DL-UNet: A Convolutional Neural Network for Median Nerve Segmentation

1st Andrian Melnic
IT Engineering student
UNIVPM, Ancona, Italy
s1098384@studenti.univpm.it

2nd Edoardo Conti
IT Engineering student
UNIVPM, Ancona, Italy
s1100649@studenti.univpm.it

3rd Lorenzo Federici
IT Engineering student
UNIVPM, Ancona, Italy
s1098086@studenti.univpm.it

***Abstract*—Carpal tunnel syndrome is the compression of median nerve, is a condition that causes numbness, tingling, or weakness in the hand. It commonly occurs in individuals working in occupations that involve use of vibrating manual tools or tasks with highly repetitive and forceful manual exertion. CTS diagnosis is done with Ultrasound imaging by monitoring the movement of the nerve. Medical US image segmentation present several challenges such as low image quality, noise, diversity and data insufficiency. Over the past few years, several interesting Deep Learning based solutions were presented to overcome these challenges. In this work, we investigate a few of them, in particular Lightweight Unet and Double Unet, as well as how they behave when applied on our particular problem, that is the localization and segmentation of the median nerve. Last, but not least, we also propose our variant called DL-Unet, that integrates some of the features from the models mentioned above. In our particular case, it achieved higher performance and generated average Dice measurement, Intersection over Union, precision and recall values of 0.9027, 0.8145, 0.9460 and 0.8499, respectively.**

***Index Terms*—Carpal tunnel syndrome, Median nerve, Ultrasound Images, Segmentation, Deep Learning, Convolutional Neural Network, U-Net, Lightweight Unet, Double Unet, Double Lightweight Unet**

I. Introduction

Carpal tunnel syndrome (CTS) is one of the most common peripheral neuropathies that causes pain, numbness, and tingling in the hand and arm. It occurs when the tunnel becomes narrowed or when tissues surrounding the flexor tendons swell, putting pressure on the median nerve. In most patients, carpal tunnel syndrome gets worse over time, so early diagnosis and treatment are important. Early on, symptoms can often be relieved with simple measures like wearing a wrist splint or avoiding certain activities[1]. If pressure on the median nerve continues, however, it can lead to nerve damage and symptoms worsening. To prevent permanent damage, surgery to take pressure off the median nerve may be recommended for some patients[1].

Women are three times more likely than men to develop carpal tunnel syndrome. People with diabetes or other metabolic disorders that directly affect the body's nerves and make them more susceptible to compression are also at high risk[2]. CTS usually occurs only in adults. Workplace factors may contribute to existing pressure on or damage to the median nerve. The risk of developing CTS is not restricted to people in a single industry or job, but may be more reported in those performing assembly line work such as manufacturing, sewing, finishing, cleaning, and meatpacking than it is among data-entry personnel[2]. As can be seen from Figure 1[3], the Carpal Tunnel, narrow where the tendons pass through the wrist to reach the hand, is bounded by the transverse carpal ligament on the volar side and eight carpal bones on the dorsal side. Inside it pass both the median nerve, responsible for the sensitivity of the thumb, index, middle and ring fingers, and the nine digital flexor tendons.

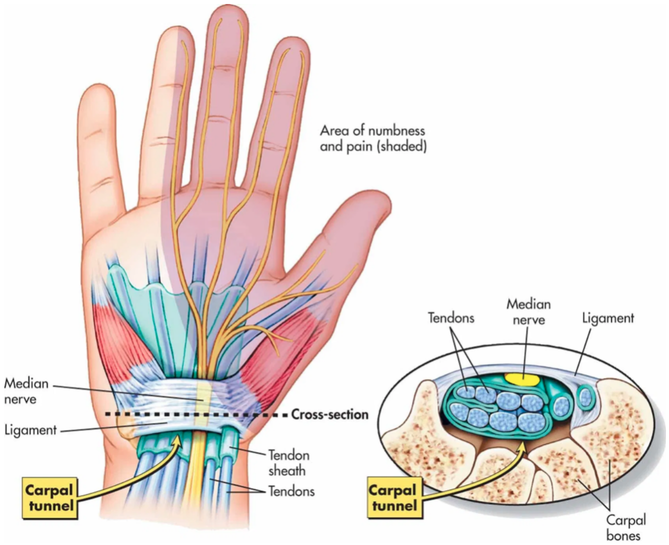


Fig. 1: Median nerve and carpal tunnel

One of the most used method to diagnosis CTS is using ultrasound imaging, this thanks to his convenience, lower cost, non-invasiveness and shorter examination times. It is used most to monitors the movement of the median nerve. Several studies have shown that the cross-sectional area and the flattening ratio of the median nerve in US are the most effective parameters for identifying the swelling of the median nerve [1]. Measurements that describe the deformation of the

[1] https://orthoinfo.aaos.org/en/diseases–conditions/carpal-tunnel-syndrome/

[2] https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Carpal-Tunnel-Syndrome-Fact-Sheet

[3] https://backblog.co.uk/conditions/carpal-tunnel-syndrome/

median nerve are: Area, Perimeter, Proportions and Circularity. Using ultrasound images and deep approaches, it is possible to study and obtain the affected area relating to the median nerve, diagnosing the presence or absence of carpal tunnel syndrome. Ultrasound image sequences are generally acquired at 24 fps. From the Figure 2 it can be seen that the median nerve (ROI) is the one marked in red.

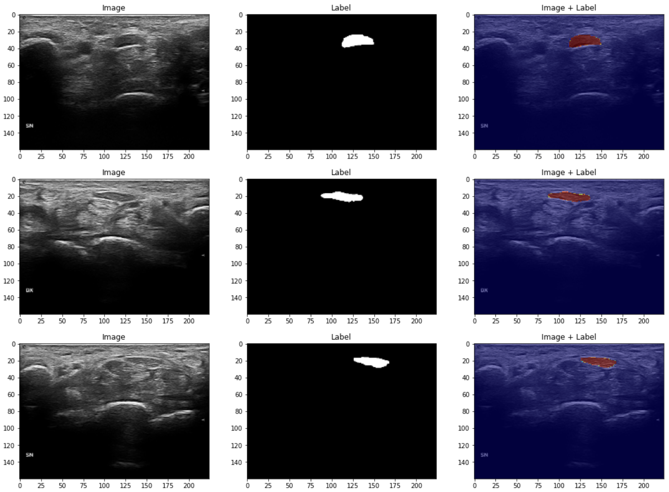


Image
Label
Image + Label

Fig. 2: Ultrasound image of Carpal Tunnel

The goal of this paper is to develop a CNN architecture for the segmentation of the median nerve from ultrasound images based on the U-Net architecture. The architecture need to be able to predict masks as close as possible to the annotation indicated in the Figure 2. We designed our model and we called it DL-Unet. It essentially is a compressed version of the DB-Unet proposed by Debesh Jha et al. [2] with lower training time but at the same with slightly better performance.

II. Related Work

The typical application of a machine learning based approach for median nerve's US image segmentation is to classify the ROI (e.g. diseased region or healthy region). The steps involved for designing such an application begins with the pre-processing stage which may involve the use of a filter to remove any noise or for the purpose of contrast enhancement. Following the pre-processing stage, the image is segmented using a segmentation technique like thresholding, clustering based approach and edge-based segmentation. Next, several features are extracted based on color information, texture, contrast and size from the ROI from which only the dominant ones are extracted using feature selection techniques like principal component analysis (PCA) or statistical analysis. Afterwards, the selected features are used as an input to the ML classifier such as Support Vector Machine (SVM) or a Neural Network [3]. The method for median nerve localization proposed by Oussama Hadjerci et al. [4] is an example of this type of metodology. The approach was based on despeckle filtering, feature engineering and SVM as classification technique and it achieved a high accuracy of 89% of the f-score measure. Juan J. Giraldo et al. [5] proposed a peripheral nerve segmentation method in medical ultrasound images, based on Non-parametric Bayesian Hierarchical Clustering.

Deep learning has recently emerged as the leading machine learning tool in various research fields, and especially in general imaging analysis and computer vision. Deep learning also shows huge potential for various automatic US image analysis tasks [6] and it's expeditiously turning into the state-of-the-art for medical image processing because of the performance improvements in diverse clinical applications [3].

A Deep Learning-based Classifier (DLC) like a Convolutional Neural Network (CNN) doesn't require segmentation and feature engineering since it can process the raw pixels values directly and the convolution kernels are learned during the training process. Although, some image pre-processing techniques (e.g.itensity normalization and contrast enhancement) may still be necessary, a DLC can have higher classification accuracy as it can avoid errors associated with erroneous feature vector or imprecise segmentation [7].

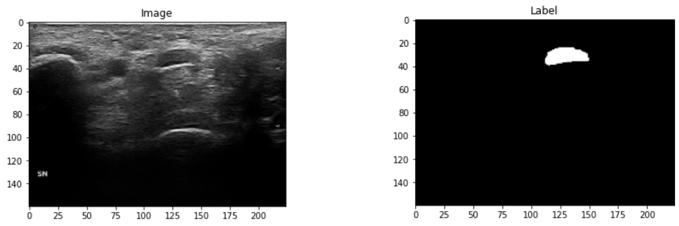
Existing modern Deep Learning approaches are based on the Unet, one of the most used Fully Convolutional Neural Network (FCNN). Ming-Huwi Horng et al. [1] proposed a new convolutional neural network for the localization and segmentation of the median nerve in ultrasound image sequences called DeepNerve, which generated an average Dice measurement of 89.75%. The segmentation results of DeepNerve are significantly higher in comparison with those of conventional active contour models. Debesh Jha et al. proposed Double Unet, a combination of two Unet architectures stacked on top of each other where the first one uses a VGG-19 pre-trained on ImageNet as the encoder. Double Unet's encouraging results, produced on various medical image segmentation datasets, show that it can be used as a strong baseline for medical image segmentation to measure the generalizability of Deep Learning (DL) models [2].

III. Materials & Methods

In this section, we illustrate the dataset, the data augmentation technique, the experimental setup configuration, the evaluation metrics and the implemented architectures.

A. Dataset

The dataset is made of 492 images (in .bmp format) with a resolution of 606x468, 246 ultrasound images of the median nerve (E.g. Figure 3a), and 246 masks annotated by professionals who have marked the ROI for each ultrasound image (E.g. Figure 3b). Diagnoses were made on 103 different patients, multiple images may come from the same patient for example because of the right and left hand. The filename describes the patient and the number of the relative ultrasound image/mask: Pz003-004 = 3rd patient, 4th image/mask.



(a) e.g. Ultrasound Image

(b) e.g. Mask

B. Data Pre-Processing

In order to train and test the neural network the images and masks were resized to a resolution of 384x288. All the images were standardized by scaling pixel values to have a zero mean and unit variance. Also, all the masks were normalized in order to rescale pixel values from the range of 0-255 to the range 0-1, which is preferred for neural network models. The dataset's splitting method proposed for our model is *patient-based* in order to avoid that images of the same patient occur simultaneously in the test and training set, therefore a possible *bias* on certain patients. The 246 images were splitted in 3 sets:

Dataset type	Total images
Training Set	164
Validation Set	38
Test Set	44

TABLE I: Nr. of images for each set

C. Data Augmentation

Medical datasets are challenging to obtain and annotate. A lot of existing datasets have only a few samples (in the hundreds), which makes the training of DL models challenging [2]. So to address this issue and improve model performance by reducing chances of overfitting, we adopted on-the-fly data augmentation techniques such as: Fixed rotations, shears, horizontal shifts and flips.

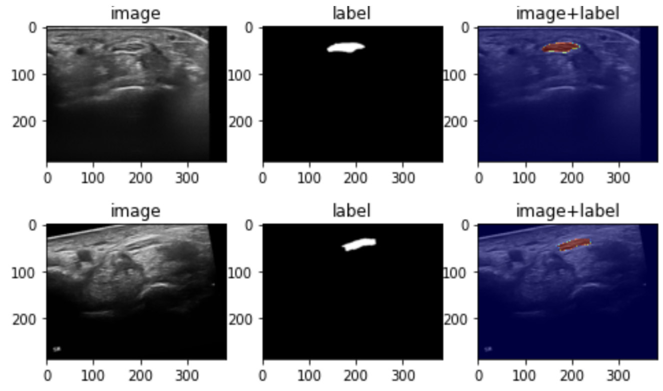


Fig. 4: E.g. of Data Augmentation

D. Hyperparameters

All models were trained for 100 epochs and the *batch size* was set to 8. The initial learning rate was set to $1e^{-3}$, but *ReduceLROnPlateau* and *Early Stopping* callbacks have also been used. *ReduceLROnPlateau* reduces the learning rate by a factor of 0.1 if no loss improvement is seen for 10 epochs and *Early Stopping* stops the training process once it stagnates for 50 epochs. The implemented loss function is BCE-Dice. It combines the Dice loss with the Binary Cross-Entropy (BCE), which is generally the default loss for segmentation models. Combining the two methods allows for some diversity in the loss, while benefitting from the stability of BCE [4].

E. Evaluation Metrics

Performance of medical images segmentation is often evaluated by the Dice Similarity Coefficient (DSC eq.1), Intersection over Union (IoU eq.2), Precision (eq.3) and Recall (eq.4). So we decided to use these as our evaluation metrics.

$$DSC = \frac{2\,|x \cap y|}{|x| + |y|} \tag{1}$$

$$IoU = \frac{|x \cap y|}{|x \cup y|} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F. Unet

The starting point is *Unet* [8], an encoder-decoder based five-layer architecture that has been effectively applied to biomedical image segmentation. The encoder consists of a repeated application of two 3x3 convolutions, each followed by a Rectified Linear Unit (ReLU) and a 2x2 Max Pooling operation for downsampling. Every step in the decoder include an upsampling of the feature map followed by a 2x2 convolution, a concatenation with the correspondingly cropped feature map and two other 3x3 convolutions with ReLU as activation function. The final 1x1 convolution is used to map each feature vector to the desired number of classes, in our case only one (the median nerve).

G. The Lightweight Unet architecture

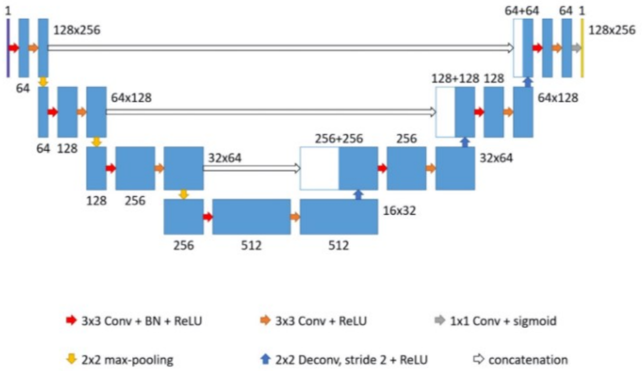


Fig. 5: The Lightweight Unet Architecture

[4] https://www.kaggle.com/bigironsphere/loss-function-library-keras-pytorch#BCE-Dice-Loss

The next architecture comes from DeepNerve, it's a compressed version of the Unet and it is called *Lightweight Unet (L-Unet)* [1]. It reduces the network's depth from 5 to 4 layers and uses batch normalization as a follow-up step to the first convolution in each layer in order to avoid premature convergence. The designed L-Unet contains much less trainable parameters compared to Unet with comparable results.

H. The DoubleU-net architecture

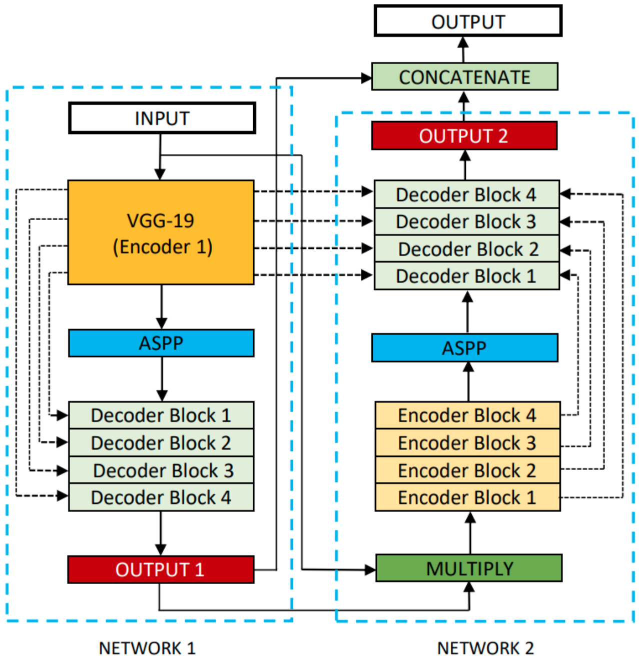


Fig. 6: The Double Unet Architecture

This architecture is proposed in *"DoubleU-net: A deep convolutional neural network for medical image segmentation"* [2] by Debesh Jha et al. As mentioned above in II, Double Unet (DB-Unet) is a combination of two Unet architectures stacked on top of each other. The first starts with a VGG-19 pre-trained on ImageNet as encoder. The output of the first Unet is multiplied with the input image and acts as input for the second Unet. The squeeze-and-excite blocks reduce the redundant information and pass the most relevant information. Also ASPP has been a popular choice for modern segmentation architectures because it helps to extract high-resolution feature maps that lead to superior performance. [2]

1) Network 2 Encoder: Each encoder block in the second Unet performs two 3 × 3 convolution operation, each followed by a batch normalization. The batch normalization reduces the internal co-variant shift and also regularizes the model. A Rectified Linear Unit (ReLU) activation function is applied, which introduces non-linearity into the model. This is followed by a squeeze-and-excitation block, which enhances the quality of the feature maps. After that, max-pooling is performed with a 2 × 2 window and stride 2 to reduce the spatial dimension of the feature maps [2].

2) Decoders: Each decoder block performs a 2 × 2 bilinear up-sampling on the input feature, which doubles the dimension of the input feature maps. Then, the appropriate skip connections feature maps from the encoder are concatenated to the output feature maps. In the first decoder, only skip connection from the first encoder are used. In the second decoder, the skip connections from both the encoders are used, which maintains the spatial resolution and enhance the quality of the output feature maps. After concatenation, two 3 × 3 convolution operations are performed, each of which are followed by batch normalization and then by a ReLU activation function. After that a squeeze-and-excitation block is used. At last, a convolution layer with a sigmoid activation function is applied, which is used to generate the mask for the corresponding modified U-Net [2].

I. The proposed Double Lightweight Unet architecture

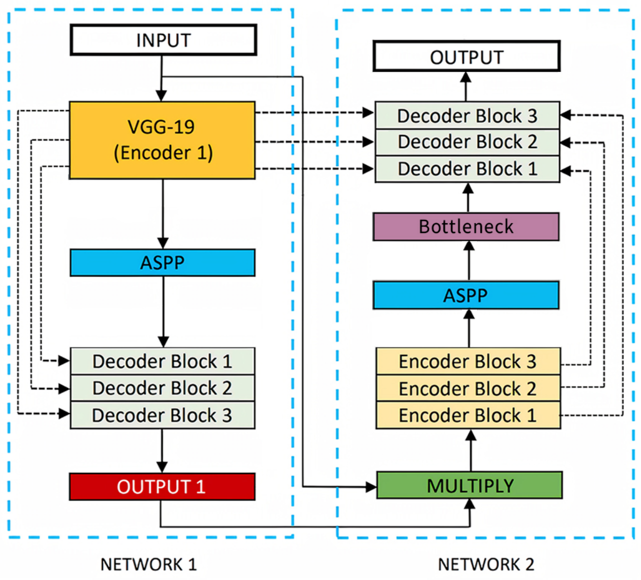


Fig. 7: The proposed Double Lightweight Unet (DL-Unet)

Lastly, with this paper, we propose a variant model that we've called *Double Lightweight Unet (DL-Unet)*. Although DB-Unet achieved remarkable results, in our particular case, we realized that as the epochs were increasing during training, and accordingly also the improvement of the network's learning, the contribution of the second network gradually became less useful in order to improve the segmentation of the first. Probably because the first network was getting better and better at segmenting correctly, leaving little room for improvements for the second part. So we thought of reducing the overall complexity of the architecture by cutting out one convolutional block from both VGG19 and the corresponding decoder and go for L-Unet as second network. With our solution (8) we achieved remarkable results compared to the two architectures mentioned above and at the same time reduced the complexity, trainable parameters and training time of DB-net.

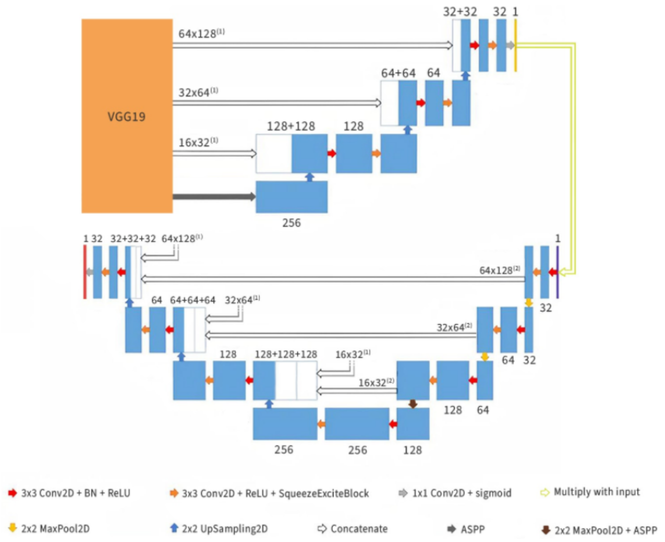


Fig. 8: A focus on DL-Unet layers

J. Experiment Environment

We ran our experiments on GPU offered by Google Colab PRO, a Jupyter notebook environment that works entirely in the cloud. It doesn't require setup, and the notebooks created can be edited simultaneously by multiple members. All models are implemented using Keras framework with Tensorflow 2.5.0 as backend. The implementation can be found at our GitHub repository[5].

IV. Results

In this section we are going to show the results and compare them. We compare the models by using the same experiment setup and configuration as described above (Section III-D). We also report some samples where Double Lightweight Unet predicted the median nerve correctly in comparison with the other two architectures, in order to prove the benefit of DL-Unet. For each model, the average values of the metrics and the related graphs are reported.

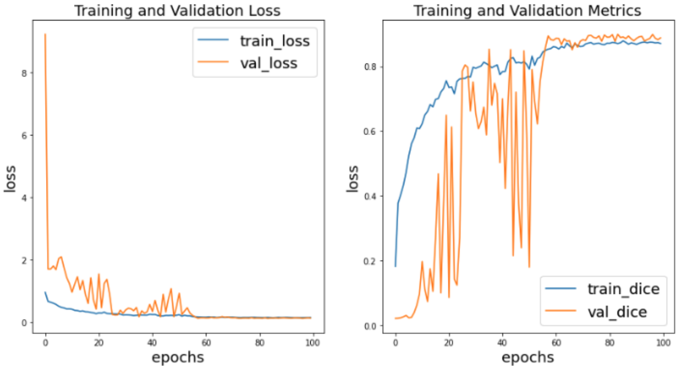


Fig. 9: Unet training loss and dice score

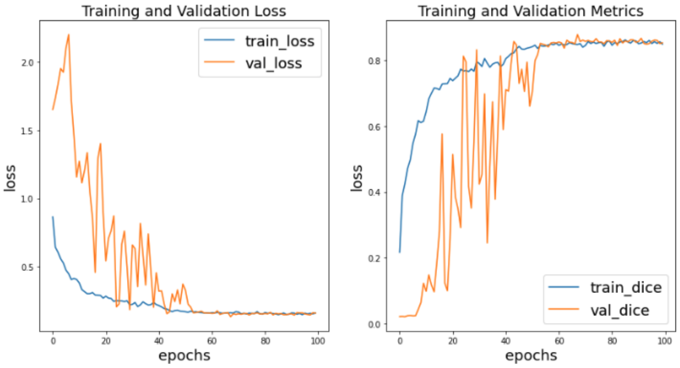


Fig. 10: L-Unet training loss and dice score

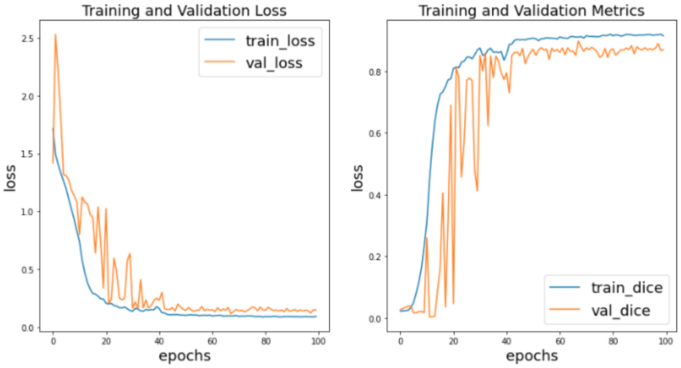


Fig. 11: DB-Unet training loss and dice score

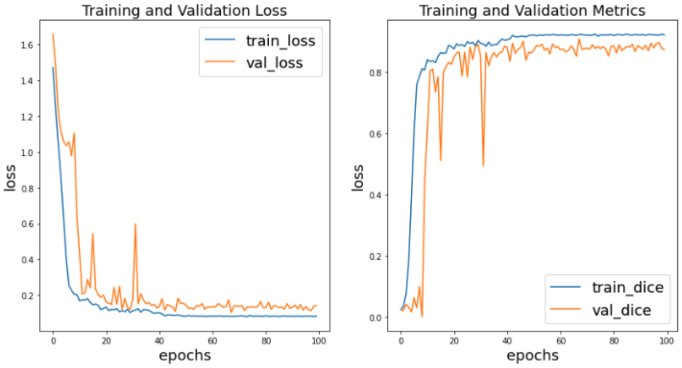


Fig. 12: DL-Unet training loss and dice score

Let's start with UNet which scores a DSC of 85% and an IoU of 73.7%. Precision and Recall amount respectively 85.43% and 80.39%. We define these results as baseline for comparison with the next architectures. Considering the dataset's size we can confirm that UNet is an excellent neural network model for the segmentation of biomedical images, achieving a respectable average DSC (Figure 9).

Lightweight Unet approaches the previous results reaching 84.09% in terms of DSC score and IoU of 73.40%, which are respectively 0.91% and 0.30% lower than the UNet's results. The Precision amount at 84.60% and 83.44% of the Recall (Figure 10).

[5] https://github.com/SasageyoOrg/cvdl-deep-mn

DB-Unet outperforms the current best result by 3.37% in terms of DSC and 5.46% in IoU. The Precision is also significantly higher than Unet, reaching a value of 89.71%. The Recall is similar to Unet but a bit lower, 85.54% (Figure 11).

Lastly DL-Unet brings further improvements over Double Unet both in terms of DSC and IoU. In fact it crosses the 90% threshold in the DSC reaching 90.27% and marks a IoU of 81.45%. The Precision is even higher than DB-Unet, with a value of 94.60%. The Recall is similar to the others. These results highlight the effectiveness of the proposed model which offers more precise predicted masks(Figure 12). The results are summarized in Table II.

Furthermore, pre-trained networks are compared in order to point out the backbone that leads to better results:

	VGG16	VGG19
DSC	0.8953 ± 0.047	**0.9027 ± 0.060**
IoU	0.7592	**0.8145**
Precision	**0.8598**	0.8499
Recall	0.8821	**0.9460**
Parameters	12.411.178	14.774.698

TABLE III: Backbones DLUnet's metrics Comparison

As anticipated we also want to report some critical examples where DL-Unet predicts the median nerve correctly and both Unet and L-Unet fail instead.

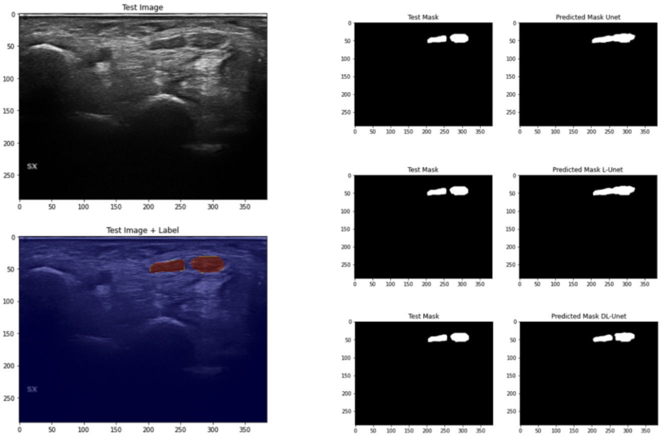


Fig. 13: Median nerve split into two regions

The Figure 13 shows the case where the ground truth of the segmentation area is divided into two parts and DL-Unet is the only neural network able to predict the median nerve correctly. Unet and L-Unet predict a single mask with a size comparable to the two true ones joined together. The reason is due to the coupling of two cascaded networks, in fact network 2 in DL-Unet succeeds in separating the segmented areas properly (Fig. 14).

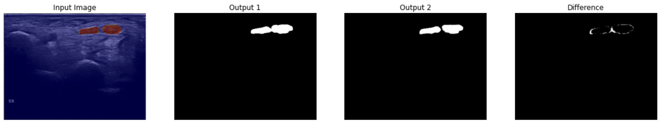


Fig. 14: Network 2 successfully differentiates the nerves

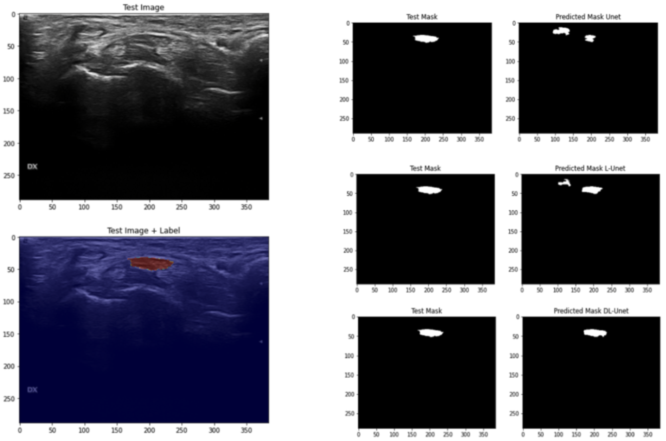


Fig. 15: Exclusion of false positive (FP) segmented areas

DL-Unet is also better at excluding outlying areas (false positives) as shown in Figure 15. The reasons should be the same as those assumed above.

V. Discussion

With the aim of showing the inner behaviour of the Double Lightweight Unet we want to focus on the Multiply Layer and on the improvements of Output 2 over Output 1 by comparing them.

The Multiply Layer is represented with a green block shown in Figure 8 which multiplies the input image with the output of the first network (Output 1). It's interesting to investigate it because it is the input of Network 2 and if it's misleading it

	Unet	L-Unet	DB-Unet	DL-Unet
DSC	0.8500 ± 0.083	0.8409 ± 0.096	0.8837 ± 0.082	**0.9027 ± 0.060**
IoU	0.7370	0.7340	0.7916	**0.8145**
Recall	0.8543	0.8344	**0.8554**	0.8499
Precision	0.8039	0.8460	0.8971	**0.9460**
Parameters	7.858.405	1.952.485	29.296.994	14.774.698

TABLE II: Models metrics comparison

would lead to poor final results (a.k.a. ”Garbage in, Garbage out”).
To visualize the layer a custom callback has been implemented which is called at the end of each epoch during training. The method consists of searching the layer under consideration across all the others and once identified collect all the previous up to it assembling a new partial model. This new smaller model will be used to display only the Multiply Layer and the Output 1 at the end of every epoch. The Figure 16 shows how the Output 1 gets better and better as the epochs increase, the same for the multiply layer which will be fed to Network 2. So its purpose is to provide the input image with an increasingly highlighted ROI to Network 2 for further refinements of the mask. The images shown (Fig. 16) were taken from epochs 5, 15, and 40 respectively.

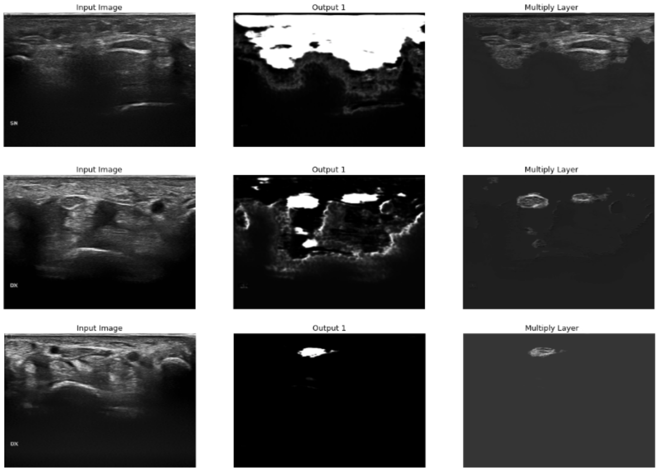


Fig. 16: Input, Output 1 and Multiply Layer

We used a similar method as the one described above to show the improvements of Output 2 over Output 1. To meet this need, it was necessary to temporarily modify the network’s output by replacing it with the concatenation of Output 1 and Output 2 as in the original architecture. Then we implemented another custom callback to display the two outputs separately from epoch to epoch. Examples are given below with the column ”Difference” (Fig. 17) showing the marked dissimilarity between the outputs. As anticipated our proposed model plays a decisive role at excluding outlying regions and filling the gaps to get closer to the ground truth.

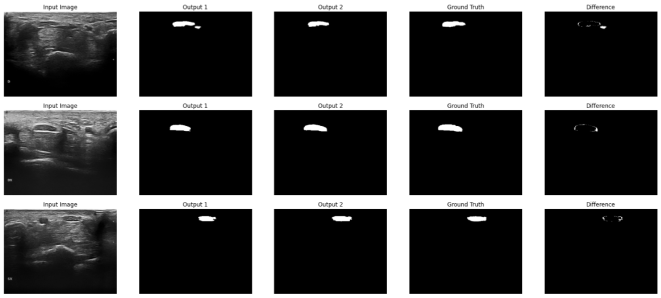


Fig. 17: Difference between Output 1 and Output 2

	Output1	Output2	Difference
Sample #1	0.8900	0.9273	**+3.73%**
Sample #2	0.8783	0.9438	**+6.55%**
Sample #3	0.9063	0.9280	**+2.17%**

TABLE IV: DL-Unet Outputs comparison

VI. Conclusion

In this paper, we have proposed a variant model called Double Lightweight U-Net simplifying the Double Unet’s architecture while maintaining its ability to predict accurate segmentation masks. Testing the architectures in our dataset, Double Lightweight Unet’s performances turned out to be better than DB-Unet’s and it improves the results of the tough examples most of all. In fact even from a visual inspection, we can see that DL-Unet is capable of producing better segmentation masks even for the most challenging images.

From the above experiments, we observed that the transfer learning from a pre-trained ImageNet network (eg. VGG19) significantly improves the results, that’s probably due to the fact to compensate for the lack of enough training data [2].

The limitation of the DB-Unet is that it counts a lot of parameters which means long training time. Our goal was to focus more on designing a compressed architecture with fewer parameters while maintaining its performance. The proposed model turned out to be both slimmer and more performing when applied to our dataset. Therefore we consider the goal as achieved.

References

[1] Ming-Huwi Horng, Cheng-Wei Yang, Yung-Nien Sun, and Tai-Hua Yang. Deepnerve: A new convolutional neural network for the localization and segmentation of the median nerve in ultrasound image sequences. *Ultrasound in Medicine & Biology*, 46(9):2439–2452, 2020.

[2] Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 558–564, 2020.

[3] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297, 2020.

[4] Oussama Hadjerci, Adel Hafiane, Donatello Conte, Pascal Makris, Pierre Vieyres, and Alain Delbos. Ultrasound median nerve localization by classification based on despeckle filtering and feature selection. 09 2015.

[5] J. J. Giraldo, M. A. Álvarez, and Á. A. Orozco. Annu Int Conf IEEE Eng Med Biol SocPeripheral nerve segmentation using Nonparametric Bayesian Hierarchical Clustering. *Annu Int Conf IEEE Eng Med Biol Soc*, 2015:3101–3104, 2015.

[6] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. Deep learning in medical ultrasound analysis: A review. *Engineering*, 5(2):261–275, 2019.

[7] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological Physics and Technology*, 10(3):257–273, Sep 2017.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015.