use of complementary masks is the most useful when the amount of the annotated ground truth is relatively small (up to 1000 images). Future work aims at improving the system to handle patient data. We will also validate the usability when using the proposed method as mask initialization to reduce time of expert's segmentation. The methodology may also benefit other type of sequential data as ultrafast US imaging, as well as other applications, where muscle images need to be segmented as in sports.

## References


1. Arbelle, A., Raviv, T.R.: Microscopy cell segmentation via convolutional lstm networks. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1008–1012. IEEE (2019)
2. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S.: Bi-directional convlstm u-net with densley connected convolutions. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019)
3. Cerrolaza, J.J., Sinclair, M., Li, Y., Gomez, A., Ferrante, E., Matthew, J., Gupta, C., Knight, C.L., Rueckert, D.: Deep learning with ultrasound physics for fetal skull segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 564–567. IEEE (2018)
4. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In: Advances in neural information processing systems. pp. 3036–3044 (2016)
5. Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1635–1643 (2015)
6. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE international conference on computer vision. pp. 2650–2658 (2015)
7. Gee, A., Prager, R., Treece, G., Cash, C., Berman, L.: Processing and visualizing three-dimensional ultrasound data. The British journal of radiology **77**(suppl_2), S186–S193 (2004)
8. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7482–7491 (2018)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2014)
10. Kolesnikov, A., Lampert, C.H.: Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In: European Conference on Computer Vision. pp. 695–711. Springer (2016)
11. Kuga, R., Kanezaki, A., Samejima, M., Sugano, Y., Matsushita, Y.: Multi-task learning using multi-modal encoder-decoder networks with shared skip connections. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 403–411 (2017)
12. Li, B., Yan, J., Wu, W., Zhu, Z., Hu, X.: High performance visual tracking with siamese region proposal network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8971–8980 (2018)

13. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T.: Deep learning in medical ultrasound analysis: a review. Engineering (2019)
14. Loram, I.D., Maganaris, C.N., Lakie, M.: Use of ultrasound to make noninvasive in vivo measurement of continuous changes in human muscle contractile length. Journal of applied physiology **100**(4), 1311–1323 (2006)
15. Lu, Z., Fu, Z., Xiang, T., Han, P., Wang, L., Gao, X.: Learning from weak and noisy labels for semantic segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(3), 486–500 (2016)
16. Petit, O., Thome, N., Charnoz, A., Hostettler, A., Soler, L.: Handling missing annotations for semantic segmentation with deep convnets. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 20–28. Springer (2018)
17. Pichiecchio, A., Alessandrino, F., Bortolotto, C., Cerica, A., Rosti, C., Raciti, M.V., Rossi, M., Berardinelli, A., Baranello, G., Bastianello, S., et al.: Muscle ultrasound elastography and mri in preschool children with duchenne muscular dystrophy. Neuromuscular Disorders **28**(6), 476–483 (2018)
18. Pillen, S., Arts, I.M., Zwarts, M.J.: Muscle ultrasound in neuromuscular disorders. Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine **37**(6), 679–693 (2008)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
20. Sener, O., Koltun, V.: Multi-task learning as multi-objective optimization. In: Advances in Neural Information Processing Systems. pp. 527–538 (2018)
21. Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J.: Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. In: Advances in neural information processing systems. pp. 2998–3006 (2015)
22. Treece, G.M., Prager, R.W., Gee, A.H., Berman, L.: Surface interpolation from sparse cross sections using region correspondence. IEEE transactions on medical imaging **19**(11), 1106–1114 (2000)
23. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. pp. 802–810 (2015)
24. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6848–6856 (2018)
25. Zlateski, A., Jaroensri, R., Sharma, P., Durand, F.: On the importance of label quality for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1479–1487 (2018)