ICPR 2020

25th INTERNATIONAL CONFERENCE
ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Which are the factors affecting the performance of audio surveillance systems?

Greco A., Roberto A., Saggese A., Vento M.

Mivia Lab - University of Salerno (ITA)

.DIEM

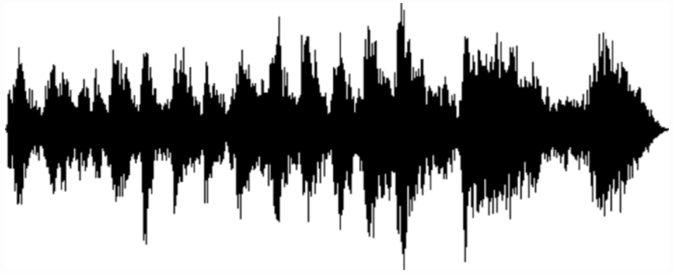
UNIVERSITÀ DEGLI STUDI
DI SALERNO

MIVIA

Outline

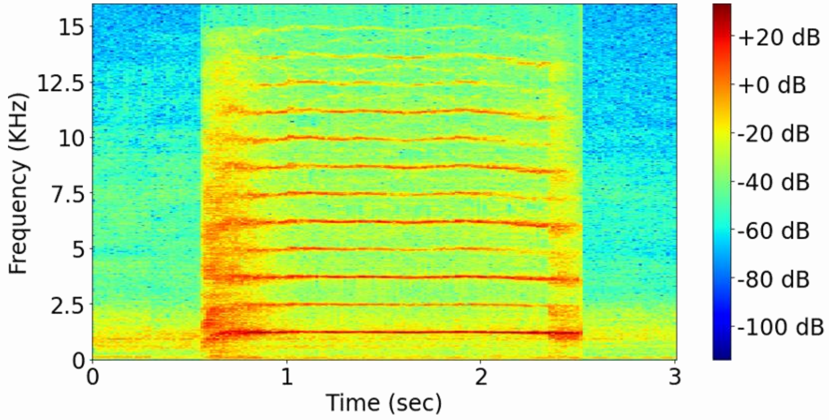
- Sound event detection as image classification
- Experimental setup
  - Design choices
  - Mivia Audio Events dataset
- Experimental results
- Useful insights
- Conclusions

Sound event detection as image classification



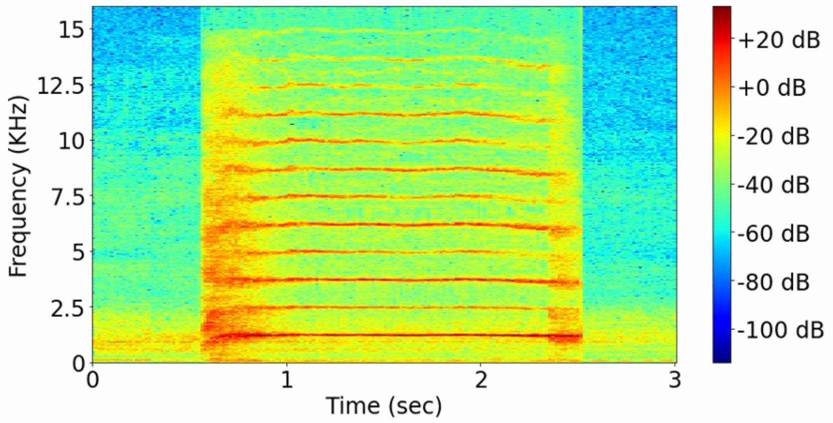
Audio Waveform

STFT



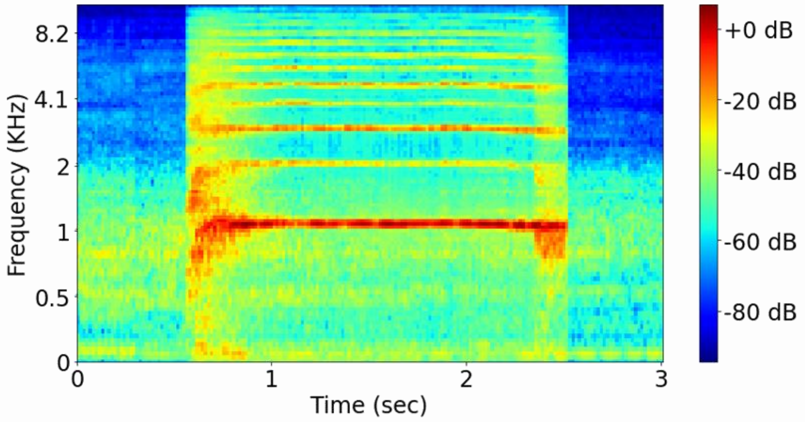
Spectrogram

Sound event detection as image classification



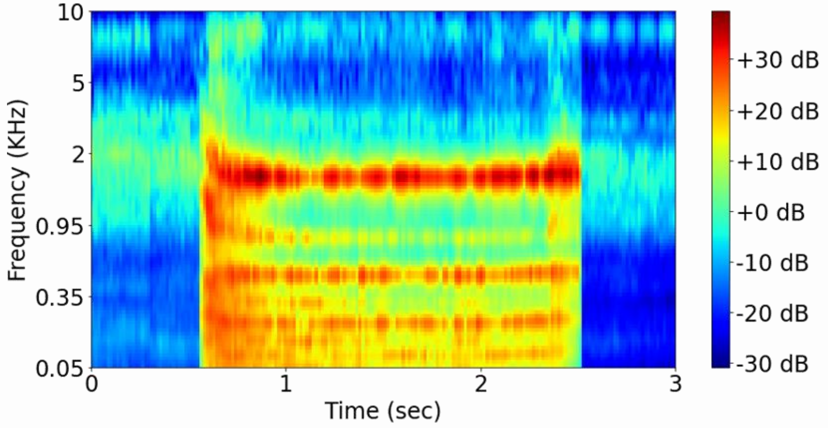
Spectrogram

Mel Filterbank



Mel-Spectrogram

Gammatone Filterbank



Gammatonegram

Design choices

Visual Representation

Spectrogram, Mel, Gammatone

Scaling range

Fixed, Dynamic

Convolutional Neural Network (CNN) architecture

MobileNet, DenseNet, ResNet, …

Weights initialization

Random, Imagenet

The dataset

MIVIA Audio Events public dataset

- Widely adopted by the scientific community for benchmarking purposes

Events of interest

- Glass breakings
- Gunshots
- Screams

The dataset

	TRAINING SET		TEST SET	
	# Events	Duration (s)	# Events	Duration (s)
Background	-	58371.6	-	25036.8
Glass Breaking	4200	6024.8	1800	2561.7
Gunshot	4200	1883.6	1800	743.5
Scream	4200	5488.8	1800	2445.4

Performance indices

- Precision
- Recall
- False Positive Rate
- F-beta score

$$\beta = 0.5$$

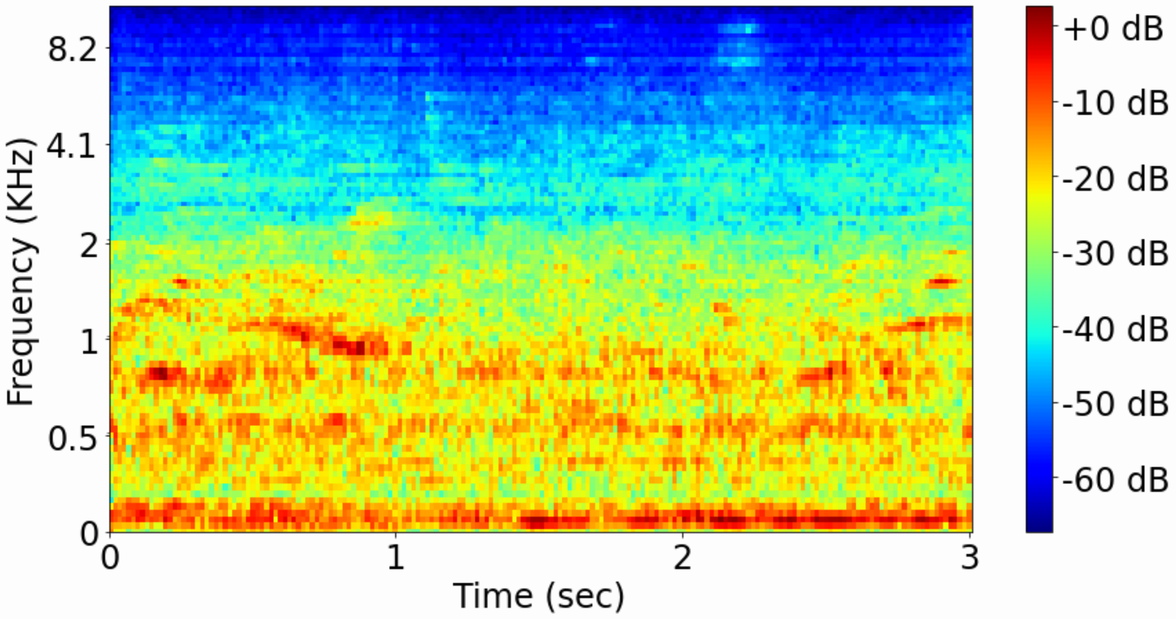
$$F_\beta = (1 + \beta^2) \frac{Precision * Recall}{\beta^2 * Precision + Recall}$$

Experimental results

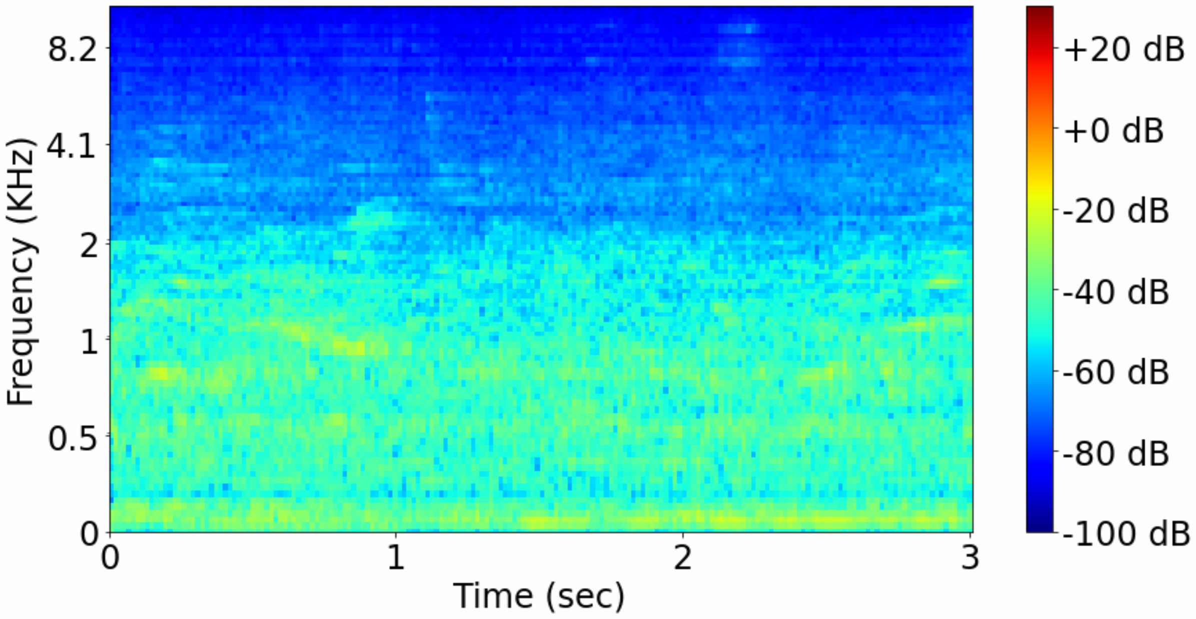
Main findings

Design choice	Worst		Best		Difference
Visual representation	Mel Spectrogram	0.9209	Gammatonegram	0.9230	0.002
Scaling range	Dynamic	0.8772	Fixed	0.9671	0.089
CNN architecture	MobileNet	0.9098	Xception	0.9310	0.021
Weights initialization	ImageNet	0.9221	Random	0.9222	<0.001

Discussion – background samples



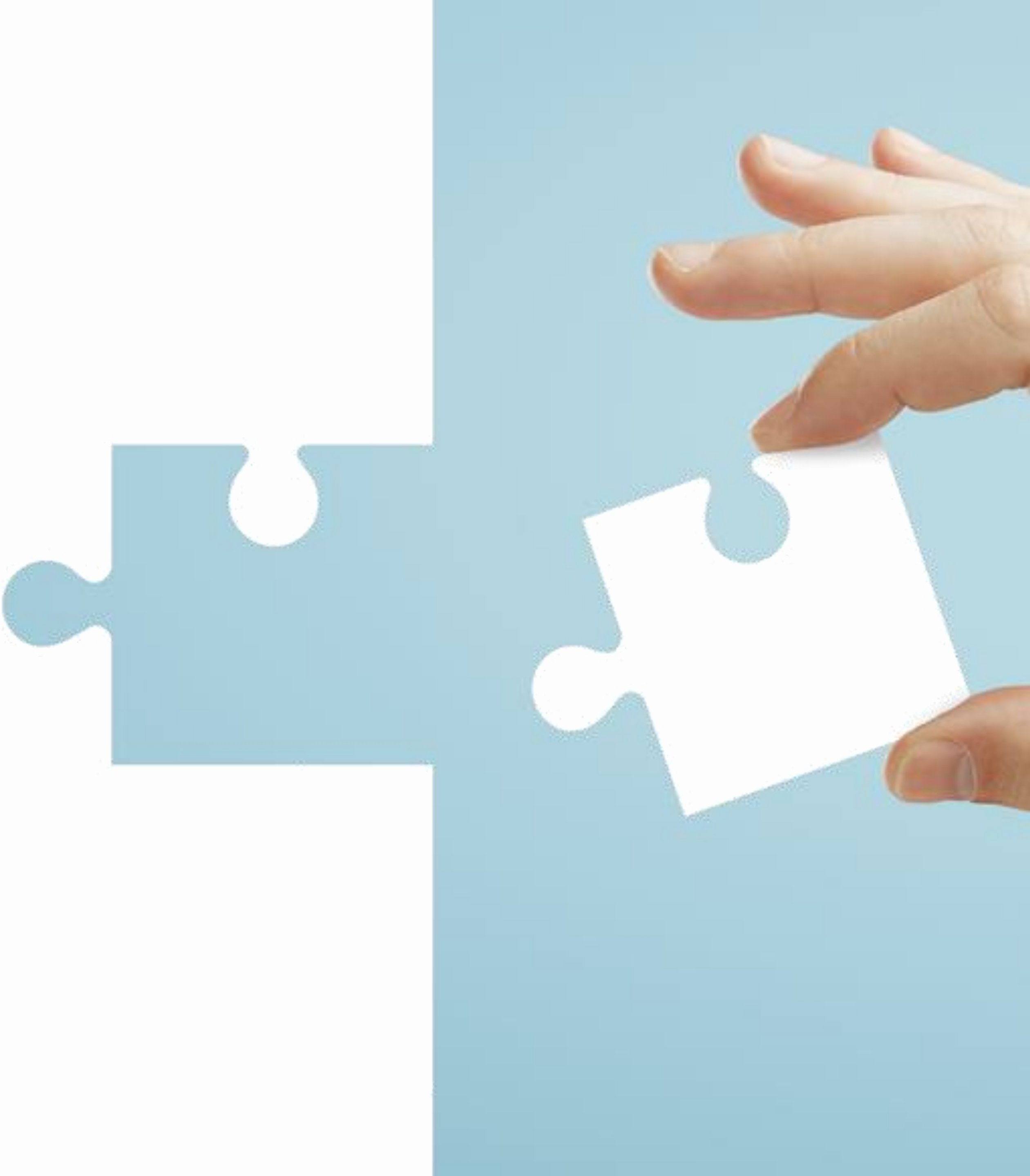
Dynamic range



Fixed range

Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices



Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices
  - Visual Representation

Spectrogram
Mel Spectrogram
Gammatonegram

🙄?

Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices
  - Visual Representation
  - Scaling Range

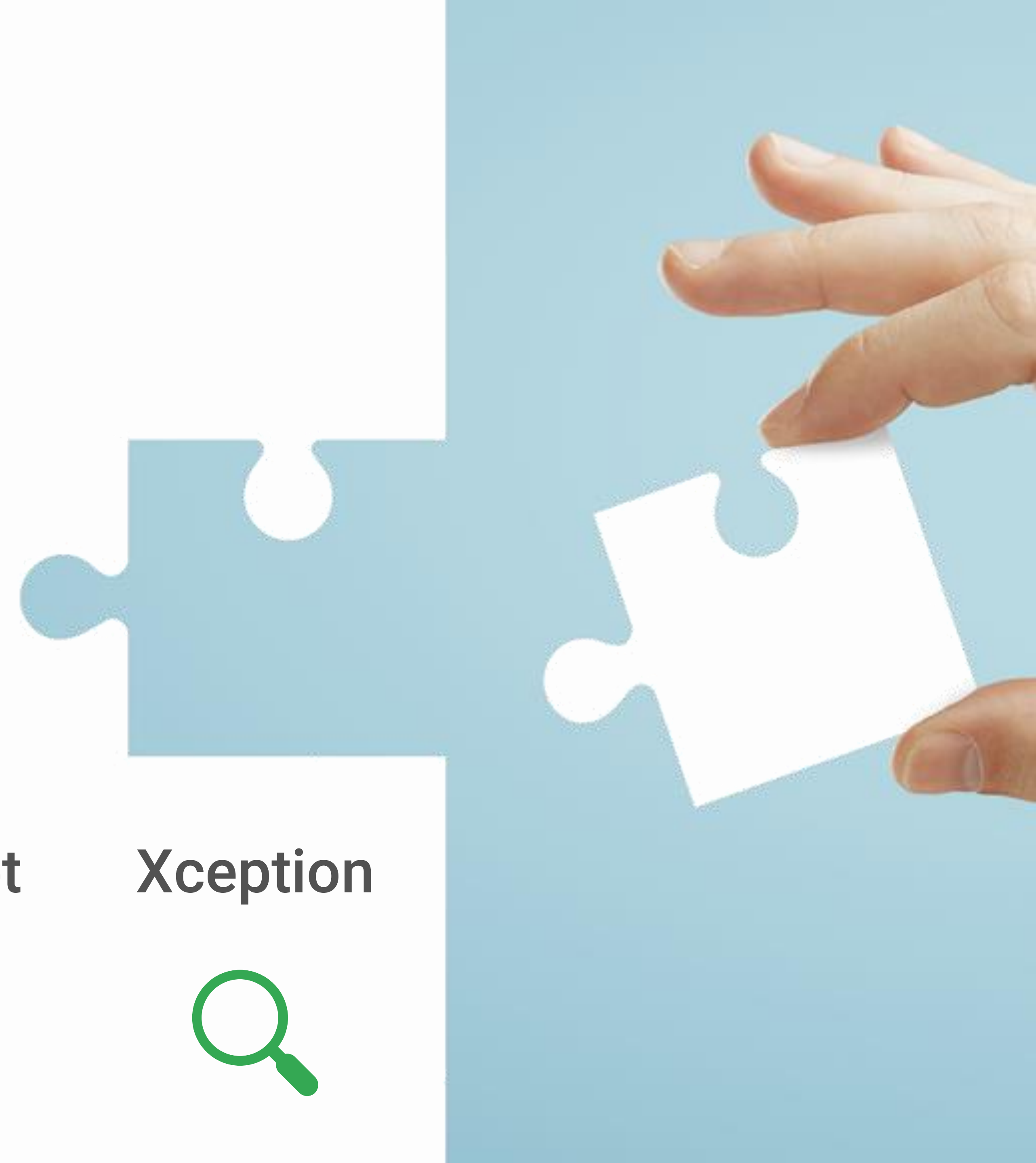
Dynamic 😟

Fixed 😎

Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices
  - Visual Representation
  - Scaling Range
  - CNN Architecture

MobileNet **Xception**



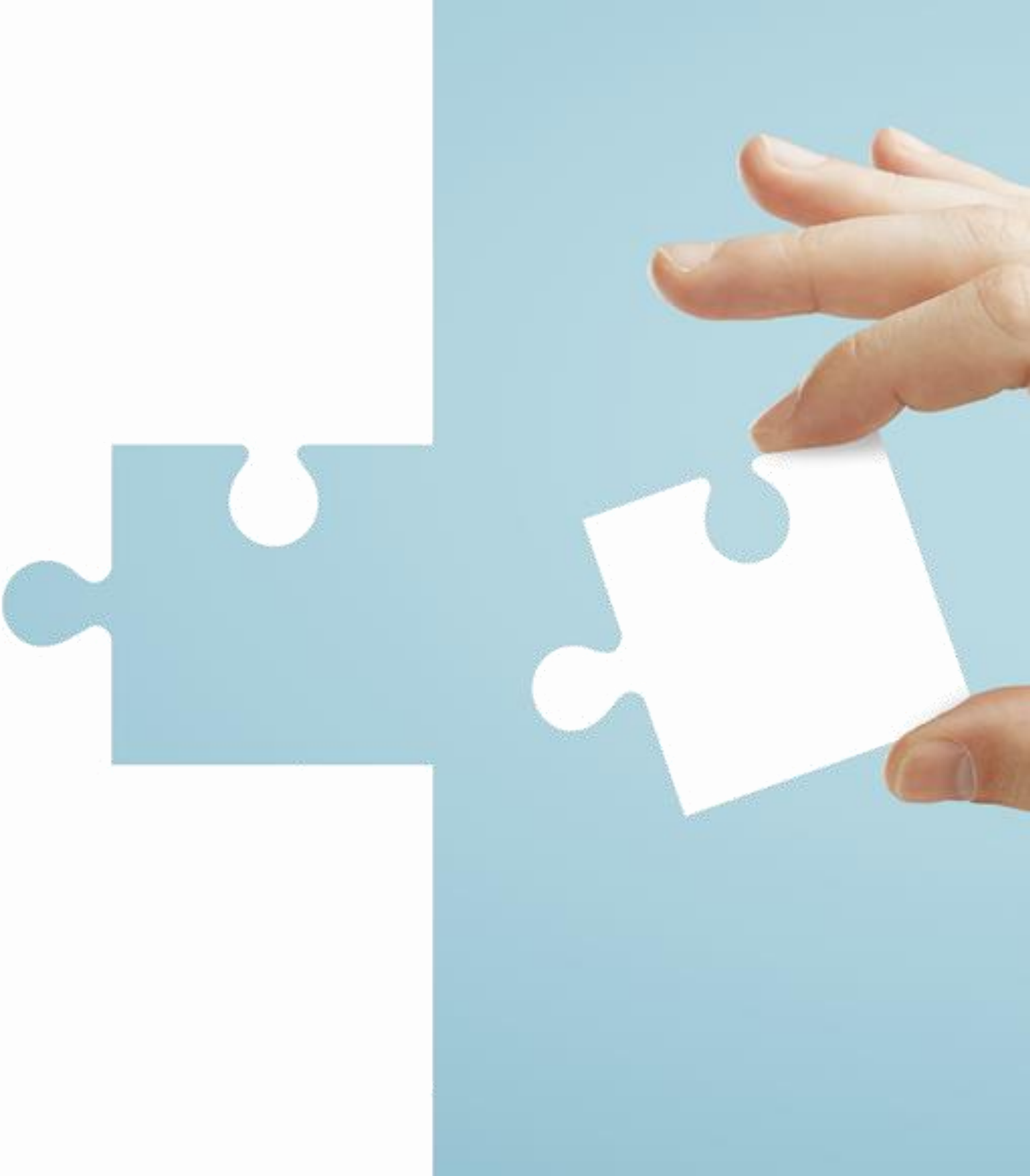
Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices
  - Visual Representation
  - Scaling Range
  - CNN Architecture
  - Weights initialization

ImageNet pre-training 🙄?

Conclusions

- Experimental evaluation for surveillance audio systems
- Analysis of several design choices
    - Visual Representation
    - Scaling Range
    - CNN Architecture
    - Weights initialization
- In-depth discussion about obtained results



ICPR 2020

25th INTERNATIONAL CONFERENCE
ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Thank you for your attention!

Questions?

For more information you can contact the authors at:

aroberto@unisa.it