Learn Features to Classify Coins from (Ultra-)Sound

Felix Meyer, Marvin Arnold, Remke Albrecht

Problem

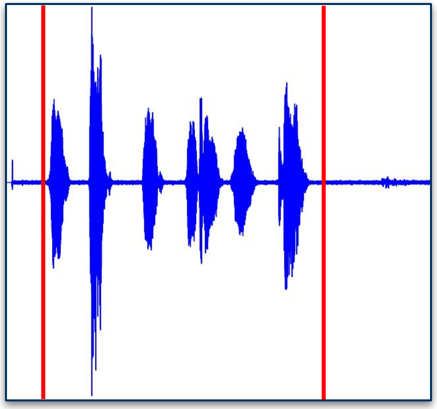
- long term goal: identify machine failures using sound
- problem: no prior knowledge of target label → unsupervised approach

- preparatory work: identify different types of coins using ultrasound
- coin types: 1ct, 2ct, 5ct, 20ct, 50ct, 1€, 2€

- traditional method: identify characteristic frequencies, check for these on new examples

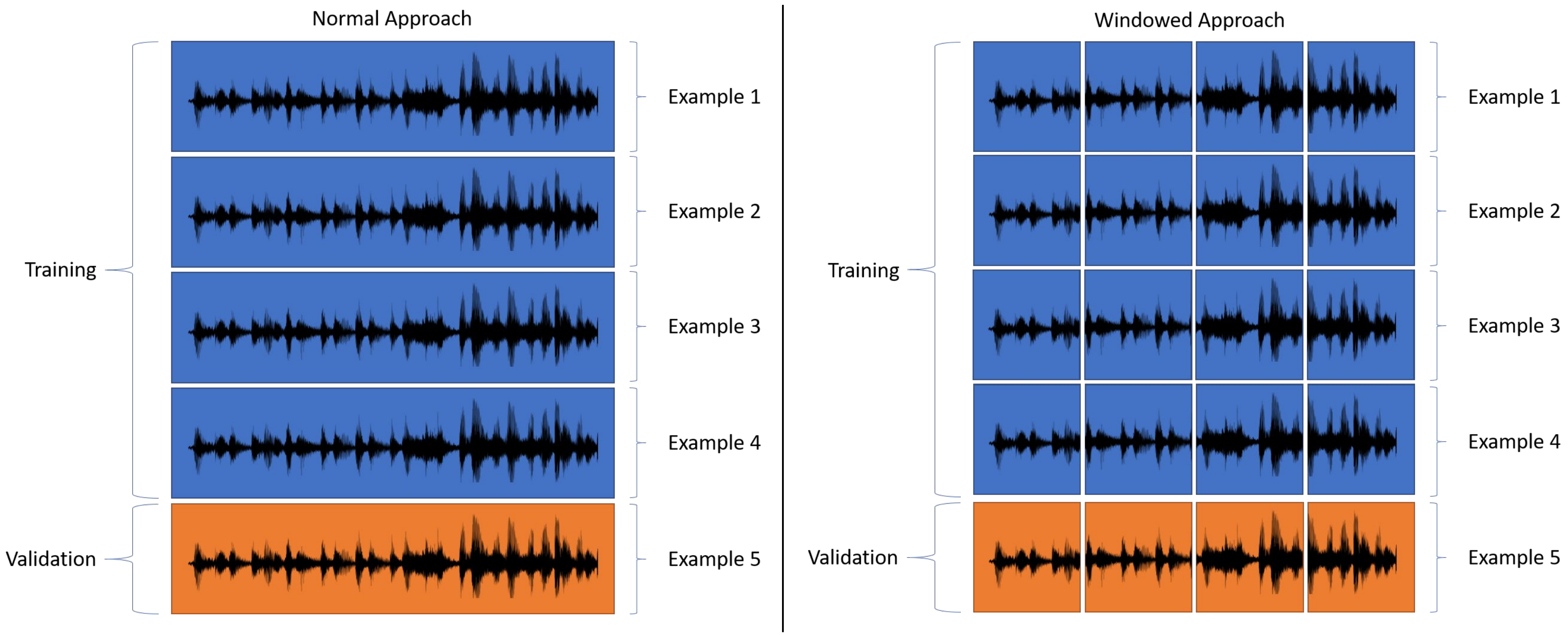
- our method: use machine learning to let the model learn the characteristic frequencies by itself (from the raw signal)

Data Preparation



- trim audio signal to cut away silence (dB threshold)
- only use every n-th value for performance
- normalize signal to value range [-1, 1]
- sample same amount of examples from each class
- random batch shuffling

Data Preparation - Unwindowed/Windowed

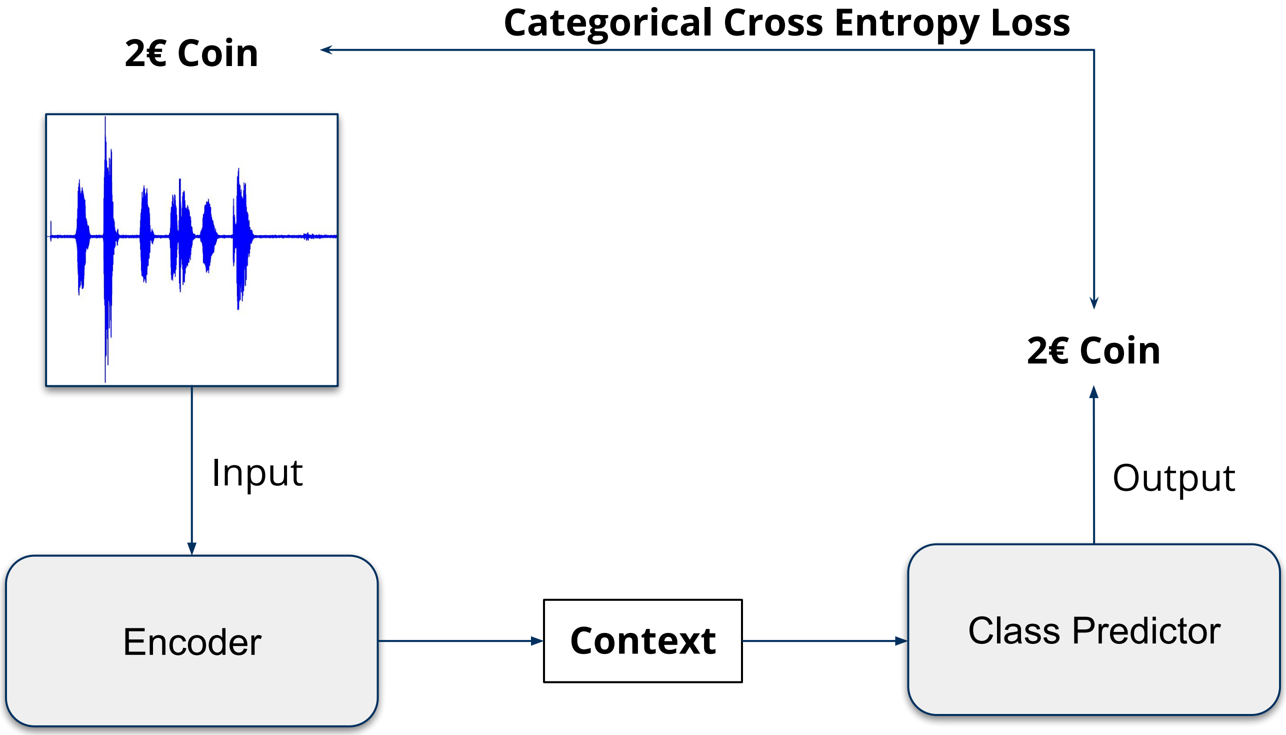


Architectures

Comparison of 2 network architectures:

RNN based	CNN based
Encoder-Decoder model (semi-supervised) Trivial model (supervised)	
Windowed/Unwindowed data	Windowed data

Architectures - Trivial Models (supervised)



Categorical Cross Entropy Loss

2€ Coin

Input

Encoder

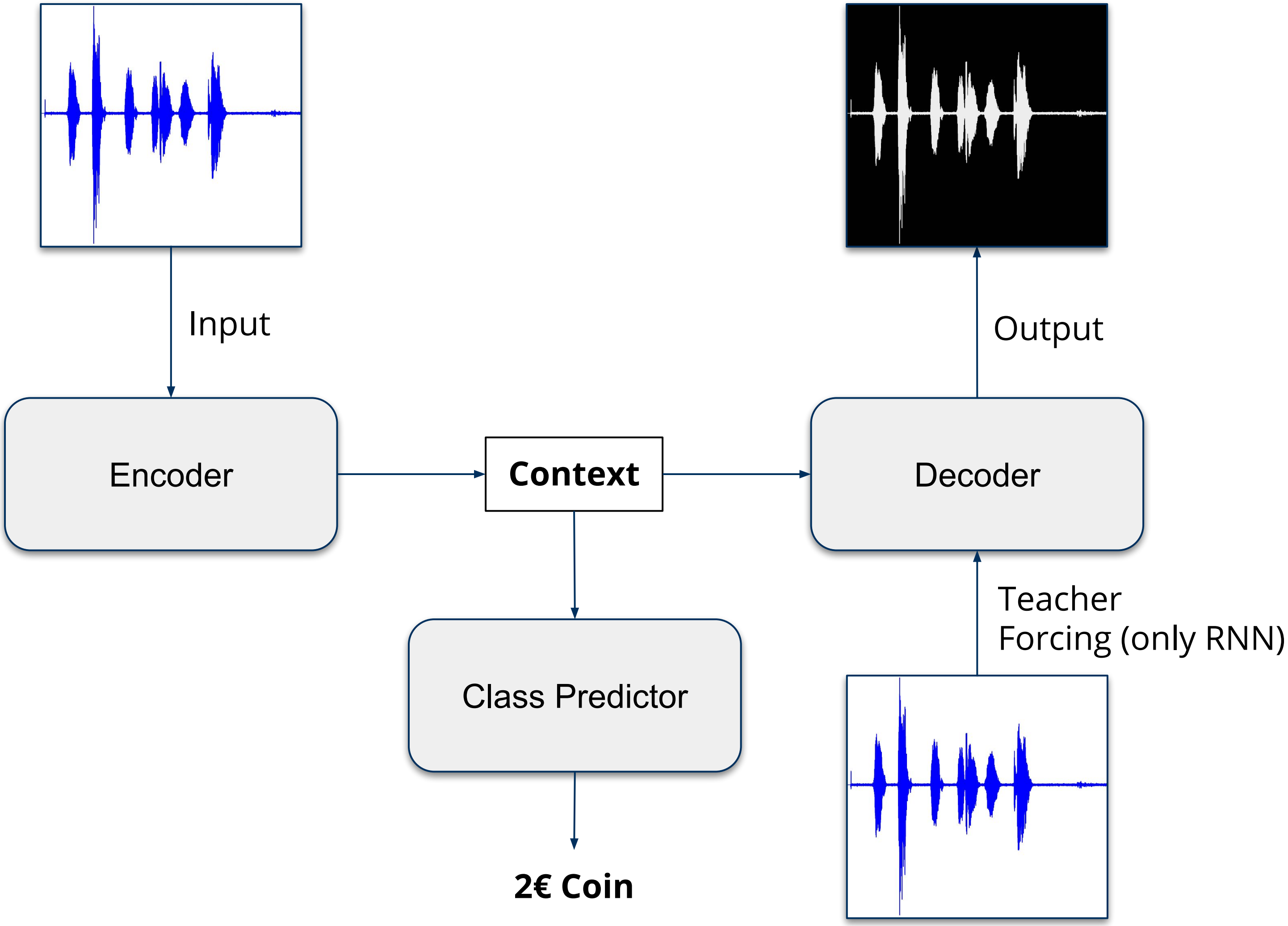
Context

Class Predictor

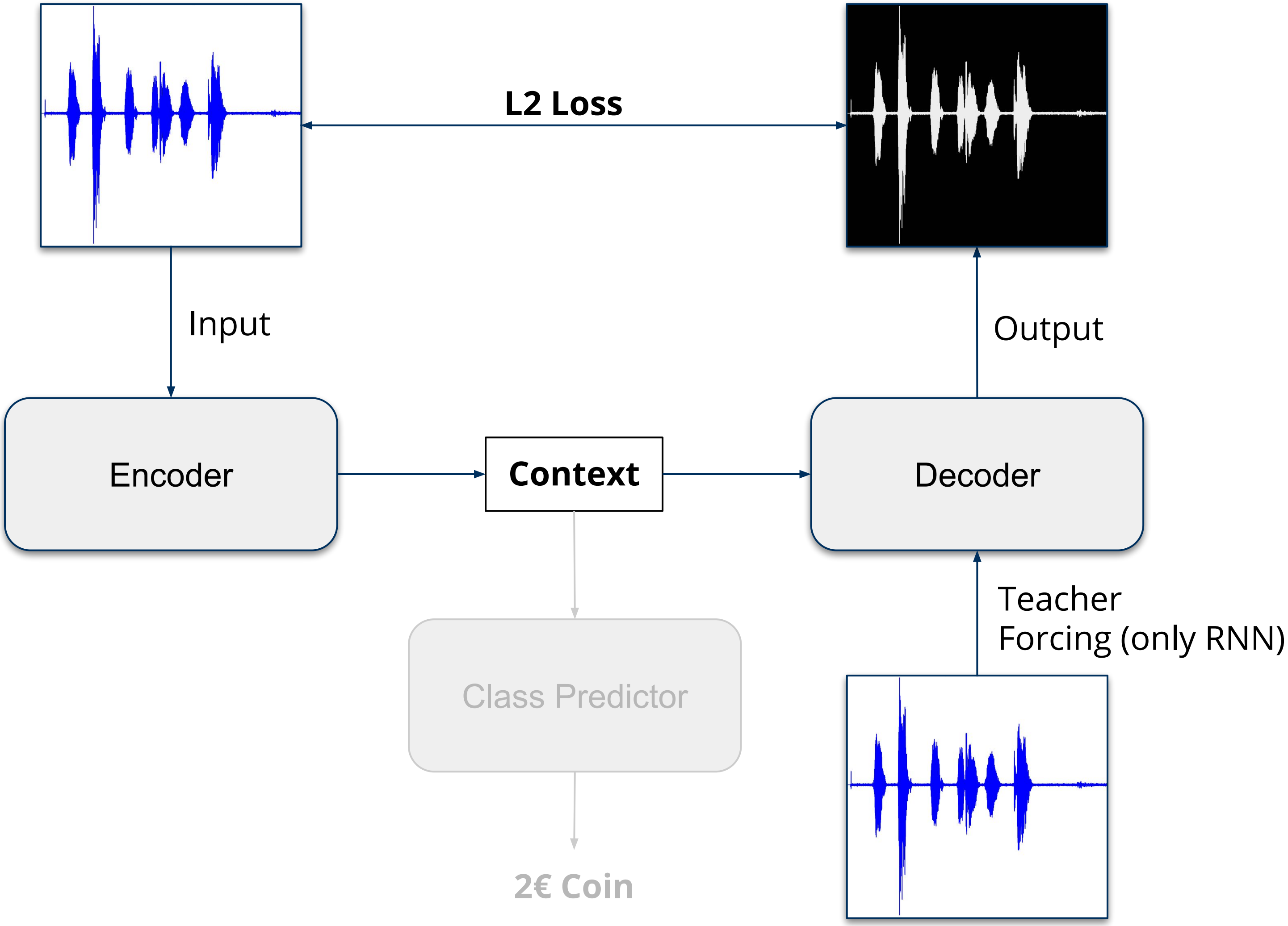
Output

2€ Coin

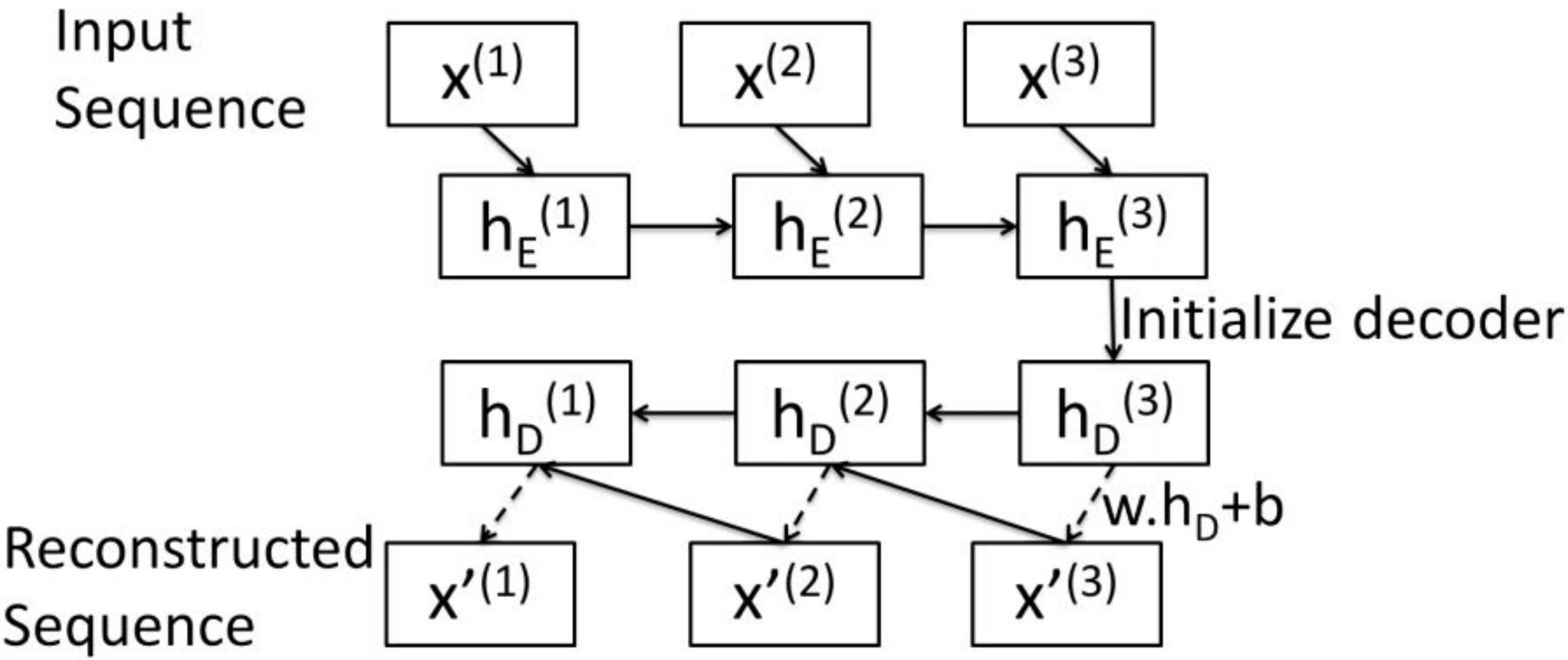
Architectures - Encoder-Decoder



Training - Step 1

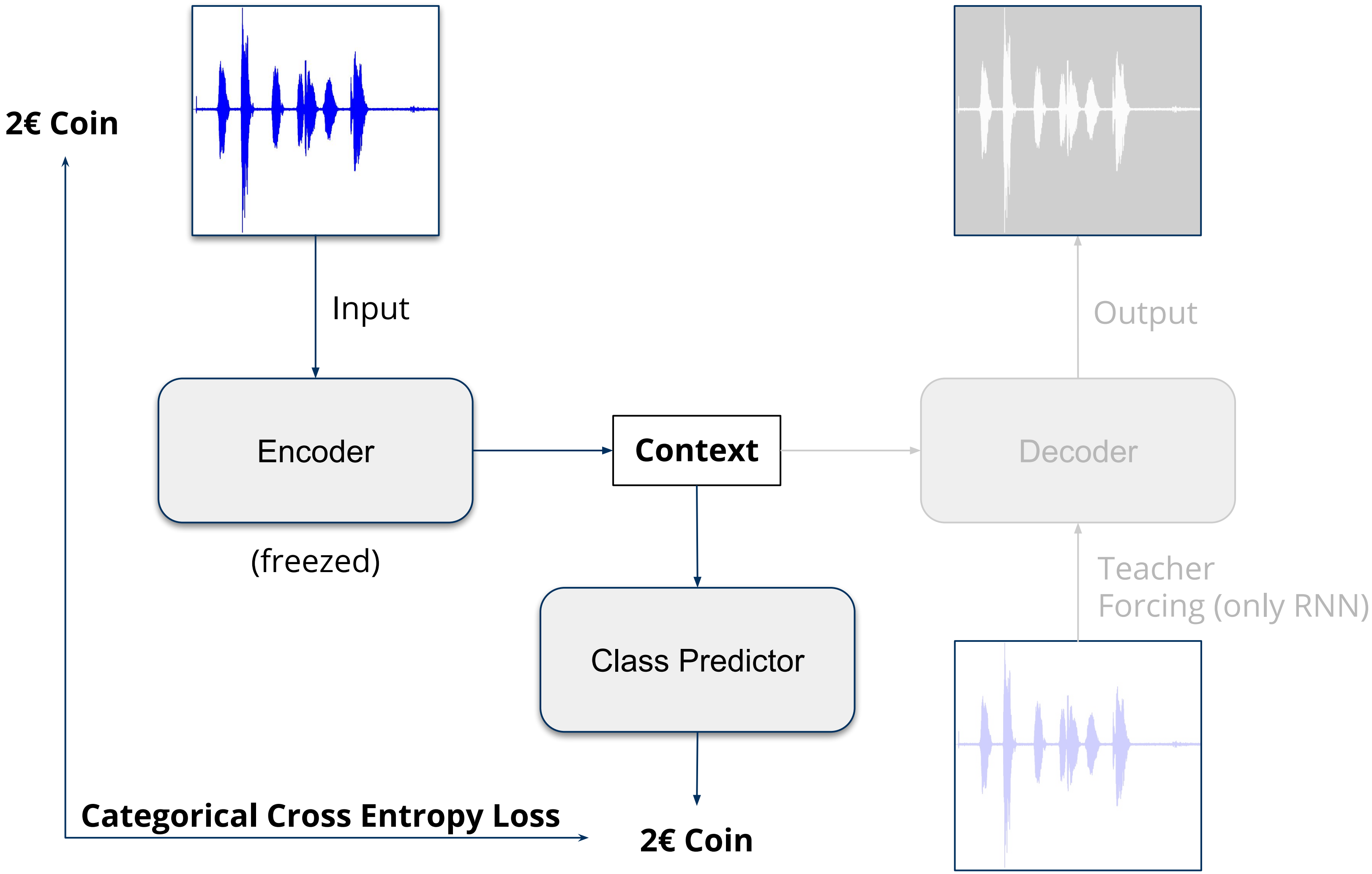


Training - Step 1 - RNN Encoder-Decoder

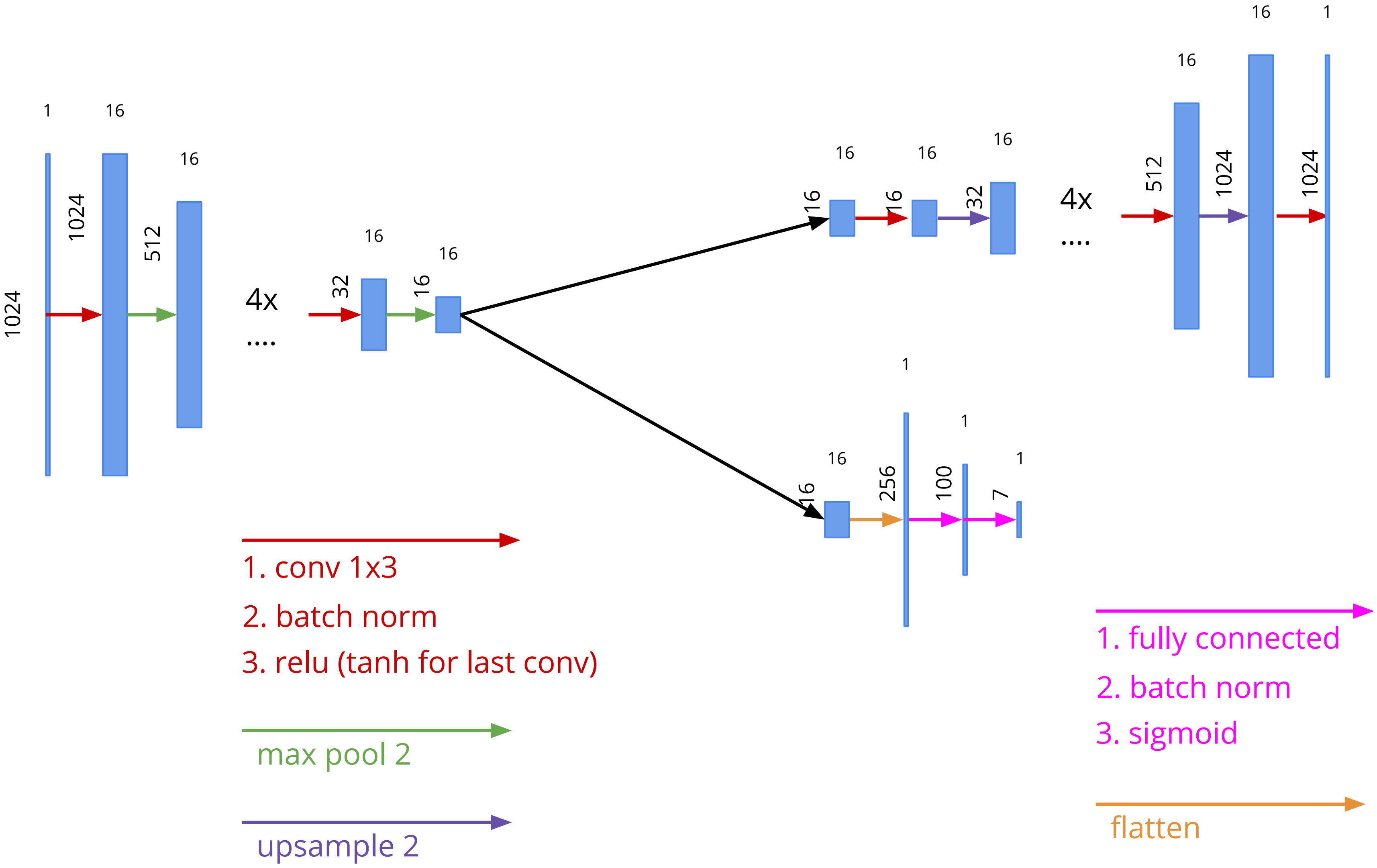


Malhotra, Pankaj, et al. "LSTM-based encoder-decoder for multi-sensor anomaly detection."

Training - Step 2



Architectures - CNN



1. conv 1x3
2. batch norm
3. relu (tanh for last conv)

max pool 2

upsample 2

1. fully connected
2. batch norm
3. sigmoid

flatten

Evaluation

TSNE (t-stochastic neighbour embedding) plot:

- *non-linear* dimensionality reduction similar to PCA
- same coins cluster together
- created on all available data (training + validation set)
- use of encoded input sequence (context)

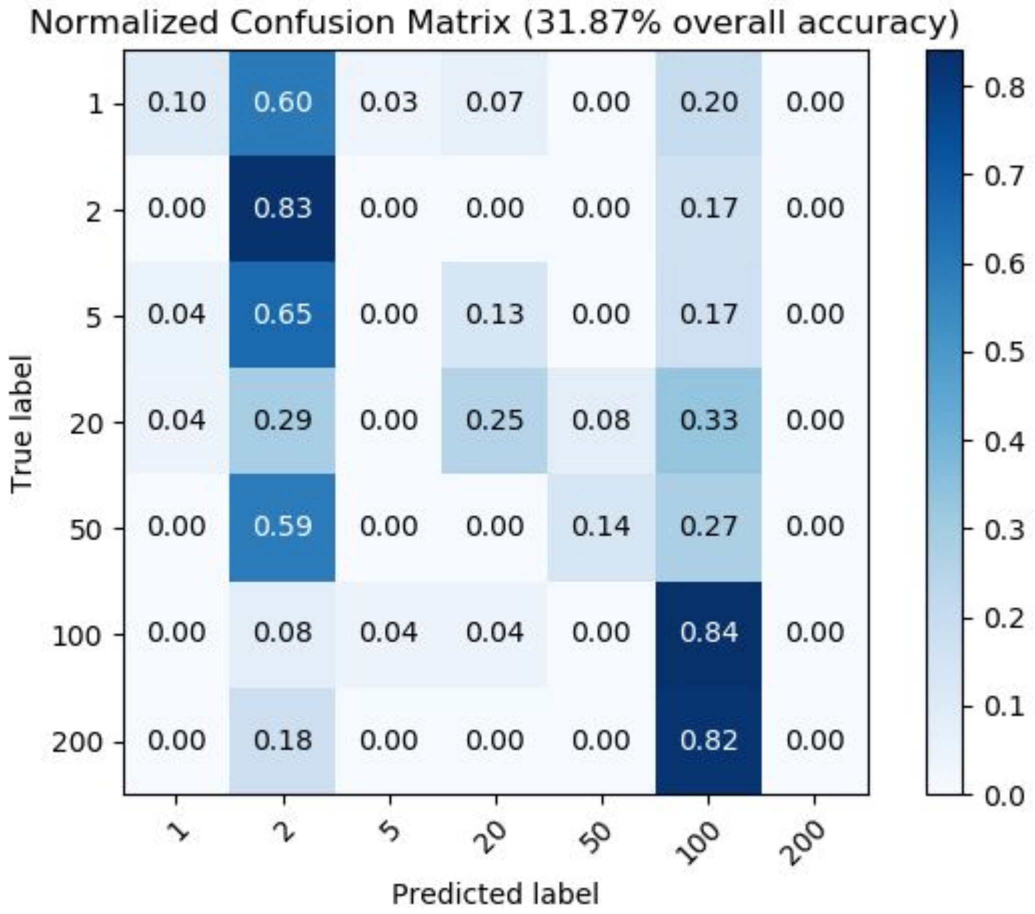
Confusion matrix:

- visualize performance on validation set
- due to data limitations no test set

Box plot of highest validation accuracy:

- compare peak performance of all architectures
- each architecture sampled 20 times

LSTM Enc-Dec Model *semi-supervised*

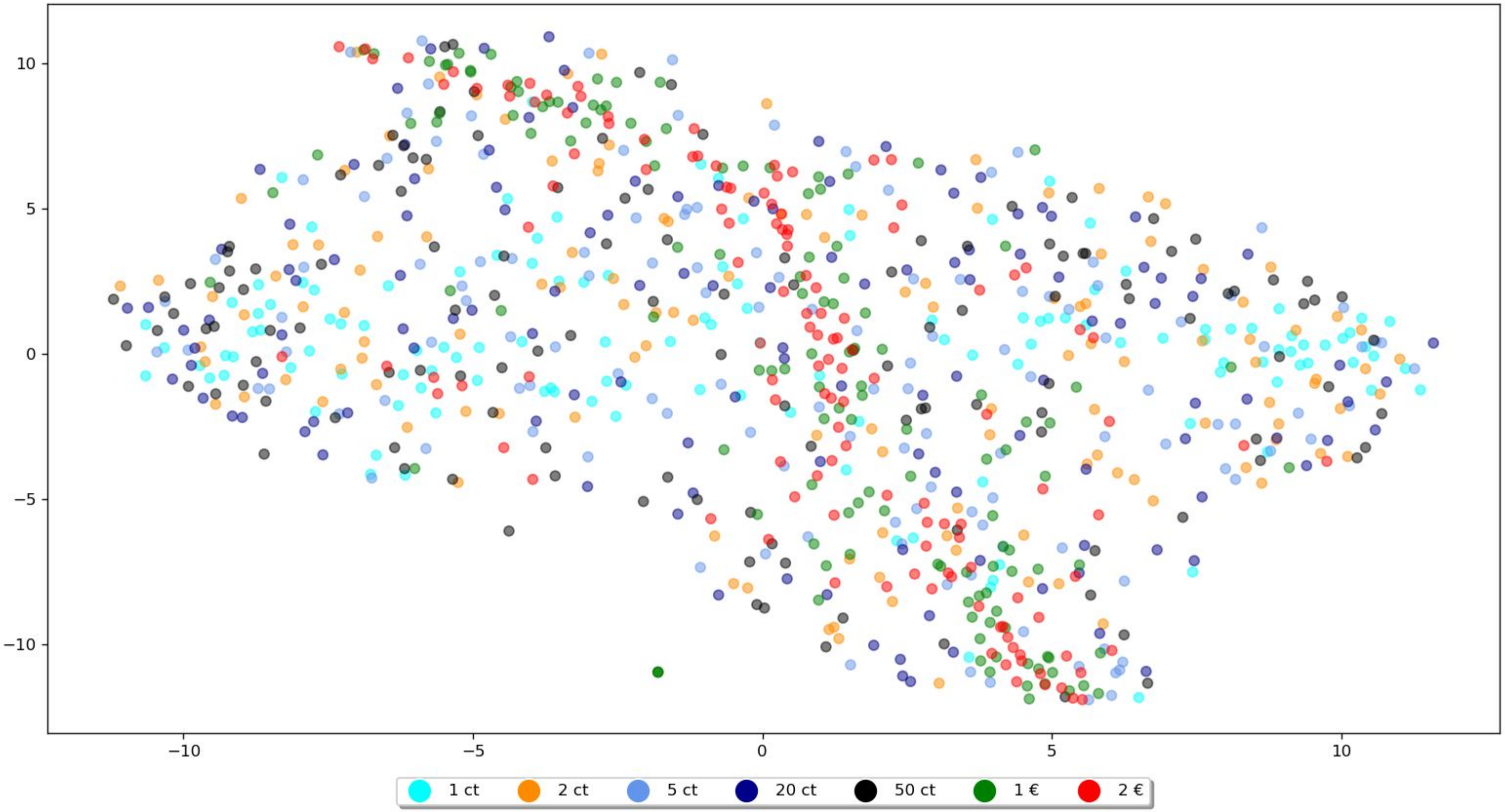


Normalized Confusion Matrix (31.87% overall accuracy)

True label \ Predicted label	1	2	5	20	50	100	200
1	0.10	0.60	0.03	0.07	0.00	0.20	0.00
2	0.00	0.83	0.00	0.00	0.00	0.17	0.00
5	0.04	0.65	0.00	0.13	0.00	0.17	0.00
20	0.04	0.29	0.00	0.25	0.08	0.33	0.00
50	0.00	0.59	0.00	0.00	0.14	0.27	0.00
100	0.00	0.08	0.04	0.04	0.00	0.84	0.00
200	0.00	0.18	0.00	0.00	0.00	0.82	0.00

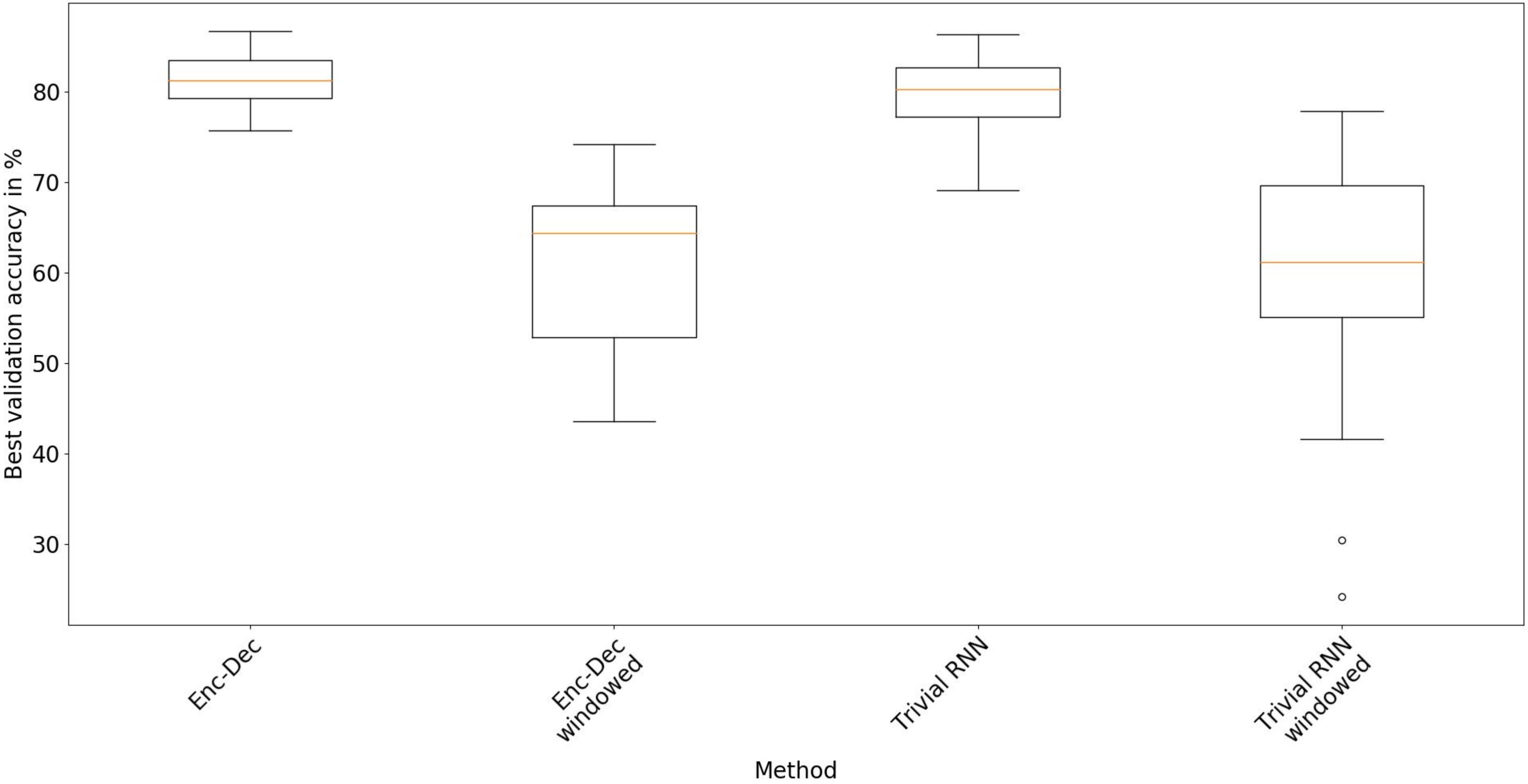
- unable to identify coins → mapping to 2ct or 1€

LSTM Enc-Dec Model *semi-supervised*

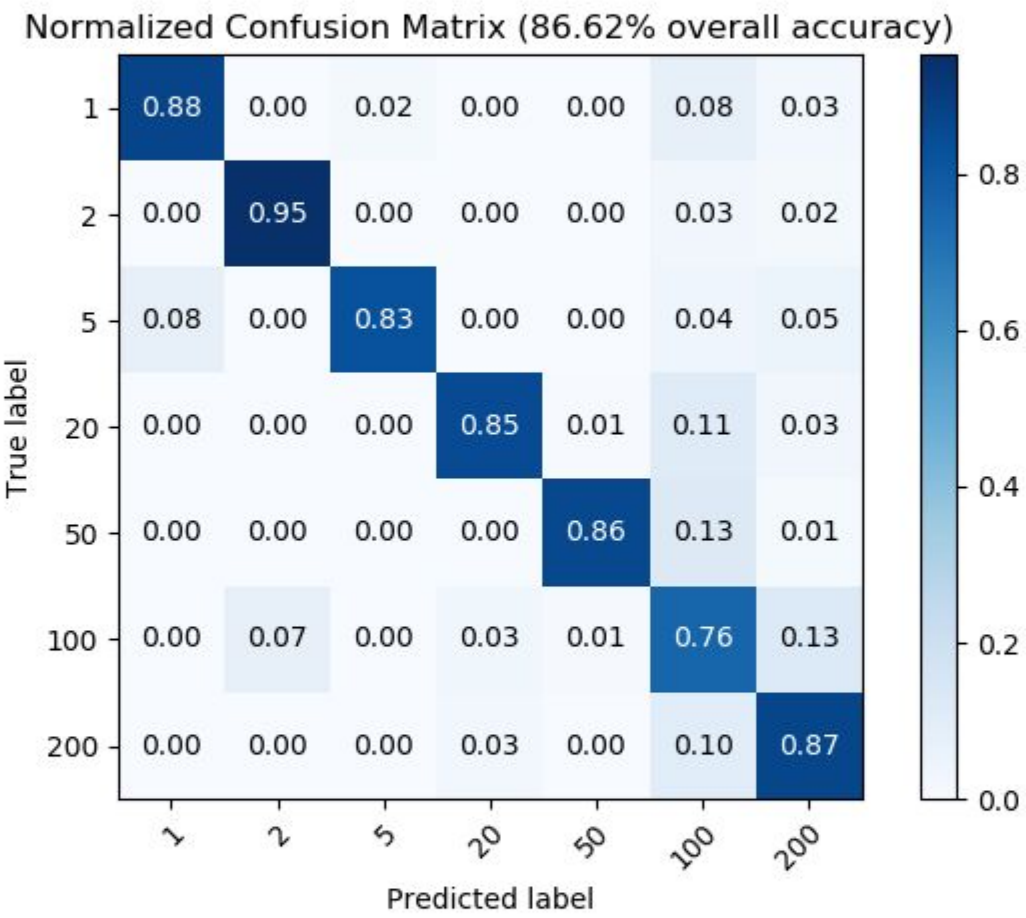


LSTM based method *supervised*

- changing to supervised improves accuracy significantly



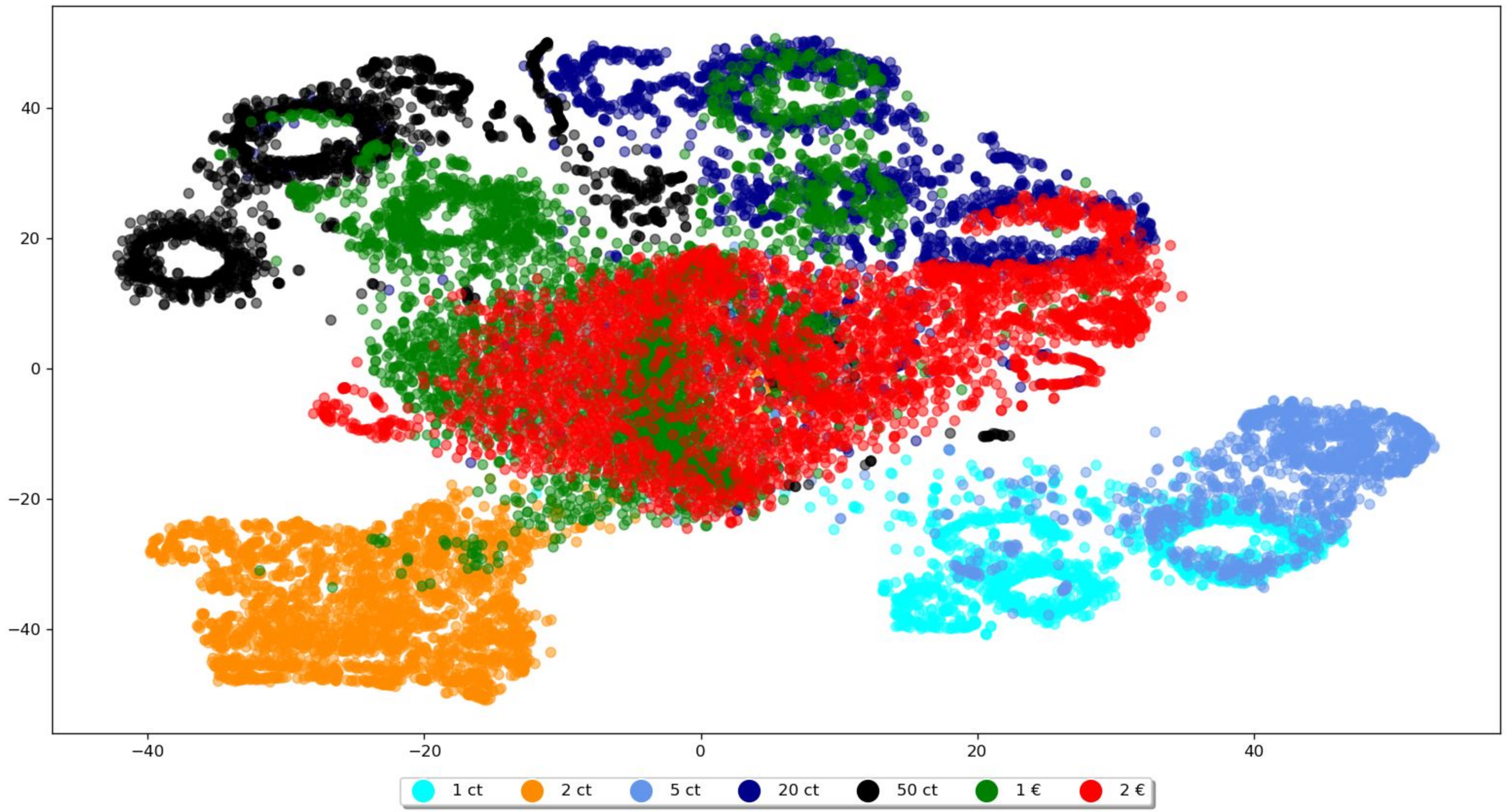
CNN Enc-Dec Model *semi-supervised*



Normalized Confusion Matrix (86.62% overall accuracy)

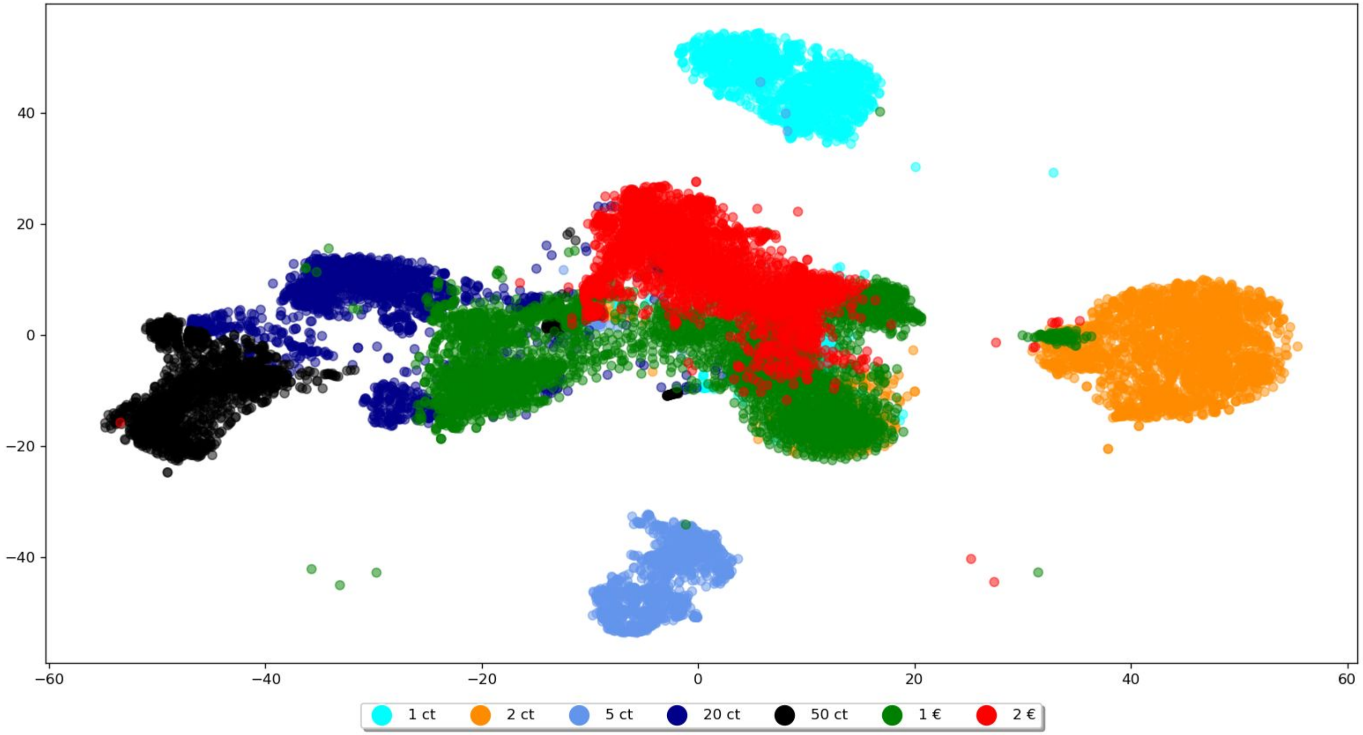
True label \ Predicted label	1	2	5	20	50	100	200
1	0.88	0.00	0.02	0.00	0.00	0.08	0.03
2	0.00	0.95	0.00	0.00	0.00	0.03	0.02
5	0.08	0.00	0.83	0.00	0.00	0.04	0.05
20	0.00	0.00	0.00	0.85	0.01	0.11	0.03
50	0.00	0.00	0.00	0.00	0.86	0.13	0.01
100	0.00	0.07	0.00	0.03	0.01	0.76	0.13
200	0.00	0.00	0.00	0.03	0.00	0.10	0.87

CNN Enc-Dec Model *semi-supervised*

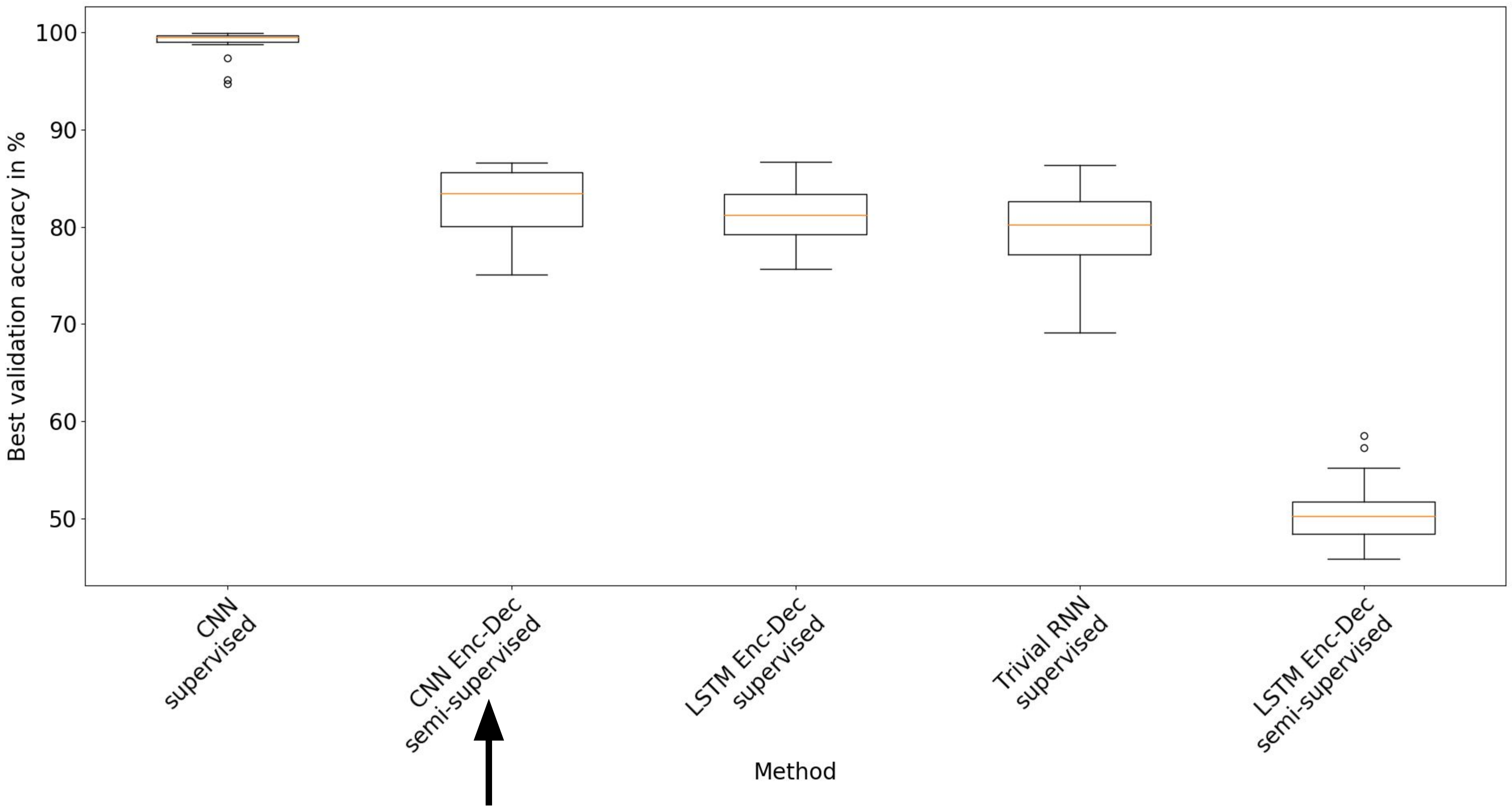


Benchmark: Perfect CNN Classification (*supervised*)

- only learn predictor path, ignore autoencoder
- achieves 99,7% accuracy on validation set
- visually very good TSNE plot



Comparison of all methods



Conclusion

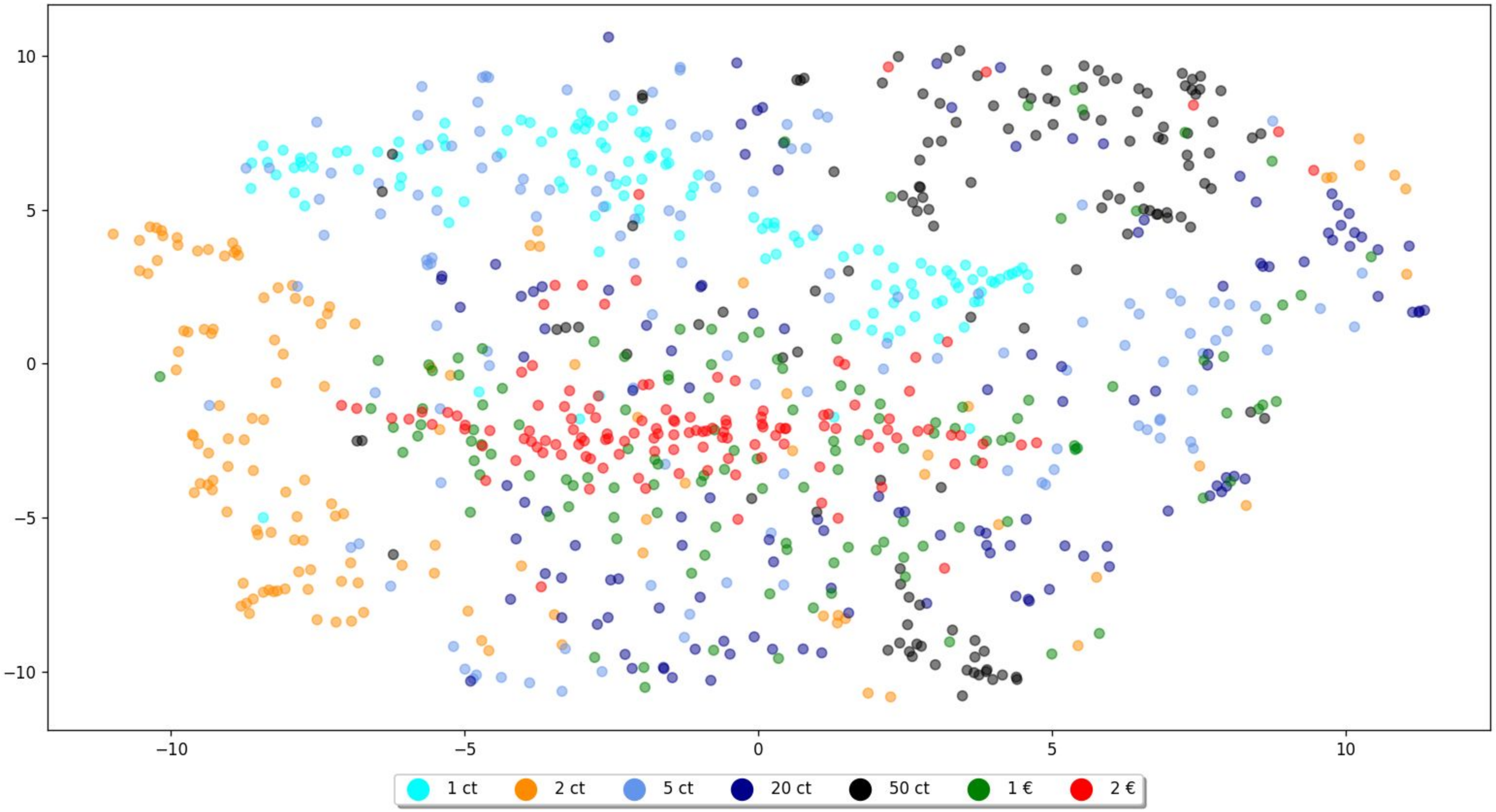
- LSTM based *semi-supervised* learning does not work
  - modification to supervised approach makes it work under same conditions

- CNN based *semi-supervised* learning works (better)
  - TSNE plots indicate self-learned clustering of coins
  - automatic feature extraction works (in this simple case)

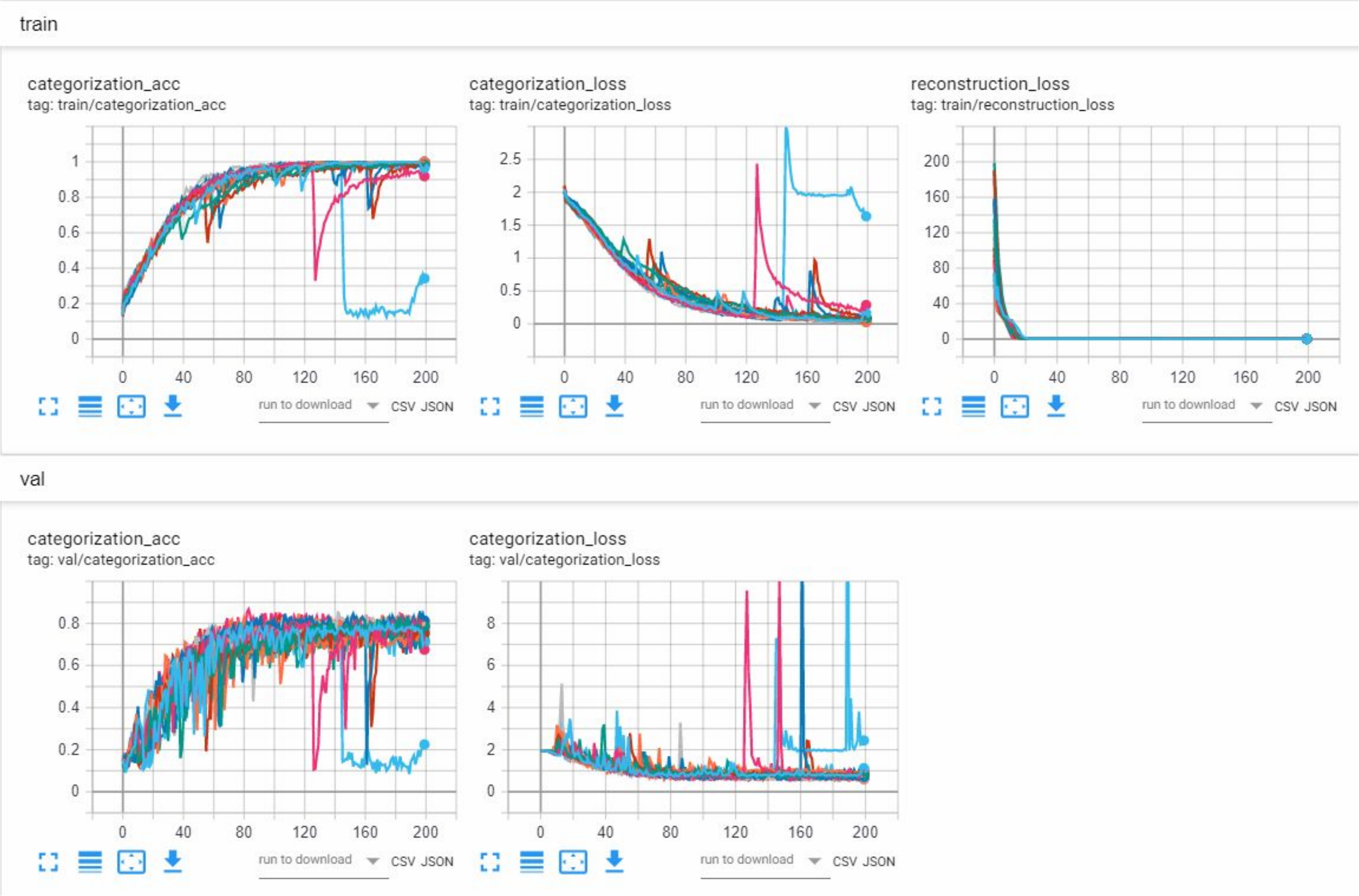
- trivial CNN (*supervised*) approach beats all methods

Danke für die Aufmerksamkeit!

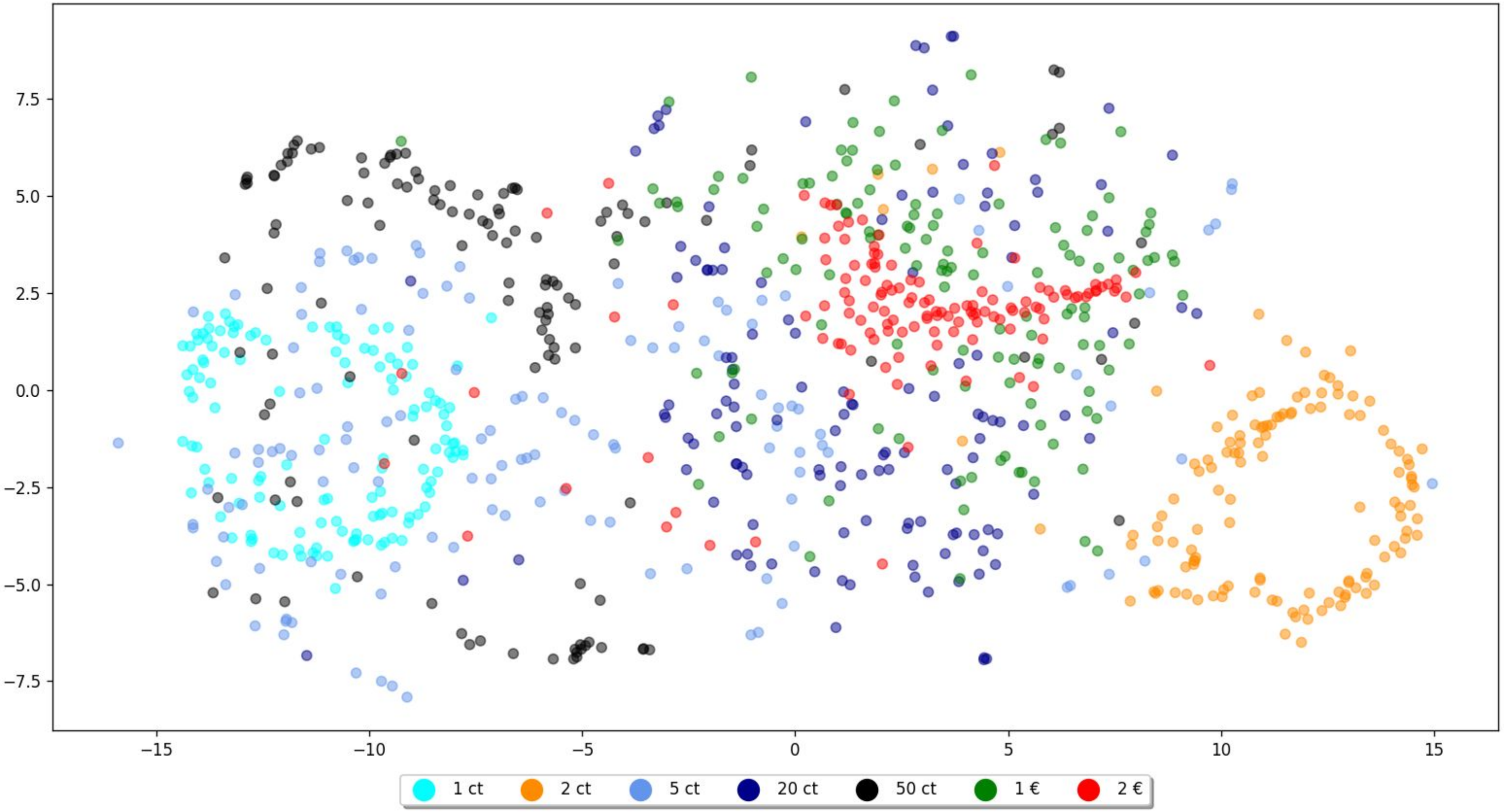
LSTM Enc-Dec *supervised*



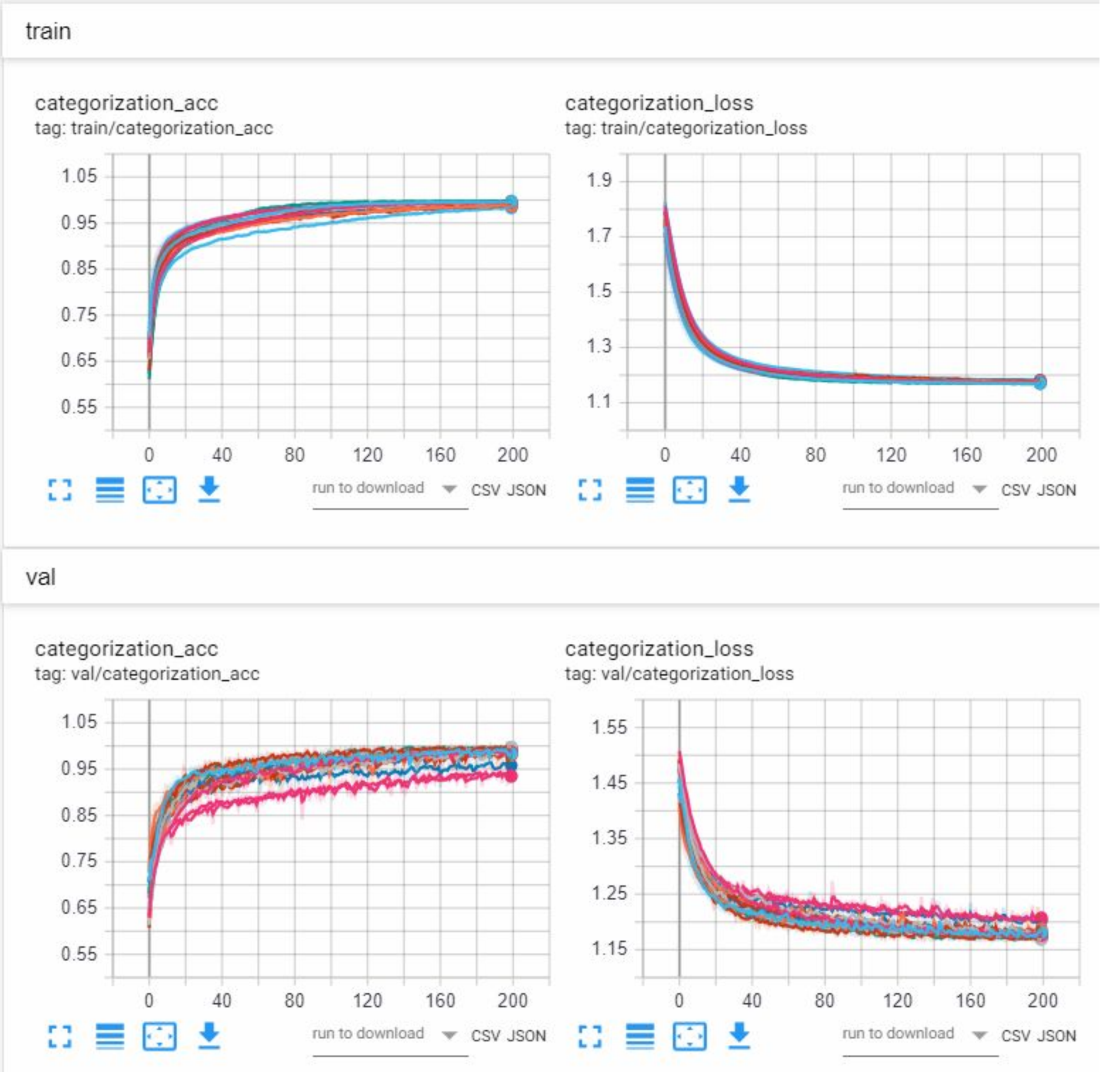
LSTM Enc-Dec *supervised*: Training Graphs



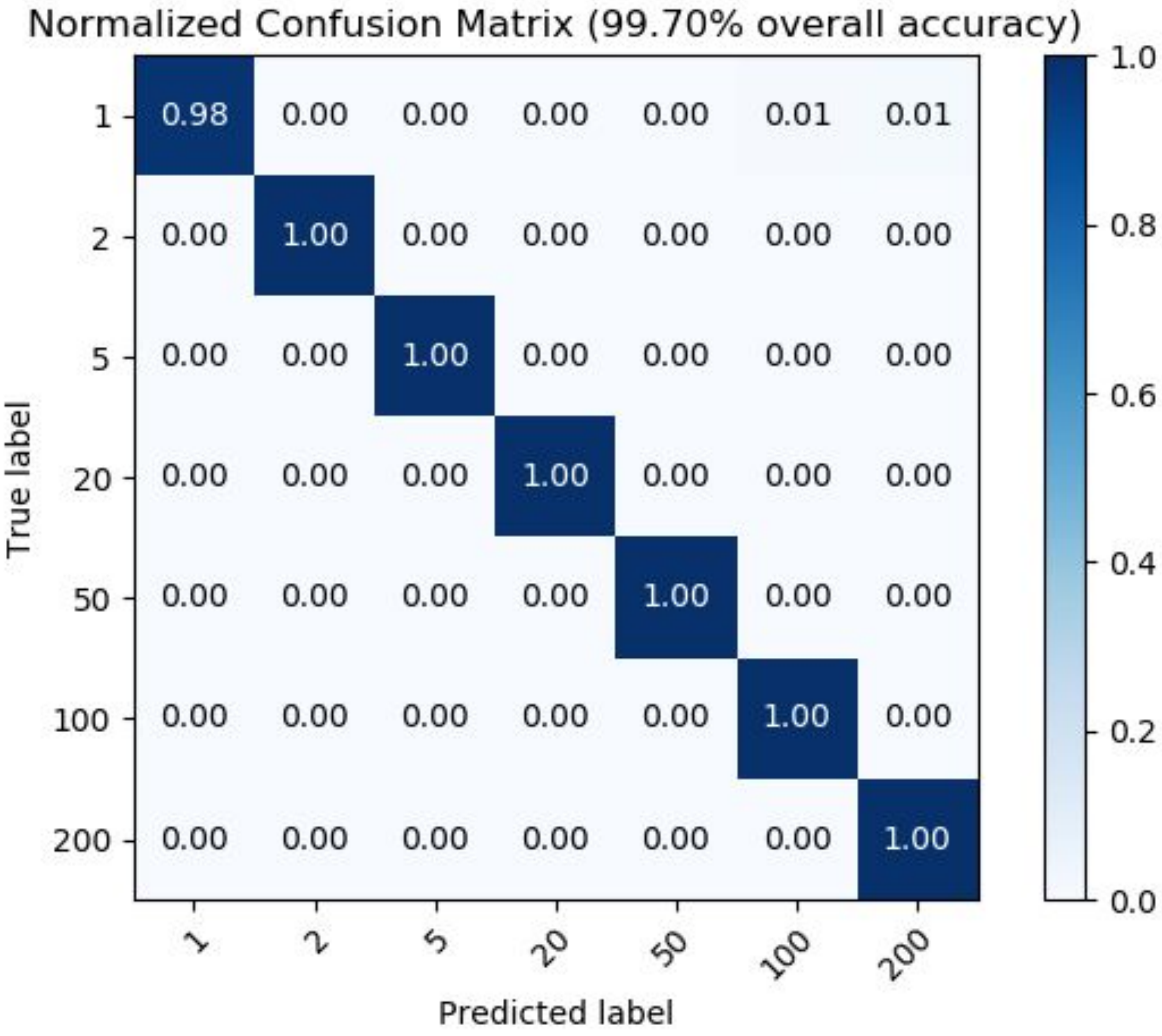
Trivial RNN *supervised*



CNN Windowed: Training Graphs



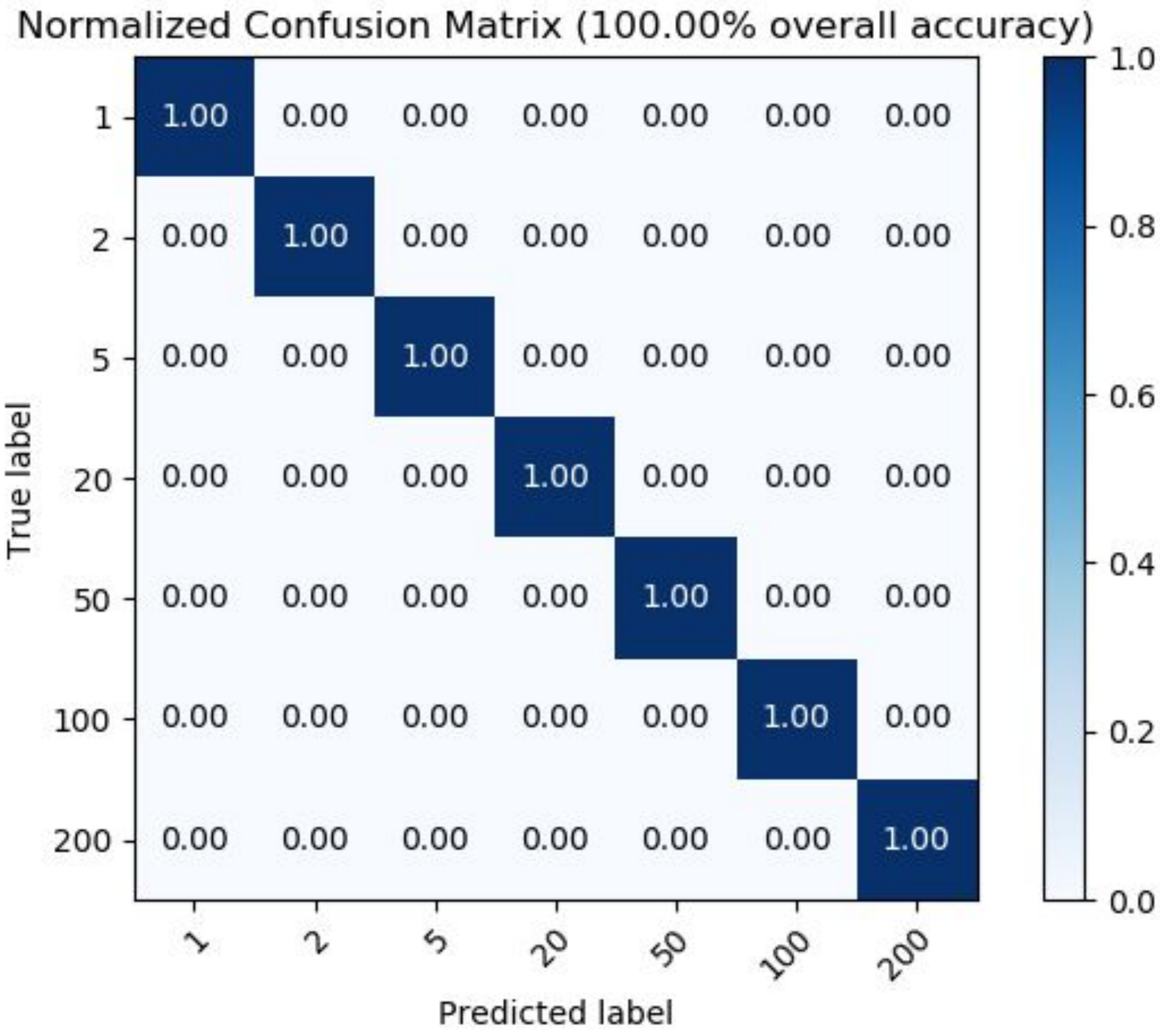
CNN Windowed: Confusion Matrix (Every window)



Normalized Confusion Matrix (99.70% overall accuracy)

True label \ Predicted label	1	2	5	20	50	100	200
1	0.98	0.00	0.00	0.00	0.00	0.01	0.01
2	0.00	1.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	1.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	1.00	0.00	0.00	0.00
50	0.00	0.00	0.00	0.00	1.00	0.00	0.00
100	0.00	0.00	0.00	0.00	0.00	1.00	0.00
200	0.00	0.00	0.00	0.00	0.00	0.00	1.00

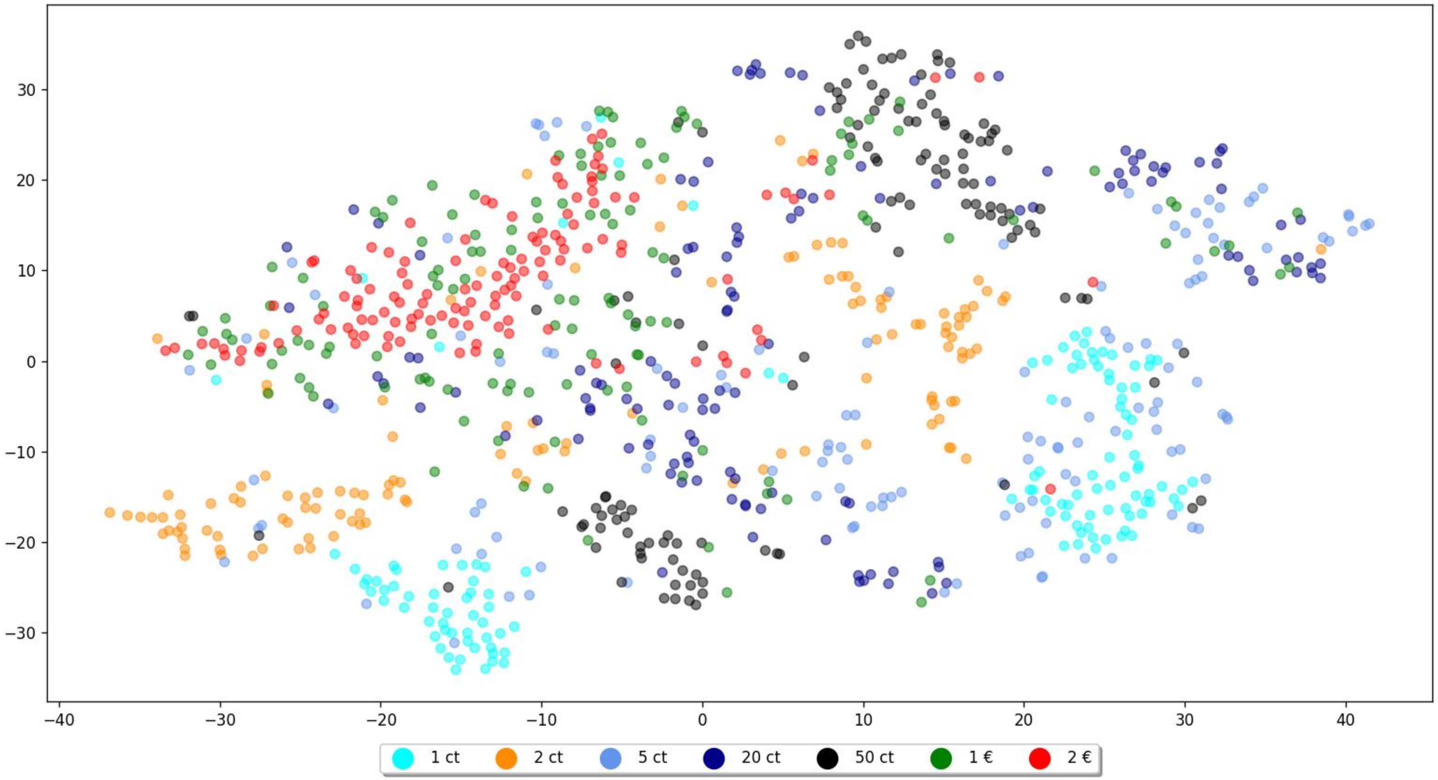
CNN Windowed: Confusion Matrix (Majority vote)



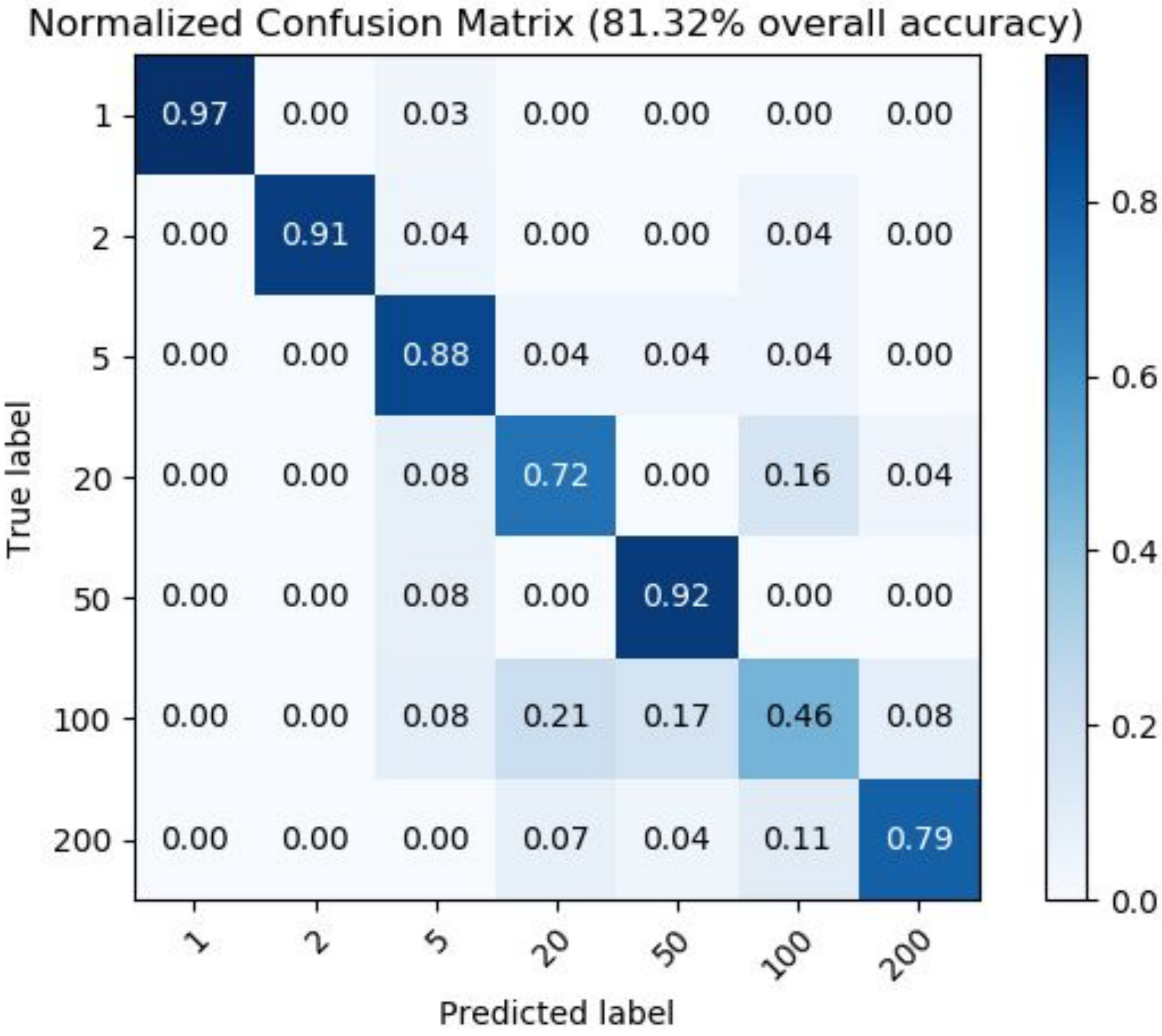
Normalized Confusion Matrix (100.00% overall accuracy)

True label \ Predicted label	1	2	5	20	50	100	200
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	1.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	1.00	0.00	0.00	0.00
50	0.00	0.00	0.00	0.00	1.00	0.00	0.00
100	0.00	0.00	0.00	0.00	0.00	1.00	0.00
200	0.00	0.00	0.00	0.00	0.00	0.00	1.00

LSTM Enc-Dec *supervised*: TSNE Plot



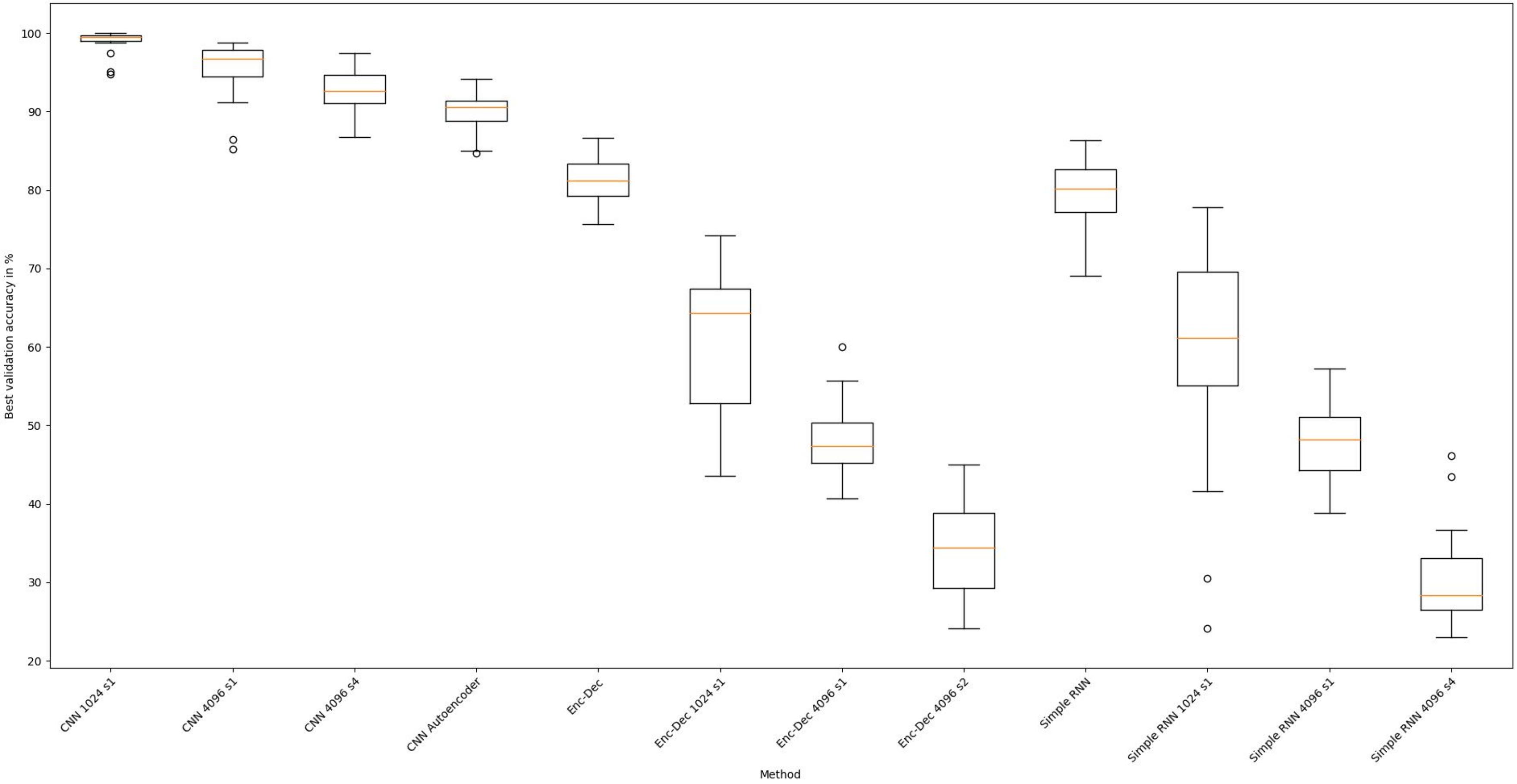
LSTM Enc-Dec *supervised*: Confusion Matrix



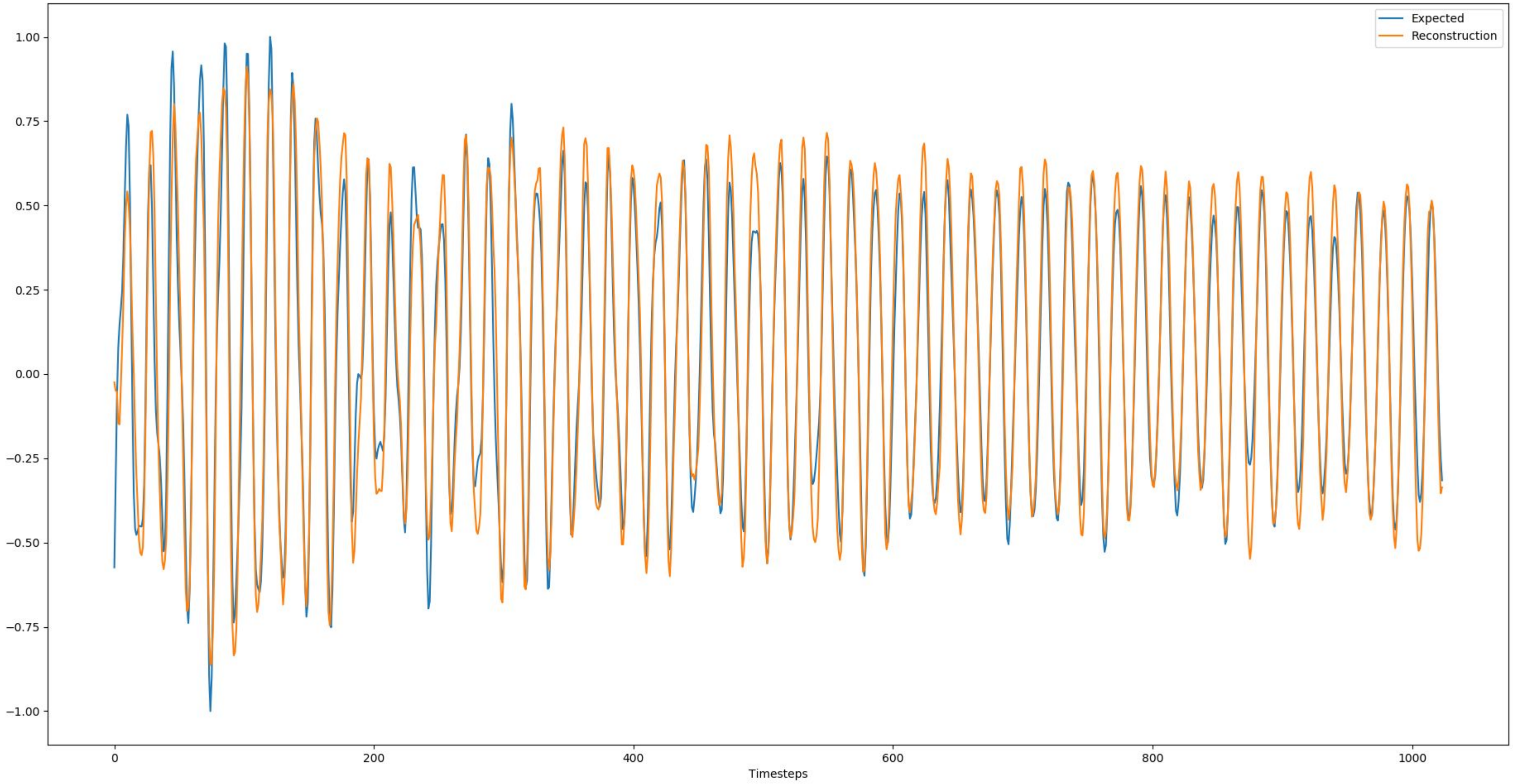
Normalized Confusion Matrix (81.32% overall accuracy)

True label \ Predicted label	1	2	5	20	50	100	200
1	0.97	0.00	0.03	0.00	0.00	0.00	0.00
2	0.00	0.91	0.04	0.00	0.00	0.04	0.00
5	0.00	0.00	0.88	0.04	0.04	0.04	0.00
20	0.00	0.00	0.08	0.72	0.00	0.16	0.04
50	0.00	0.00	0.08	0.00	0.92	0.00	0.00
100	0.00	0.00	0.08	0.21	0.17	0.46	0.08
200	0.00	0.00	0.00	0.07	0.04	0.11	0.79

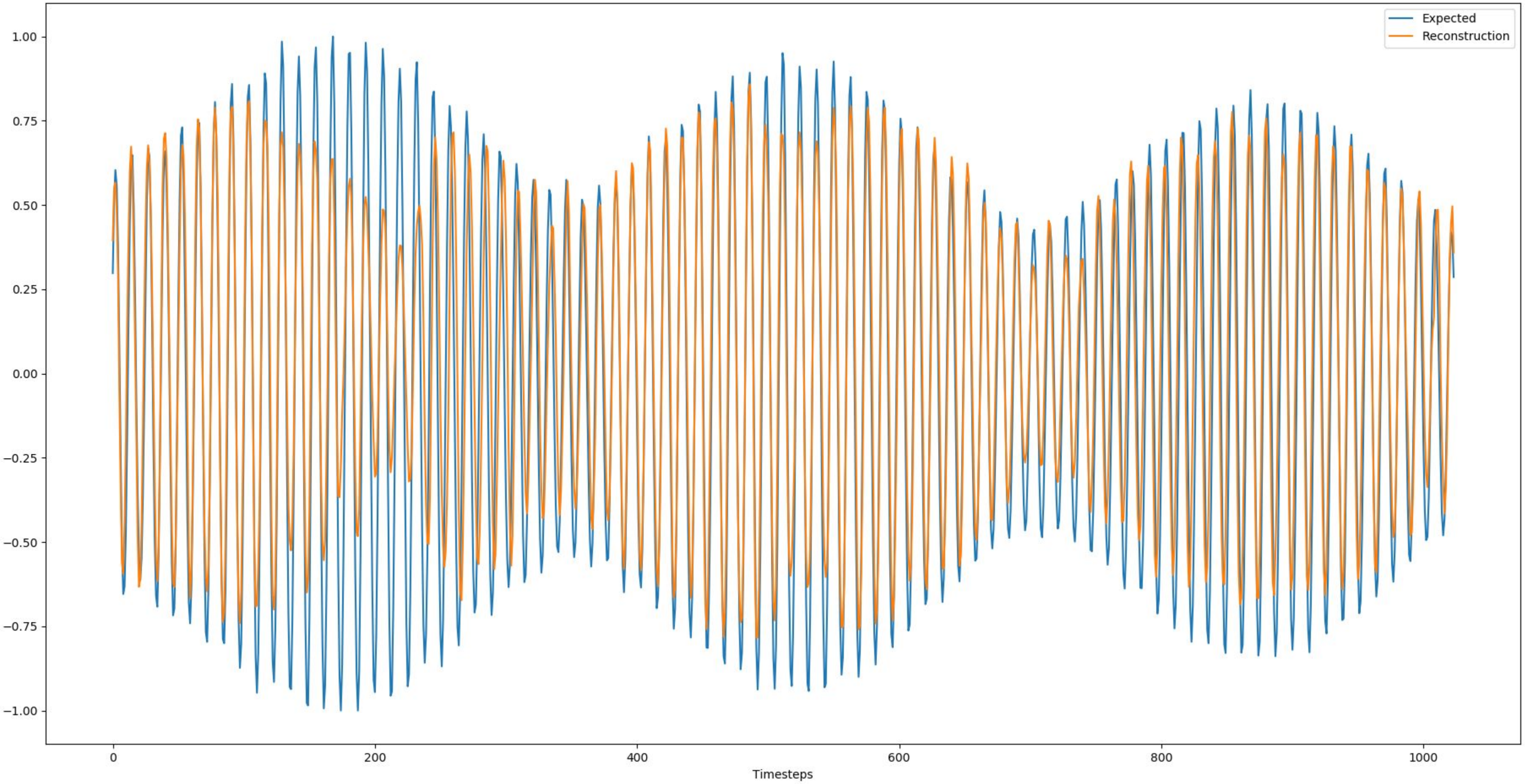
Box Plot *supervised*



CNN Enc-Dec (*semi-supervised*) reconstruction:



CNN Enc-Dec (*semi-supervised*) reconstruction:


Expected
Reconstruction
Timesteps