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 a compression of the median nerve, 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 presents many 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. With the proposed model, called DL-Unet, we achieved high performance over the test set with a Dice measurement, Intersection over Union, Precision and Recall equal to 0.9027±0.060, 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. CTS usually occurs only in adults. Workplace factors may contribute to existing pressure on or damage to the median nerve. One of the most used methods to diagnosis CTS is using ultrasound imaging, this thanks to convenience, lower cost, non-invasiveness and shorter examination times. 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]. 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.

The goal of this paper is to segment the median nerve from ultrasound images. Segmentation is useful to automatically isolate the median nerve from ultrasound scans in order to perform further analysis, such as the identification of median nerve's compression or swelling.

II. Related Work

Deep learning (DL) has recently emerged as the leading machine learning tool in various research fields, and especially in general imaging analysis and computer vision. DL also shows huge potential for various automatic US image segmentation tasks [2] 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].

Most of the modern DL approaches are based on the Unet [4], a Fully Convolutional Neural Network (FCNN).

Ming-Huwi Horng et al. [1] proposed a CNN 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. The proposed model is designed to use frame sequences as training data, therefore it's not suitable to our problem. From this work we considered only the Unet and Lightweight Unet, base structure of DeepNerve.

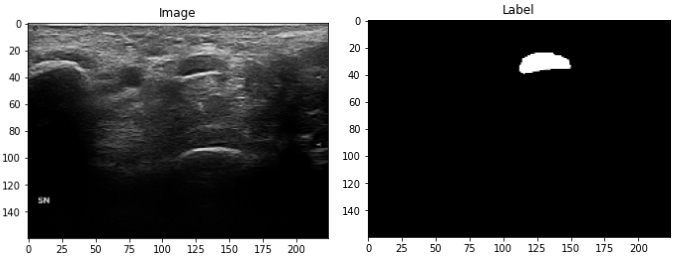
A contribution for this work is given by Debesh Jha et al. who proposed DoubleU-net, a combination of two Unet architectures stacked on top of each other where the first one uses a VGG-19 pre-trained on ImageNet as the encoder. DoubleU-net's encouraging results, produced on various medical image segmentation datasets, show that it can be used as a strong baseline for medical image segmentation to measure the generalizability of DL models [5]. This model was not tested on US images of the median nerve, so we wanted to find out how DoubleU-Net would perform on our dataset.

III. Materials & Methods

In this section, we illustrate the dataset, the data augmentation techniques, the experimental setup configuration, the hyperparameters of the experiments 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 1a), and 246 masks annotated by professionals who have marked the median nerve for each ultrasound image (E.g. Figure 1b).



(a) Ultrasound Image (b) Ground Truth Mask

B. Data Pre-Processing

In order to train and test the models, 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 into 3 sets: 164 are used for training, 38 as validation set and 44 are used for testing.

C. Data Augmentation

Medical datasets are often 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 [5]. To 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.

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 benefiting from the stability of BCE[1]. Performance of medical images segmentation is often evaluated by the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision and Recall. So we decided to use these as our evaluation metrics.

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

Hyperparameters	
Loss	BCE-DSC
Metrics	Dice, IoU, Recall, Precision
Batch Size	8
Learning Rate	$1e^{-3}$
Nr. Epoches	100

TABLE I: Recap. Hyperparameters

E. The Unet architecture

The starting point is *Unet* [4], 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).

F. The Lightweight Unet architecture

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.

G. The DoubleU-net architecture

This architecture is proposed in *"DoubleU-net: A deep convolutional neural network for medical image segmentation"* [5] by Debesh Jha et al. As mentioned above in Section 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. [5]

1) Network 2 Encoder: Each encoder block in the second Unet performs two 3×3 convolution operations, each followed by a batch normalization and a Rectified Linear Unit (ReLU). This is followed by a squeeze-and-excitation block and then max-pooling is performed with a 2×2 window and stride 2 to reduce the spatial dimension of the feature maps [5].

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

connections 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 [5].

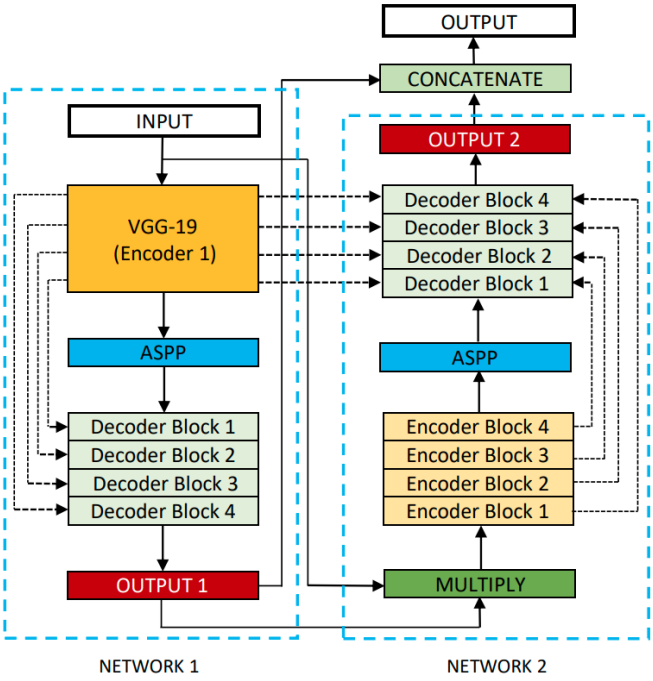


Fig. 2: The Double Unet architecture

H. The proposed Double Lightweight Unet architecture

We propose a novel model that we've called *Double Lightweight Unet (DL-Unet)* [2], which is a DB-Unet's variant. Although DB-Unet achieved remarkable results, in our particular case, we realized that as the epochs were increasing during training, 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 we achieved remarkable results compared to the two architectures mentioned above and at the same time reduced the complexity, the number of trainable parameters and training time.

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

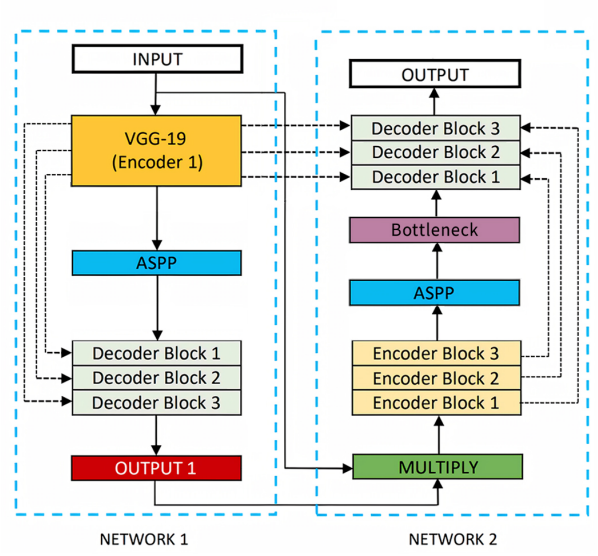


Fig. 3: The Double Lightweight Unet architecture

IV. Results

In this section we compare the models by using the same experiment setup and configuration as described above (Subsection 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 implemented, the mean values of DSC, IoU, Precision and Recall and reported in the Table II. Mean DSC is the average of all Dice coefficients calculated from the test set's images. IoU, Recall and Precision are obtained with the Keras evaluation method. In Table III Pre-trained networks are also compared in order to point out the backbone that leads to better results.

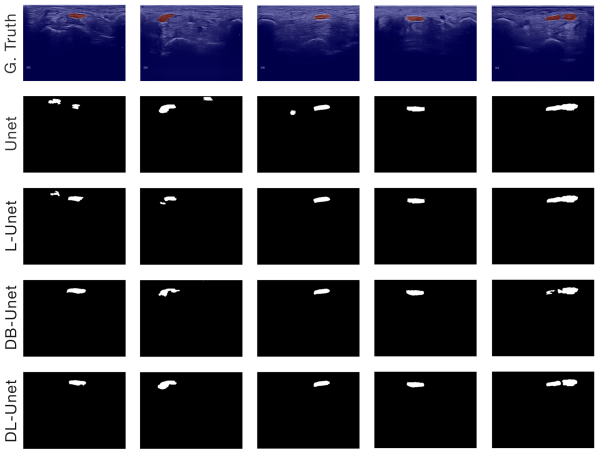


Fig. 4: Qualitative results

The Figure 4 shows different cases and the respective architectures segmentations. 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

	Unet	L-Unet	DB-Unet	DL-Unet
mDSC	0.8500 ± 0.083	0.8409 ± 0.096	0.8837 ± 0.082	$\mathbf{0.9027 \pm 0.060}$
mIoU	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

correctly. The reason is due to the coupling of two cascaded networks, in fact we checked that DL-Unet's Network 2 succeeds in separating the segmented areas properly.

	VGG16	VGG19
mDSC	0.8953 ± 0.047	$\mathbf{0.9027 \pm 0.060}$
mIoU	0.7592	**0.8145**
Precision	**0.8598**	0.8499
Recall	0.8821	**0.9460**
Parameters	12.411.178	14.774.698

TABLE III: DLUnet's backbones metrics comparison

V. Discussion

With the aim of showing the inner behaviour of the Double Lightweight Unet we want to focus a bit on the Multiply Layer and on the improvements of Output over Network 1's Output. The Multiply Layer is represented with a green block shown in Figure 3 which multiplies the input image with the output of the first network (Output 1). It's interesting to investigate it because it's the input of Network 2 and if its output is misleading it would lead to poor final results ("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 layers 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.

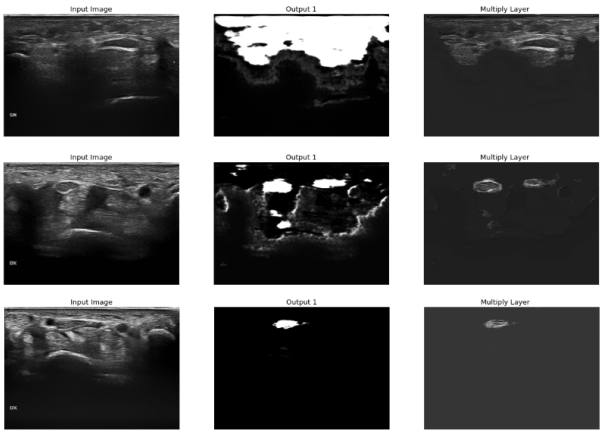


Fig. 5: Input, Output 1 and Multiply layer

The Figure 5 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 region of interest to Network 2 for further enhancement of the mask. 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. As anticipated our proposed model plays a decisive role at excluding outlying regions and filling the gaps to get closer to the ground truth.

From the above experiments, we observed that the transfer learning from a pre-trained ImageNet network significantly improves the results, that's probably because it compensates for the lack of enough training data [5]. About the Double U-Net based architectures, Network 1 acts as a pilot "denoiser" in order to provide a better input for Network 2 that will output an enhanced prediction, getting even closer to the ground truth. During the architectures testing with our dataset, DL-Unet's performances turned out to be better than DB-Unet's. 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.

VI. Conclusion

In this paper, we have proposed a CNN variant for the segmentation of the median nerve called Double Lightweight U-Net. It simplifies DB-Unet's architecture potentially reducing the training time while maintaining its ability to predict accurate segmentation masks.

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] 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.

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

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

[5] 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.