Muscle Ultrasound Image Segmentation using Deep Learning: A Comparative Study of FCN, U-Net and its variations

Maria Morandini, Antonio Roberto Ventura, Bryan Ranger

Anonymous CVPR submission

Paper ID *****

1. Introduction

Human body composition measurements such as fat and muscle mass play a critical role in qualifying health and nutritional status, the impact of disease, and change due to nutritional, therapeutic, or behavioral intervention on patients.

Specifically, due to the low market price these technologies have reached in the past decade, they are progressively becoming game-changer in global public health, particularly in resource-constrained regions.

These kind of measures have a variety of possible applications, we list here a few of them:

- Forecast of musculoskeletal diseases such as osteoporosis, arthritis and muscular dystrophy [2];
- Help identifying muscle atrophy or hypertrophy in response to injuries or diseases [1];
- Train practitioners in interpreting ultrasound images accurately;

Among all of these possible applications, advised by professor Bryan Ranger, we identified the space for a research that connects the growth pace of the children with the thickness of their muscles.

In order to achieve our purposes, we will be working with different pre-existing Deep Learning models for segmentation. We will implement minor changes to reach the highest scores on the specific task of muscle measurements.

2. Related Work

The literature on segmentation models is extensive, and we reviewed both general computer vision models and those tailored to ultrasound images.

The models we decided to explore more are Fully Convolutional Neural Networks [4] on the side of purely segmentation models and Unet [5] and its variations for ultrasound images segmentation models.

However, it needs to be pointed out that there are not resources on this specific segmentation task, being our data set private. Undertaking research that combines computer vision with growth analysis necessitates a new proposal. This proposal must account for the unique characteristics of our specific dataset, requiring modifications to pre-existing models to effectively train them to analyze growth patterns.

3. Method

Our work centered around the implementation of different models with the scope of comparing their performances and deliver the best one to professor Bryan Ranger to continue his research (in pending due to the completion of privacy agreements). Then, on top of our ”chosen” model, we built a simple algorithm to get the average height of the segmented area.

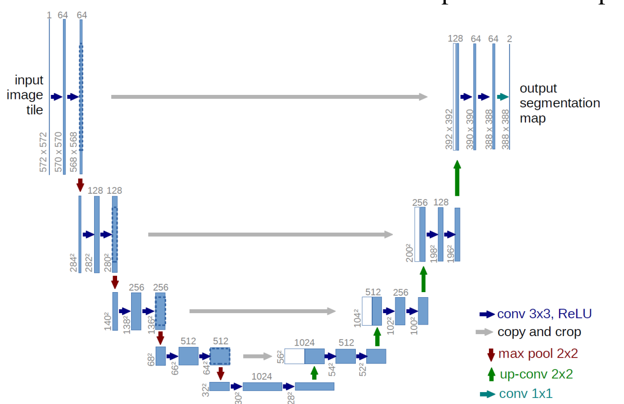
Before stepping on the discussion of the methods we need to point out that all of our models are evaluated through the IOU metric. This choice was done due to the segmentation nature of the task: a high IOU value indicates a good overlap between the predicted and ground truth regions, while a low value indicates a poor overlap. Typically, a threshold IOU value is set to determine if an object has been accurately detected or segmented.

3.1. Fully Convolutional Neural Network

FCNs are an architectures used mainly for semantic segmentation. They employ solely locally connected layers, such as convolution, pooling and upsampling. Avoiding the use of dense layers means less parameters (making the networks faster to train). It also means an FCN can work for variable image sizes given all connections are local. [4] The network we built consists downsampling path, used to extract and interpret the context, and an upsampling path, which allows for localization. FCNs also employ skip connections to recover the fine-grained spatial information lost in the downsampling path.

3.2. Unet

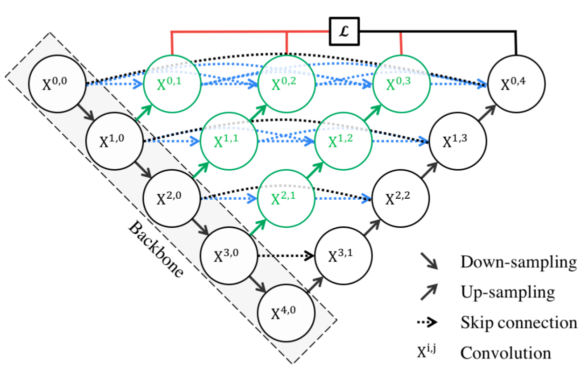
Unet architecture was first introduced for the ISBI challenge for segmentation of neuronal structures. It originated as a modification of FCNN to make it work with fewer training images and to give more accurate predictions. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. The name of the architecture comes from its particular shape.



The main difference with FCN is the employement of skip connections between downsampling and upsampling paths. In addition to this, instead of sampling only once in the decoding layer multiple upsampling layers are added. This newtwork works only on square images.

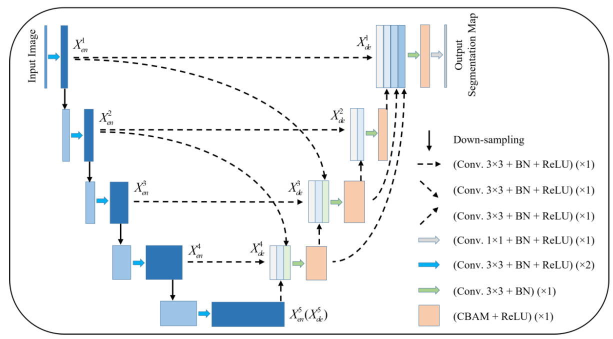
3.3. Unet++

We implemented also Unet ++ [6], wich relying on dense layers improves the performance of UNet. Convolution layers and Deep Supervision are added as well. Deep supervision in our case is on the accurate mode accurate mode wherein the outputs from all segmentation branches are averaged.



3.4. Unet-3plus

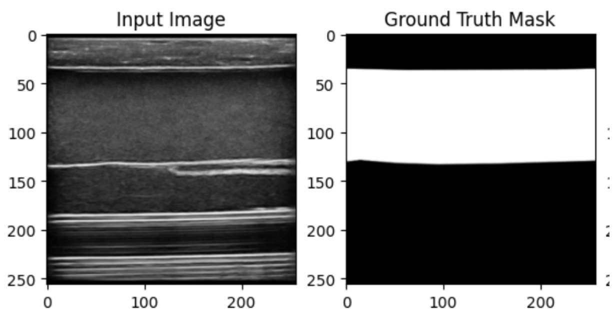
The last variation of Unet we implemented is Unet 3 - plus [3]. Initially idealized by Zhejiang University, with the main scope of medical image segmentation. The main variation is the usage of full-scale skip connections, which allows the combination of low-level details and high-level semantics from feature maps in multiple scales. The deep supervision approach (same as UNET ++) learns hierarchical representations from the feature maps that have been aggregated at full scale.



4. Experiments

4.1. Data preprocessing

The dataset we were given from the lab of professor Bryan ranger, is composed of 300 images of size 415 x 287. Each image has its own mask, namely a binary image, highlighting the segmentation task we need to complete. Due to privacy reasons, the request to have real clinical data from the Boston Children Hospital is still pending. The images where thus obtained using a Clarius probe to scan phantoms. These phantoms that mimic muscle tissues were built in the lab to mimic two pieces of the human quadriceps's muscle: the vastus intermedius and the rectus femoris. The creation of the training and testing sets was realized through Photoshop, by manually segmenting the images. We uploaded the dataset and we reshaped the images to have square images of 256 x 256. This is requested in some of the models we employ such as Unet. The images we were given are RGB even if they are in gray scale, as usual for ultrasound images. In the data preparation we discarded two dimensions to ease the training in the future steps. We then split the data into training set and test set, with proportions of 80 % and 20 %.
Here we have an example of the ultrasound image and its manually segmented mask



4.2. Training

When training deep learning models for image segmentation tasks, we started with the U-Net architecture and gradually progressed to more advanced models such as U-Net++, U-Net 3+. We finished by implementing FCN. For each of these models, we initially used implementations that were available from other researchers and adapted them to suit our specific task and dataset. We decided to utilize the Binary Cross Entropy as the loss function and the stochastic gradient descent as the optimizer.

One important modification that we made to all of the models was to ensure that the input and output dimensions for the color channels were consistent and reduced to a single dimension. This allowed us to process grayscale images more efficiently, which were the input images for our segmentation task. To achieve this, we adjusted the number of input and output channels for the first and last layers of the models to match the number of grayscale channels.

Due to the computational cost of training deep learning models for image segmentation, we were limited to using a single epoch and a small batch size for each model. However, even with these limitations, we were able to obtain great results on our segmentation task. As a result, we did not find it necessary to increase the batch size for further training, as this would only increase the computational requirements without significant improvement in accuracy.

4.3. Testing and Results

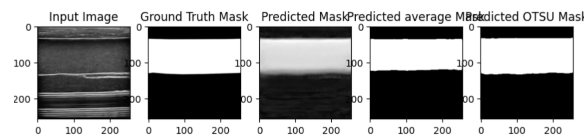
We observed that the output images generated by all the models did not have binary values of only 0 or 255. Therefore, we opted to apply a post-processing step of thresholding to obtain the final segmentation results with only black or white pixels. We opted for two methods, manual segmentation with a threshold value, found using a grid search on a range of values on a validation set and taking the one that gave the best result, and Otsu thresholding. After comparing the two, we discarded the Otsu one, in favour of the manual, since it gave us worse results.

After having created the final masks, we computed the IOU value of each model, measuring it on every image in the test set and taking the average. Table 1 shows the results.

Table 1. average IOU value on the test set

Model	IOU
FCNN	0.904
U-Net	0.99468
U-Net++	0.99471
U-Net 3-Plus	0.99469

Based on the results obtained, it can be seen that all four models performed well in segmenting muscle ultrasound images. The FCN model achieved an IOU score of 0.904, while the U-Net, U-Net++, and U-Net 3+ models achieved IOU scores of 0.99468, 0.99471, and 0.99469, respectively.



It is worth noting that the IOU scores obtained by all the models are quite high, indicating that they were successful in accurately segmenting the images. The U-Net model and its variations achieved the highest IOU scores, which may suggest that their architectures are particularly well-suited to the task of segmenting muscle ultrasound images. Additionally, the original masks, since manually segmented, may present a very slightly error, suggesting that the models errors may be amplified.

Finally, we have developed a simple algorithm that measures the thickness of the muscle. It first measure for each column of the image the longest consecutive list of white pixels and, then, it takes its average. The decided to implement this instead of just computing the average of white pixels per column, to minimize the possibility of mistake. Here we present the pseudocode.

```
set avg_length1 to 0
for each column x in mask:
  find indices where consecutive
  sequences start and end in column x

  set arr to the indices where the value
  in column x is 255

  set diff to the difference between
  adjacent values in arr

  set diff_idx to the indices where diff
  is not equal to 1, incremented by 1
```

```
  insert 0 at the beginning of diff_idx

  append len(arr) at the end of diff_idx

  set seq_len to the differences between
  consecutive elements of start_idx and
  end_idx

  set max_len to the maximum value in
  seq_len
  add max_len to avg_length1

set avg_length1 to avg_length1 divided
by the number of columns in mask
```

5. Conclusions

Thanks to our experiments we observed, as expected, that U-net models and its variations, perform way better than FCNN. As mentioned above, these models were born as an improvement of FCNN.

U-net ++ is the model we decided to retain for our further analysis, since it is faster than U-net 3-plus and slightly better than U-net. Our next step in the project is the employment of this trained model on real clinical data, as soon as they will be given to the lab.

6. Contributions

- Introduction: Maria
- Related work: Maria and Antonio
- Methods: Maria and Antonio
- Experiments:
  - 4.1 Maria
  - 4.2 Maria and Antonio
  - 4.3 Antonio
- Conclusions: Maria and Antonio

REPOSITORY

References

[1] Hashem E. Zikry Carly A. Creelman Rachel Welbel Alfred C. Gellhorn, Jordan M. Stumph. Ultrasound measures of muscle thickness may be superior to strength testing in adults with knee osteoarthritis: a cross-sectional study, 2018. 1

[2] Van der Paelt. Automated muscle segmentation of the lower leg in mri for the diagnosis and follow-up of inflammatory myopathies, 2019. 1

[3] Ruofeng Tong1 *Hongjie Hu2 Qiaowei Zhang2 Yutaro Iwamoto3 Xianhua Han3 *Yen-Wei Chen3 4 1 Jian Wu1 Huimin Huang 1, *Lanfen Lin1. Unet 3+: A full-scale connected unet for medical image segmentation, 2020. 2

[4] Trevor Darrell Jonathan Long, Evan Shelhamer. Fully convolutional networks for semantic segmentation, 2015. 1

[5] V. B. Shereena G. Raju. Medical ultrasound image segmentation using u-net architecture, 2022. 1

[6] Nima Tajbakhsh Jianming Liang Zongwei Zhou, Md Mahfuzur Rahman Siddiquee. Unet++: A nested u-net architecture for medical image segmentation, 2018. 2