DCASE 2016

CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION

Michele Valenti[1] (valenti.michele.w@gmail.com),

Aleksandr Diment[2], Giambattista Parascandolo[2],

Stefano Squartini[1], Tuomas Virtanen[2]

[1]Università Politecnica delle Marche, Italy

[2]Tampere University of Technology, Finland

DCASE 2016

CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION

Michele Valenti[1] (valenti.michele.w@gmail.com),

Aleksandr Diment[2], Giambattista Parascandolo[2],

Stefano Squartini[1], Tuomas Virtanen[2]

[1]Università Politecnica delle Marche, Italy

[2]Tampere University of Technology, Finland

Outline

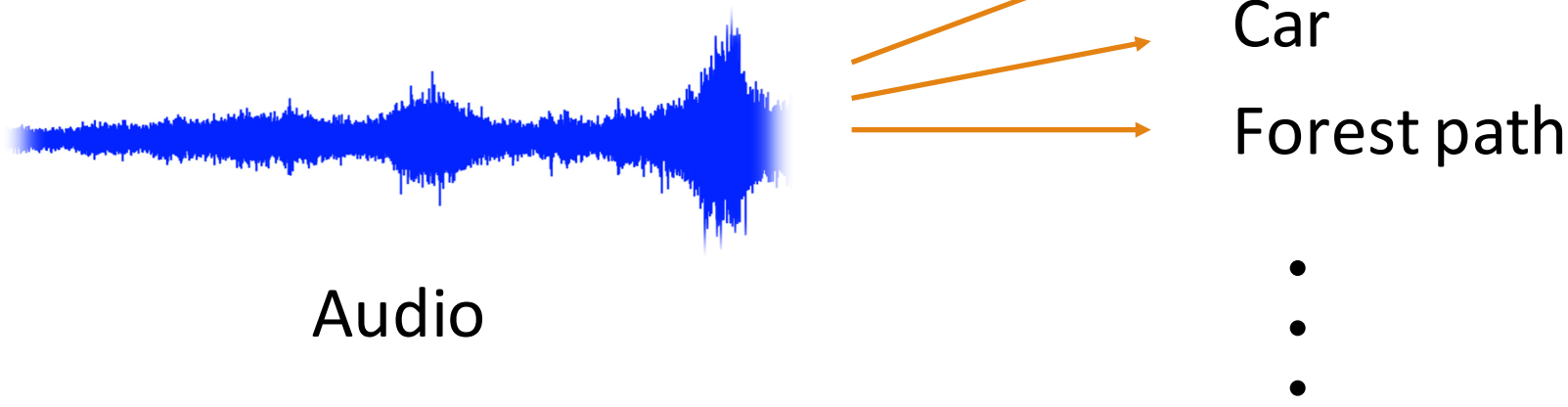
- Introduction
- Our system
- Training modes
- Results
- Challenge ranking

Introduction

What is “acoustic scene classification”?

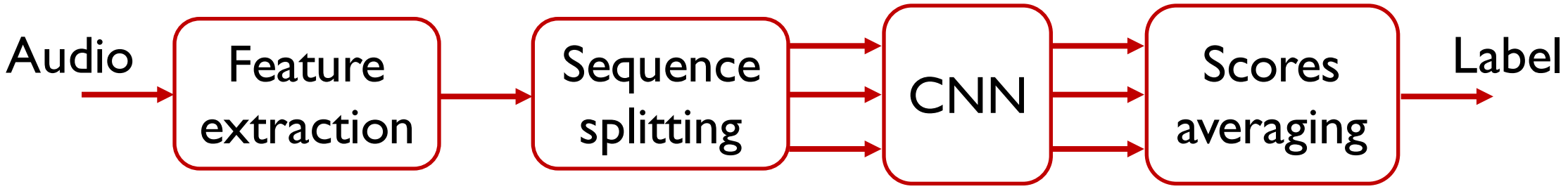
Introduction

What is “acoustic scene classification”?

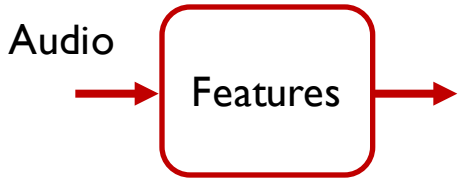


Our system

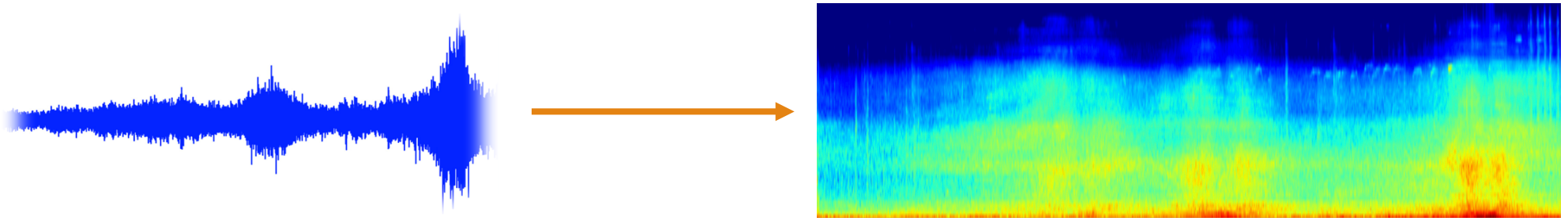
Overview



Our system



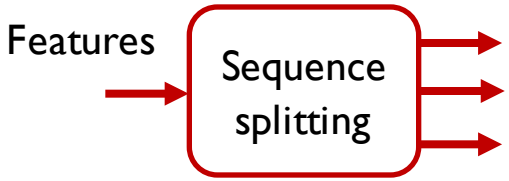
Features



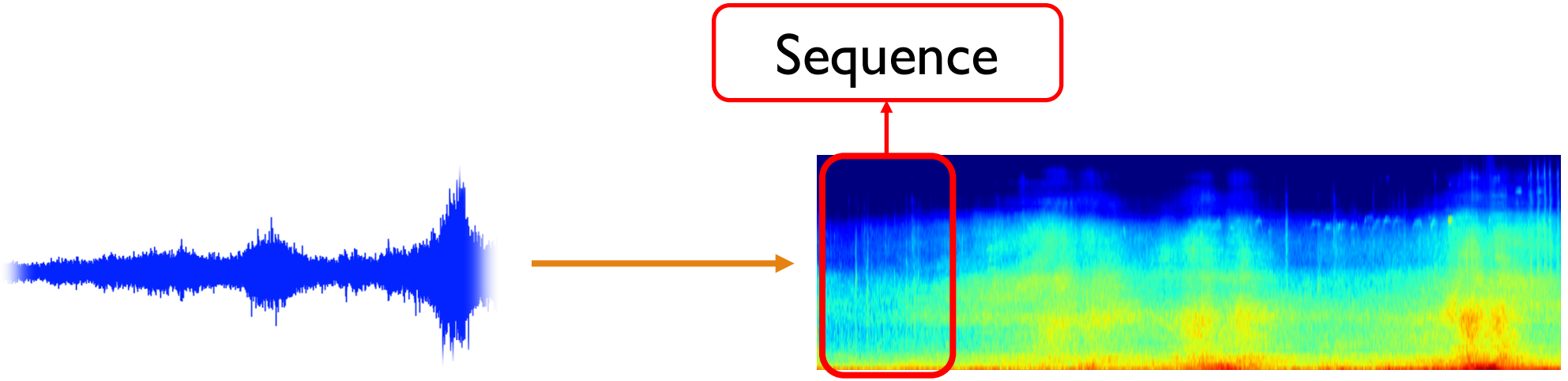
Raw audio

Log-mel spectrogram

Our system



Sequence splitting

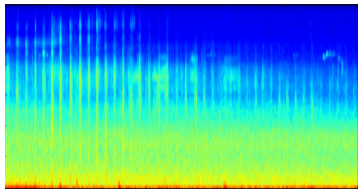


Raw audio segment

Log-mel spectrogram

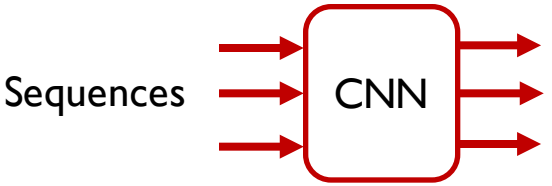
Our system

Convolutional neural network

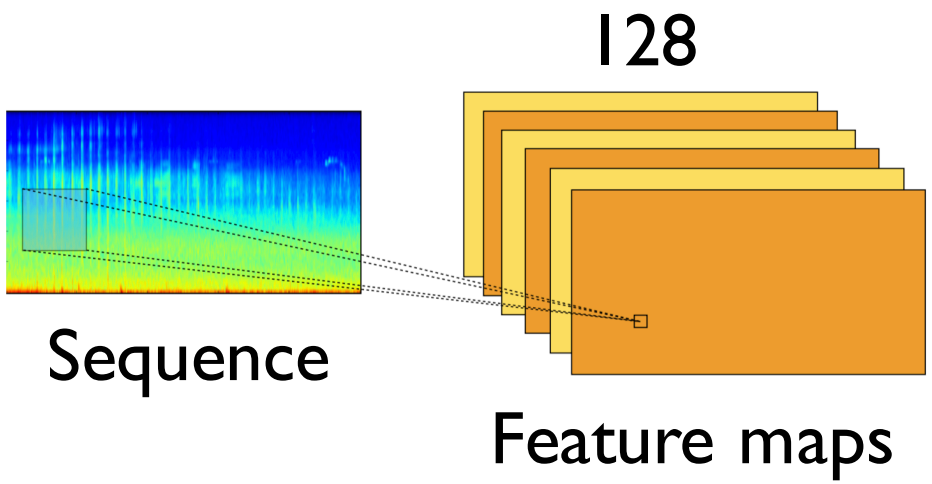


Sequence

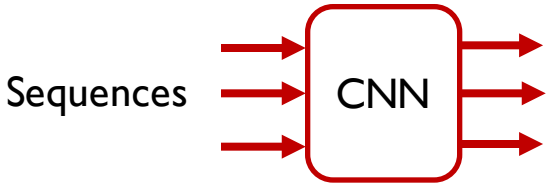
Our system



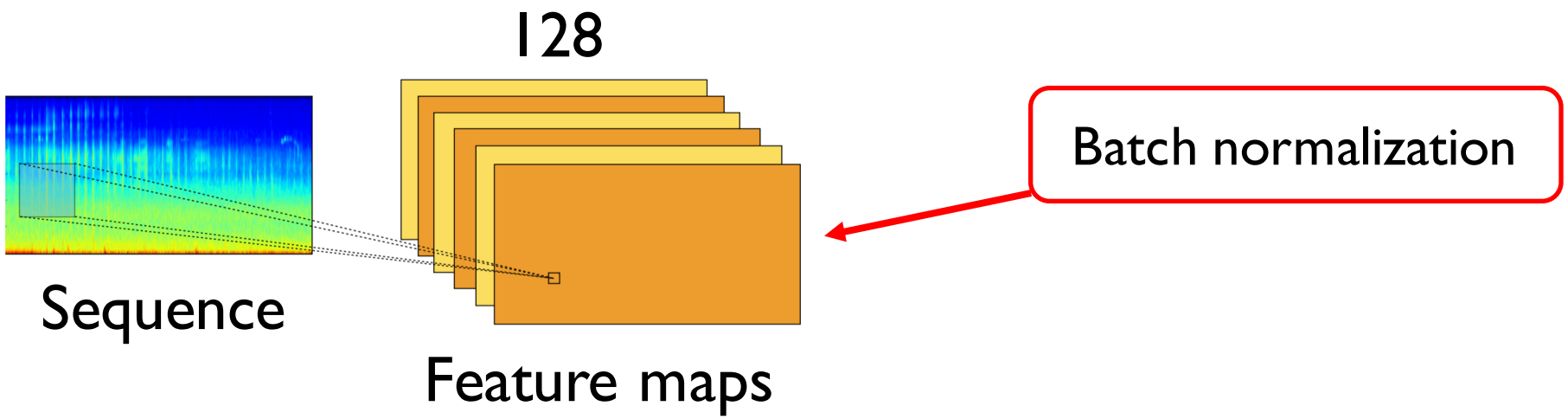
Convolutional neural network



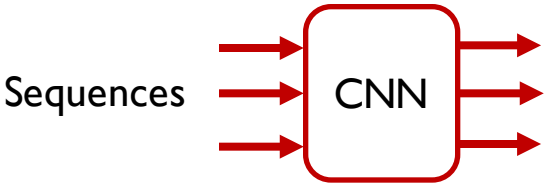
Our system



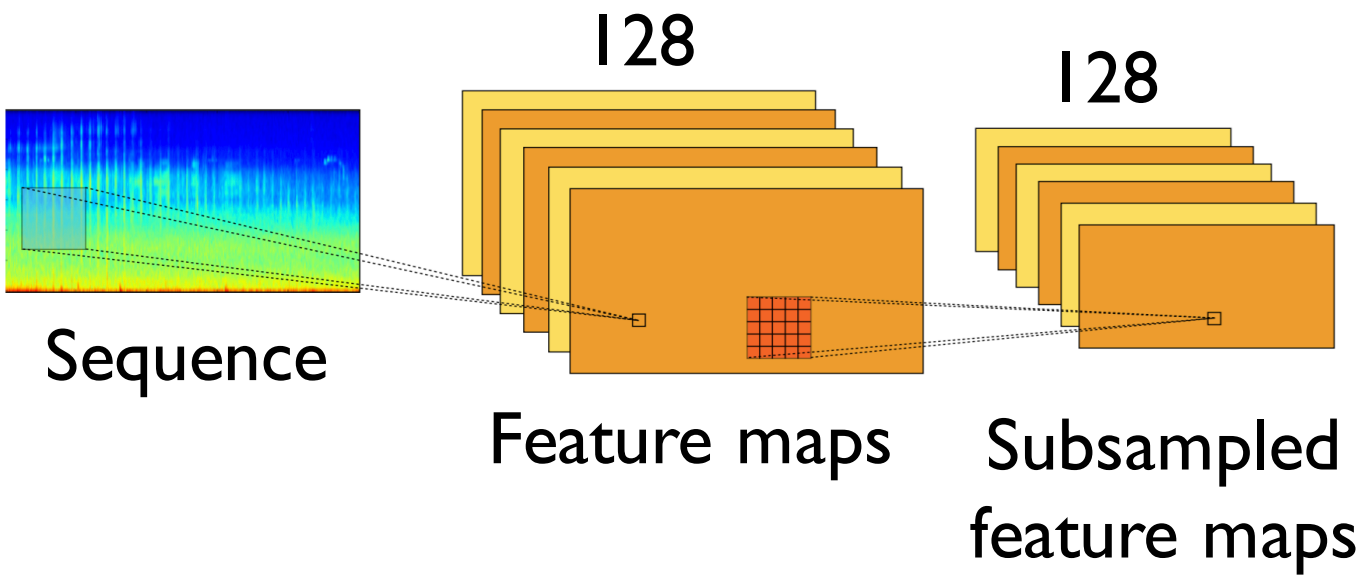
Convolutional neural network



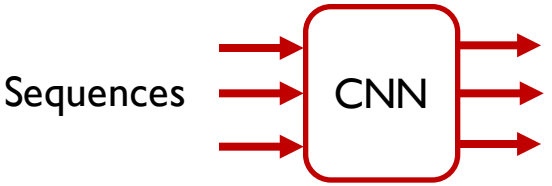
Our system



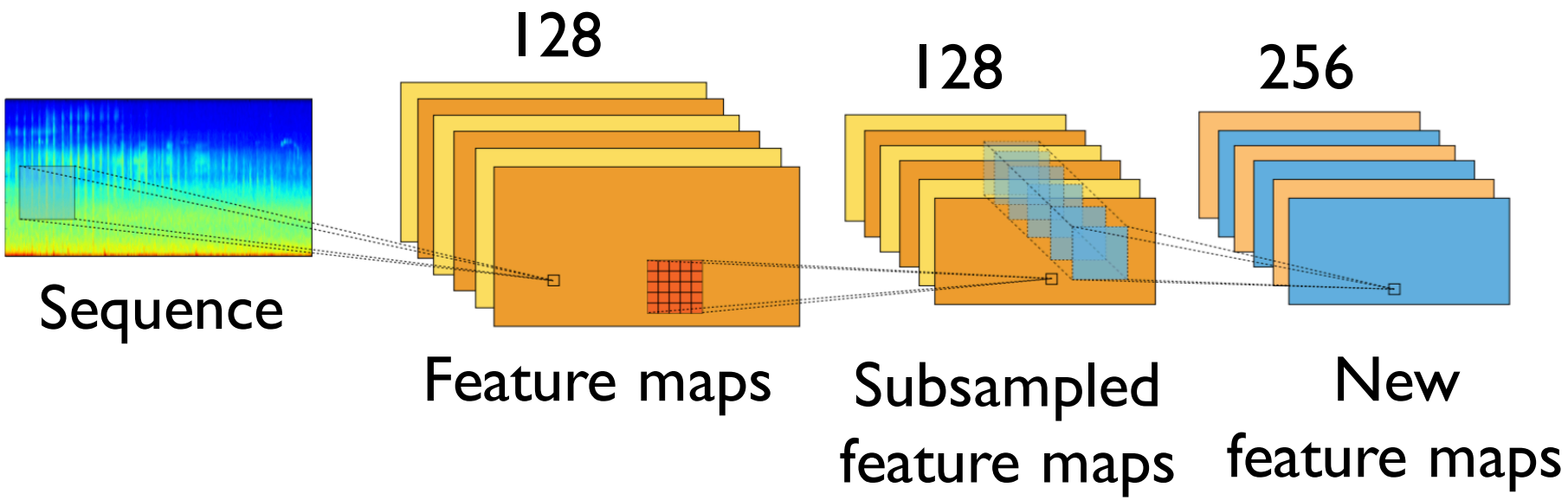
Convolutional neural network



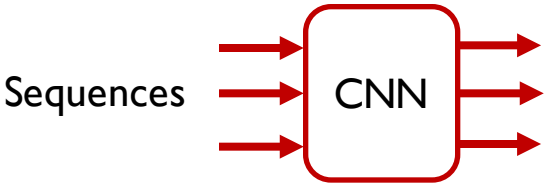
Our system



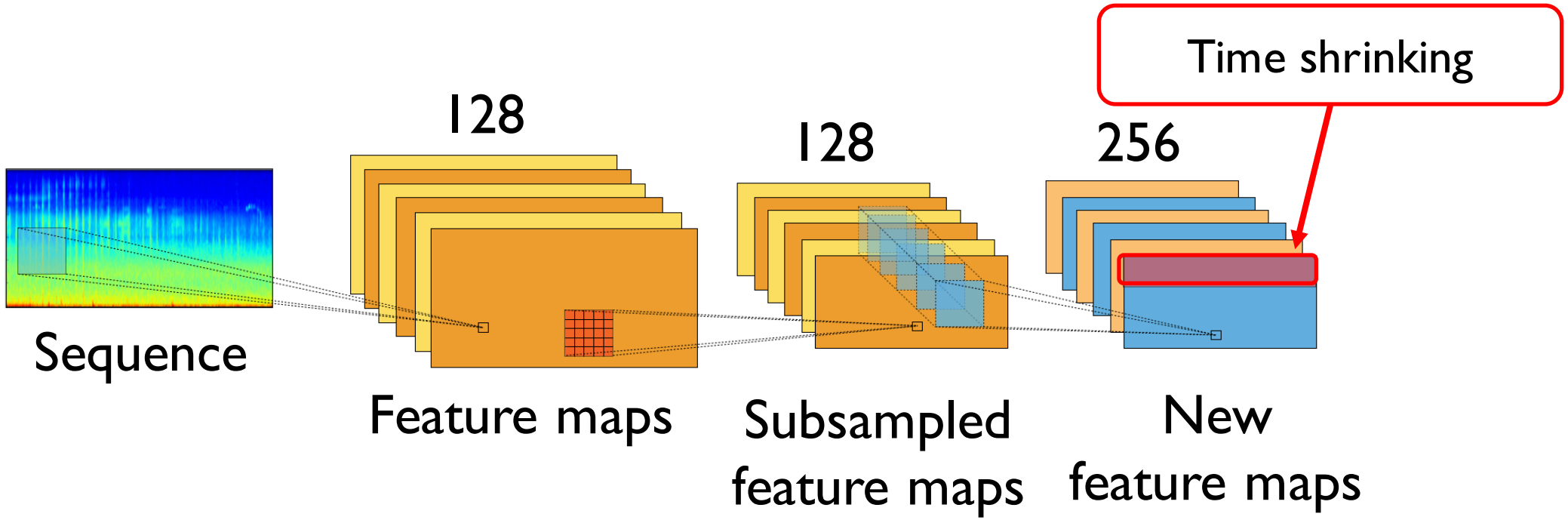
Convolutional neural network



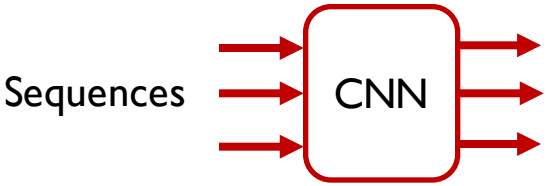
Our system



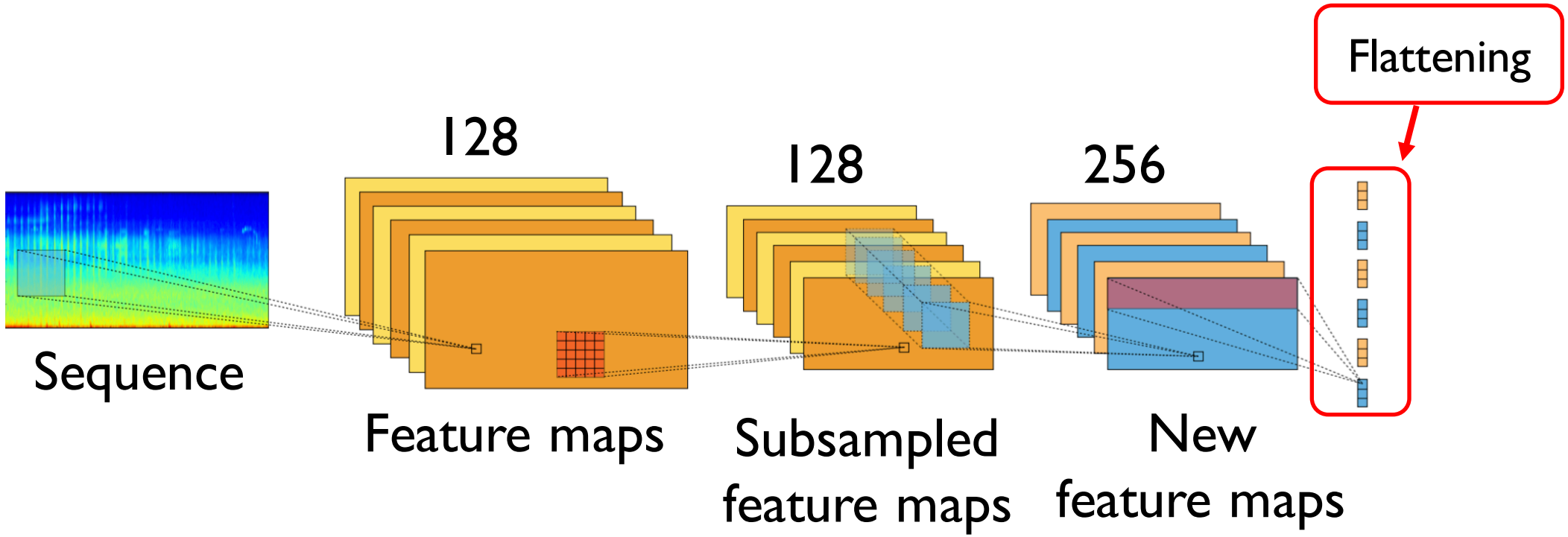
Convolutional neural network



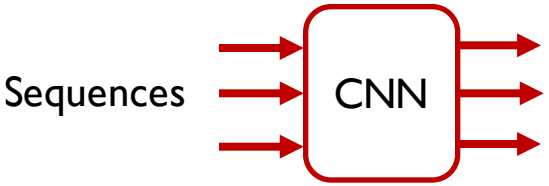
Our system



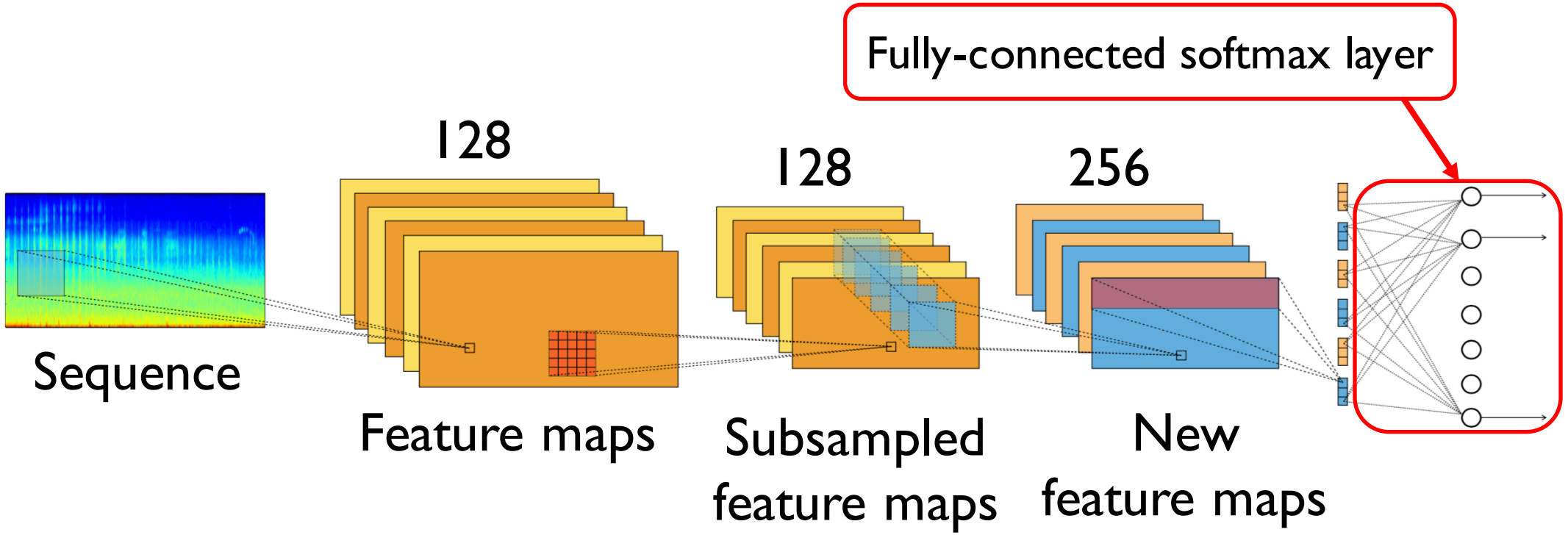
Convolutional neural network



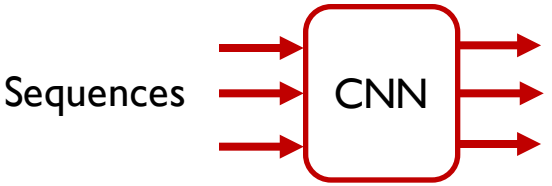
Our system



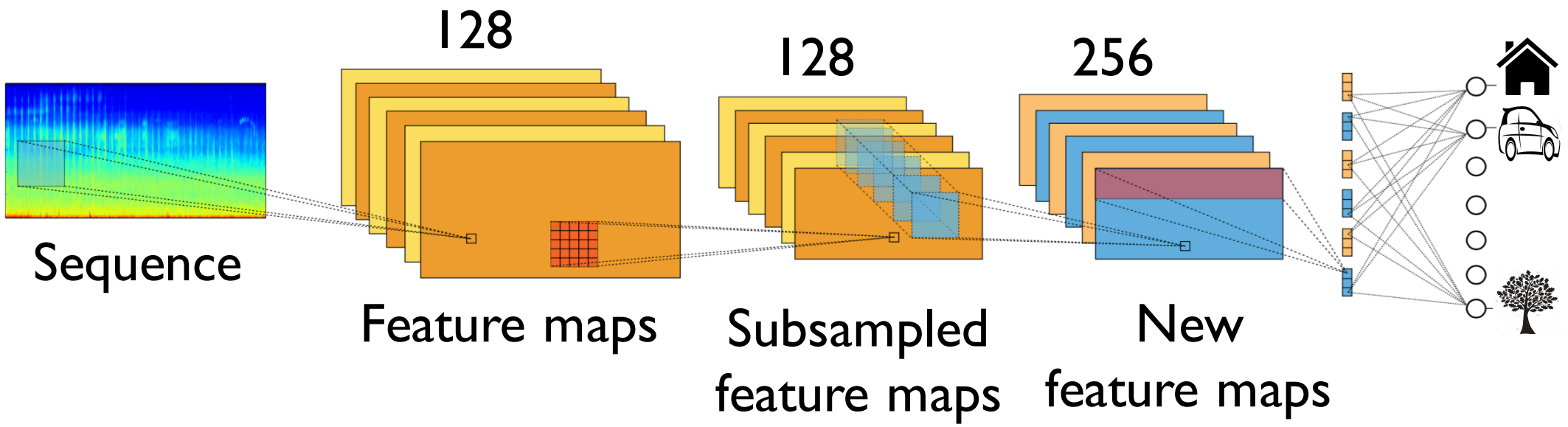
Convolutional neural network



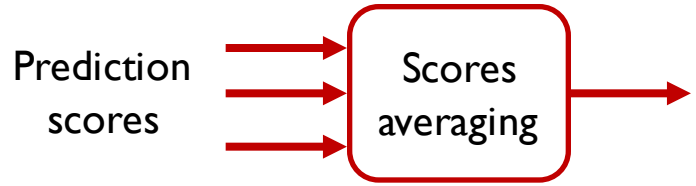
Our system



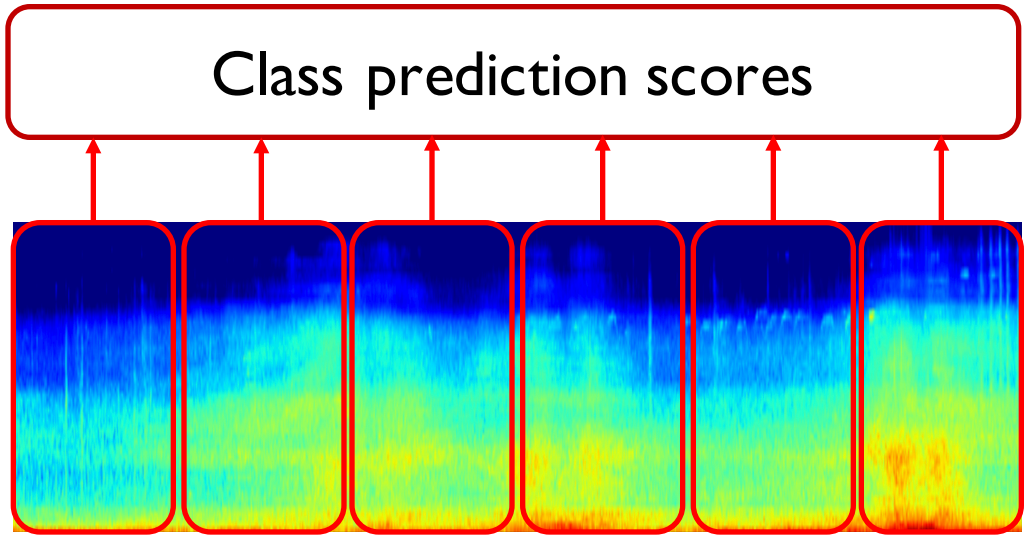
Convolutional neural network



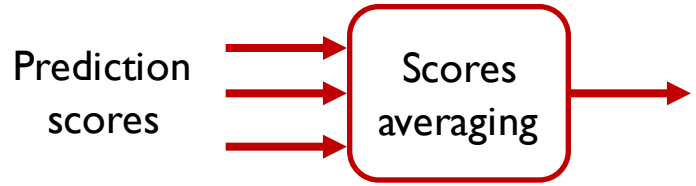
Our system



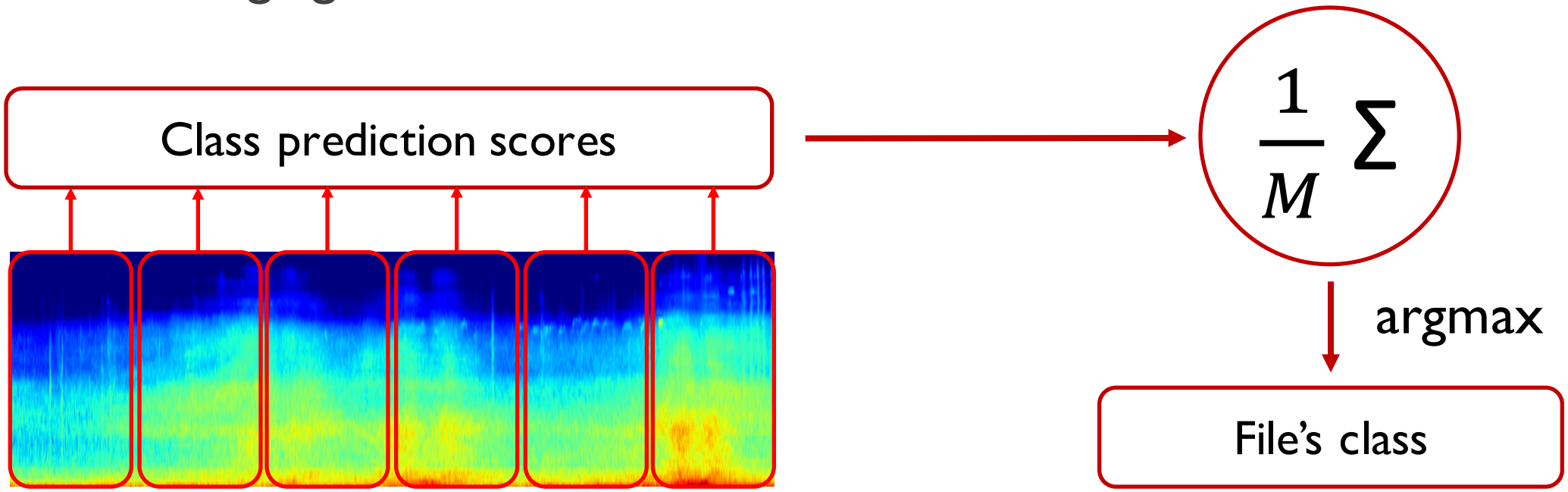
Scores averaging



Our system



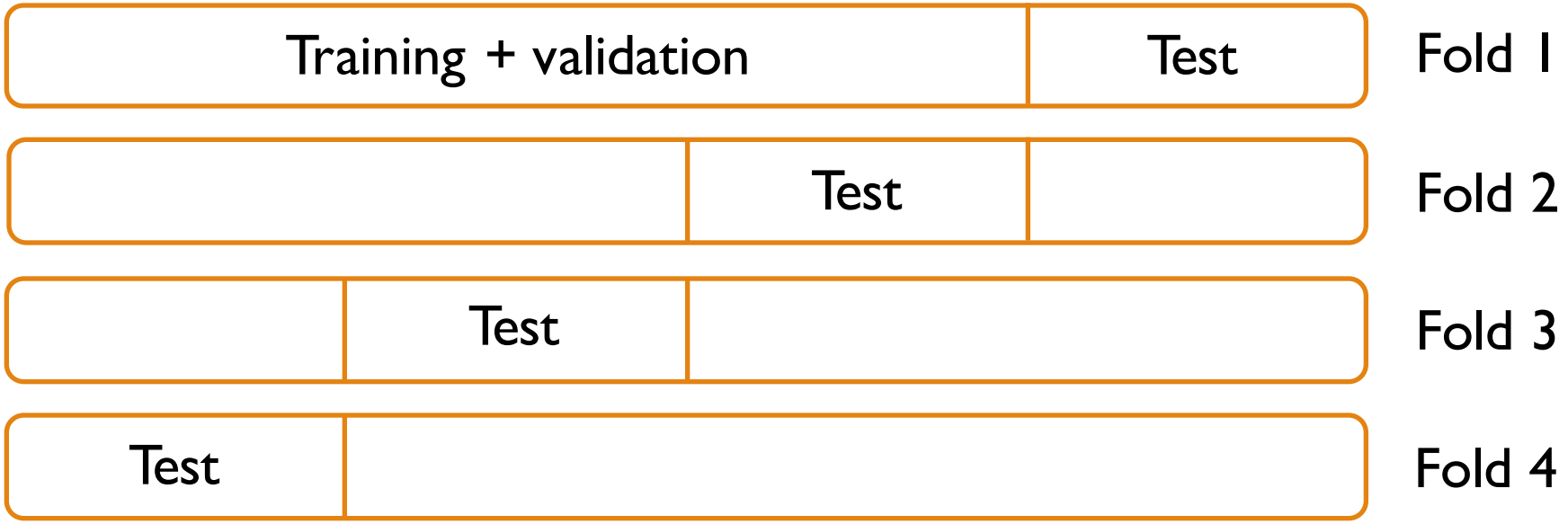
Scores averaging



Training

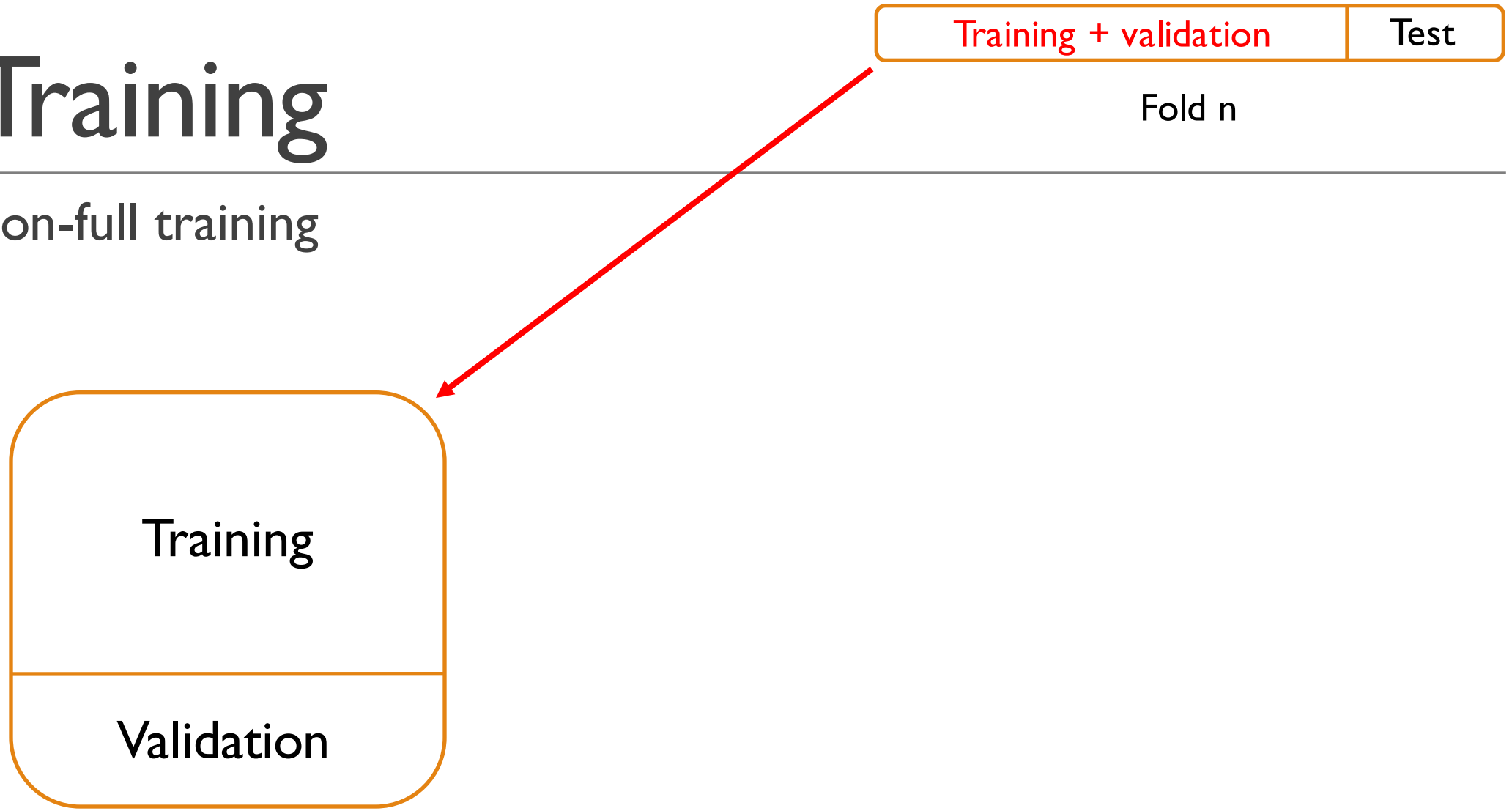
Training

Cross-validation setup



Training

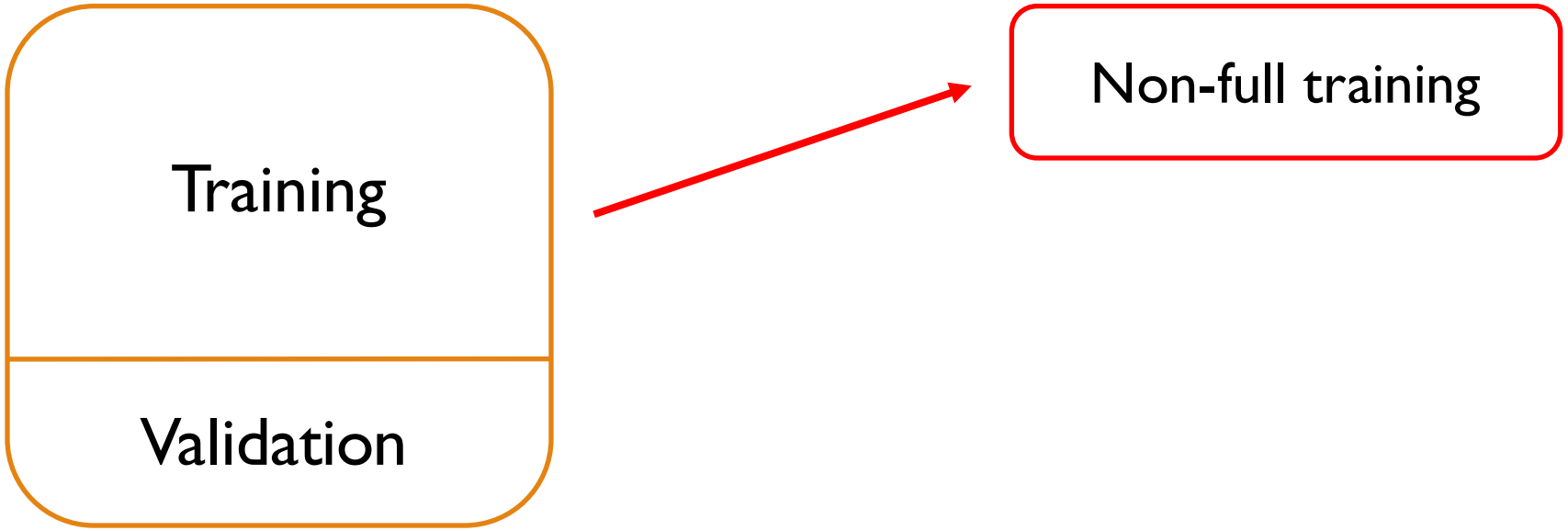
Non-full training



Training

Fold n

Non-full training

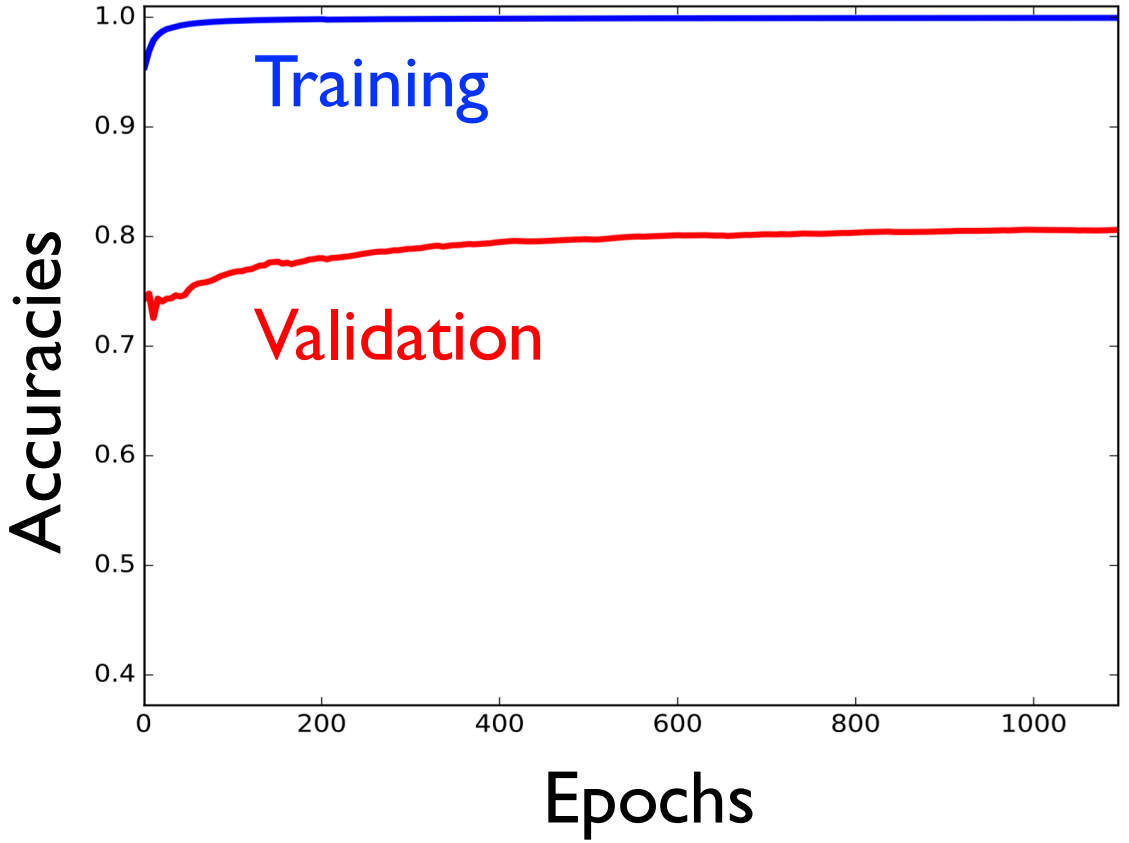


Training

Training + validation	Test

Fold n

Non-full training

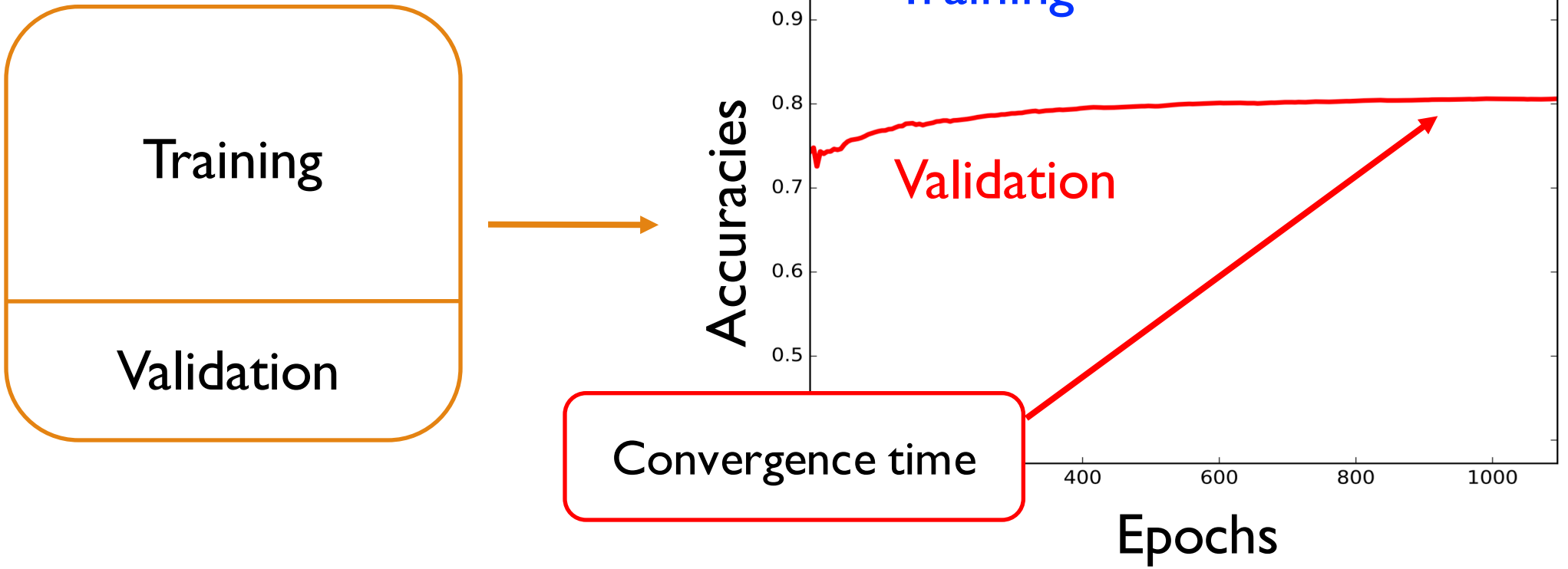


Training

Training + validation	Test

Fold n

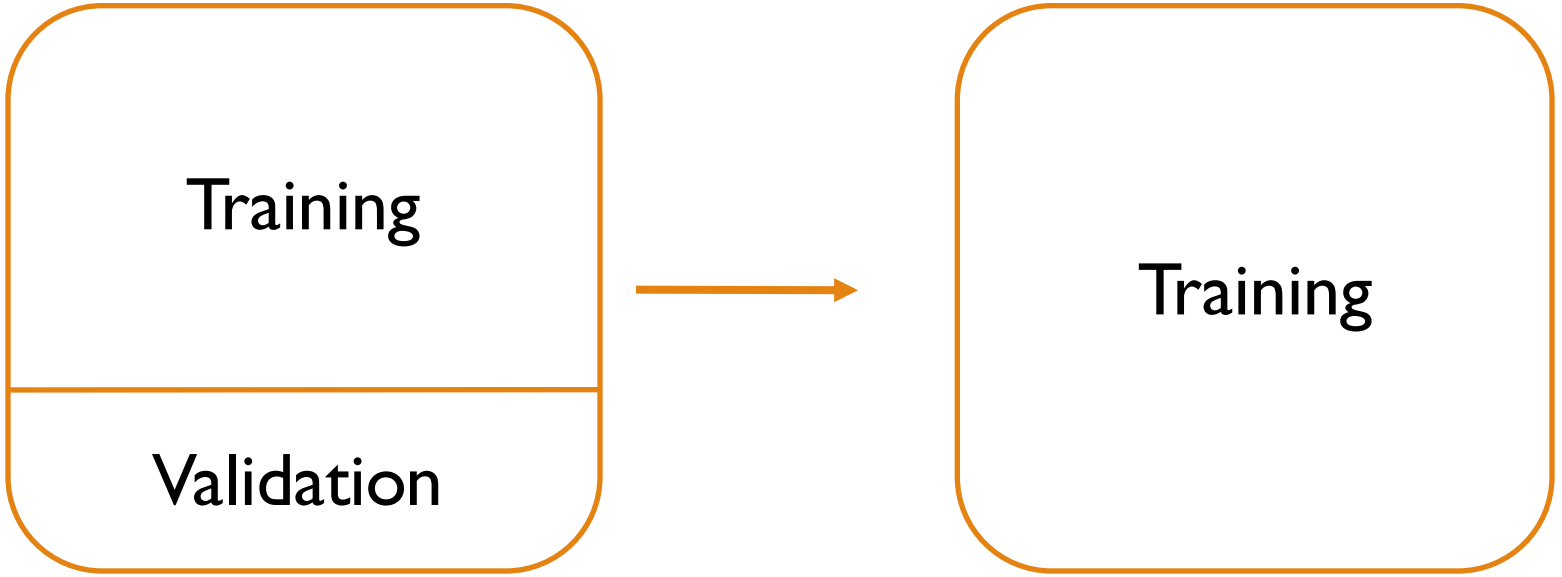
Non-full training



Training

Fold n

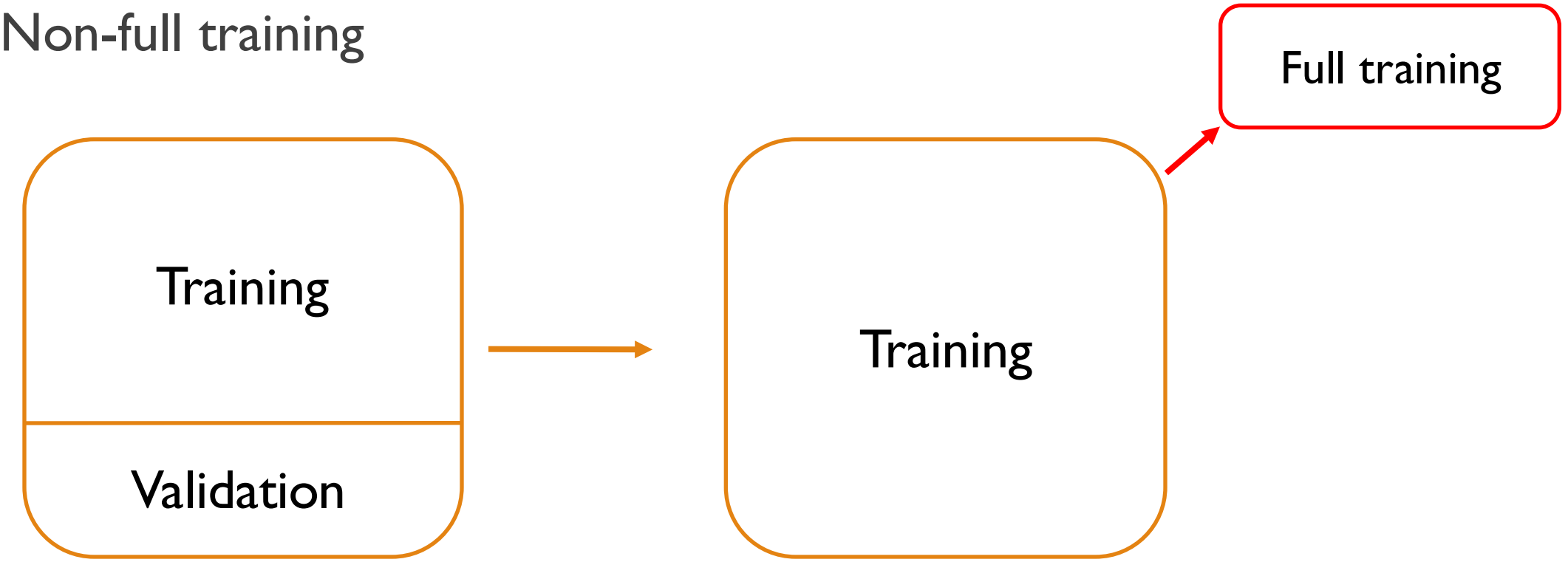
Non-full training



Training

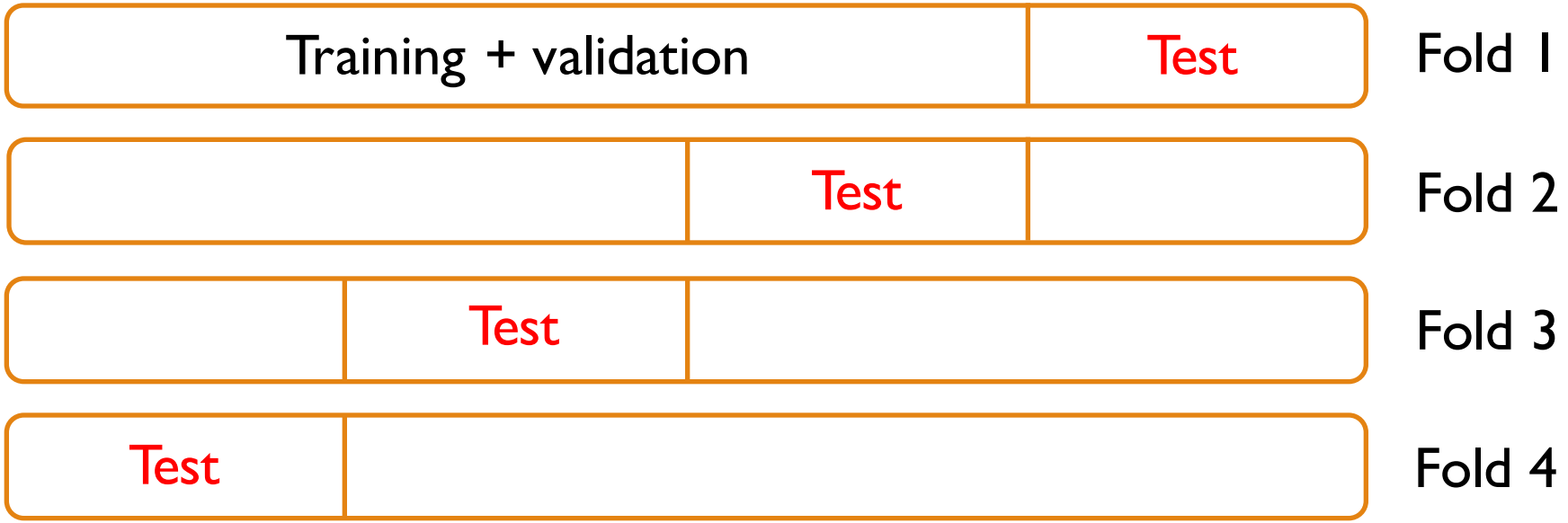
Fold n

Non-full training



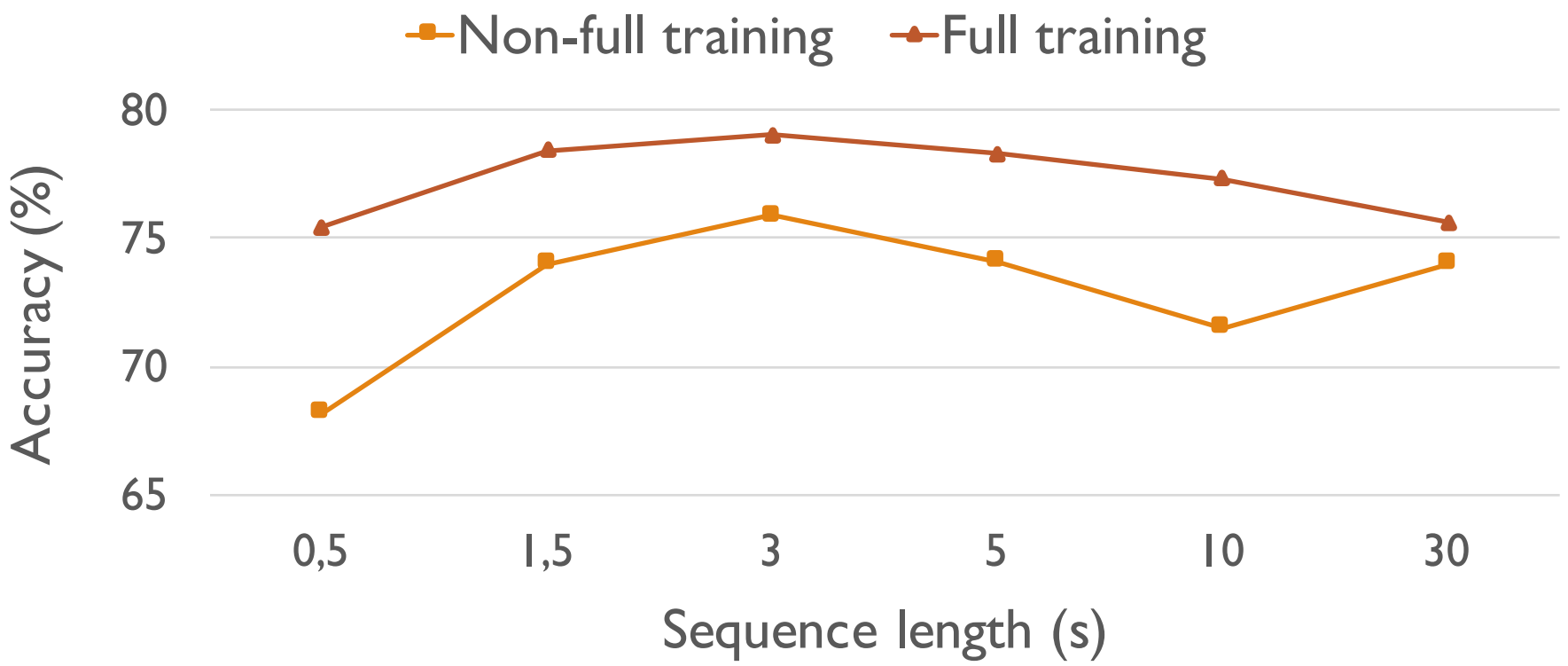
Results

Test data



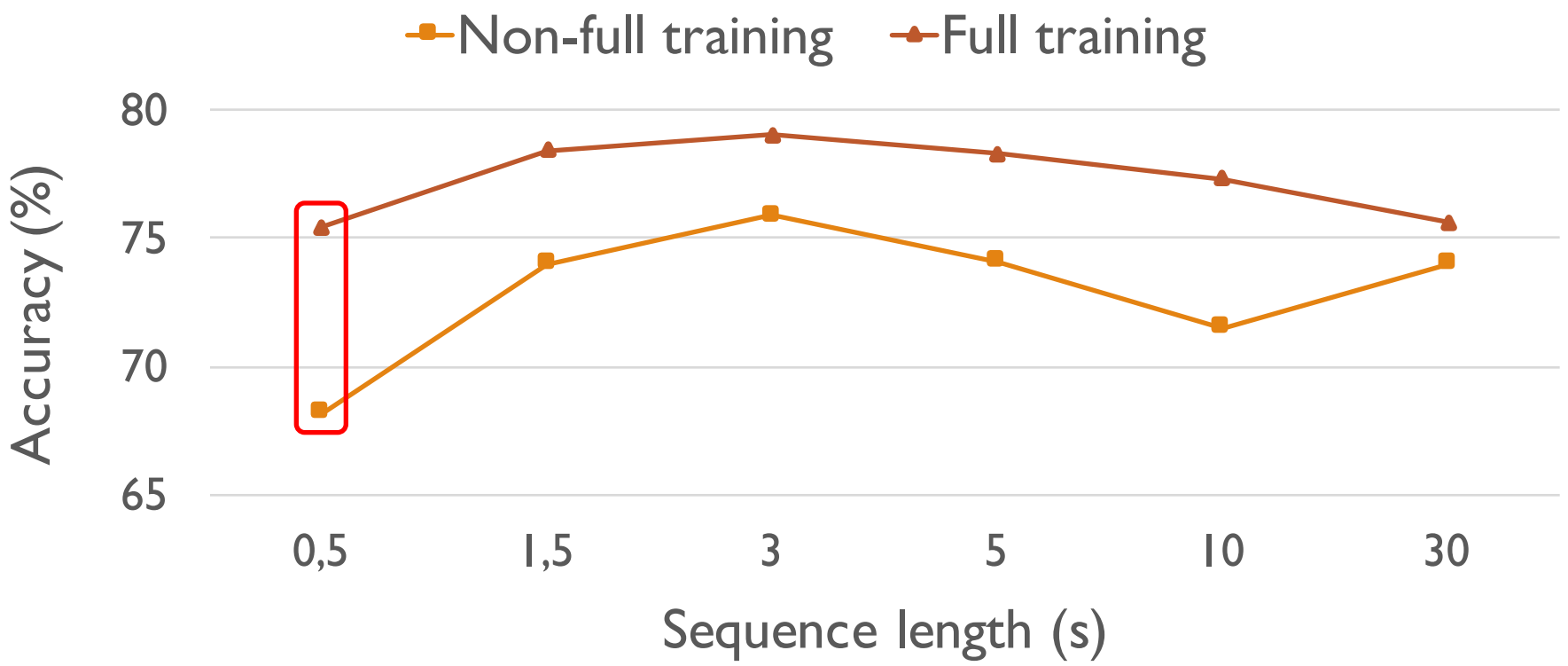
Results

Sequence length



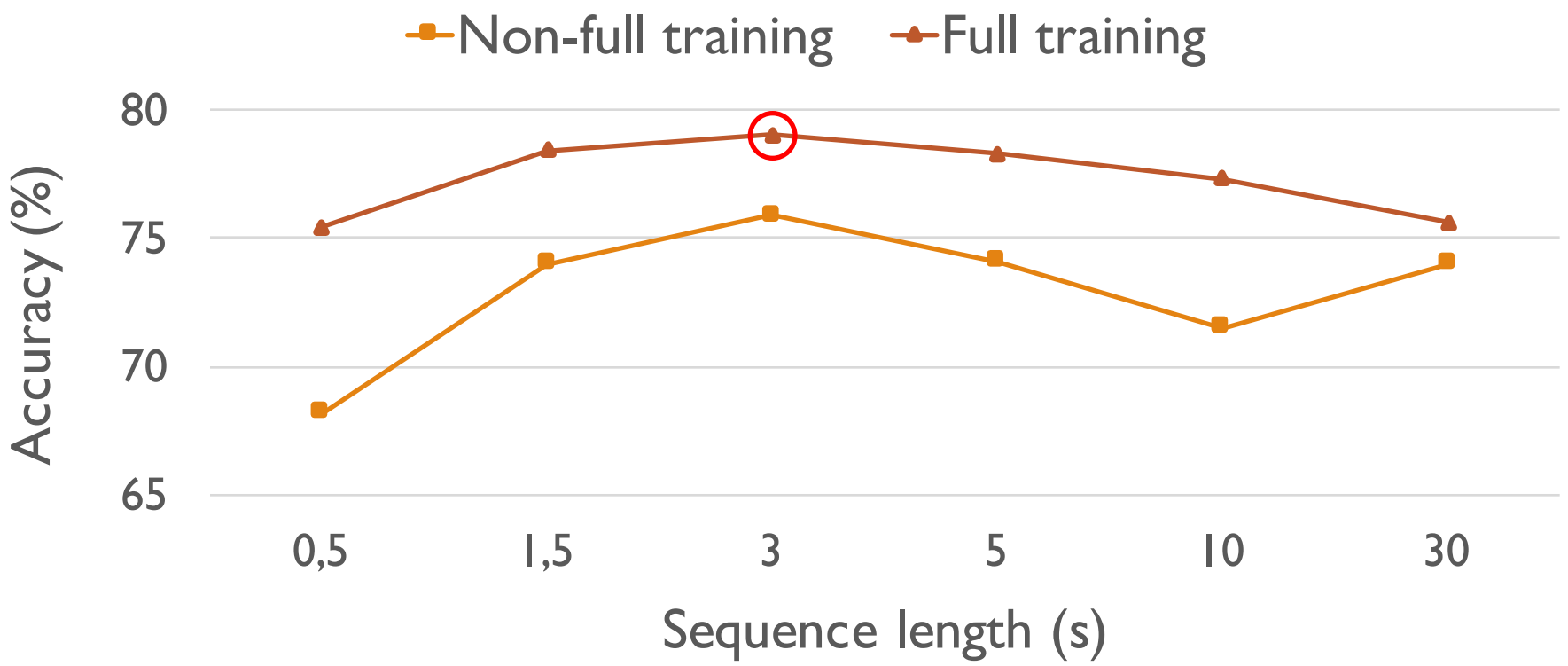
Results

Sequence length



Results

Sequence length



Results

Class accuracies

Class	Accuracy (%)
Beach	75.6
Bus	76.9
Café/Restaurant	74.4
Car	91.0
City center	93.6
Forest path	96.2
Grocery store	88.5
Home	80.8

Class	Accuracy (%)
Library	66.6
Metro station	96.2
Office	97.4
Park	59.0
Residential area	73.1
Train	46.2
Tram	78.2

Results

Class accuracies

Class	Accuracy (%)
Beach	75.6
Bus	76.9
Café/Restaurant	74.4
Car	91.0
City center	93.6
Forest path	96.2
Grocery store	88.5
Home	80.8

Class	Accuracy (%)
Library	66.6
Metro station	96.2
Office	34.6% Residential area
Park	59.0
Residential area	73.1
Train	46.2
Tram	29.5% Bus

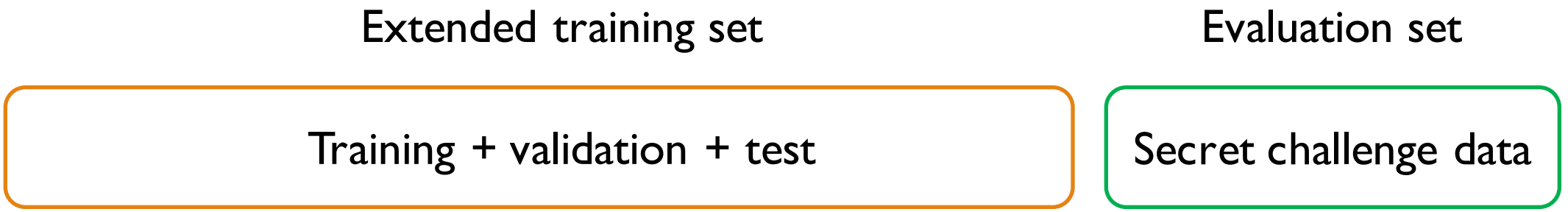
Results

Other classifiers

System	Sequence length (s)	Accuracy (%)	
		Non-full training	Full training
Baseline GMM (MFCC)	-	-	72.6
Two-layer CNN (MFCC)	5	67.7	72.6
Two-layer MLP (log-mel)	-	66.6	69.3
One-layer CNN (log-mel)	3	70.3	74.8
Two-layer CNN (log-mel)	**3**	**75.9**	**79.0**

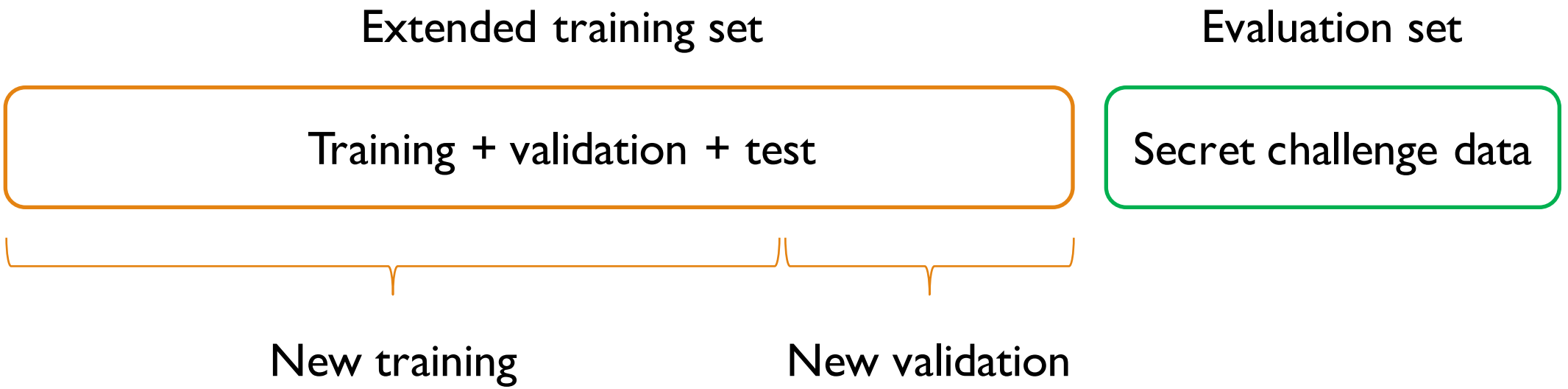
Challenge ranking

Final training



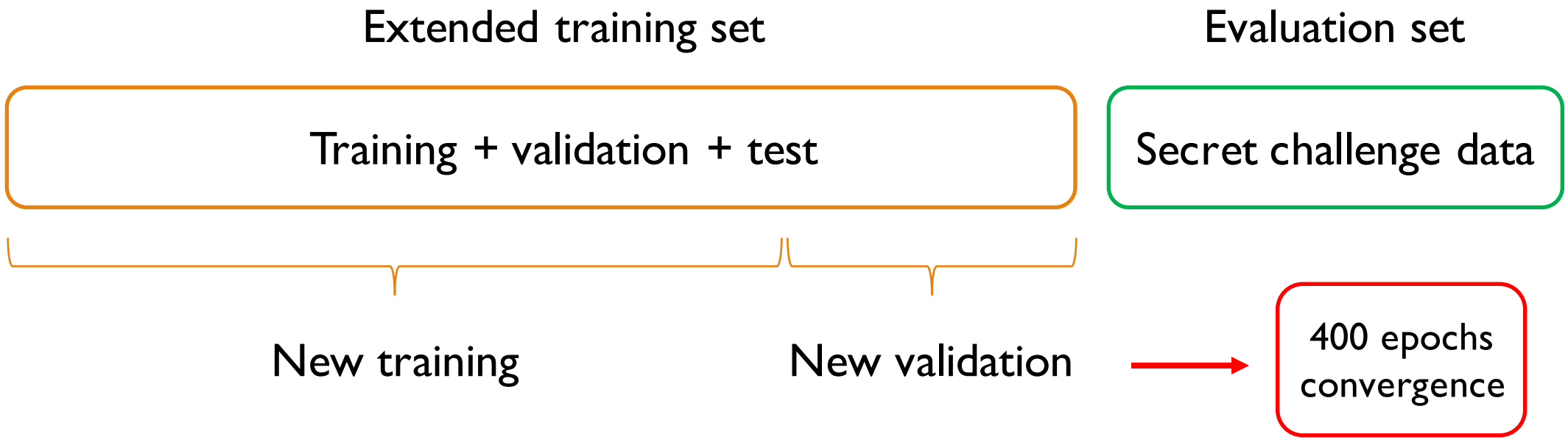
Challenge ranking

Final training



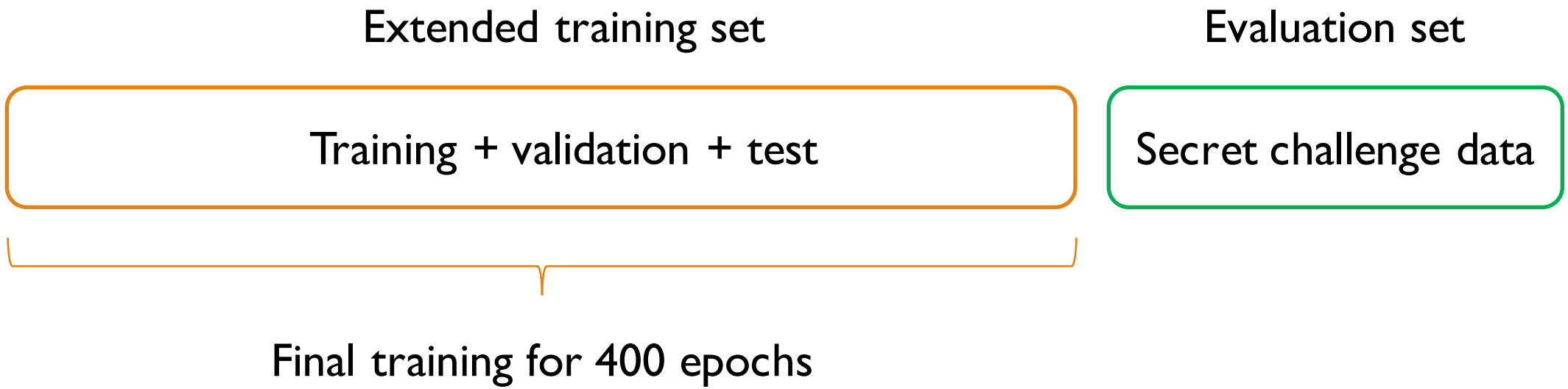
Challenge ranking

Final training

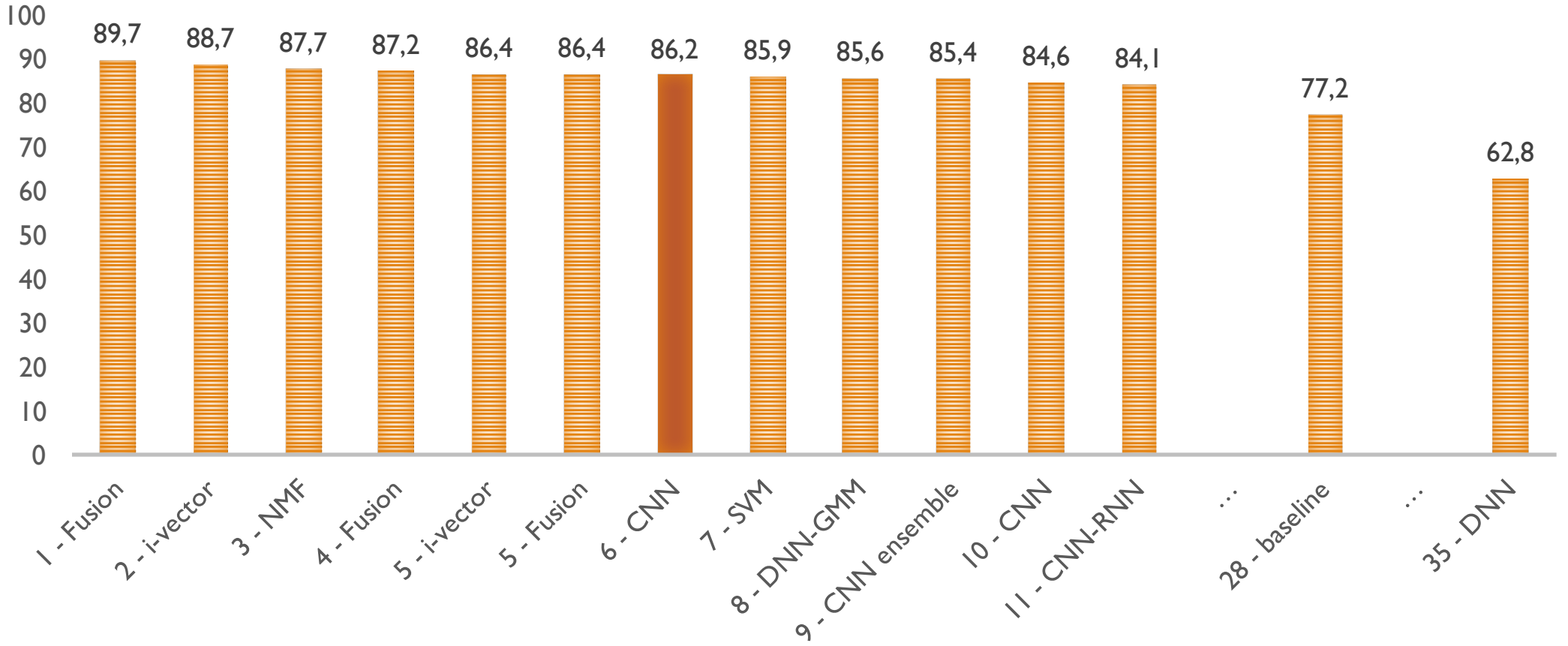


Challenge ranking

Final training



Challenge ranking



DCASE 2016

CONVOLUTIONAL NEURAL NETWORKS FOR ACOUSTIC SCENE CLASSIFICATION

Michele Valenti[1] (valenti.michele.w@gmail.com),

Aleksandr Diment[2], Giambattista Parascandolo[2],

Stefano Squartini[1], Tuomas Virtanen[2]

[1]Università Politecnica delle Marche, Italy

[2]Tampere University of Technology, Finland

Results

Feature comparison

System	Sequence length (s)	Accuracy (%)	
		Non-full training	Full training
Two-layer CNN (MFCC)	5	67.7	72.6
Two-layer CNN (log-mel)	5	74.1	78.3