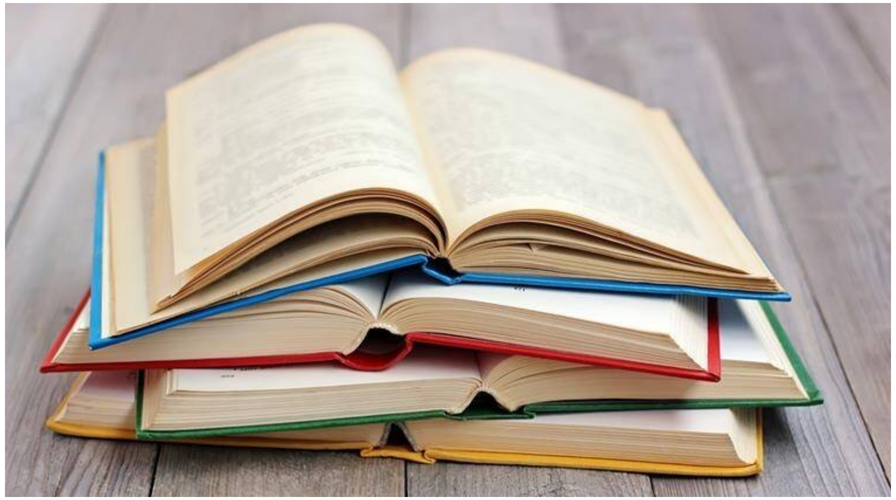
Not Enough Data: Transfer Learning

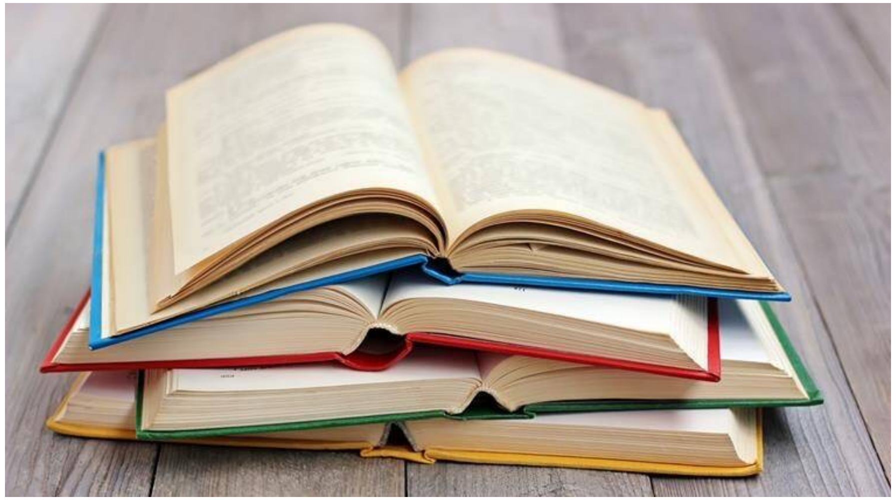
Zuzana Koščová

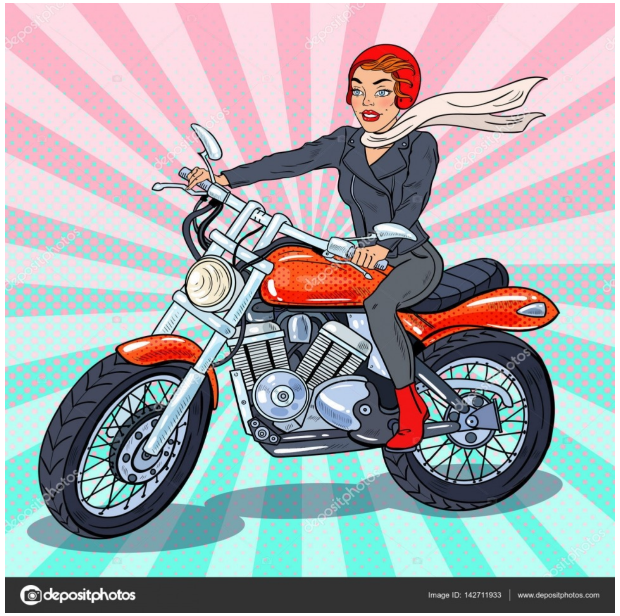
ISI of the CAS – AIMT, Brno, Cz

Transfer Learning in a real world

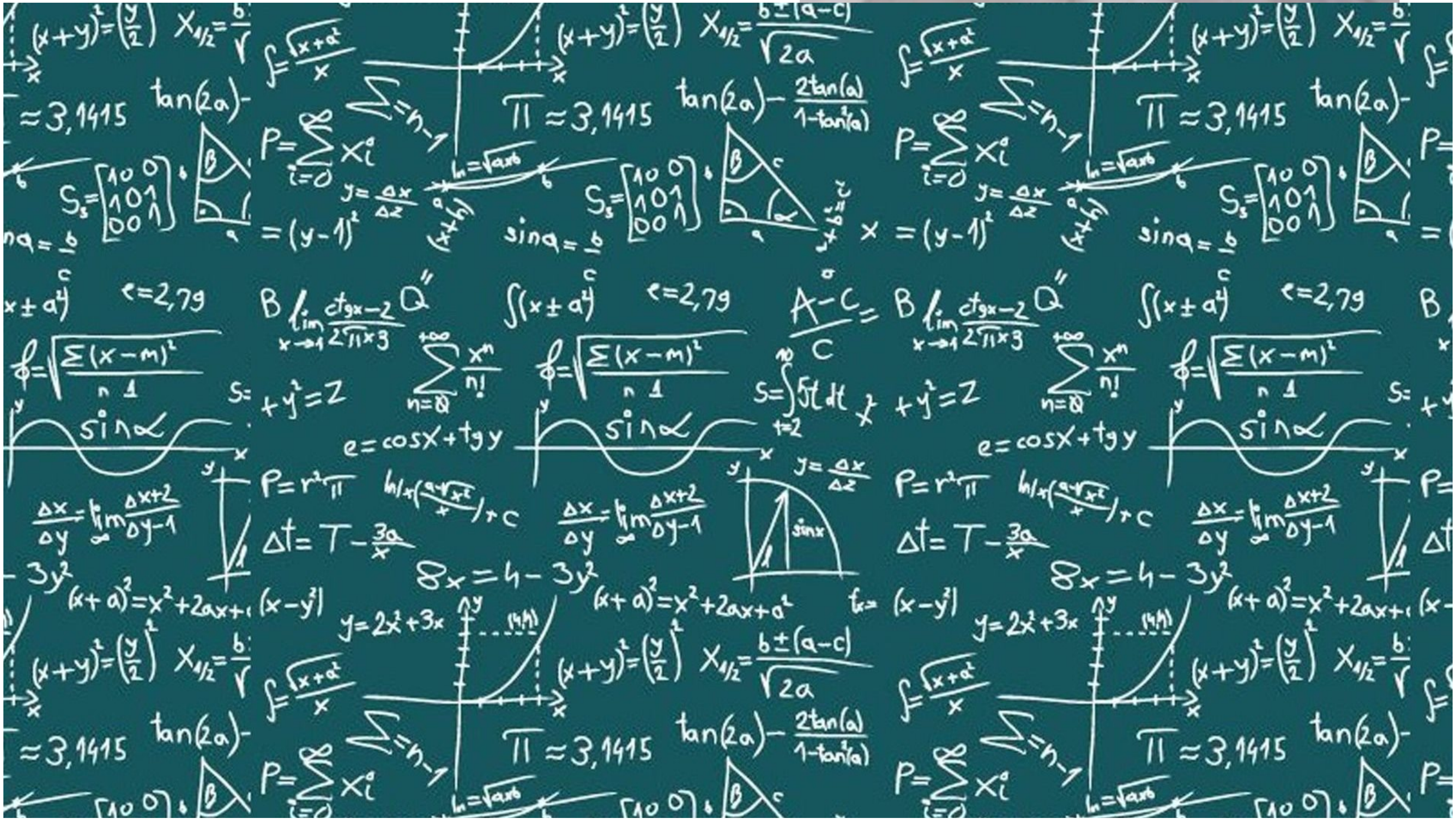


Transfer Learning in a real world

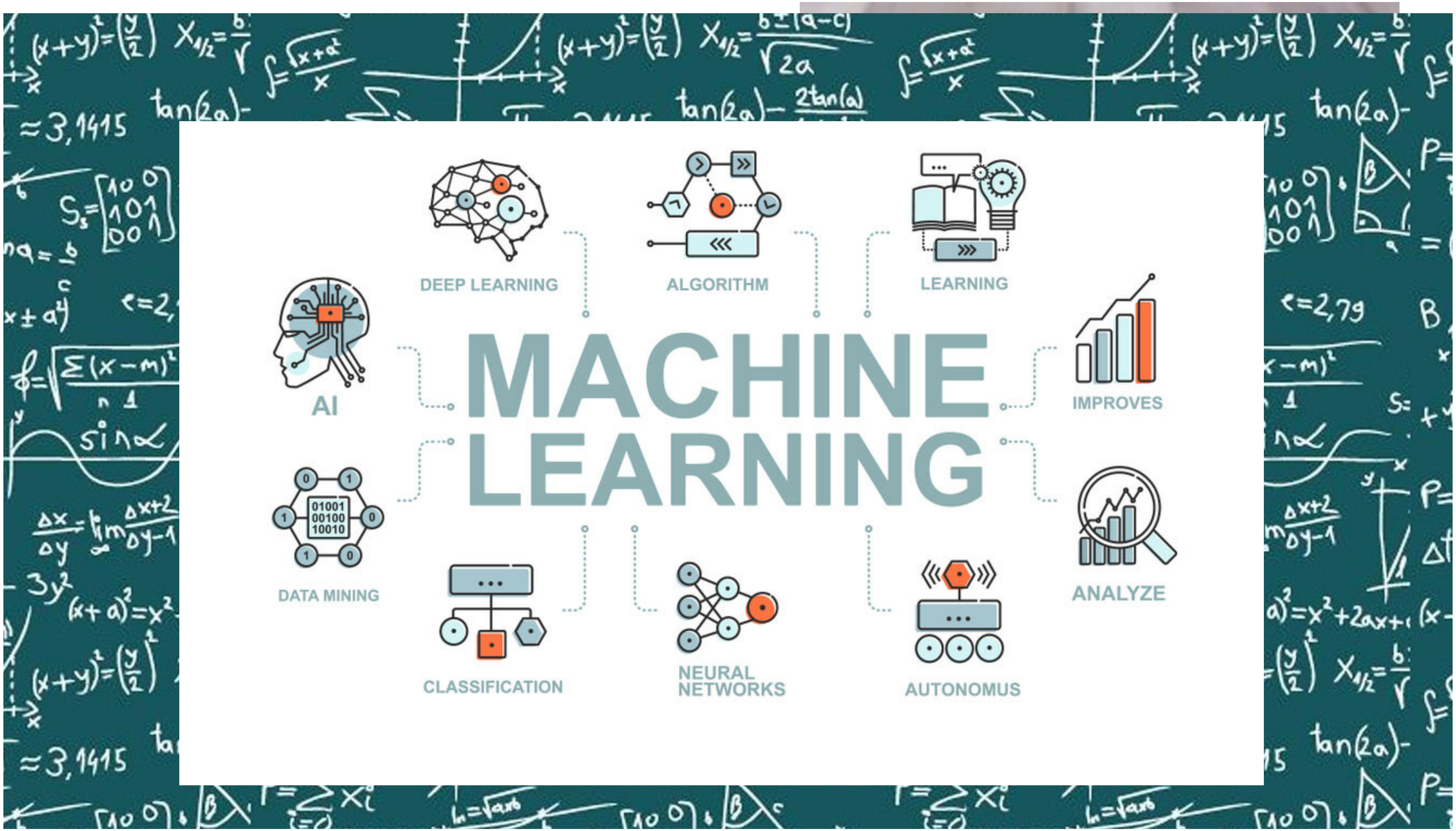




Transfer Learning in a real world



Transfer Learning in a real world

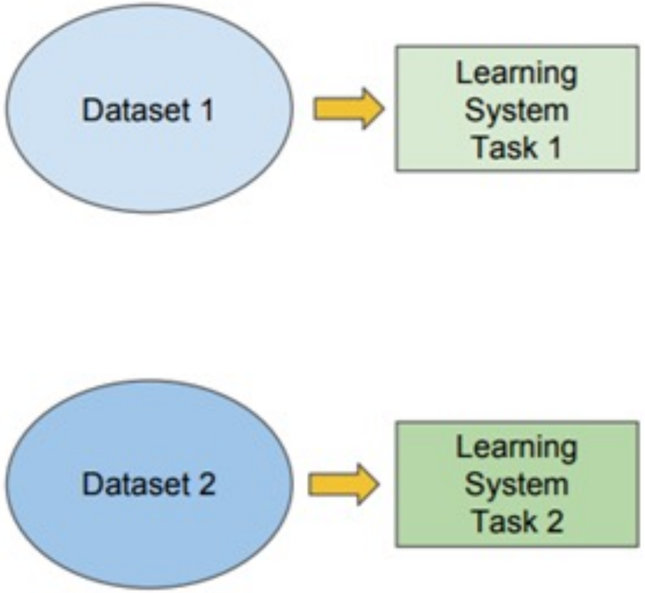


Transfer Learning Intro

Traditional ML vs Transfer Learning

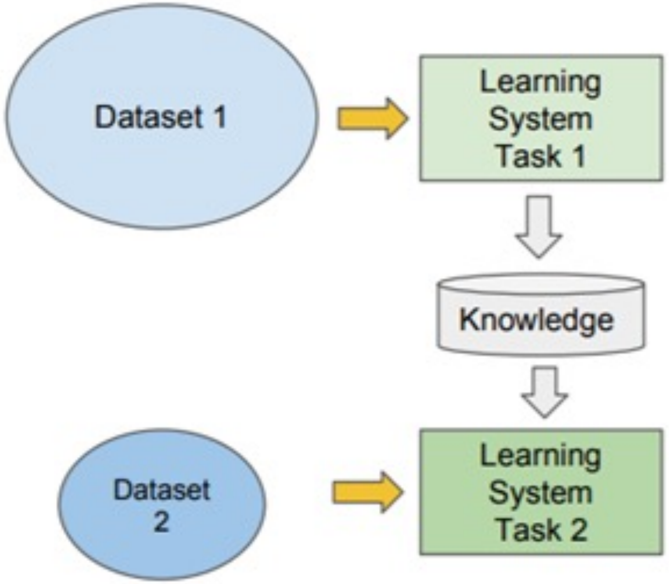
Traditional ML

- Isolated, single task learning:
  - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks



Transfer Learning

- Learning of a new tasks relies on the previous learned tasks:
  - Learning process can be faster, more accurate and/or need less training data



Freeze or fine-tune?

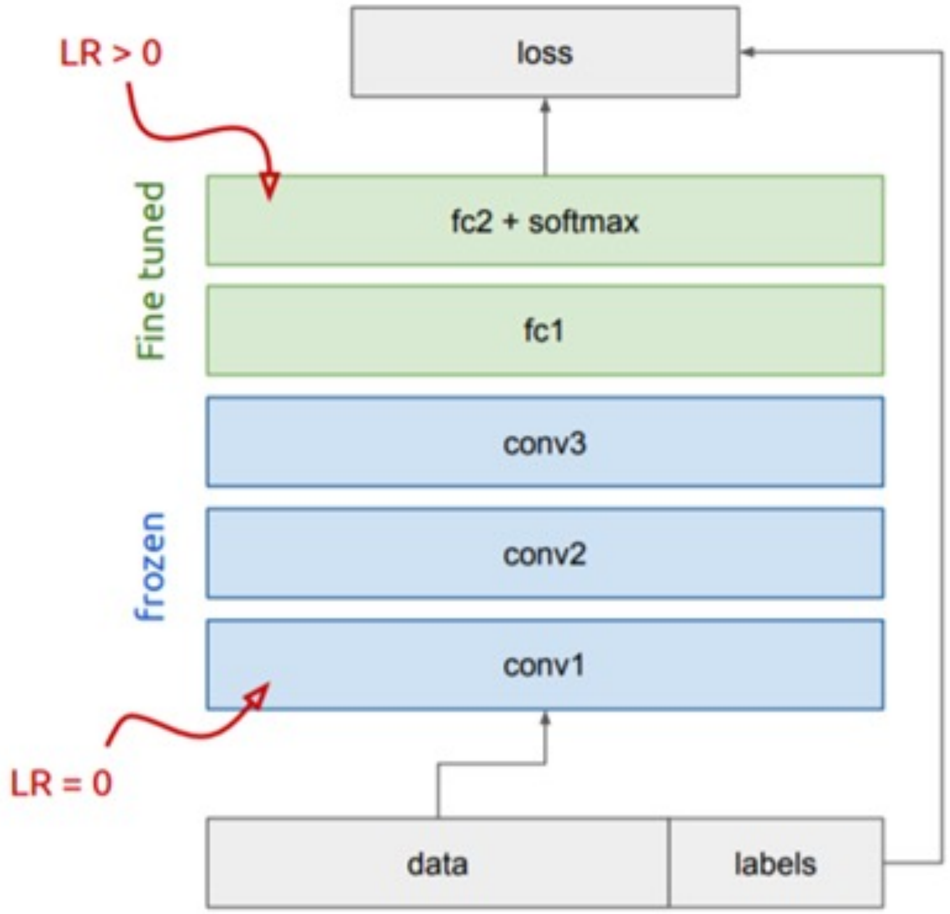
Bottom *n* layers can be frozen or fine tuned.

- **Frozen**: not updated during backprop
- **Fine-tuned**: updated during backprop

Which to do depends on target task:

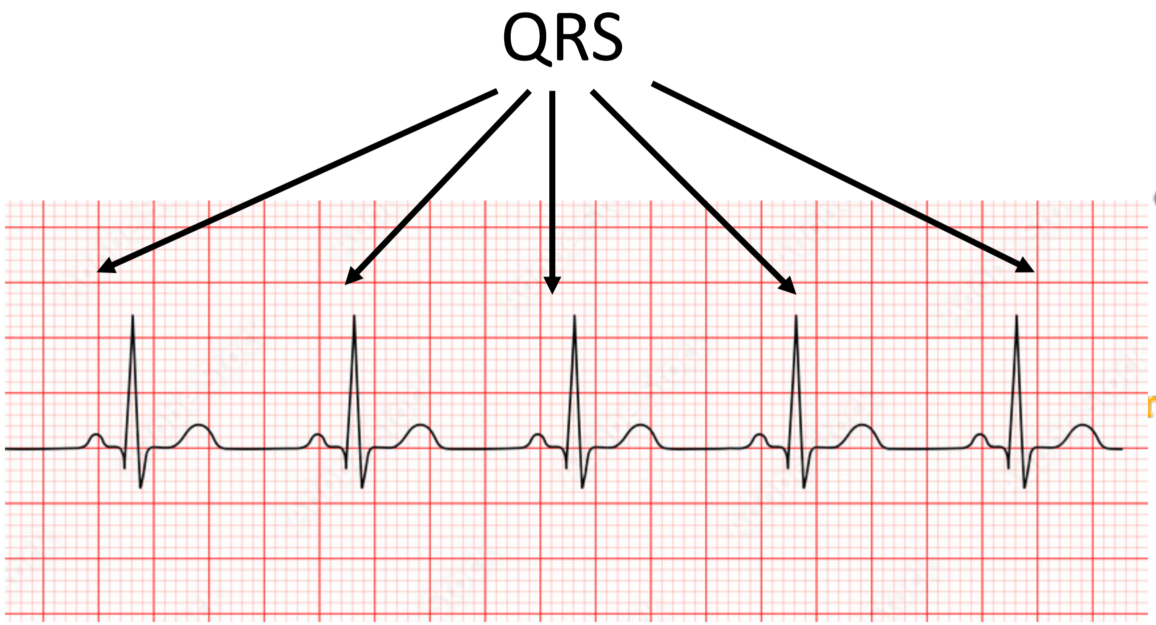
- **Freeze**: target task labels are scarce, and we want to avoid overfitting
- **Fine-tune**: target task labels are more plentiful

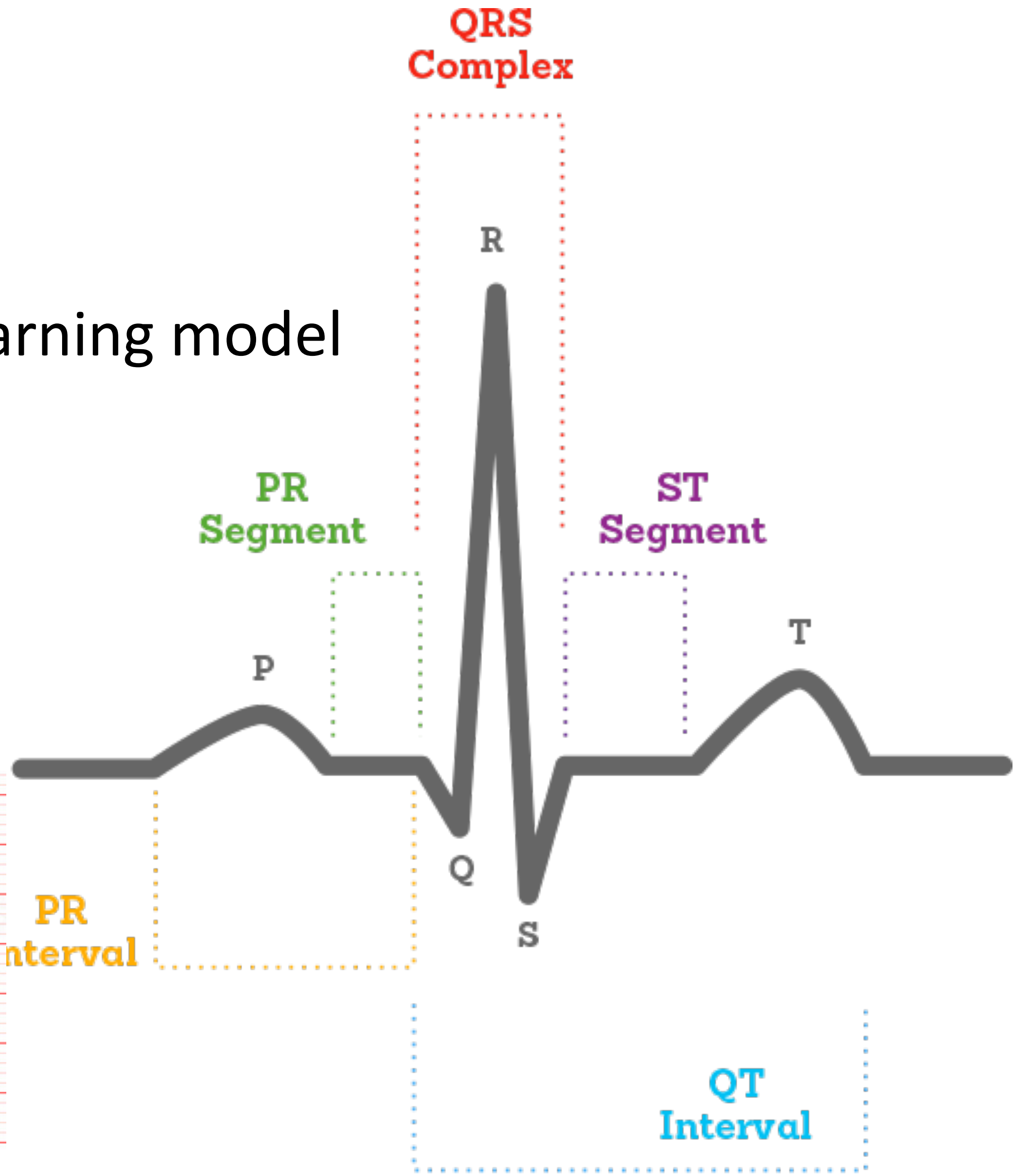
In general, we can set learning rates to be different for each layer to find a tradeoff between freezing and fine tuning



Task

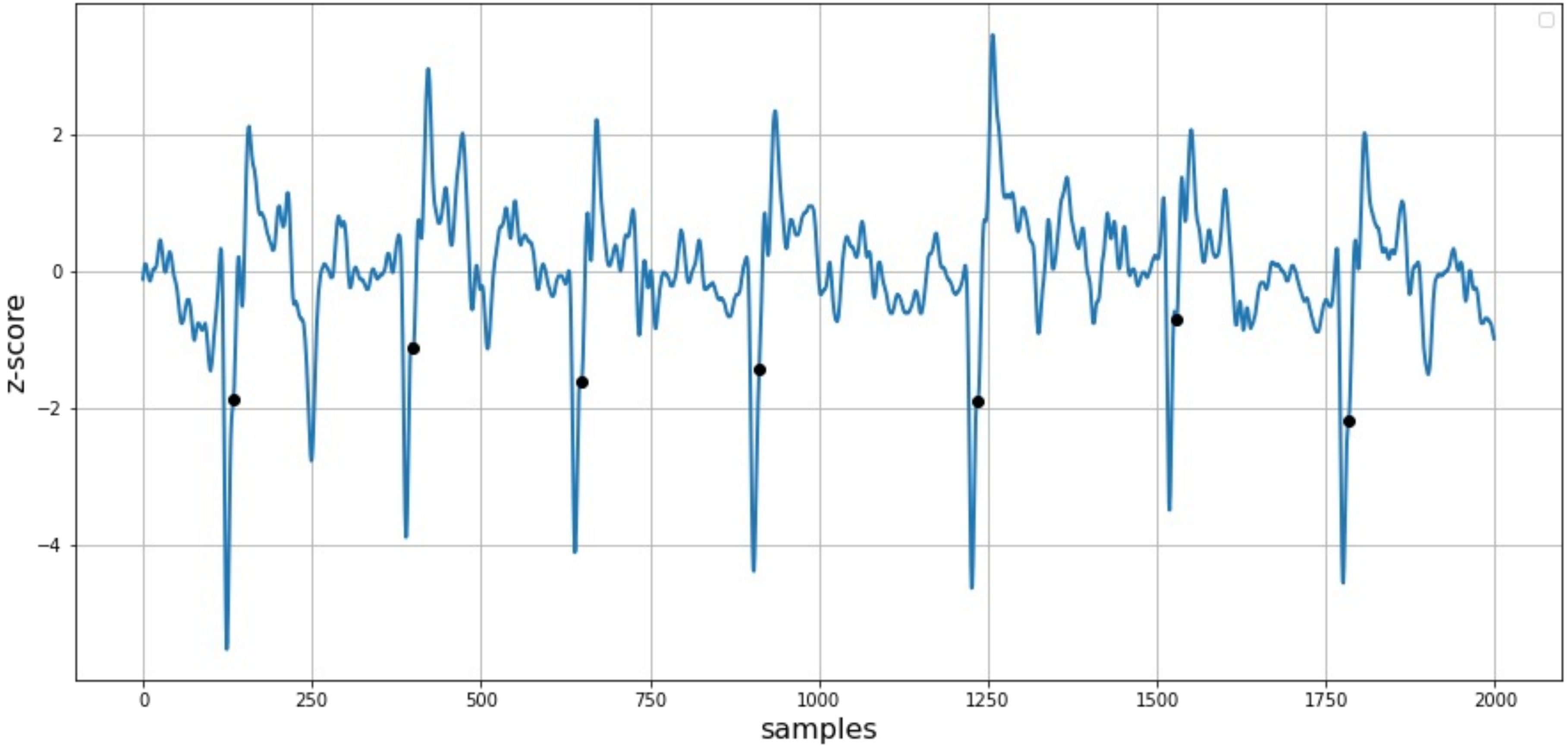
- Heart beat detection using deep learning model
- ECG (electrocardiam) recordings





Dataset

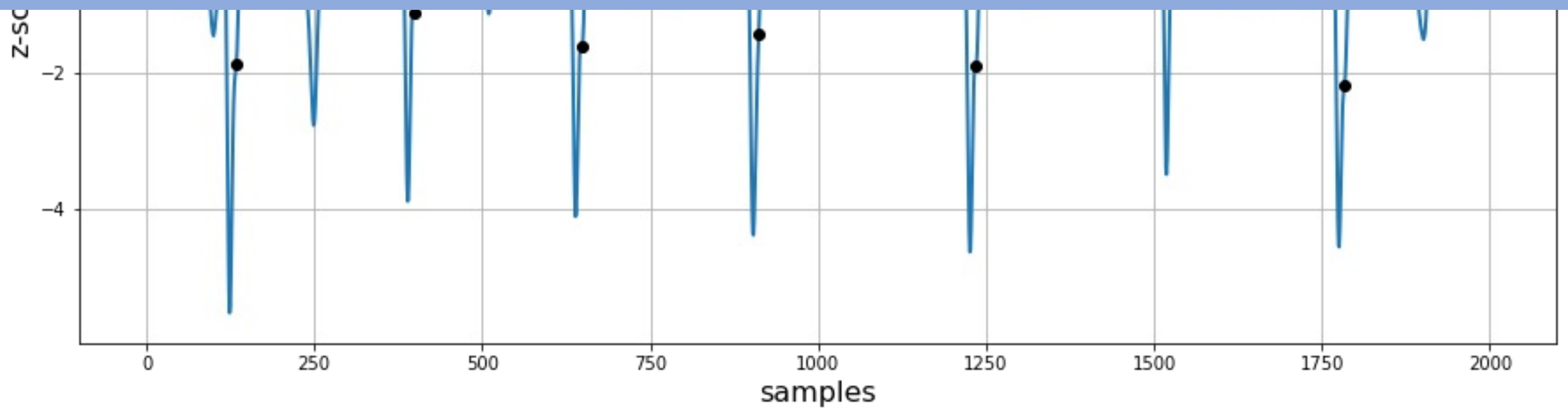
- 300 ECG recordings corrupted by MRI noise
- 250 Hz, 50s, 1 lead
- Manually prepared heart beat annotations



Dataset

- 300 ECG recordings corrupted by MRI noise
- 250 Hz, 50s, 1 lead
- Manually prepared heart beat annotations

- Is 300 records enough for the deep learning?



Old task- Dataset used for training

- UHF-ECG recordings
- Training- 2,250 records
- QRS marks: VDI vision
- 12 lead,
  5 kHz

Old task- Dataset used for training

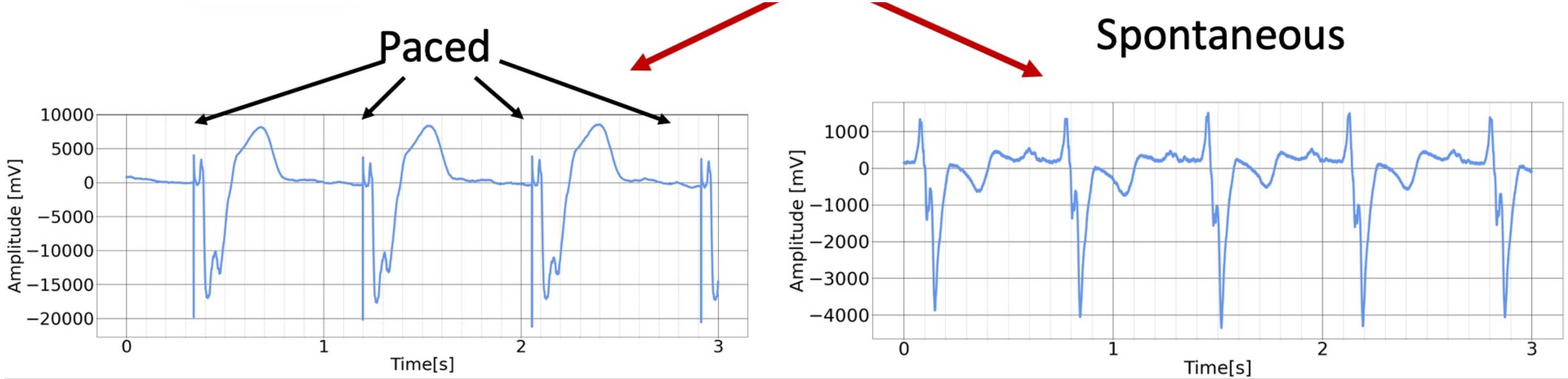
UHF-ECG recordings

Training- 2,250 records

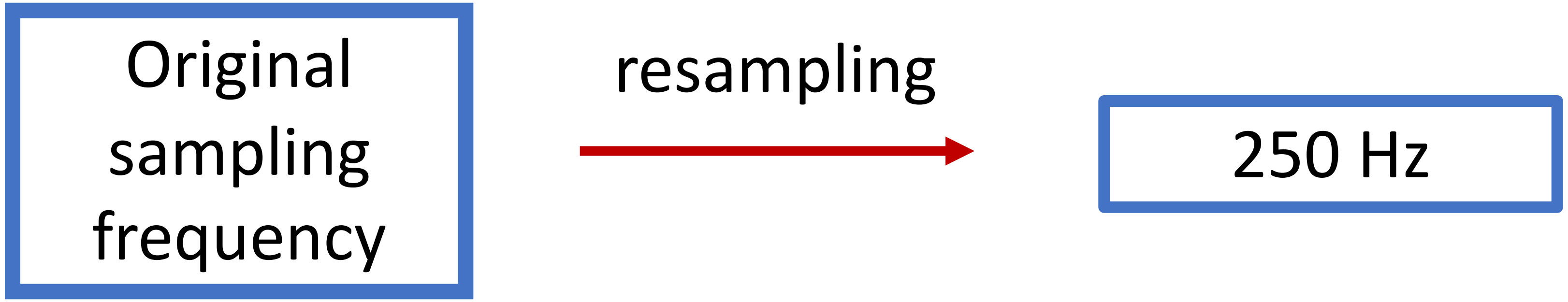
QRS marks: VDI vision

12 lead,

5 kHz



Preprocessing

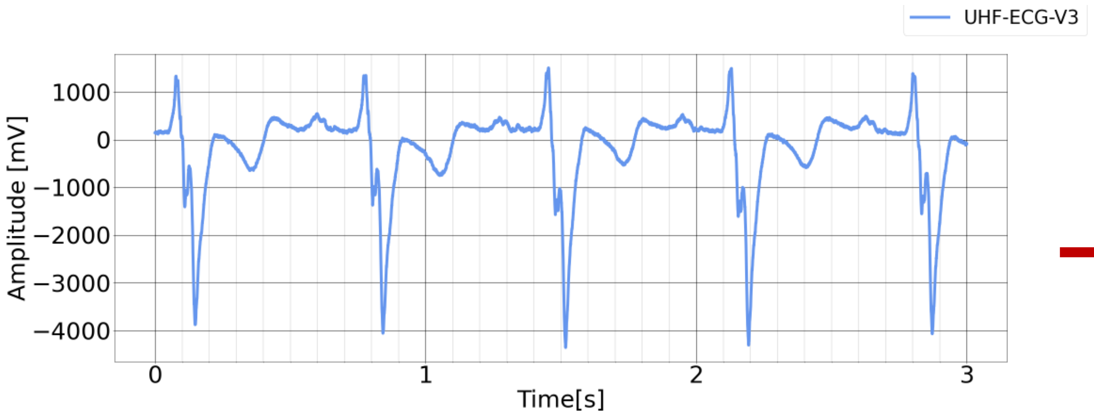


Preprocessing

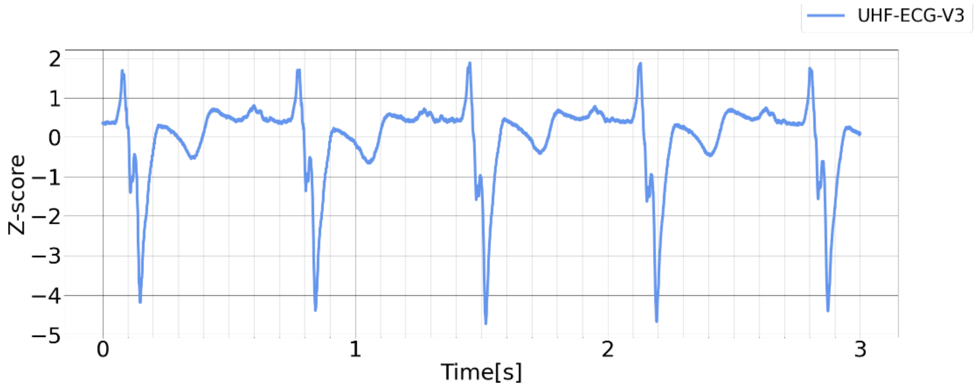
Original sampling frequency

resampling

250 Hz



Z - score

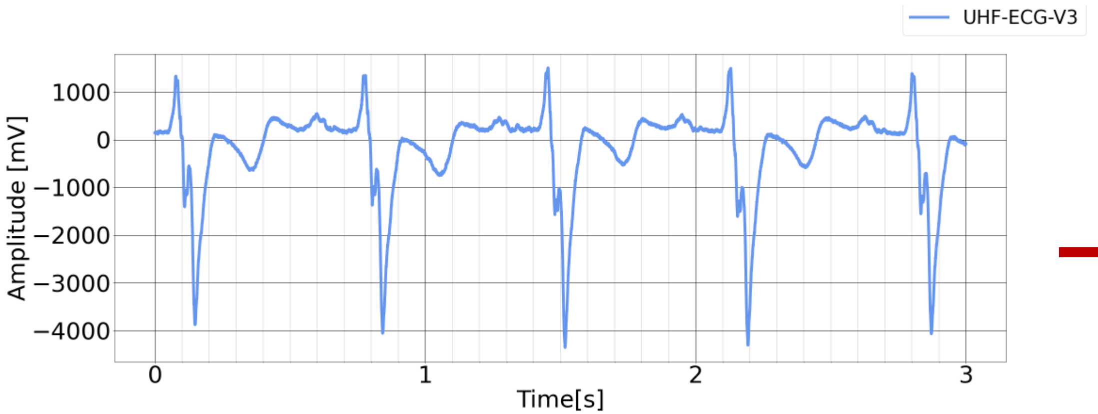


Preprocessing

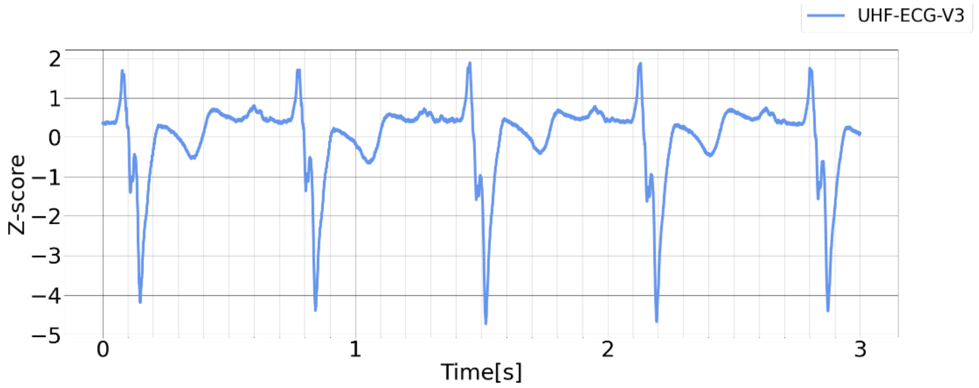
Original sampling frequency

resampling

250 Hz



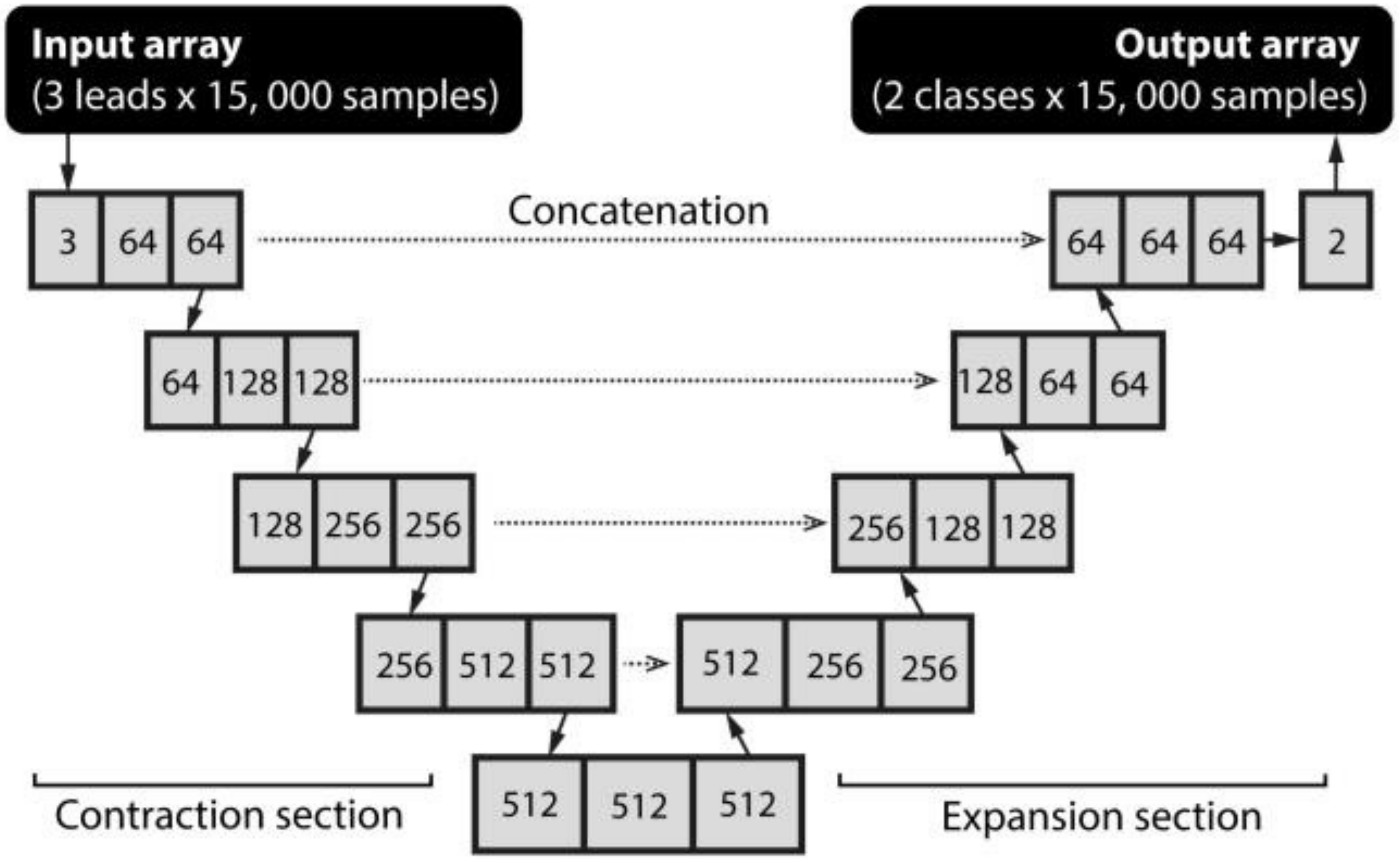
Z - score



Using only 1 lead

CNN architecture- UNet

- 1D CNN
- **kernel size=12, stride=6 or 5**



Method

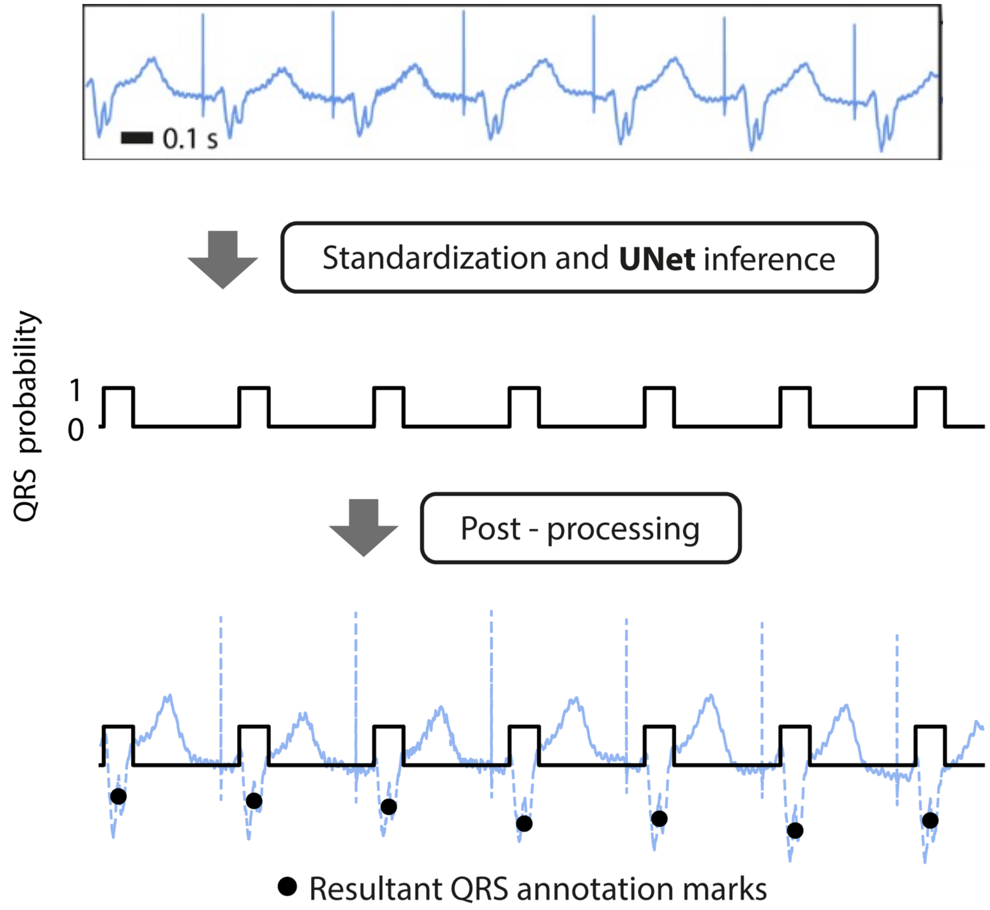
1 lead

3s x 250 Hz

Softmax,

QRS vs nonQRS probability

QRS position distance and probability criterion



Method

1 lead

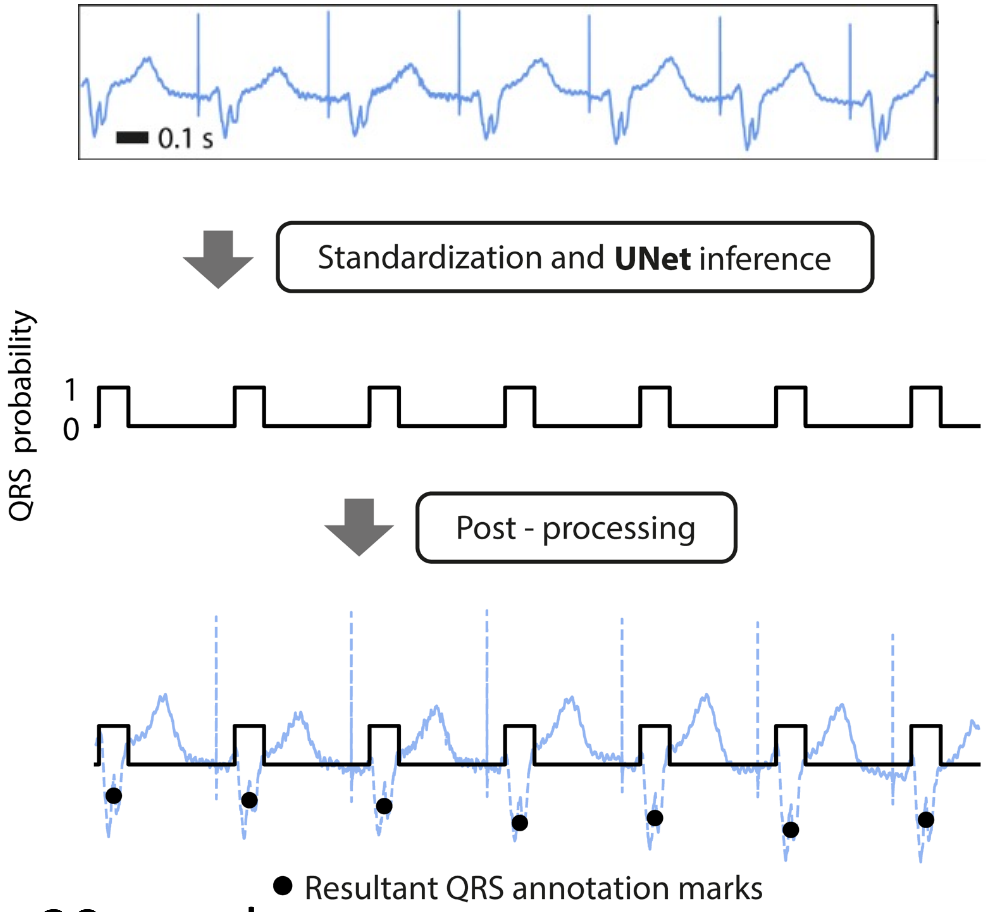
3s x 5,000 Hz

Softmax,

QRS vs nonQRS probability

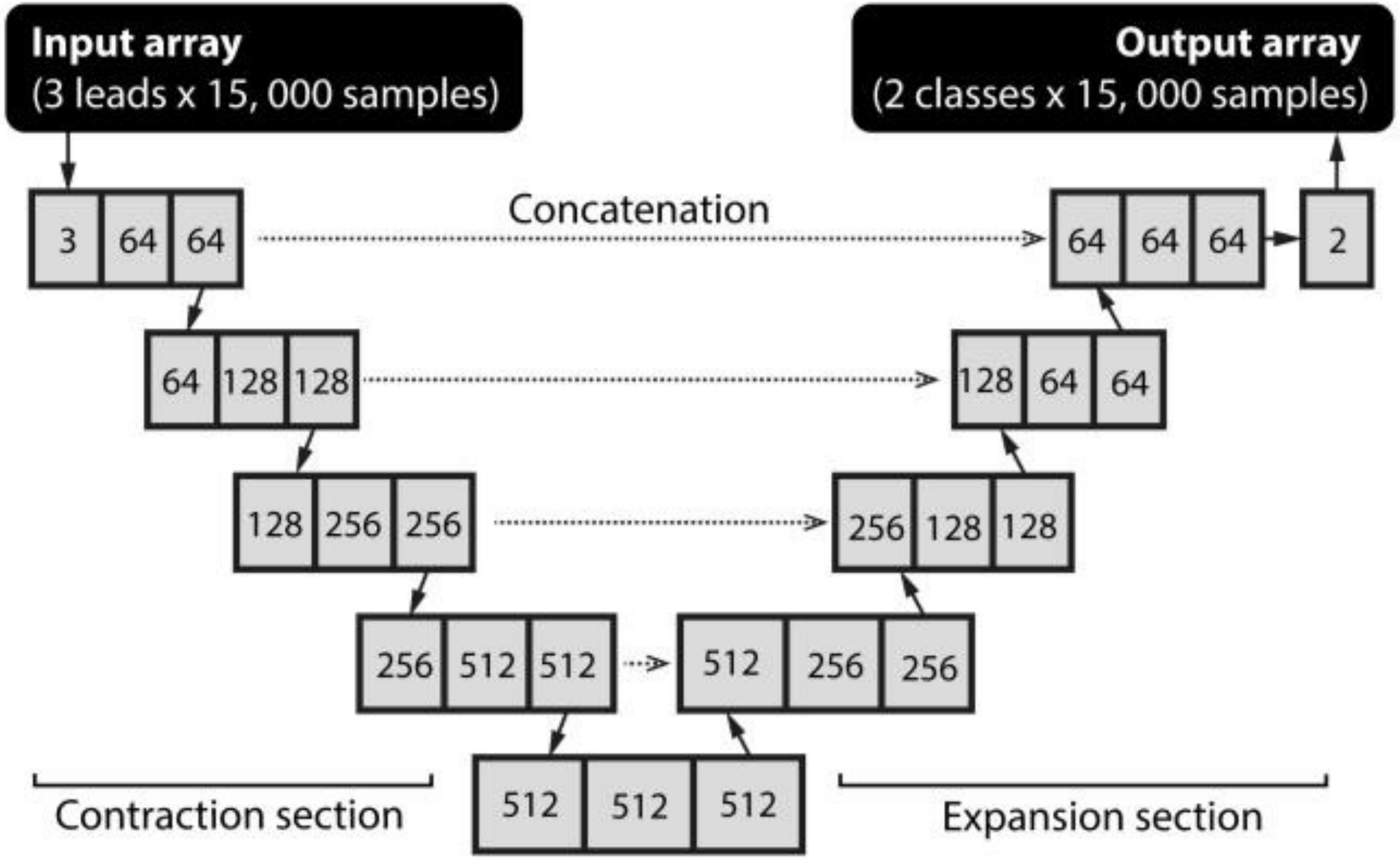
2 x 15,000

QRS position Distance and probability criterion



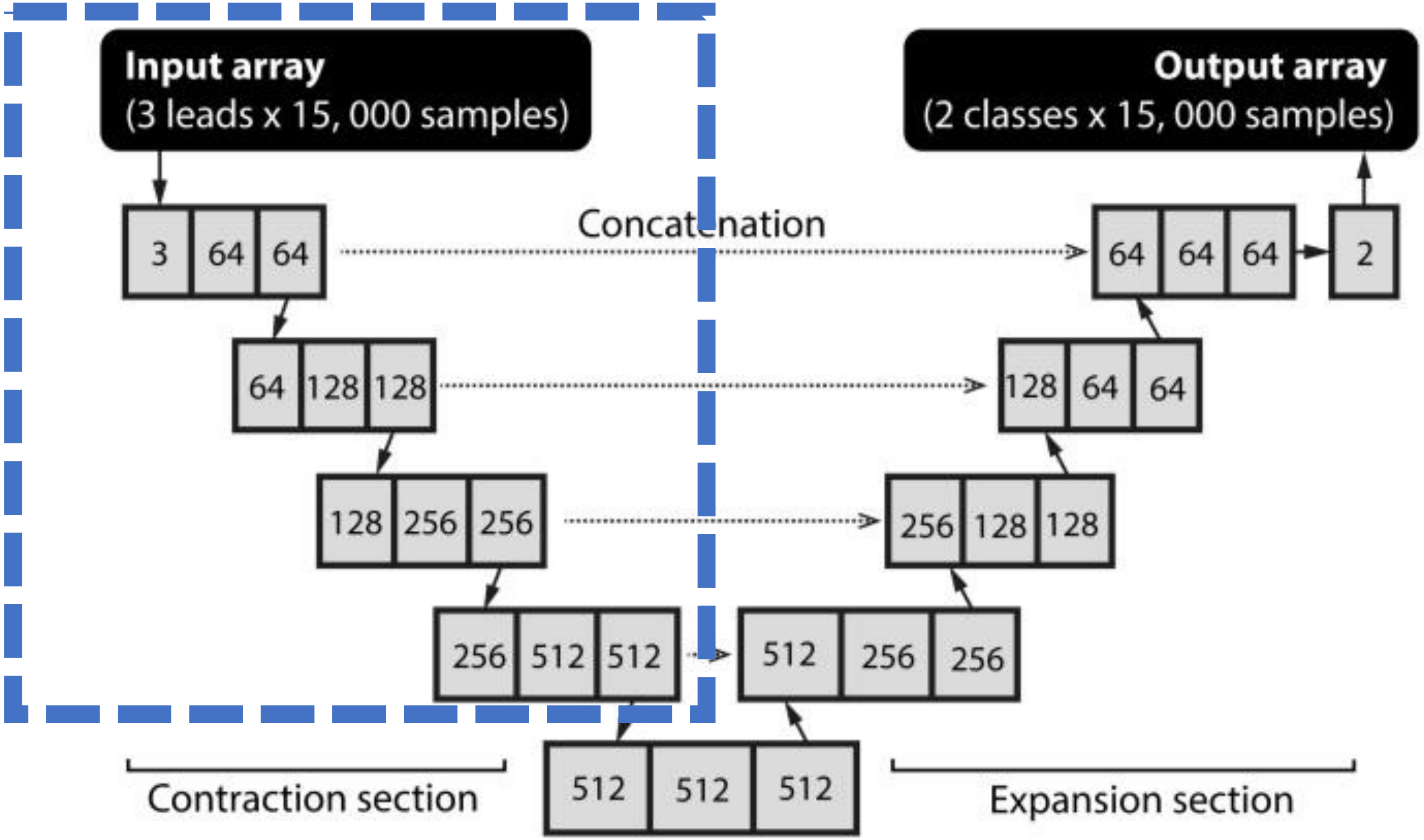
- Training 30 epochs
- LR 0.001
- Adam optimization
- Weighted cross enthropy

Transfer of knowledge



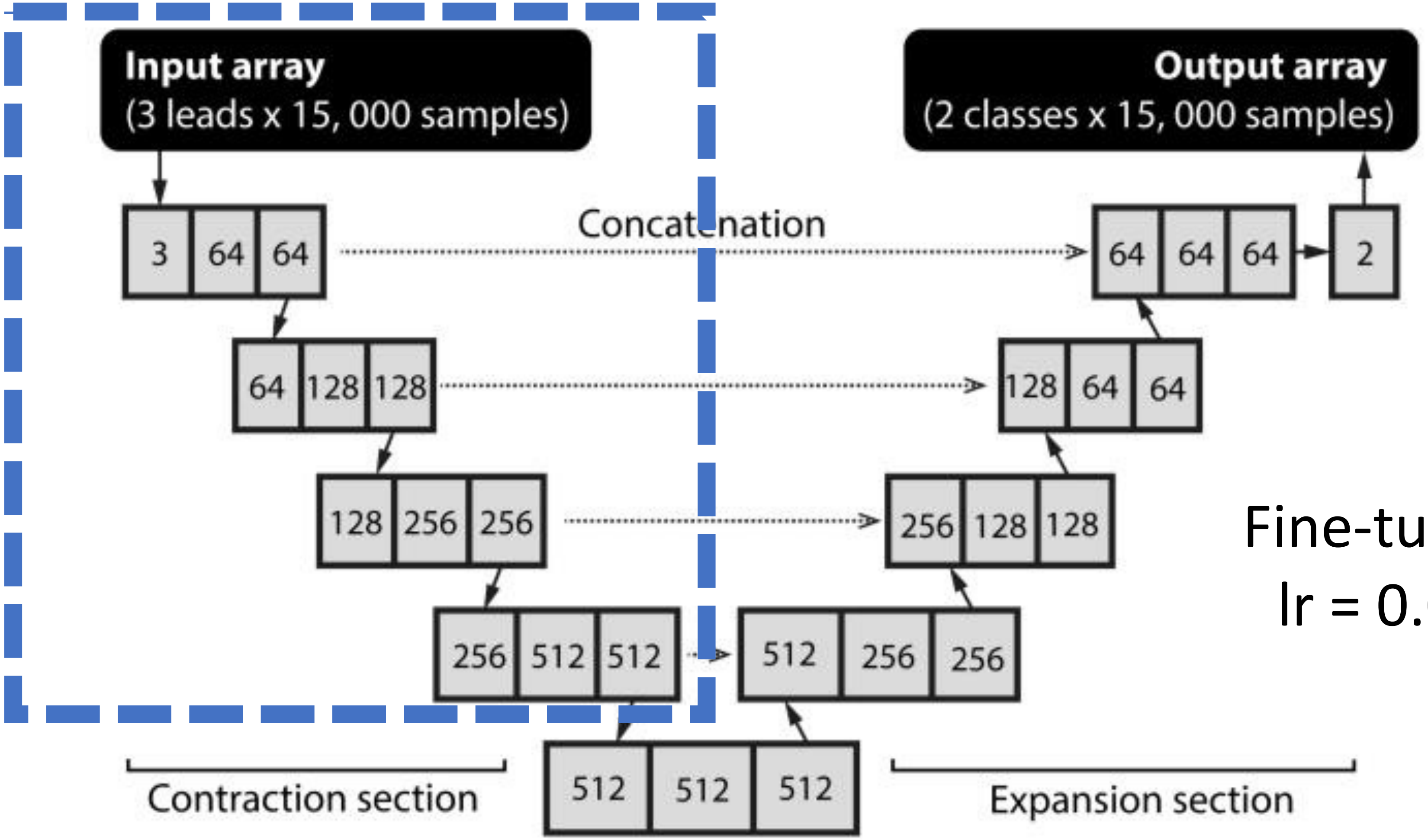
Transfer of knowledge

Frozen →
lr = 0



Transfer of knowledge

Frozen →
lr = 0



Fine-tuned→
lr = 0.0001

Results

- Division of 300 signals – 250 fine-tuning, 50 validation
- Test → 5 manually annotated records – 8 min long

Results

- Division of 300 signals – 250 fine-tuning, 50 validation
- Test → 5 manually annotated records – 8 min long

Original model (before transfer learning)			
	Sensitivity [%]	PPV [%]	F-score [%]
Validation	57.80	69.93	63.29
Test	63.40	75.36	68.86

Results

- Division of 300 signals – 250 fine-tuning, 50 validation
- Test → 5 manually annotated records – 8 min long

Original model (before transfer learning)			
	Sensitivity [%]	PPV [%]	F-score [%]
Validation	57.80	69.93	63.29
Test	63.40	75.36	68.86

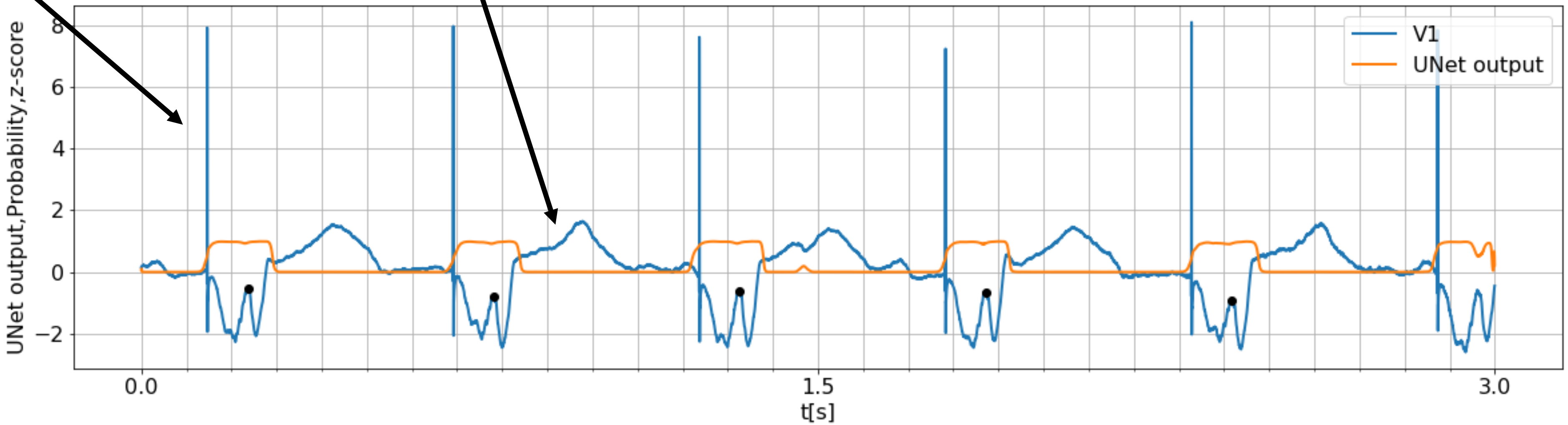
After transfer learning			
	Sensitivity [%]	PPV [%]	F-score [%]
Validation	97.11	98.24	97.67
Test	97.62	95.12	96.36

Examples

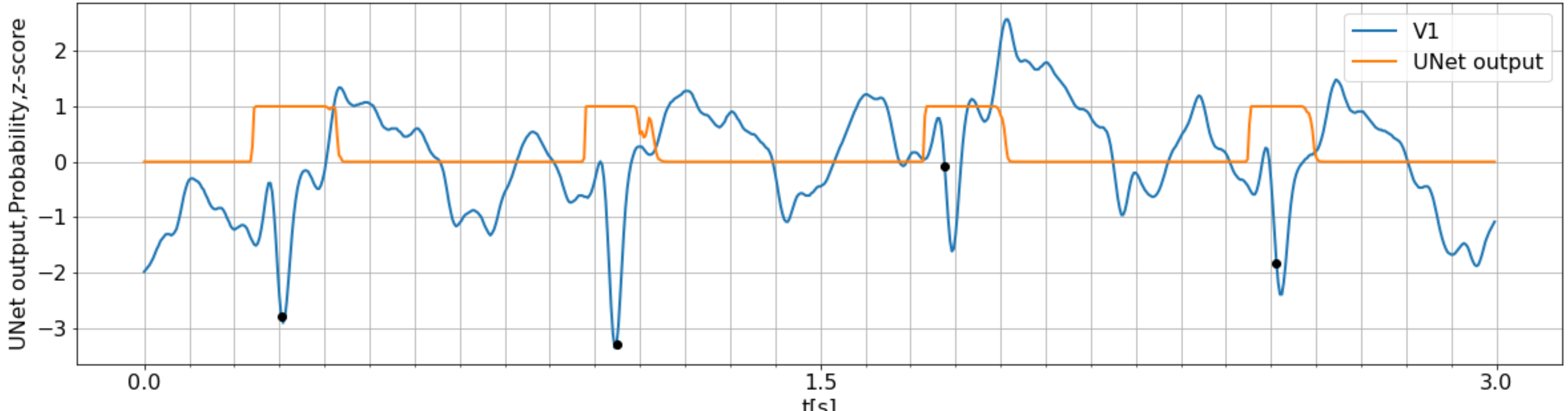
Pacing stimuli

T wave

Original model



Data used for transfer learning



Conclusion

- Ability of transfering the knowledge
- Less data needed (2250 training original model, 250 fine-tuning)
- Fine tuning is faster than training from scratch
- Working on a different type of data

Thank you for your attention