Segmentation of the VOCAL TRACT

Neuroengineering project (PW 2)
Group 3
2023-2024

Students:

Francesco Benedetto
Angelica Bombacigno
Livia Fabiano
Federico Pivetta
Letizia Santini

Professor:
Pietro Cerveri

Tutor:
Matteo Cavicchioli

TABLE OF CONTENTS

01

Context

Angelica

Clinical overview

MRI

Magnetic resonance is a medical imagine technique and, as it is non-invasive, it can be used to visualize human features to extract information

VOCAL TRACT

In this case, using segmentation, information from the vocal tract were extracted. The aim of is to perform some quantitative analysis

Angelica

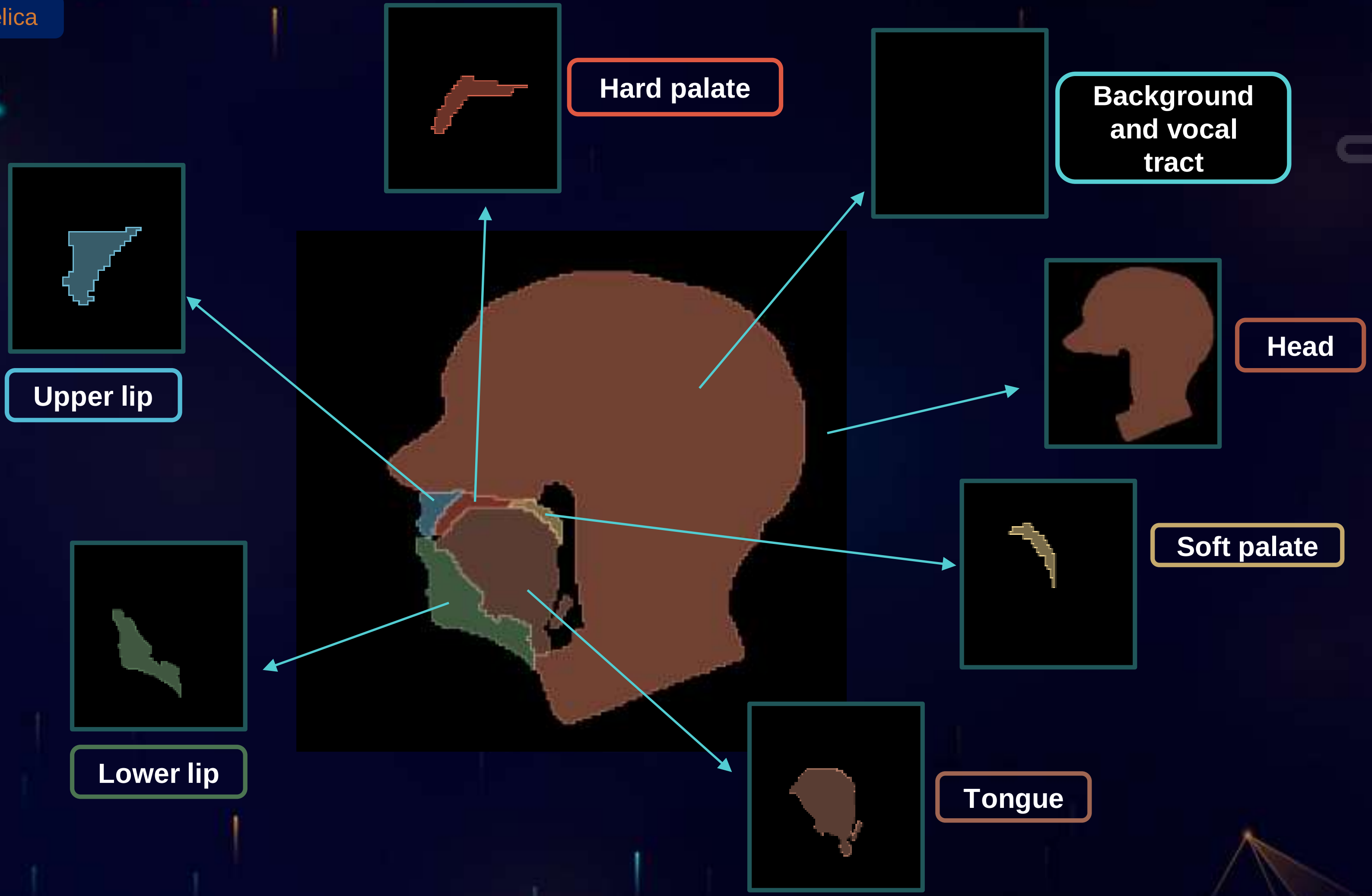

Hard palate
Background and vocal tract
Upper lip
Head
Soft palate
Lower lip
Tongue

TABLE OF CONTENTS

02

Data preparation

Dataset composition

S00001
280 images

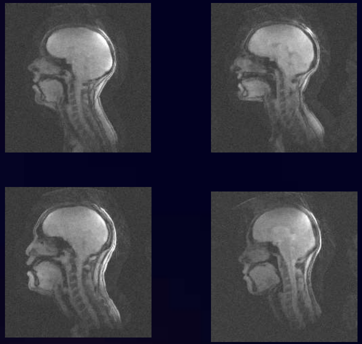
S00002
240 images

S00004
150 images

S00005
150 images

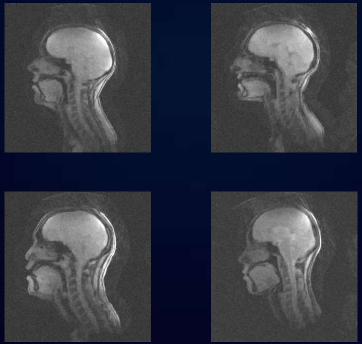
Dataset division

Training



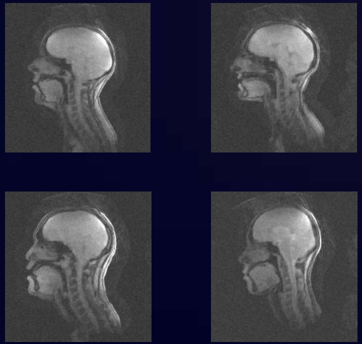
70%

Validation



20%

Test



10%

Pre-processing

Gamma transformation

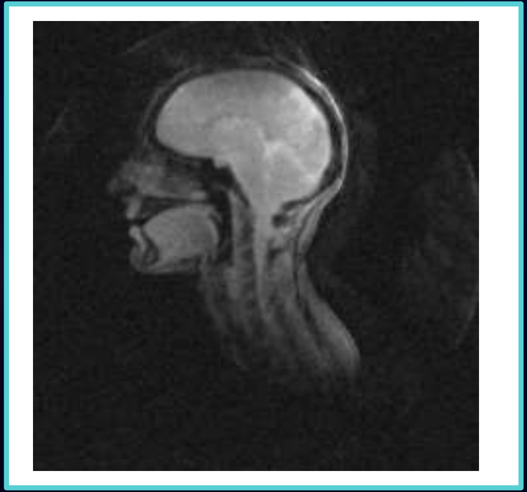
- $\gamma = 1.5$
- Amplify the grey levels on the darker part of the spectrum

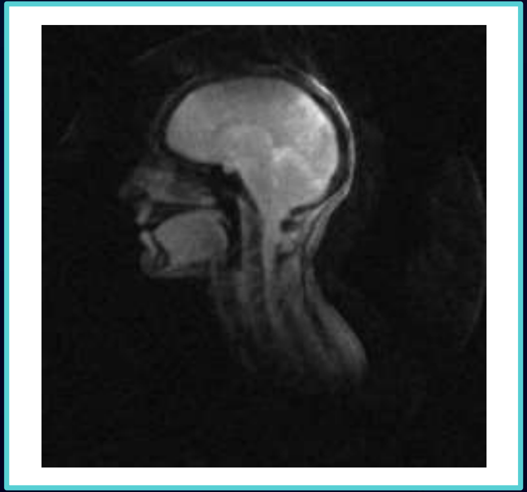
Gaussian filter

- $\sigma = 0.4$
- Eliminate noisy pixels

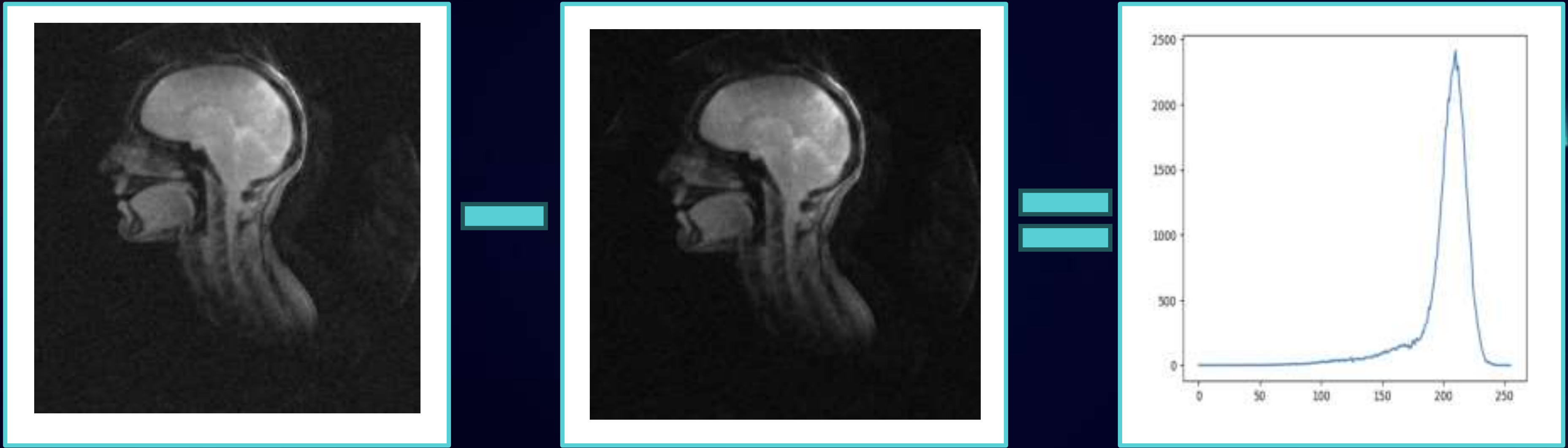
Saturation

- Saturation to black pixels that are below a predefined threshold
- Remove dots from the background

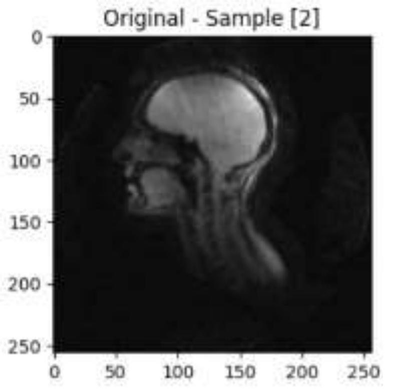


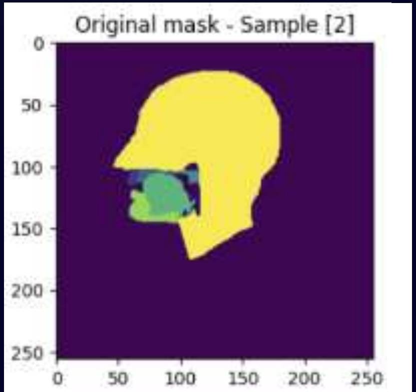


Pre-processing

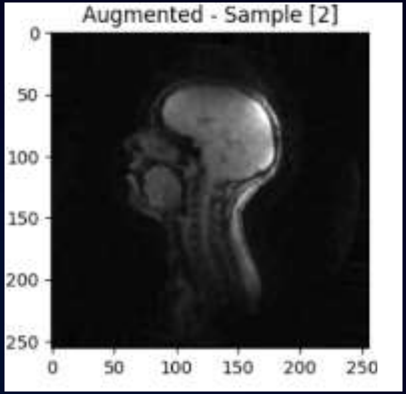


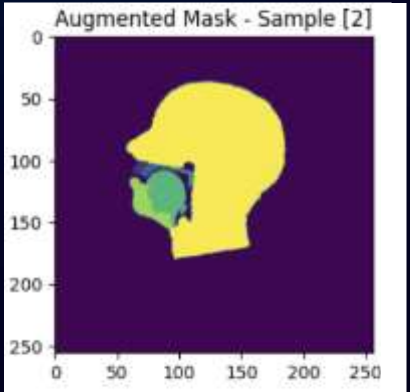
Data augmentation





10° rotation
5 side translation
5 height translation
0.1 zoom



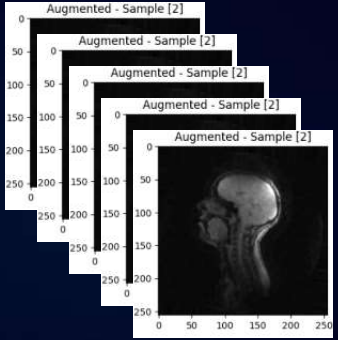


- Increase the overall amount of data
- Control data in case of rotation and translation of patients
- Increase the generalizing capability of the network
- Images concatenated to the training set

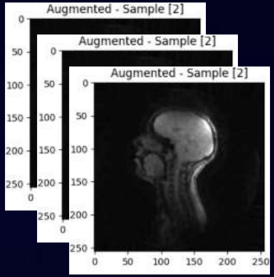
Final dataset for training

500 images

400 images



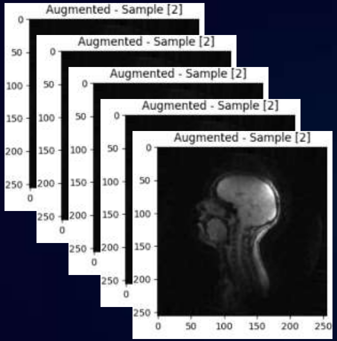
300 images



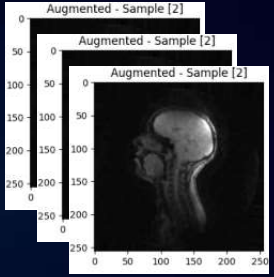
Final dataset for training

500 images

400 images

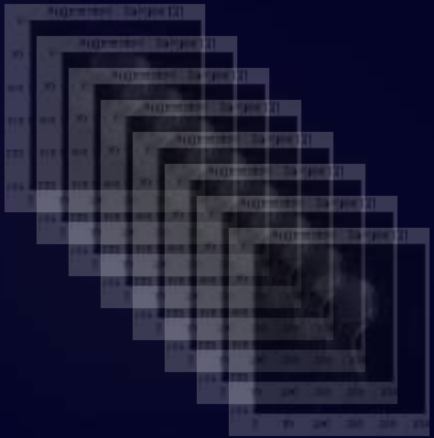


300 images

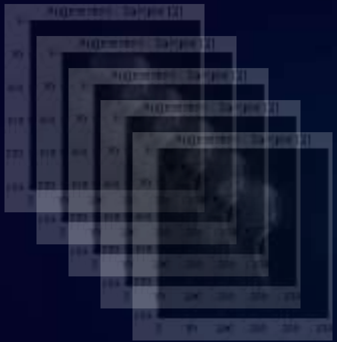


Final dataset for training

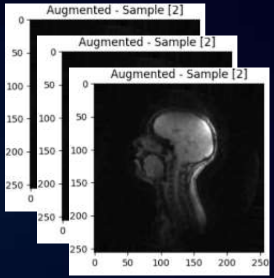
500 images



400 images



300 images



Training

Federico

TABLE OF CONTENTS

Federico

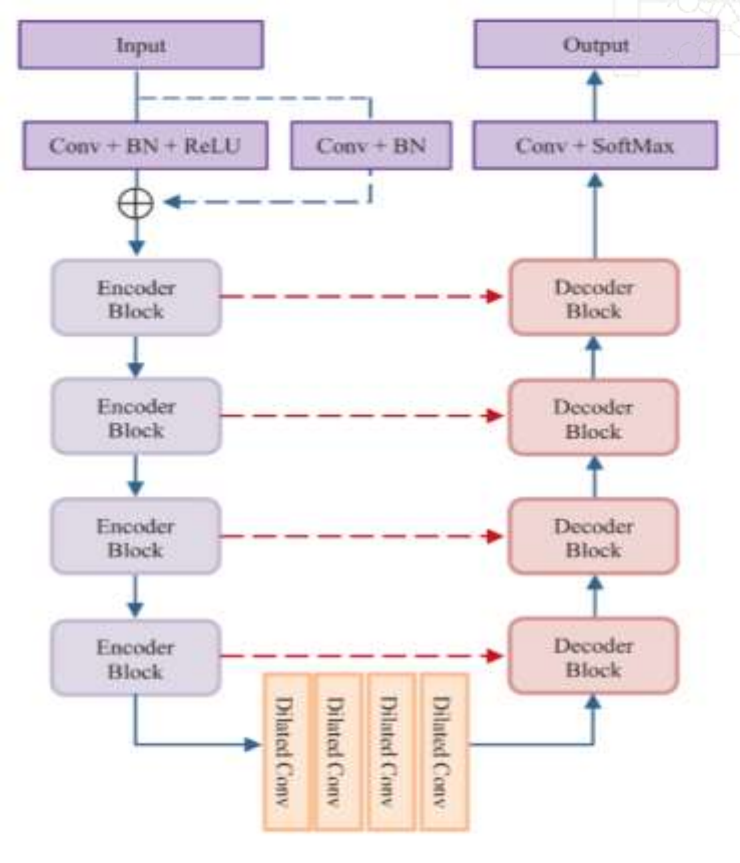
03

Network

IMproved U-Net [1]

- Improved:
  → residual connections
  → larger bottleneck section

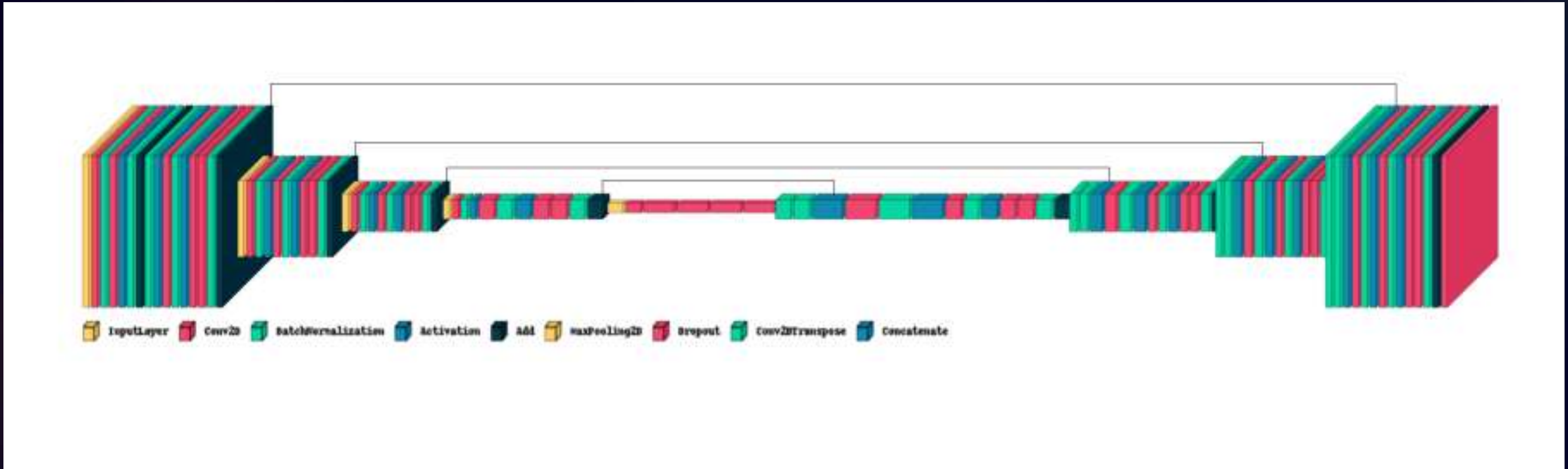
- Composed by:
  → encoder section
  → decoder section



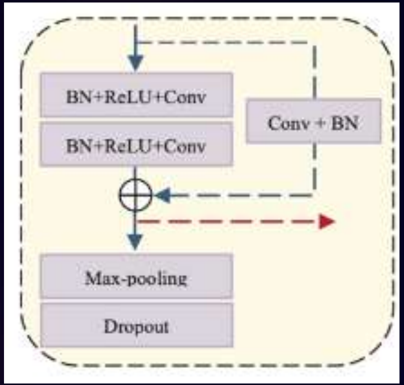
[1] https://doi.org/10.1016/j.cmpb.2021.105946

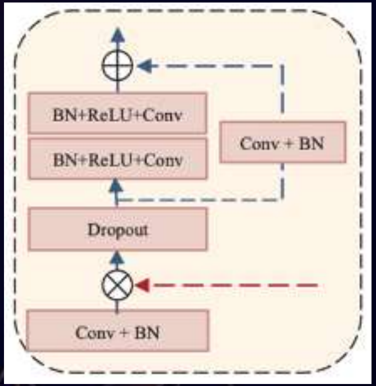
Federico

Network implemented in the project

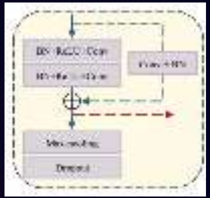


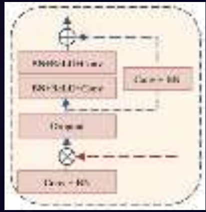
Network implemented

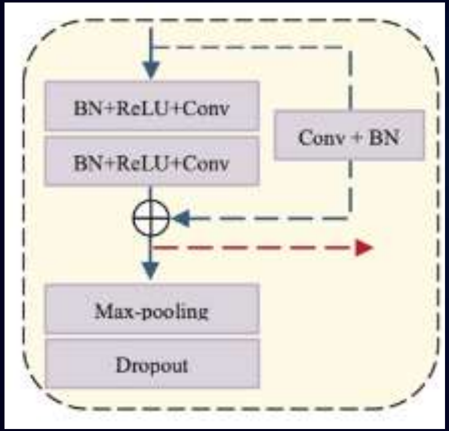




Network implemented





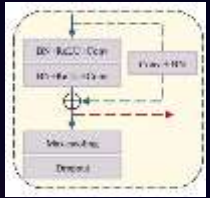


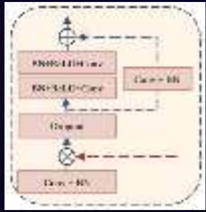
Encoding

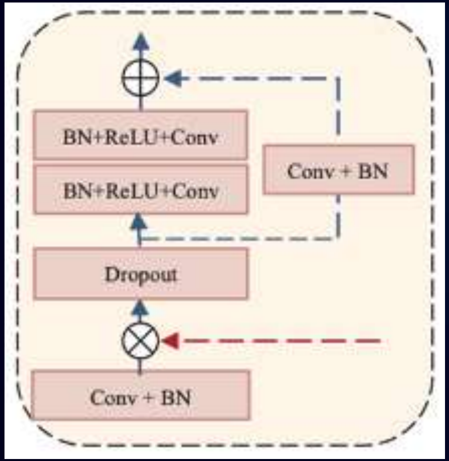
- Kernel_size=3
- Number of activation maps: double block by block
- MaxPool: pool_size=2, strides=2
- Dropout: rate=0.5

Federico

Network implemented







Decoding

- Transpose Conv: kernel_size=1, strides=2
- Kernel_size=3
- Number of activation maps: half block by block

Federico

Training methodology

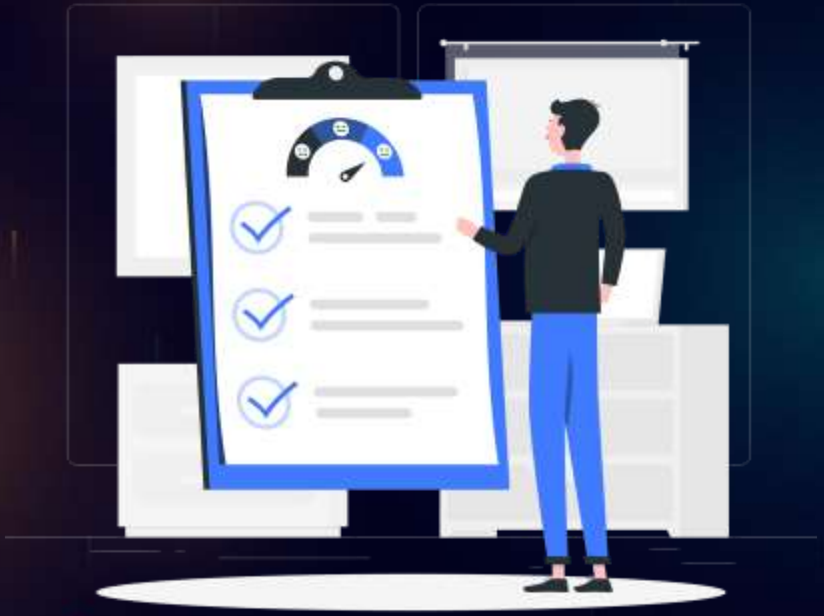
Weighted cross entropy	→	Loss function
Adam optimizer	→	Loss optimizer
DICE coefficient Accuracy Precision	→	Monitor the training
Recall Mean_IoU	→	Evaluation of final model

- batch_size = 8
- epochs = 50
- learning_rate = 0.001
- early_stopping_patience = 10
- LR_scheduler_patience = 4
- LR_scheduler_factor = 0.5

Callbacks:

- Early stopping
- LR scheduler

TABLE OF CONTENTS



04 Validation and Test

Choice of dataset

Dataset division

70% Training
20% Validation
10% Test

Cross-Validation

Combination of:
- 3 patients' datasets used for training
- 1 for validation

We performed trials using different approaches to the choice of the dataset

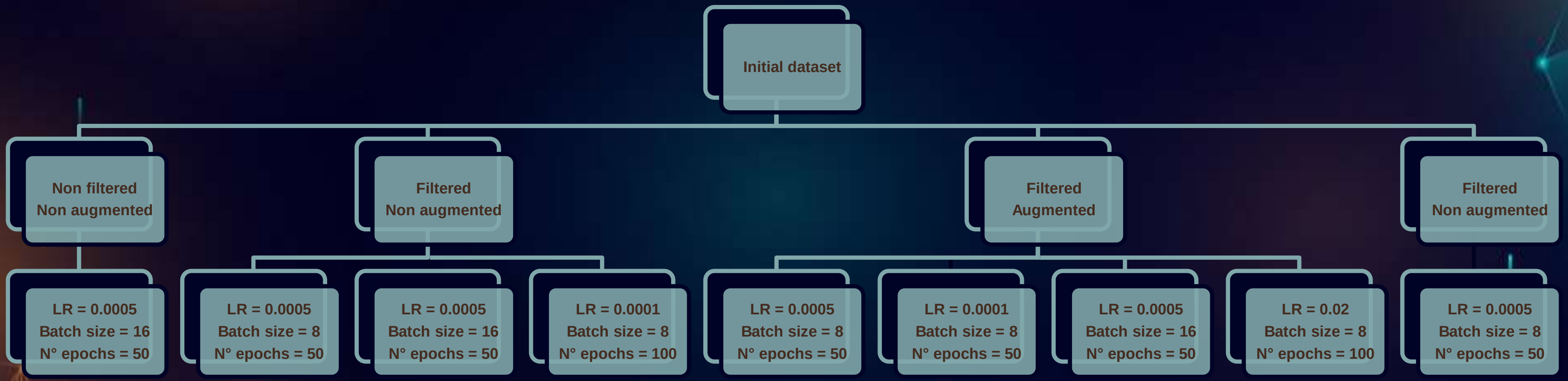
Choice of dataset

Dataset division

Cross-Validation

Livia

Combination of parameters



Results obtained by each model

	Training				Validation			
	Precision	DICE	Accuracy	Loss	Precision	DICE	Accuracy	Loss
1	0.9845	0.8468	0.9843	0.0497	0.9856	0.8482	0.9855	0.0737
2.1	0.9847	0.8576	0.9845	0.0463	0.9819	0.8533	0.9817	0.0930
2.2	0.9836	0.8394	0.9833	0.0511	0.9870	0.8522	0.9868	0.0769
2.3	0.9835	0.8453	0.9833	0.0497	0.9870	0.8648	0.9868	0.0830
3.1	0.9855	0.8839	0.9853	0.0401	0.9848	0.8637	0.9877	0.1123
3.2	0.9852	0.8340	0.9850	0.0441	0.9885	0.8787	0.9887	0.0988
3.3	0.9834	0.8241	0.9831	0.0508	0.9811	0.8173	0.9808	0.0927
3.4	0.9818	0.8318	0.9815	0.0579	0.9840	0.8155	0.9738	0.1207
4	0.9504	0.6505	0.9492	0.2250	0.9447	0.6578	0.9441	0.1339

Livia

Best model

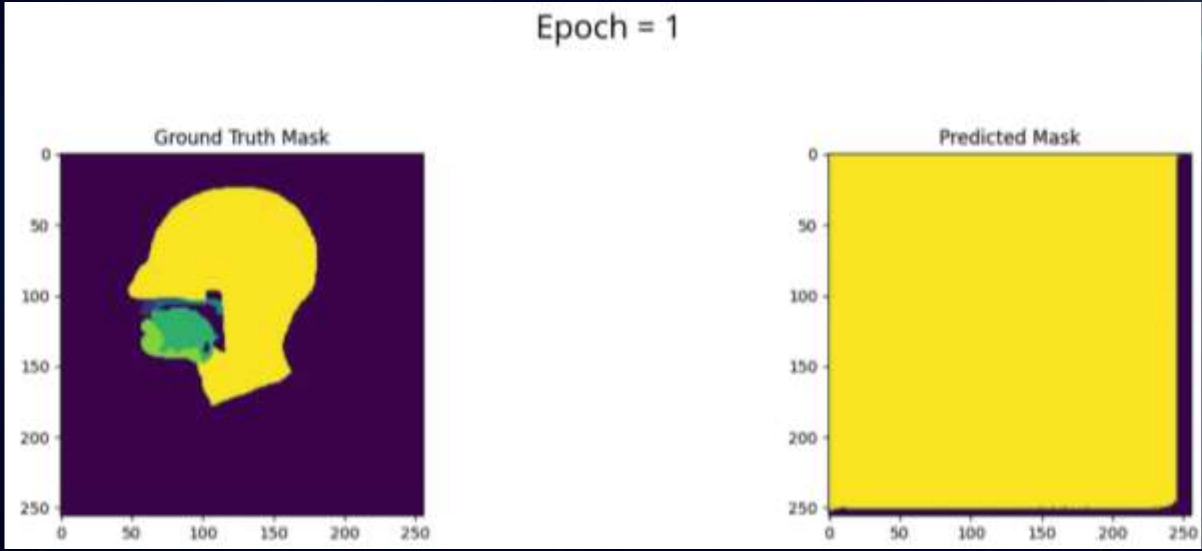
	Training				Validation			
	Precision	DICE	Accuracy	Loss	Precision	DICE	Accuracy	Loss
1	0.9845	0.8468	0.9843	0.0497	0.9856	0.8482	0.9855	0.0737
2.1	0.9847	0.8576	0.9845	0.0463	0.9819	0.8533	0.9817	0.0930
2.2	0.9836	0.8394	0.9833	0.0511	0.9870	0.8522	0.9868	0.0769
2.3	0.9835	0.8453	0.9833	0.0497	0.9870	0.8648	0.9868	0.0830
3.1	0.9855	0.8839	0.9853	0.0401	0.9848	0.8637	0.9877	0.1123
3.2	0.9852	0.8340	0.9850	0.0441	0.9885	0.8787	0.9887	0.0988
3.3	0.9834	0.8241	0.9831	0.0508	0.9811	0.8173	0.9808	0.0927
3.4	0.9818	0.8318	0.9815	0.0579	0.9840	0.8155	0.9738	0.1207
4	0.9504	0.6505	0.9492	0.2250	0.9447	0.6578	0.9441	0.1339

$$Validation\ sum = \frac{Precision_{validation} + DICE_{validation} + Accuracy_{validation}}{3}$$

$$Validation\ sum_{3.2} = \frac{0.9885 + 0.8787 + 0.9887}{3} = 0.952$$

Best model

	Training				Validation			
	Precision	DICE	Accuracy	Loss	Precision	DICE	Accuracy	Loss
1	0.9845	0.8468	0.9843	0.0497	0.9856	0.8482	0.9855	0.0737
2.1	0.9847	0.8576	0.9845	0.0463	0.9819	0.8533	0.9817	0.0930
2.2	0.9836	0.8394	0.9833	0.0511	0.9870	0.8522	0.9868	0.0769
2.3	0.9835	0.8453	0.9833	0.0497	0.9870	0.8648	0.9868	0.0830
3.1	0.9855	0.8839	0.9853	0.0401	0.9848	0.8637	0.9877	0.1123
3.2	0.9852	0.8340	0.9850	0.0441	0.9885	0.8787	0.9887	0.0988
3.3	0.9834	0.8241	0.9831	0.0508	0.9811	0.8173	0.9808	0.0927
3.4	0.9818	0.8318	0.9815	0.0579	0.9840	0.8155	0.9738	0.1207
4	0.9504	0.6505	0.9492	0.2250	0.9447	0.6578	0.9441	0.1339



Best model

Post-processing

Test-phase

Best model

Post-processing

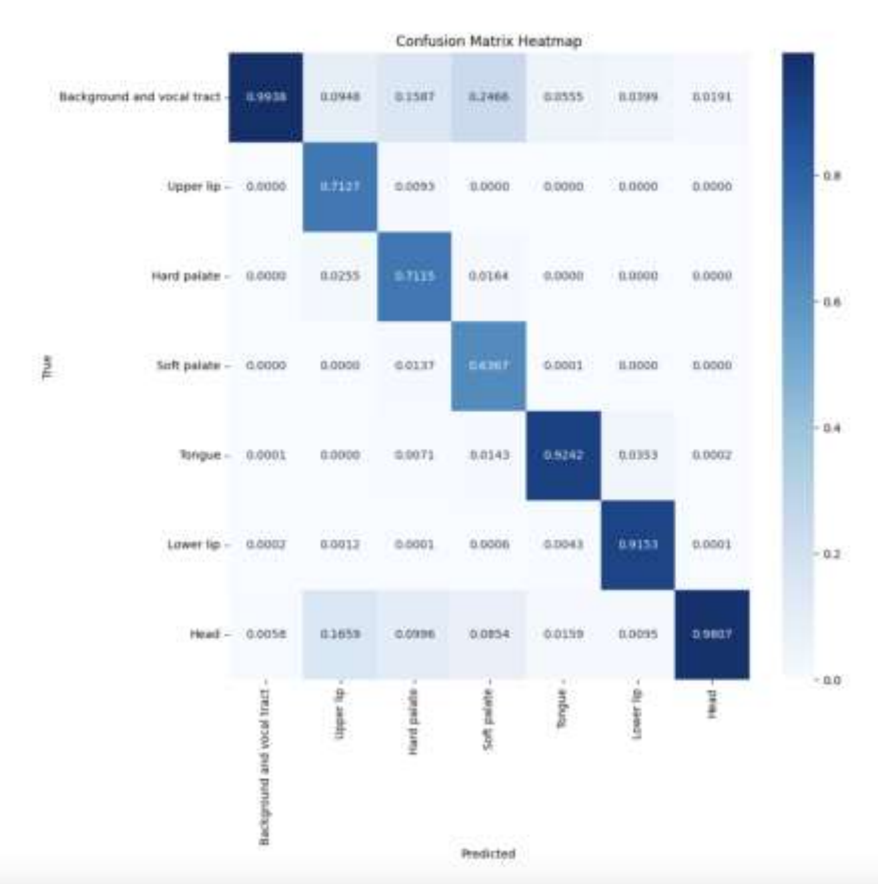
Argmax on the predicted images

Test-phase

- Precision, Recall, mean DICE in each class
- Confusion Matrix

Best model

Class	Precision	Recall	Mean DICE
0: background and vocal tract	0.9938	0.9926	0.9948
1: upper limb	0.7128	0.9770	0.8145
2: hard palate	0.7115	0.9554	0.7961
3: soft palate	0.6367	0.9610	0.7744
4: tongue	0.9242	0.9720	0.9585
5: lower lip	0.9153	0.9681	0.9532
6: head	0.9807	0.9670	0.9823



Francesco

Choice of dataset

Dataset division

70% Training
20% Validation
10% Test

Cross-Validation

Combination of:

- 3 patients' datasets used for training
- 1 for validation

We performed trials using different approaches to the choice of the dataset

Choice of dataset

Dataset division

Cross-Validation

Francesco

Cross - validation

S00001 S00002 S00004 S00005

Combination A

S00001 S00002 S00004 S00005

Combination B

S00001 S00002 S00004 S00005

Combination C

S00001 S00002 S00004 S00005

Combination D

S00001 S00002 S00004 S00005

Cross - validation

PURPOSE

The aim of this kind of analysis was to understand and give a numerical evaluation of the model's capability to generalize target features

- ✓ Prevent Overfitting
- ✓ Generalization capability on external dataset

Francesco

Cross - validation

	Comb. A	Comb. B	Comb. C	Comb. D	Final mean
DICE without argmax	0.865	0.9635	0.848	0.8132	0.8723
DICE with argmax	0.8744	0.8698	0.8549	0.8249	0.8560
Accuracy	0.9809	0.9794	0.9836	0.9690	0.9782
Precision	0.8152	0.8145	0.7863	0.7774	0.7983
Recall	0.9617	0.9492	0.9637	0.9263	0.9502
IoU	0.7467	0.8474	0.6727	0.7819	0.7620

Francesco

Cross - validation

[redacted]	Comb. A	Comb. B	Comb. C	Comb. D	Final mean
DICE without argmax	0.865	0.9635	0.848	0.8132	0.8723
DICE with argmax	0.8744	0.8698	0.8549	0.8249	0.8560
Accuracy	0.9809	0.9794	0.9836	0.9690	0.9782
Precision	0.8152	0.8145	0.7863	0.7774	0.7983
Recall	0.9617	0.9492	0.9637	0.9263	0.9502
IoU	0.7467	0.8474	0.6727	0.7819	0.7620

Final mean
0.8723
0.8560
0.9782
0.7983
0.9502
0.7620

$$MEAN_{Accuracy} = \frac{Accuracy_{Comb\ A} + Accuracy_{Comb\ B} + Accuracy_{Comb\ C} + Accuracy_{Comb\ D}}{4}$$

Cross - validation

Final mean
0.8723
0.8560
0.9782
0.7983
0.9502
0.7620

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

As the **recall** is higher than the **precision**

Under segmentation

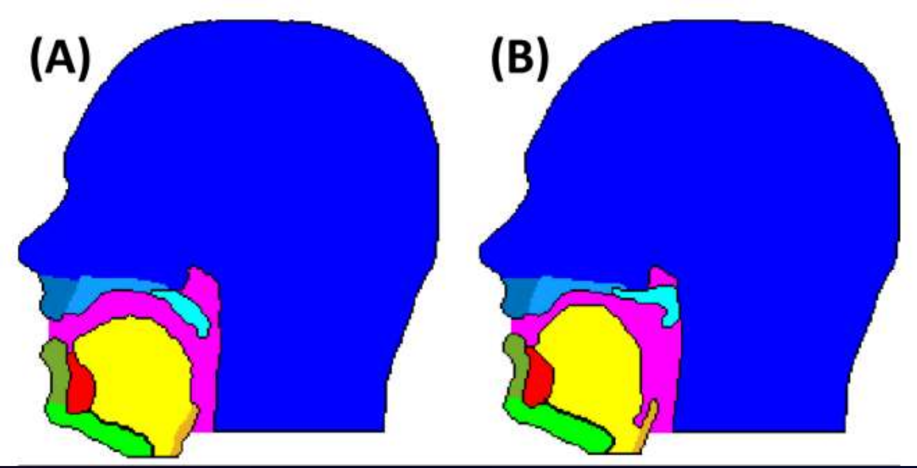
TABLE OF CONTENTS

Francesco

05

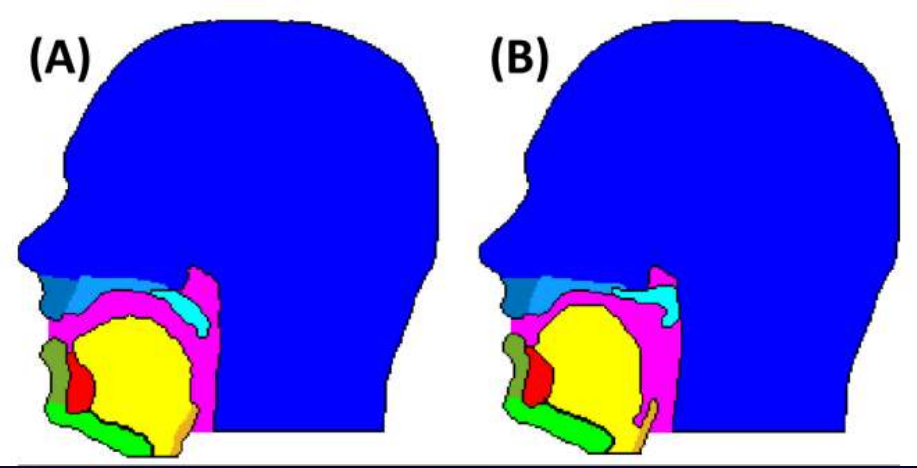
Conclusions

Clinical evaluation

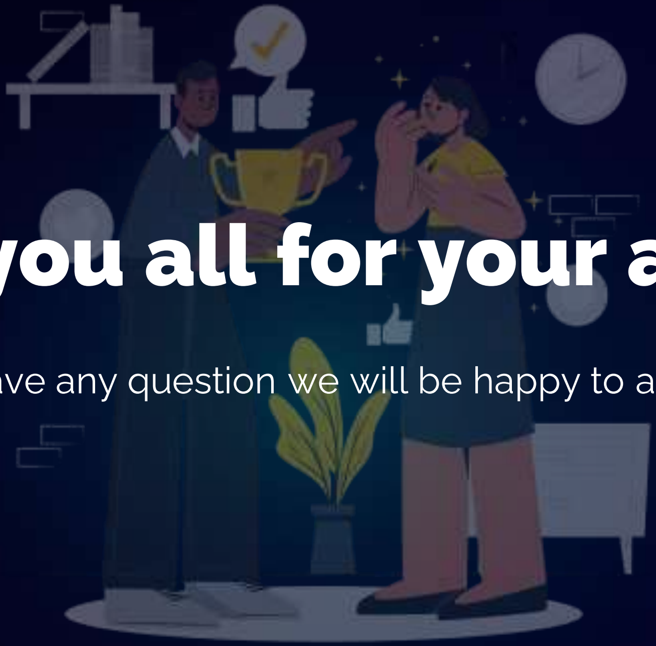


Many speech pathologies are caused by the absence of the soft palate closure

Clinical evaluation



Under Segmentation is preferable



Thank you all for your attention

If you have any question we will be happy to answer