UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Simone Milani
Room 216
DEI A

Phone: 049 827 7641
E-mail:
simone.milani@dei.unipd.it

BIOMETRICS

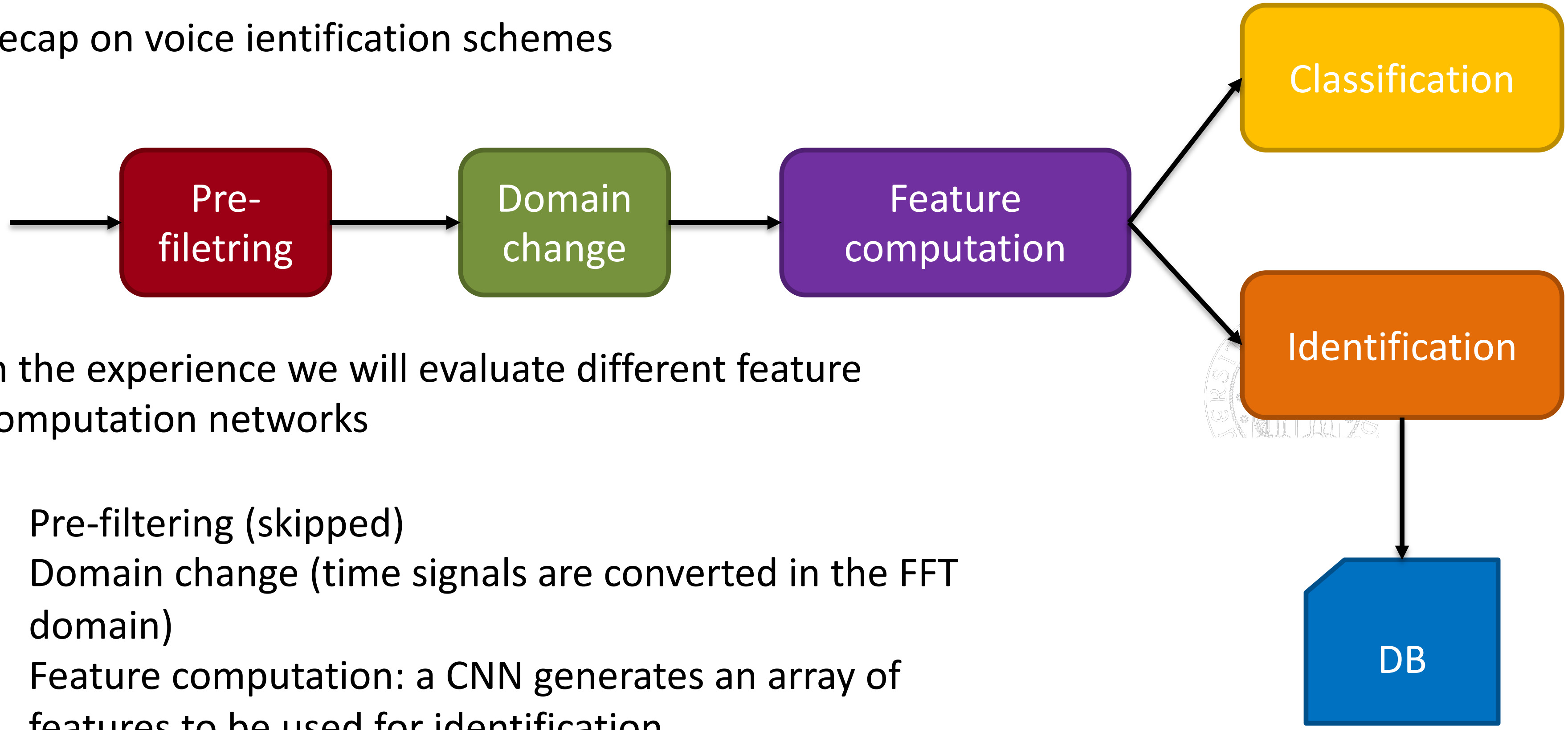
A.A. 2020/2021

LECTURE 23

LAB SESSION 2 – VOICE RECOGNITION

RECAP

Recap on voice ientification schemes



In the experience we will evaluate different feature computation networks

- Pre-filtering (skipped)
- Domain change (time signals are converted in the FFT domain)
- Feature computation: a CNN generates an array of features to be used for identification
- Classification: identifies the speaker in a pool of 5 persons

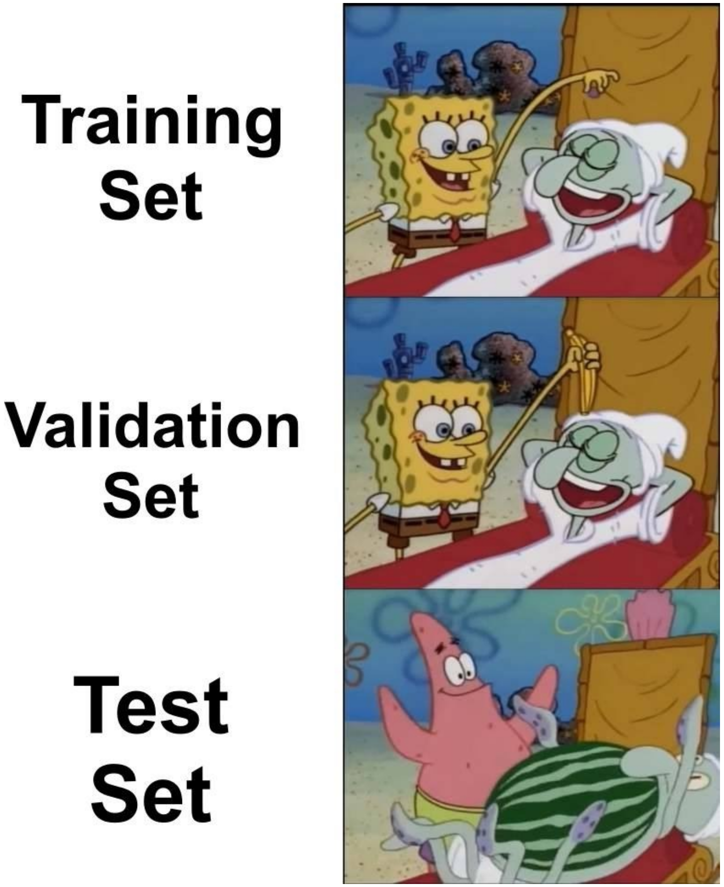
DATASET CREATION

Dataset for each speaker is divided into three sets:

- Training set: used to optimize parameters
- Validation set: used to monitor the training process
- Test set: used to verify the final accuracy

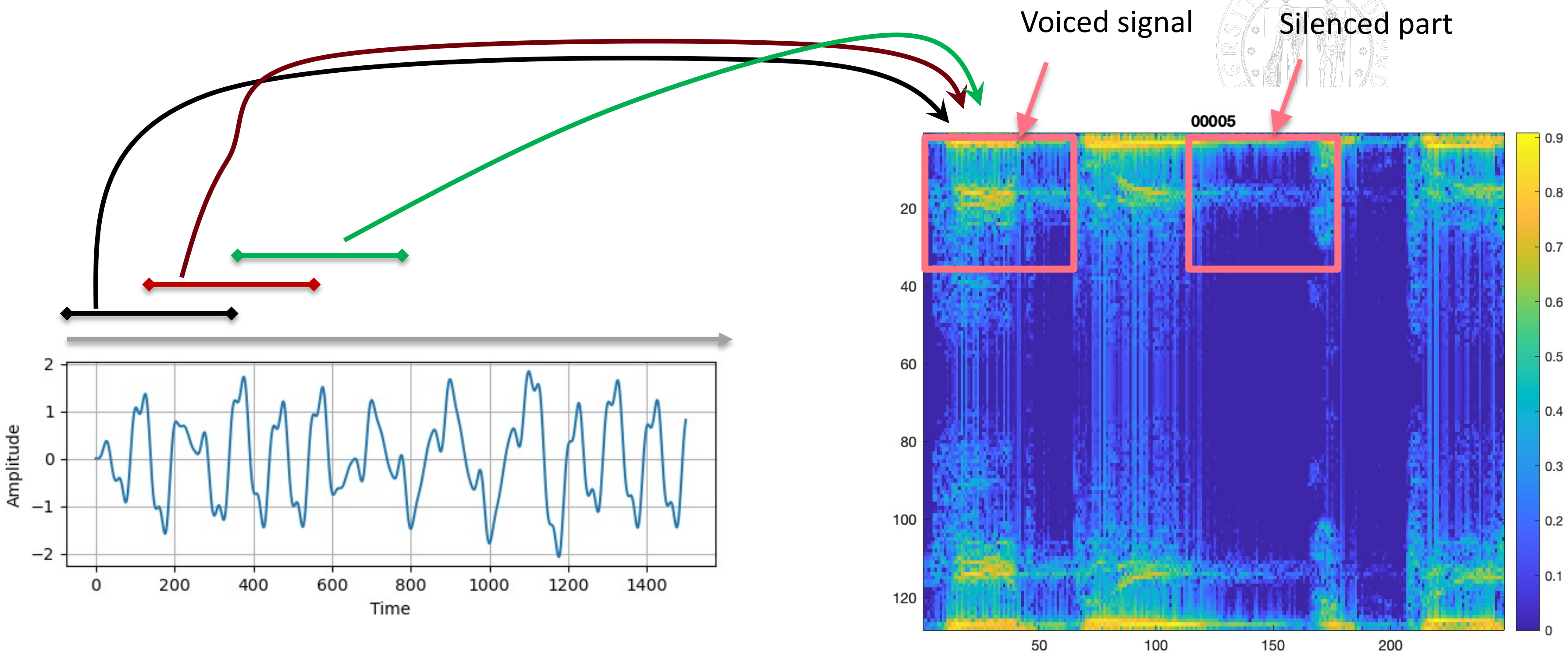
Each set includes different files.

With different cardinalities.



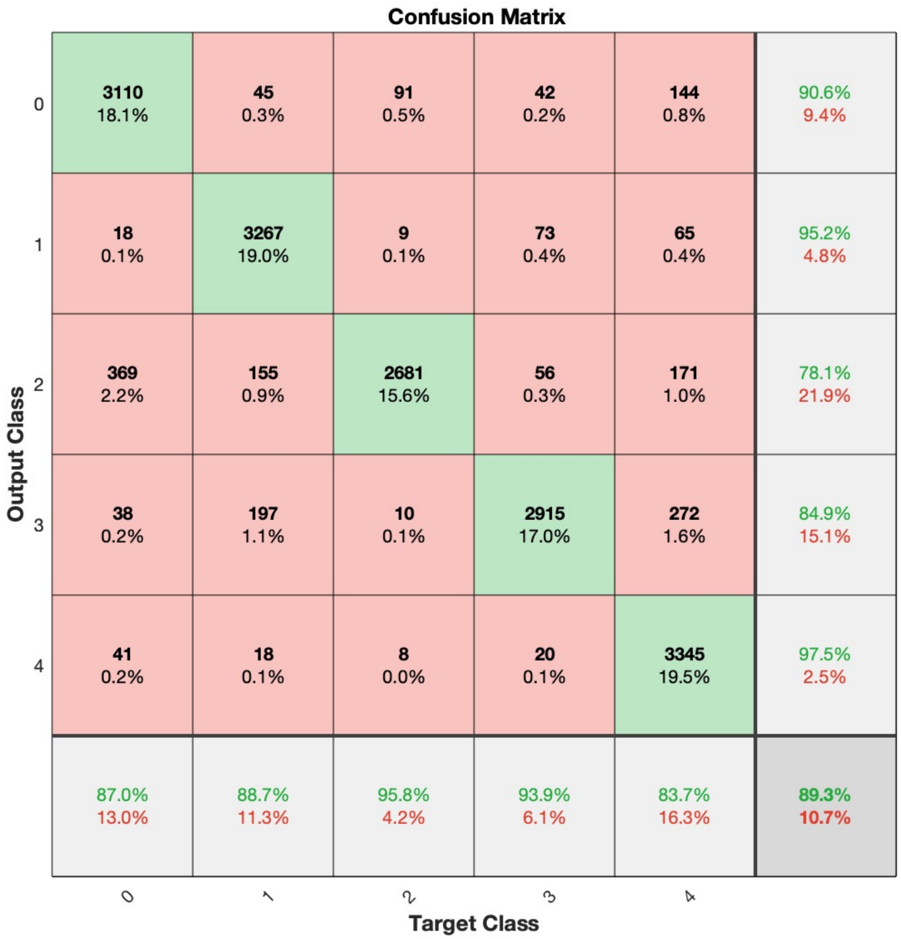
SPECROGRAM CREATION

- Spectrogram is created by windowing the input audio signal.
- Overlapping window are seletced
- FFT is computed on each window
- Results are included in the columns of the spectrogram image
- 32x32 relevant windows are selected (with voiced signal)

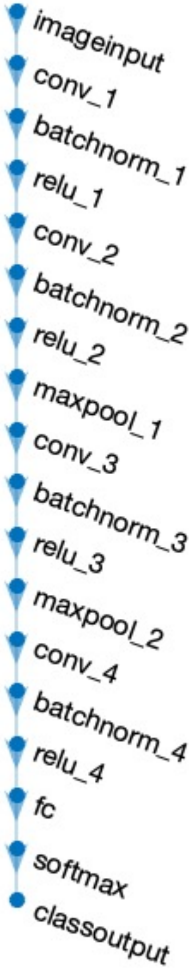


NETWORK CREATION CREATION

- Features and classification is performed using a CNN
- Progressive set of convolutional, batch normalization, activation layers
- Final layer is a fully connected
- Final accuracy visualized with confusion matrix


Confusion Matrix

	0	1	2	3	4	
0	3110 18.1%	45 0.3%	91 0.5%	42 0.2%	144 0.8%	90.6% 9.4%
1	18 0.1%	3267 19.0%	9 0.1%	73 0.4%	65 0.4%	95.2% 4.8%
2	369 2.2%	155 0.9%	2681 15.6%	56 0.3%	171 1.0%	78.1% 21.9%
3	38 0.2%	197 1.1%	10 0.1%	2915 17.0%	272 1.6%	84.9% 15.1%
4	41 0.2%	18 0.1%	8 0.0%	20 0.1%	3345 19.5%	97.5% 2.5%
	87.0% 13.0%	88.7% 11.3%	95.8% 4.2%	93.9% 6.1%	83.7% 16.3%	89.3% 10.7%



TO DO 1

Use Cepstral values instead of spectrum values.

What is the difference in spectrograms?

How does the performance change?

TO DO 2

Optimize the network in order to maximize the performance.

Which is the top accuracy you can get?