

# Multiframe neural networks in speech recognition

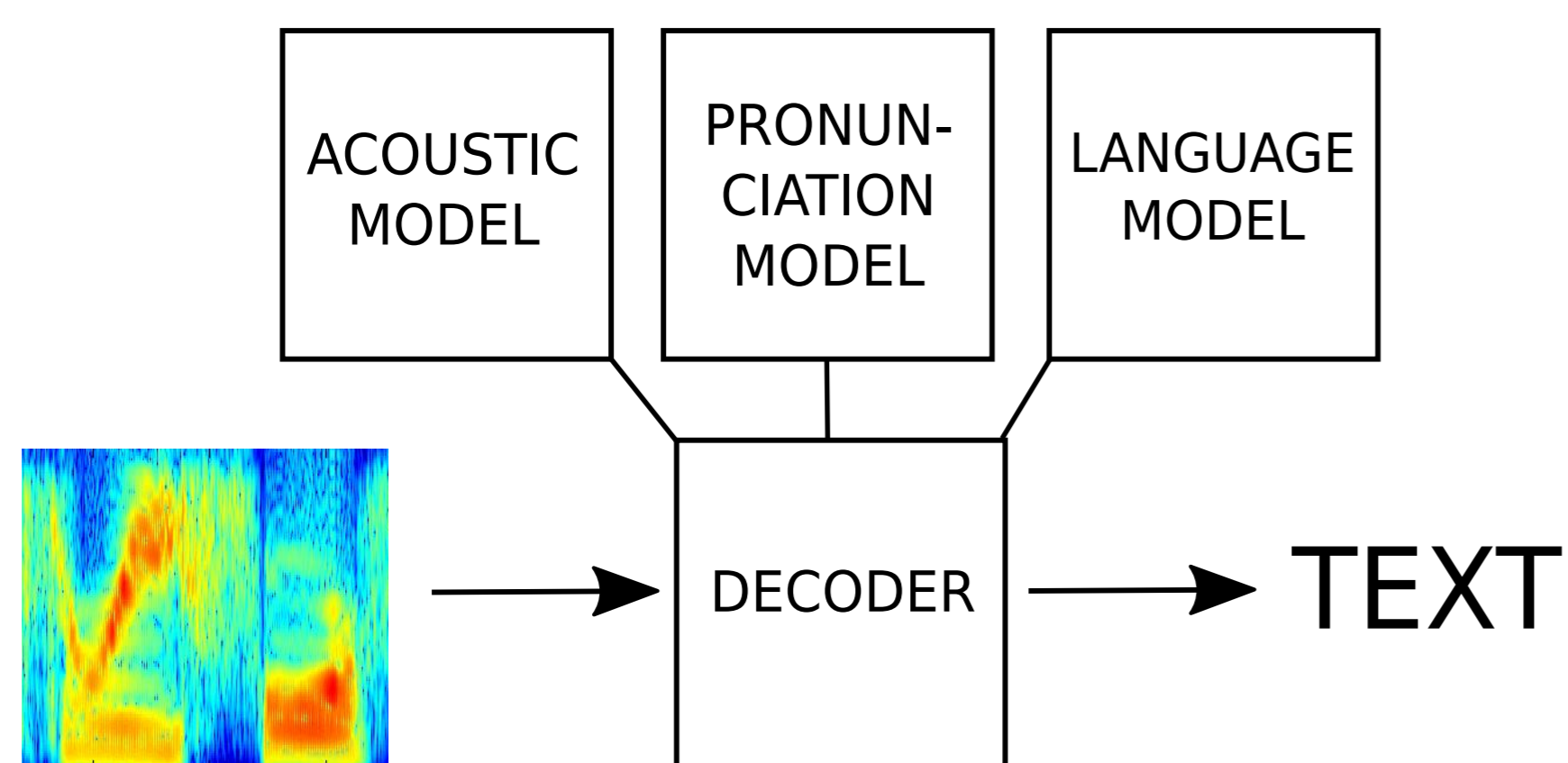
## Speech recognition

### WHAT IS THE GOAL?

Goal of speech recognition is to provide transcriptions for given speech signal.

### HOW IS IT DONE?

The speech recognition system uses knowledge about language and acoustics to choose the most probable sentence.



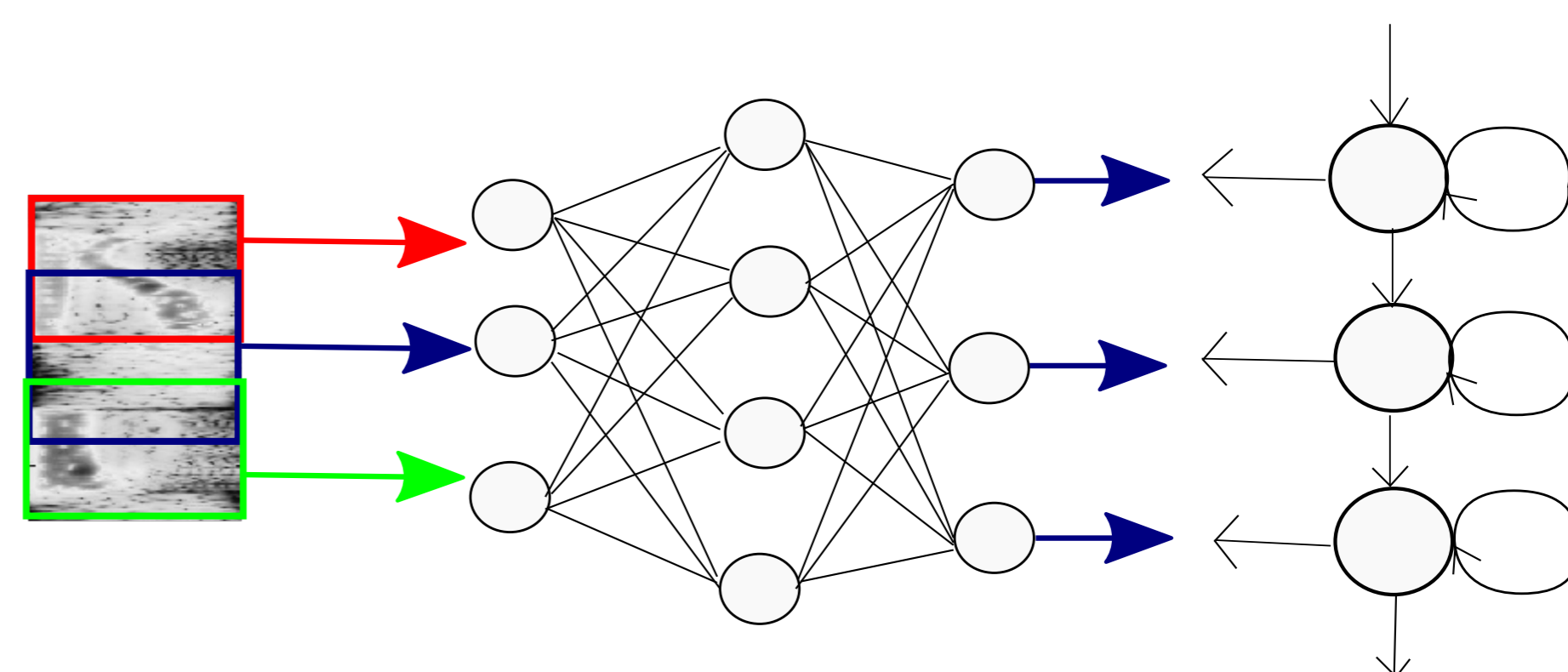
## Neural networks

### WHAT IS NEURAL NETWORK?

Neural network is mathematical model representing mapping from input to output. It can be trained to compute arbitrary function.

### HOW IS IT USED IN SPEECH RECOGNITION?

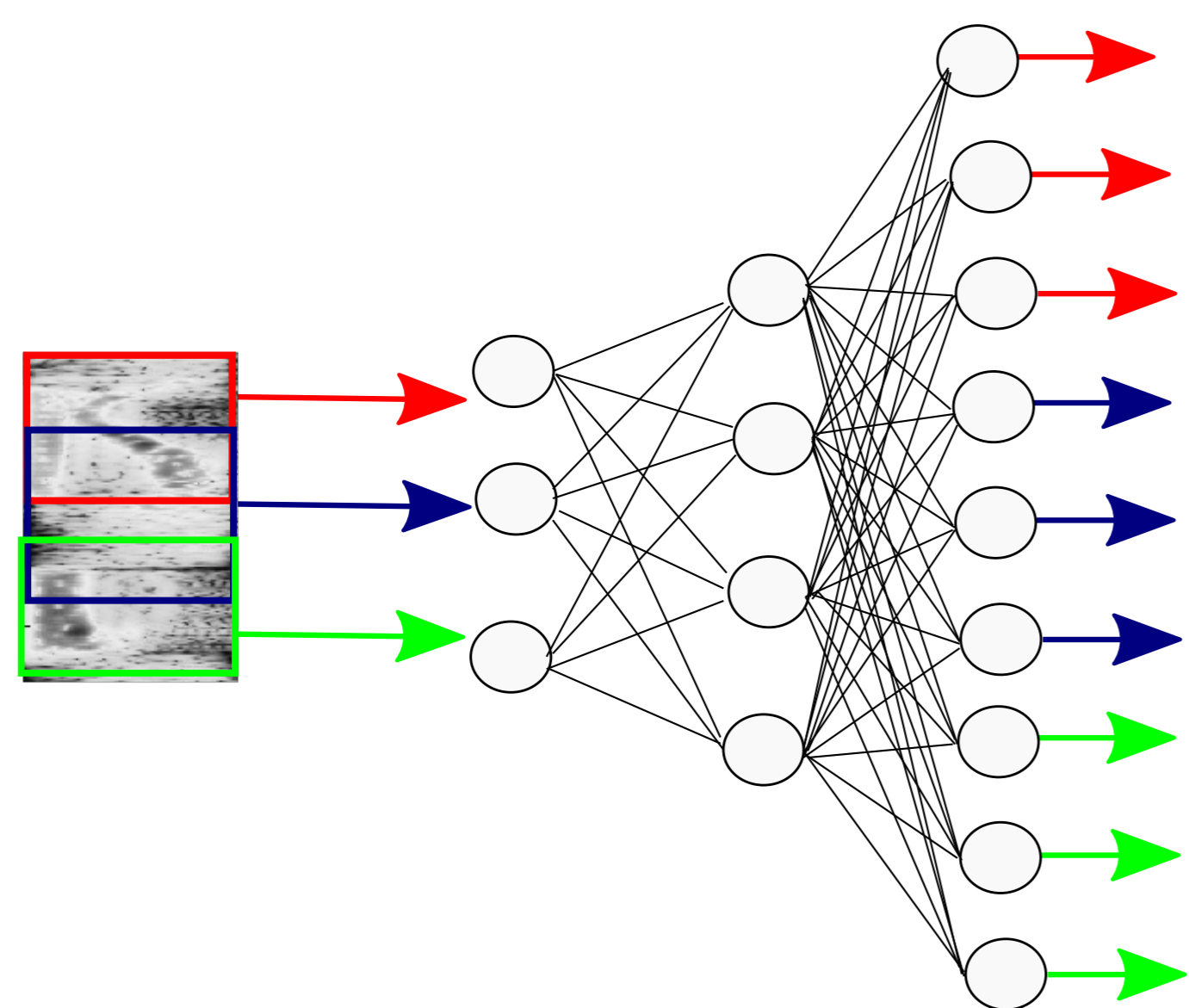
Neural network is part of acoustic model. It classifies speech frames into classes representing states of Hidden Markov Model.



## Multiframe approach

### WHAT WAS CHANGED?

The network typically has multiple speech frames on its input but predicts HMM state only for the central one. Here the output of the network is extended to predict labels for multiple frames.



## Results

### WHAT WE EXPERIMENTED WITH?

We tested the new approach on Wall Street Journal dataset experimenting with different sizes of input and output context sizes.

### WHAT WERE THE RESULTS?

Results of the experiments showed 7% and 12% relative improvement for input size 5 and output size 9.

IN	OUT	WER dev93 [%]	WER eval92 [%]
15	1	8.39	4.86
3	7	7.98	4.52
<b>5</b>	<b>9</b>	<b>7.77</b>	<b>4.24</b>
7	11	7.87	4.45
9	13	7.98	4.45
11	15	7.97	4.54
13	17	8.29	4.68
15	15	8.20	4.73