

# Temporal Smoothing in 2D Human Pose Estimation for Bouldering

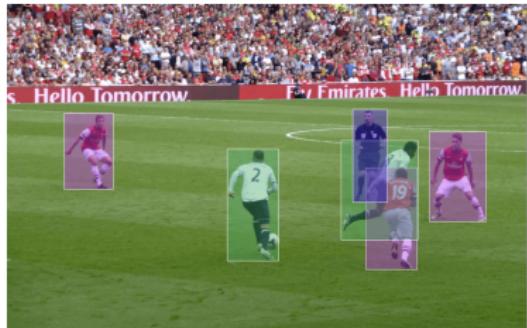
André Oskar Andersen  
wpr684

Institution of Computer Science, University of Copenhagen

2023

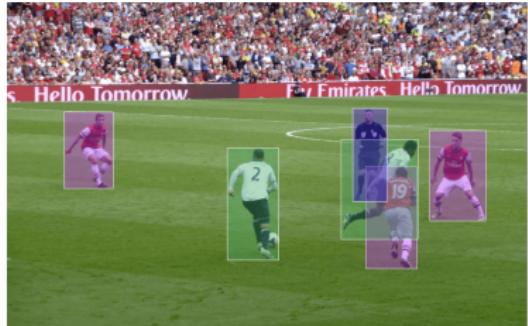
# Introduction

- ▶ Increased usage of video analysis in sports.
  - ▶ Help referee
  - ▶ Improve techniques



# Introduction

- ▶ Increased usage of video analysis in sports.
- ▶ Often requires the position of the players.
  - ▶ Already developed for popular sports.
  - ▶ Missing for the less popular sports.



# Introduction

- ▶ Increased usage of video analysis in sports.
- ▶ Often requires the position of the players.
- ▶ Problems with the data
  - ▶ Methods require a lot of data
  - ▶ Unusual poses/movements



# Introduction

- ▶ Increased usage of video analysis in sports.
- ▶ Often requires the position of the players.
- ▶ Problems with the data
- ▶ ClimbAlong at NorthTech ApS
  - ▶ Frame-independent pose-detector for bouldering - suboptimal results



# Introduction

- ▶ Increased usage of video analysis in sports.
- ▶ Often requires the position of the players.
- ▶ Problems with the data
- ▶ ClimbAlong at NorthTech ApS
  - ▶ Frame-independent pose-detector for bouldering - suboptimal results
  - ▶ Proposition: Incorporate temporal information



## Introduction

- ▶ Aim: extend the ClimbAlong pose-detector to use temporal information.

# The Models

- ▶ Generally, three approaches
  1. Convolutional layer
  2. Recurrent neural network (RNN)
  3. Transformer

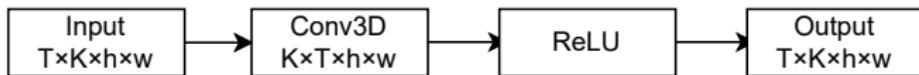
# The Models

- ▶ Generally, three approaches
  1. Convolutional layer
  2. Recurrent neural network (RNN)
  3. Transformer
- ▶ One of each approach

# The Models

## Convolutional layer

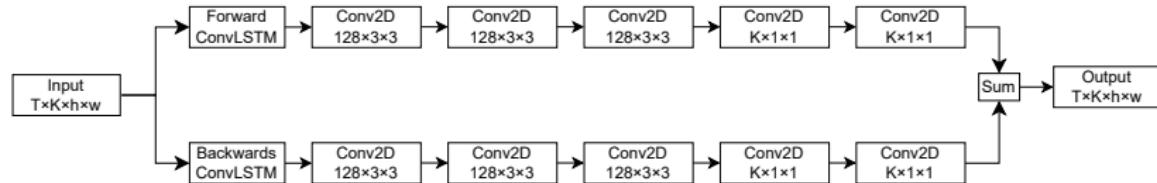
- ▶ Name: 3DConv
- ▶ 3-dimensional conv. layer + ReLU



# The Models

## RNN-based 1:

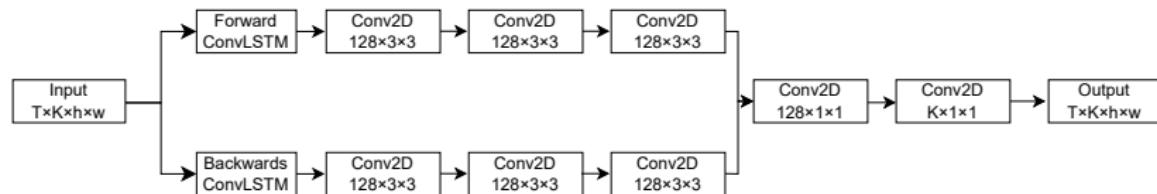
- ▶ Name: bi-ConvLSTM Model S
- ▶ Adaptation of Unipose-LSTM by Artacho and Savakis
- ▶ Bidirectional convolutional LSTM + 2D-conv. layers and ReLUs
- ▶ Processing directions summed together



# The Models

## RNN-based 2:

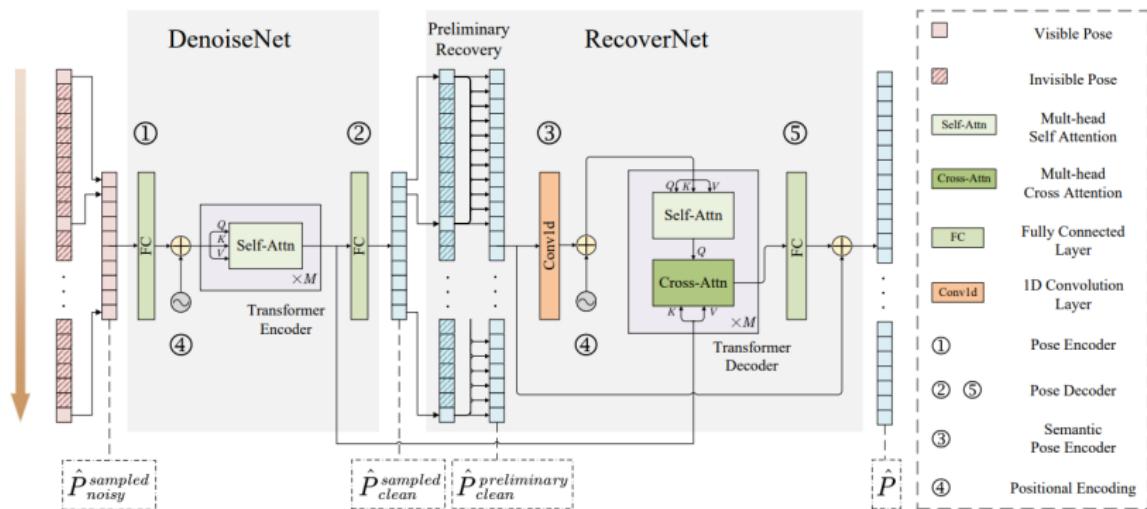
- ▶ bi-ConvLSTM Model C
- ▶ Problem: Prioritization of processing direction
- ▶ Solution: Using convolution
- ▶ Otherwise, very similar to bi-ConvLSTM Model S



# The Models

## DeciWatch by Zeng *Et al.*

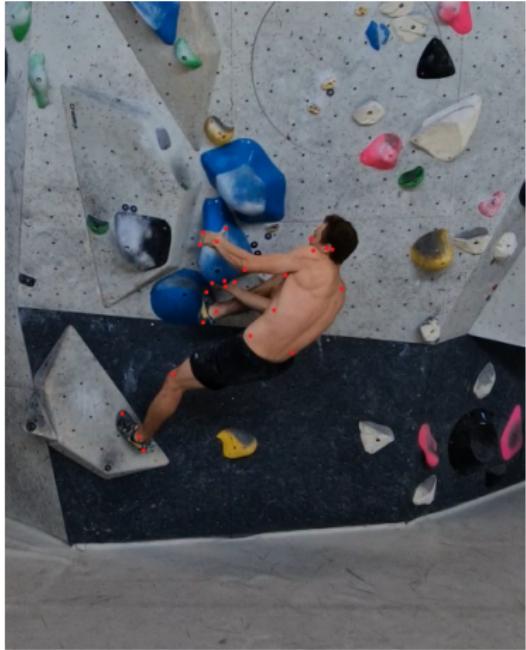
- ▶ Transformer-based
- ▶ Samples every  $n$ th frame
- ▶ DenoiseNet + RecoverNet



# The Data

## ClimbAlong

- ▶ Fully annotated videos of climbers



# The Data

## ClimbAlong

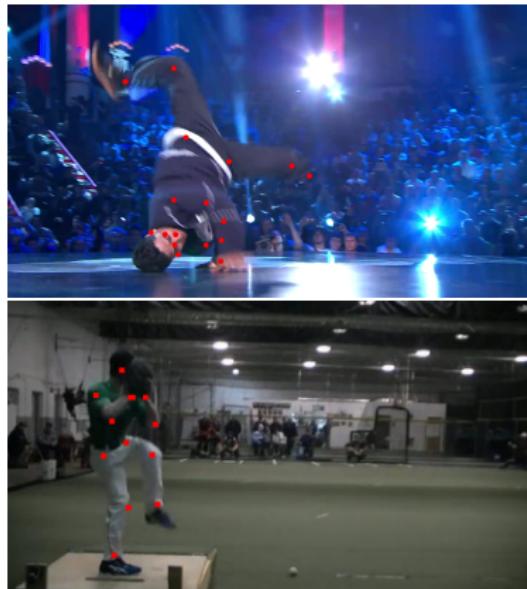
- ▶ Fully annotated videos of climbers
- ▶ Problem: very small dataset



# The Data

## ClimbAlong

- ▶ Fully annotated videos of climbers
- ▶ Problem: Very small dataset
- ▶ Solution: pretrain on related datasets and finetune on ClimbAlong
  - ▶ BRACE
  - ▶ Penn Action

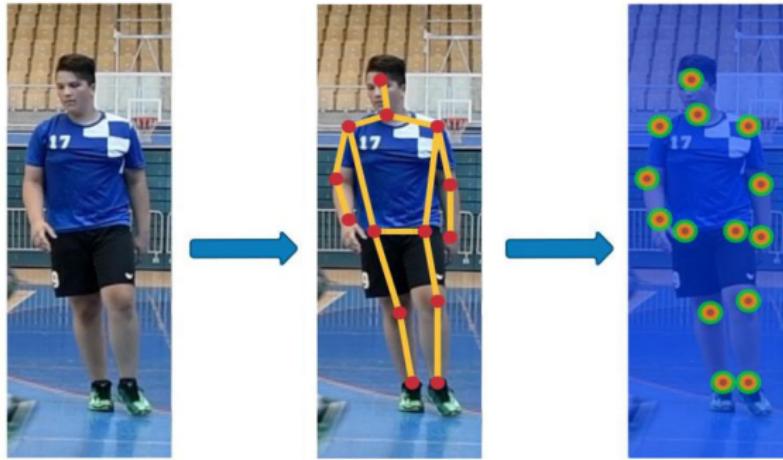


# Data Configuration

- ▶  $s = 5$  frames, as
  - 1. Suggested by Artacho *Et al.*
  - 2. Less memory
  - 3. Model-hyperparameter based on video length

# Data Configuration

- ▶  $s = 5$  frames, as
  1. Suggested by Artacho *Et al.*
  2. Less memory
  3. Model-hyperparameter based on video length
- ▶ Creation of heatmaps



# Pretraining

## Procedure

- ▶ Not using already-developed pose-detector
- ▶ Different input images = Bad representation of pose-detector predictions

# Pretraining

## Procedure

- ▶ Not using already-developed pose-detector
  - ▶ Different input images = Bad representation of pose-detector predictions
- ▶ Instead, simulate pose-detector output by shifting input keypoints = learn to denoise by temporal smoothing

# Pretraining

## Finding optimal setting of models

- ▶ Three experiments
  - 1. Various smearing standard deviations: pose-detector does not use fixed standard deviation

# Pretraining

## Finding optimal setting of models

- ▶ Three experiments
  - 1. Various smearing standard deviations: pose-detector does not use fixed standard deviation
  - 2. Fixed smearing standard deviation: removed some of randomness from experiment 1

# Pretraining

## Finding optimal setting of models

- ▶ Three experiments
  1. Various smearing standard deviations: pose-detector does not use fixed standard deviation
  2. Fixed smearing standard deviation: removed some of randomness from experiment 1
  3. Various smearing standard deviations + decreased frame rate: increased context

# Pretraining

## Finding optimal setting of models

- ▶ Three experiments
  1. Various smearing standard deviations: pose-detector does not use fixed standard deviation
  2. Fixed smearing standard deviation: removed some of randomness from experiment 1
  3. Various smearing standard deviations + decreased frame rate: increased context
- ▶ Two different shifting-scalars

## Finetuning

Freezing already-developed pose-detector

1. Quicker fitting
2. Greater understanding of effects of temporal smoothing

## Accuracy metric

PCK: percentage of correct keypoints

- ▶ Correct if  $dist(pred, gt) \leq d$
- ▶ PCK@0.05, PCK@0.1 and PCK@0.2

## Test results

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	95.7	95.7	95.8	99.2	99.3	99.3
DeciWatch	76.6	76.7	68.1	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	99.7	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	99.8	99.7	99.7

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	47.3	95.5	95.5	95.8	99.2	99.3	99.2
DeciWatch	76.0	75.9	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	99.5	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	99.6	99.3	99.6

## Test results

Only minor difference:

- ▶ Shifting-scalar: shifting-scalar  $s = 1$  better simulates the pose-detector

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	95.7	95.7	95.8	99.2	99.3	99.3
DeciWatch	76.6	76.7	68.1	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	99.7	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	99.8	99.7	99.7

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	47.3	95.5	95.5	95.8	99.2	99.3	99.2
DeciWatch	76.0	75.9	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	99.5	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	99.6	99.3	99.6

## Test results

Only minor difference:

- ▶ Shifting-scalar: shifting-scalar  $s = 1$  better simulates the pose-detector
- ▶ Translation + scaling vs only translation: standard deviations of peaks in ClimbAlong data not changing as much

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	95.7	95.7	95.8	99.2	99.3	99.3
DeciWatch	76.6	76.7	68.1	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	99.7	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	99.8	99.7	99.7

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	47.3	95.5	95.5	95.8	99.2	99.3	99.2
DeciWatch	76.0	75.9	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	99.5	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	99.6	99.3	99.6

# Test results

Decreased frame rate:

- ▶ 3DConv: benefit
- ▶ DeciWatch: drawback
- ▶ bi-ConvLSTM: benefit for less noise

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	95.7	95.7	95.8	99.2	99.3	99.3
DeciWatch	76.6	76.7	68.1	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	99.7	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	99.8	99.7	99.7

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	47.3	95.5	95.5	95.8	99.2	99.3	99.2
DeciWatch	76.0	75.9	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	99.5	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	99.6	99.3	99.6

## Test results

bi-ConvLSTM: Model S vs Model C:

- ▶ Not as a big of a concern

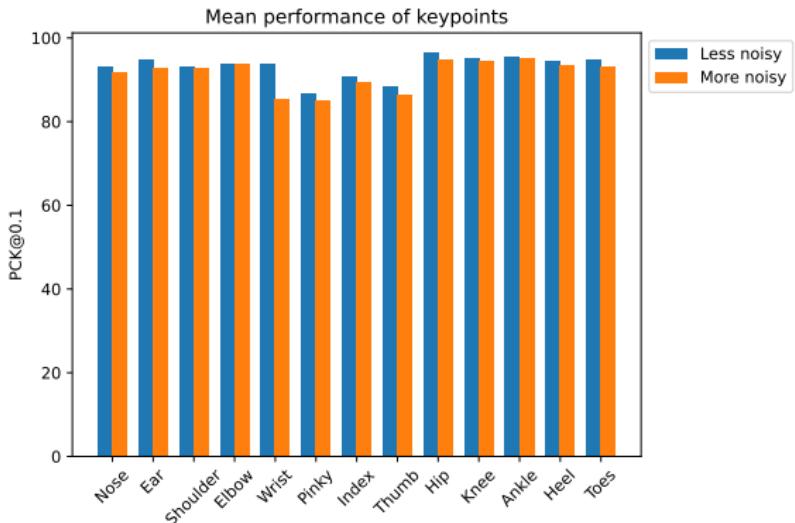
Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	95.7	95.7	95.8	99.2	99.3	99.3
DeciWatch	76.6	76.7	68.1	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	99.7	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	99.8	99.7	99.7

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	47.3	95.5	95.5	95.8	99.2	99.3	99.2
DeciWatch	76.0	75.9	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	99.5	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	99.6	99.3	99.6

# Test results

Worst performing keypoints:

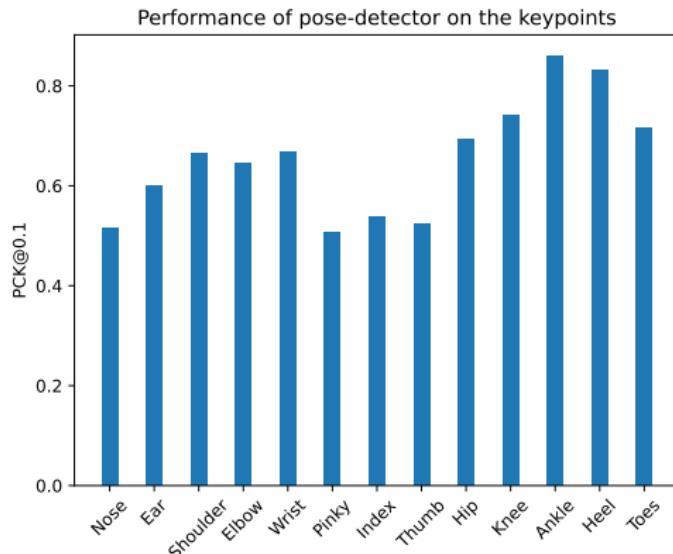
- ▶ Pinkies, index fingers and thumbs
- ▶ Not included during pretraining (minor effect - heels and toes)
- ▶ A lot of movement



# Test results

Worst performing keypoints:

- ▶ Pinkies, index fingers and thumbs
- ▶ Not included during pretraining (minor effect - heels and toes)
- ▶ A lot of movement
- ▶ Extra: performance of pose-detector



## Which model is the optimal choice?

- ▶ Most accurate predictions: DeciWatch shifting-scalar  $s = 1$ , full frame rate
- ▶ Greatest rough estimation: bi-ConvLSTM Model C, shifting-scalar  $s = 1$ , experiment 1
- ▶ Speed and memory: 3DConv, shifting-scalar  $s = 1$ , experiment 3

	Mean prediction time (ms)	Standard deviation of prediction time (ms)	Number of parameters
3DConv	0.39	$9.29 \times 10^{-2}$	78,150
DeciWatch	14.2	0.11	1,708,388
bi-ConvLSTM - Model S	10.6	$1.68 \times 10^{-2}$	1,875,666
bi-ConvLSTM - Model C	20.1	$6.93 \times 10^{-2}$	1,872,313

## General reflections

Potential mistakes we have made

- ▶ Pretraining
  - ▶ Should have estimated parameters of data

## General reflections

Potential mistakes we have made

- ▶ Pretraining
  - ▶ Should have estimated parameters of data
  - ▶ Same video sequence across subsets = could carry some bias

# General reflections

Potential mistakes we have made

- ▶ Pretraining
  - ▶ Should have estimated parameters of data
  - ▶ Same video sequence across subsets = could carry some bias
- ▶ Finetuning
  - ▶ Groundtruth outside of bbox



## Conclusion

Successfully developed and tested models that perform temporal smoothing in 2D human pose estimation for bouldering.

- ▶ Pretraining and finetuning
- ▶ Multiple experiments to find the optimal setting of the models
- ▶ The optimal model depends on ones needs

## Extras: Mistakes Were Made!

### Pretraining

- ▶ Very few overlapping samples in train/test subset

# Extras: Mistakes Were Made!

Misimplemented evaluation-function

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.3	33.4	32.8	72.5	72.4	73.1	85.8	85.8	86.0
DeciWatch	32.8	33.8	30.9	68.0	68.1	62.7	85.1	84.9	82.8
bi-ConvLSTM - Model S	31.7	30.1	31.6	71.5	68.3	71.3	86.3	82.5	86.2
bi-ConvLSTM - Model C	32.0	32.2	31.8	72.2	72.2	71.4	86.6	86.5	86.5

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.1	33.3	32.7	72.1	72.3	72.6	85.6	85.7	85.6
DeciWatch	21.3	26.4	19.4	51.1	58.7	48.1	77.1	81.0	77.0
bi-ConvLSTM - Model S	30.9	31.6	30.6	71.2	71.6	70.7	86.0	86.1	86.2
bi-ConvLSTM - Model C	31.5	32.2	30.8	71.8	72.0	71.9	86.4	86.4	86.5

# Extras: Mistakes Were Made!

## Misimplemented evaluation-function

### 1. Decreased accuracies

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.3	33.4	32.8	72.5	72.4	73.1	85.8	85.8	86.0
DeciWatch	32.8	33.8	30.9	68.0	68.1	62.7	85.1	84.9	82.8
bi-ConvLSTM - Model S	31.7	30.1	31.6	71.5	68.3	71.3	86.3	82.5	86.2
bi-ConvLSTM - Model C	32.0	32.2	31.8	72.2	72.2	71.4	86.6	86.5	86.5

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.1	33.3	32.7	72.1	72.3	72.6	85.6	85.7	85.6
DeciWatch	21.3	26.4	19.4	51.1	58.7	48.1	77.1	81.0	77.0
bi-ConvLSTM - Model S	30.9	31.6	30.6	71.2	71.6	70.7	86.0	86.1	86.2
bi-ConvLSTM - Model C	31.5	32.2	30.8	71.8	72.0	71.9	86.4	86.4	86.5

# Extras: Mistakes Were Made!

## Misimplemented evaluation-function

- Decreased accuracies
- Models do not improve the rough estimates

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.3	33.4	32.8	72.5	72.4	73.1	85.8	85.8	86.0
DeciWatch	32.8	33.8	30.9	68.0	68.1	62.7	85.1	84.9	82.8
bi-ConvLSTM - Model S	31.7	30.1	31.6	71.5	68.3	71.3	86.3	82.5	86.2
bi-ConvLSTM - Model C	32.0	32.2	31.8	72.2	72.2	71.4	86.6	86.5	86.5

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.1	33.3	32.7	72.1	72.3	72.6	85.6	85.7	85.6
DeciWatch	21.3	26.4	19.4	51.1	58.7	48.1	77.1	81.0	77.0
bi-ConvLSTM - Model S	30.9	31.6	30.6	71.2	71.6	70.7	86.0	86.1	86.2
bi-ConvLSTM - Model C	31.5	32.2	30.8	71.8	72.0	71.9	86.4	86.4	86.5

# Extras: Mistakes Were Made!

## Misimplemented evaluation-function

- Decreased accuracies
- Models do not improve the rough estimates
- bi-ConvLSTM: Model C is now always better than Model S

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.3	33.4	32.8	72.5	72.4	73.1	85.8	85.8	86.0
DeciWatch	32.8	33.8	30.9	68.0	68.1	62.7	85.1	84.9	82.8
bi-ConvLSTM - Model S	31.7	30.1	31.6	71.5	68.3	71.3	86.3	82.5	86.2
bi-ConvLSTM - Model C	32.0	32.2	31.8	72.2	72.2	71.4	86.6	86.5	86.5

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.1	33.3	32.7	72.1	72.3	72.6	85.6	85.7	85.6
DeciWatch	21.3	26.4	19.4	51.1	58.7	48.1	77.1	81.0	77.0
bi-ConvLSTM - Model S	30.9	31.6	30.6	71.2	71.6	70.7	86.0	86.1	86.2
bi-ConvLSTM - Model C	31.5	32.2	30.8	71.8	72.0	71.9	86.4	86.4	86.5

# Extras: Mistakes Were Made!

## Misimplemented evaluation-function

- Decreased accuracies
- Models do not improve the rough estimates
- bi-ConvLSTM: Model C is now always better than Model S
- 3DConv is generally the best performing model

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.3	33.4	32.8	72.5	72.4	73.1	85.8	85.8	86.0
DeciWatch	32.8	33.8	30.9	68.0	68.1	62.7	85.1	84.9	82.8
bi-ConvLSTM - Model S	31.7	30.1	31.6	71.5	68.3	71.3	86.3	82.5	86.2
bi-ConvLSTM - Model C	32.0	32.2	31.8	72.2	72.2	71.4	86.6	86.5	86.5

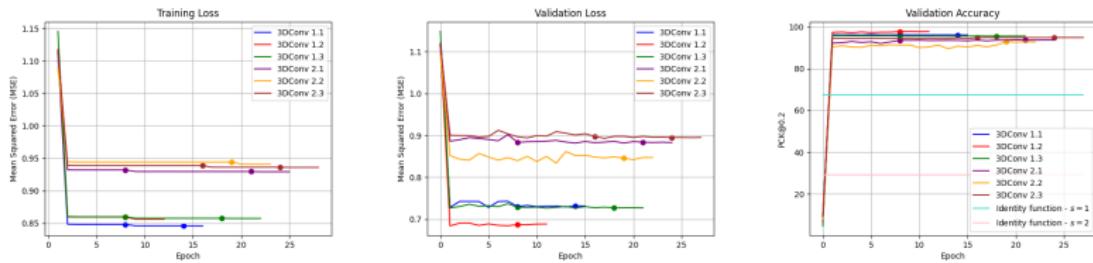
Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	33.1	33.3	32.7	72.1	72.3	72.6	85.6	85.7	85.6
DeciWatch	21.3	26.4	19.4	51.1	58.7	48.1	77.1	81.0	77.0
bi-ConvLSTM - Model S	30.9	31.6	30.6	71.2	71.6	70.7	86.0	86.1	86.2
bi-ConvLSTM - Model C	31.5	32.2	30.8	71.8	72.0	71.9	86.4	86.4	86.5

## Thanks to

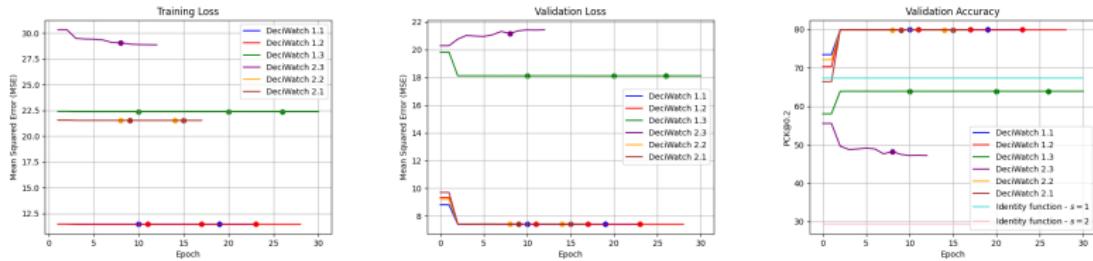
- ▶ My supervisor Kim Steenstrup Pedersen
- ▶ The team at ClimbAlong at NorthTech ApS
- ▶ My parents and my sister

## Appendix

# Pretraining evolution

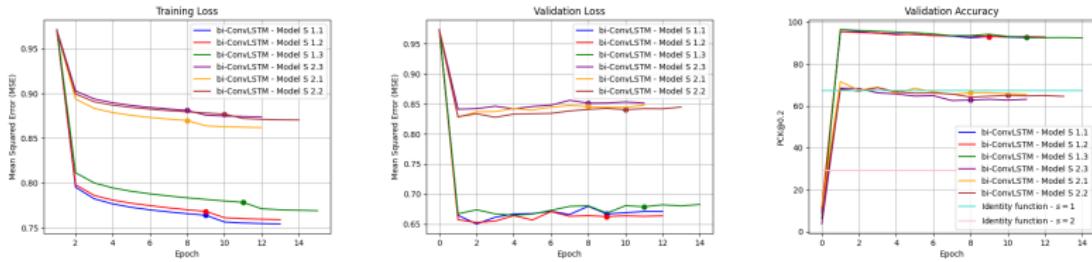


(a) Pretraining results of 3DConv.

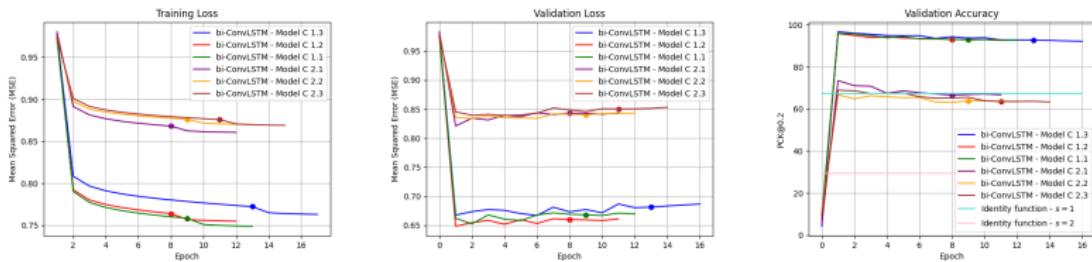


(b) Pretraining results of DeciWatch.

# Pretraining evolution



(a) Pretraining results of the bi-ConvLSTM Model S.



(b) Pretraining results of the bi-ConvLSTM Model C.

# Pretraining test results

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance (px)*	1.11			2.23			4.46		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	6.95	6.95	6.95	25.7	25.7	25.7	67.4	67.4	67.4
3DConv	1.84	20.5	0.67	30.6	58.5	23.8	<b>96.6</b>	<b>98.0</b>	96.3
DeciWatch	<b>51.4</b>	<b>51.4</b>	<b>44.5</b>	64.6	64.6	51.5	80.2	80.2	64.2
bi-ConvLSTM Model S	27.8	31.1	33.8	68.4	71.0	<b>72.5</b>	95.7	95.8	96.7
bi-ConvLSTM Model C	29.5	31.7	31.8	<b>69.5</b>	<b>71.3</b>	71.8	96.1	96.1	<b>96.9</b>

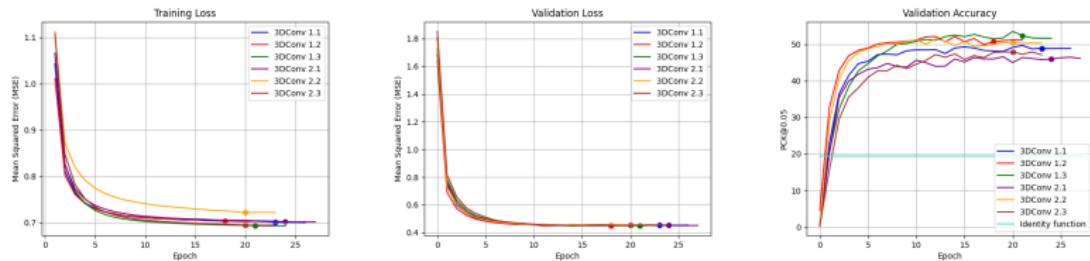
Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance (px)*	1.11			2.23			4.46		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	1.84	1.84	1.84	7.75	7.75	7.75	29.3	29.3	29.3
3DConv	0.80	2.40	0.11	21.6	27.6	16.8	<b>94.2</b>	<b>91.8</b>	<b>95.5</b>
DeciWatch	<b>51.2</b>	<b>51.2</b>	<b>10.3</b>	<b>64.4</b>	<b>64.4</b>	24.4	80.2	80.2	50.3
bi-ConvLSTM Model S	10.5	11.6	10.1	31.1	33.4	29.5	68.5	67.8	69.6
bi-ConvLSTM Model C	12.5	12.0	9.63	36.4	32.5	<b>29.8</b>	74.6	65.5	70.0

# Pretraining keypoint test results

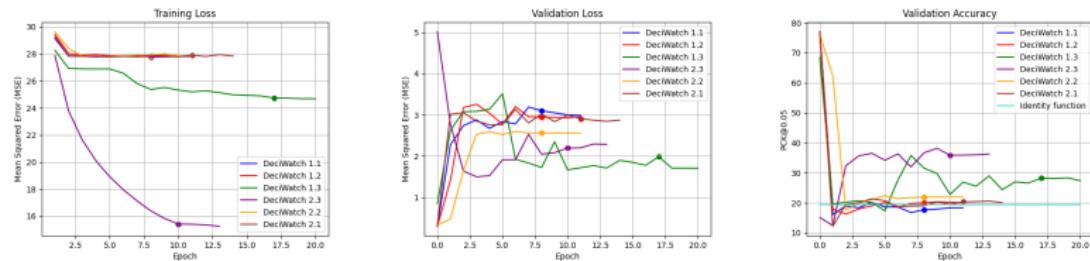
	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	1	2	3	1	2	3	1	2	3	1	2	3	
Experiment													
Nose	30.4	44.3	26.5	68.9	68.0	54.1	80.6	67.7	81.4	74.9	73.3	70.1	61.7
Ear	29.4	43.8	24.9	69.6	69.6	55.2	76.5	73.2	76.2	74.9	75.2	71.5	61.7
Shoulder	34.0	71.3	29.4	68.1	68.1	53.5	65.2	72.0	71.4	70.9	77.0	69.4	62.5
Elbow	31.1	68.3	24.0	62.7	62.7	49.4	71.9	79.0	68.0	67.6	71.9	83.6	61.7
Wrist	26.5	45.3	19.8	59.8	59.8	48.3	70.5	70.4	66.8	68.6	70.7	74.6	56.8
Hip	34.3	84.3	24.4	63.6	63.6	50.1	65.2	65.6	57.4	67.1	62.1	56.3	57.8
Knee	33.7	65.2	25.1	59.5	59.5	48.0	73.9	69.2	66.8	61.9	69.0	73.2	58.8
Ankle	25.4	37.2	17.6	58.9	58.9	48.3	80.1	69.4	66.1	73.1	72.2	75.2	56.9
Total	30.6	58.5	23.8	64.6	64.6	51.5	68.4	71.0	72.5	69.5	71.3	71.8	

	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	1	2	3	1	2	3	1	2	3	1	2	1.3	
Experiment													
Nose	20.5	20.5	17.3	67.9	22.2	41.8	75.8	36.2	29.1	32.8	35.2	25.5	35.4
Ear	22.5	29.7	18.8	69.1	18.6	47.9	80.3	46.9	42.7	47.9	49.4	42.4	43.0
Shoulder	22.4	30.5	17.0	68.1	13.3	32.2	74.7	26.3	23.3	32.8	31.5	21.0	32.8
Elbow	22.5	32.0	16.5	62.7	14.1	19.3	58.2	33.9	21.1	27.2	31.1	27.1	30.5
Wrist	21.5	26.9	17.1	59.7	34.2	35.5	69.4	37.1	32.3	39.8	33.1	33.5	36.7
Hip	22.4	27.9	16.3	63.5	16.8	19.9	64.7	26.5	32.1	35.2	22.2	30.3	31.5
Knee	19.9	25.5	15.4	59.5	24.2	23.4	60.1	28.2	16.1	26.0	21.0	13.2	27.7
Ankle	21.0	24.5	16.2	58.8	47.5	35.6	69.4	34.1	40.0	48.6	39.4	43.7	39.9
Total	21.6	27.6	16.8	64.4	64.4	24.4	31.1	33.4	29.5	36.4	32.5	29.8	

# Finetuning evolution

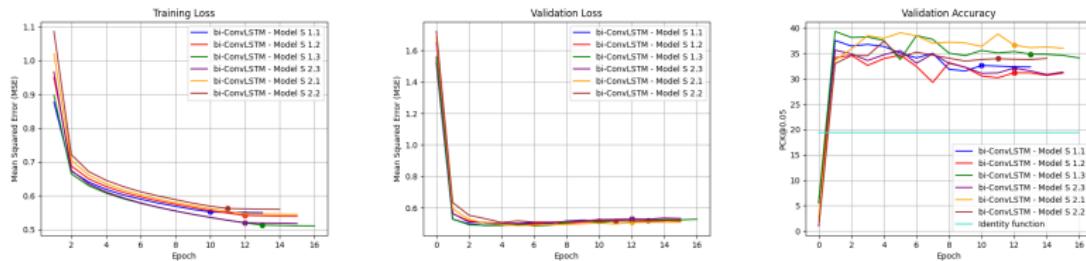


(a) Finetuning results of 3DConv.

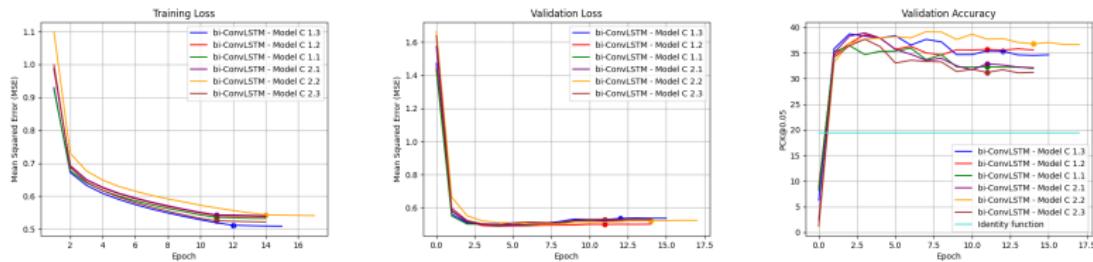


(b) Finetuning results of DeciWatch.

# Finetuning evolution

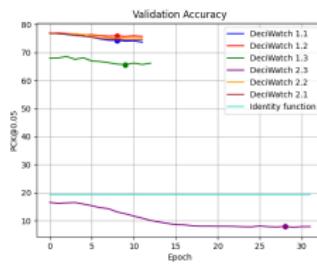
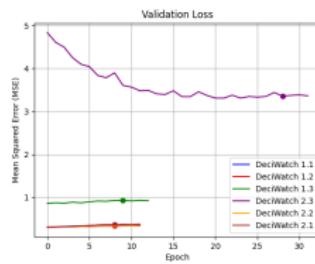
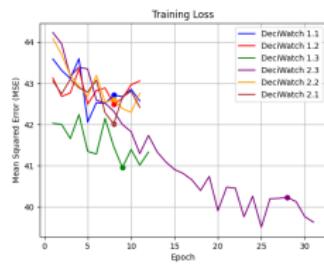


(a) Finetuning results of bi-Conv LSTM Model S.



(b) Finetuning results of bi-Conv LSTM Model C.

# Finetuning evolution with regularization



# Finetuning test results

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance (px)*	0.80			1.60			3.21		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	49.7	52.3	53.1	<b>95.7</b>	<b>95.7</b>	<b>95.8</b>	99.2	99.3	99.3
DeciWatch	<b>76.6</b>	<b>76.7</b>	<b>68.1</b>	94.4	94.3	87.3	99.2	99.2	96.1
bi-ConvLSTM - Model S	37.8	34.9	39.0	91.8	92.1	92.2	99.4	<b>99.7</b>	99.2
bi-ConvLSTM - Model C	35.9	39.0	38.5	93.1	93.6	92.6	<b>99.8</b>	<b>99.7</b>	<b>99.7</b>

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance (px)*	0.80			1.60			3.21		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	19.4	19.4	19.4	66.1	66.1	66.1	85.2	85.2	85.2
3DConv	46.5	51.6	<b>47.3</b>	<b>95.5</b>	<b>95.5</b>	<b>95.8</b>	99.2	99.3	99.2
DeciWatch	<b>76.0</b>	<b>75.9</b>	36.8	94.2	94.2	74.9	99.2	99.2	92.8
bi-ConvLSTM - Model S	38.8	37.4	35.9	92.7	92.1	91.2	99.4	<b>99.5</b>	99.3
bi-ConvLSTM - Model C	39.2	39.5	37.1	92.5	92.9	92.6	<b>99.6</b>	99.3	<b>99.6</b>

## Finetuning additional test results

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance*	0.87			1.77			3.55		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	21.2	21.2	21.2	65.5	65.5	65.5	84.7	84.7	84.7
3DConv	58.4	61.4	61.7	<b>98.7</b>	<b>98.9</b>	<b>99.0</b>	<b>99.6</b>	<b>99.8</b>	<b>99.7</b>
DeciWatch	<b>82.6</b>	<b>82.4</b>	<b>74.6</b>	96.2	96.1	92.3	99.1	99.1	97.4
bi-ConvLSTM - Model S	45.7	45.0	47.6	97.3	96.9	97.0	<b>99.6</b>	99.6	99.1
bi-ConvLSTM - Model C	44.5	46.1	48.5	97.4	97.9	97.9	99.6	99.5	99.6

Accuracy metric	PCK@0.05			PCK@0.1			PCK@0.2		
Mean threshold distance*	0.87			1.77			3.55		
Experiment	1	2	3	1	2	3	1	2	3
Identity function	21.2	21.2	21.2	65.5	65.5	65.5	84.7	84.7	84.7
3DConv	56.2	60.0	<b>56.6</b>	<b>98.9</b>	<b>98.8</b>	<b>98.8</b>	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>
DeciWatch	<b>81.6</b>	<b>81.8</b>	37.5	96.0	96.0	73.3	99.1	99.1	90.7
bi-ConvLSTM - Model S	44.8	46.2	45.0	96.9	95.9	97.1	99.5	99.6	99.5
bi-ConvLSTM - Model C	45.9	47.9	46.7	96.7	97.1	98.1	99.6	99.4	99.6

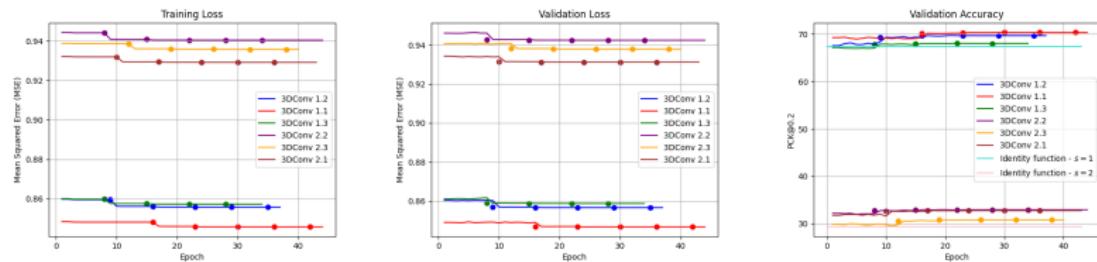
# Finetuning keypoint testing results

Experiment	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	1	2	3	1	2	3	1	2	3	1	2	3	
Nose	95.5	97.6	95.7	94.9	94.9	88.5	93.5	89.7	95.4	93.4	88.8	90.9	93.2
Ear	96.5	96.7	96.8	95.5	95.5	88.6	94.3	91.9	94.2	96.3	95.6	96.1	94.8
Shoulder	95.6	97.4	97.1	95.4	95.4	88.4	91.2	90.6	87.2	91.9	95.1	92.9	93.2
Elbow	97.6	97.2	97.5	94.5	94.6	86.5	93.2	93.7	94.9	95.1	96.7	95.2	94.7
Wrist	96.4	96.4	96.3	93.9	93.9	85.8	94.2	92.7	94.1	94.5	93.2	94.3	93.8
Pinky	86.6	85.4	85.4	92.5	92.2	84.0	84.5	86.5	86.6	83.4	88.4	87.0	86.9
Index finger	92.7	92.5	93.1	92.2	92.1	84.3	90.3	89.1	88.2	89.9	92.3	92.1	90.7
Thumb	91.6	91.0	91.4	92.8	92.7	84.5	81.0	89.3	84.9	89.6	85.5	86.9	88.4
Hip	99.4	99.0	99.8	96.4	96.4	89.9	92.5	96.5	98.0	94.6	97.7	96.1	96.4
Knee	97.7	98.1	98.4	95.0	95.0	88.4	96.1	92.5	96.5	96.0	93.9	93.6	95.1
Ankle	98.6	98.4	99.1	95.0	95.0	87.6	94.3	93.6	96.0	97.2	98.1	94.3	95.6
Heel	98.6	98.1	98.5	95.2	95.1	86.2	94.2	93.4	94.9	92.5	92.6	93.0	94.4
Toes	97.2	97.5	98.0	94.2	94.0	86.8	94.9	95.3	95.5	93.8	95.7	94.4	94.8
Total	95.7	95.7	95.8	94.4	94.3	87.3	91.8	92.1	92.2	93.1	93.6	92.6	

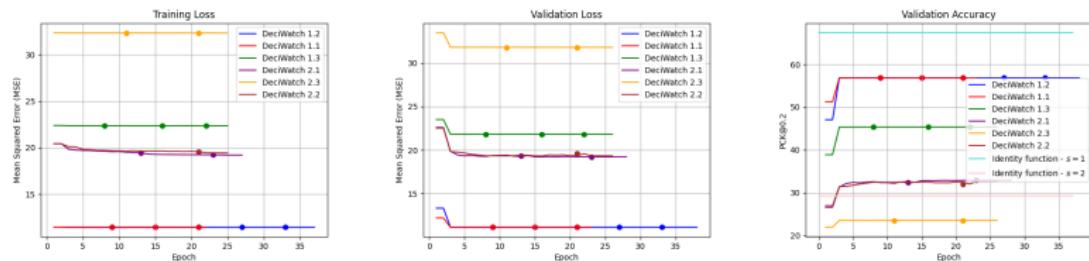
# Finetuning keypoint testing results

	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	1.1	1.2	1.3	1.1	1.2	1.3	1.1	1.2	1.3	1.1	1.2	1.3	
Experiment													
Nose	95.6	97.0	95.6	94.1	94.3	80.8	90.9	89.6	93.2	88.1	91.1	90.8	91.76
Ear	96.8	95.7	96.9	93.3	93.4	80.8	94.2	92.7	92.6	91.9	93.2	94.9	93.0
Shoulder	95.1	97.1	95.7	95.3	95.4	77.5	93.5	93.4	90.1	90.0	97.7	95.6	93.0
Elbow	97.2	97.1	97.5	94.4	94.5	72.0	96.0	96.7	90.9	97.1	97.0	94.7	93.8
Wrist	96.4	96.5	96.6	93.8	93.8	77.4	89.2	93.8	91.9	93.2	94.3	9.33	85.5
Pinky	85.0	85.3	87.0	92.3	92.4	66.1	84.5	82.6	84.0	87.4	83.6	89.7	85.0
Index finger	93.1	92.8	93.0	92.5	92.5	65.3	92.7	89.9	92.4	92.1	88.0	91.4	89.6
Thumb	91.1	90.8	91.6	93.0	92.9	71.4	85.3	83.5	81.1	82.6	84.6	88.9	86.4
Hip	99.3	98.8	99.6	96.3	96.4	78.6	97.1	96.2	96.8	97.8	97.4	84.5	94.9
Knee	97.6	97.8	97.9	95.1	95.0	77.9	94.9	93.3	91.3	95.9	98.4	97.5	94.4
Ankle	98.6	98.5	99.3	94.5	94.6	79.3	96.2	96.0	93.8	98.5	96.8	96.2	95.2
Heel	98.2	97.8	98.8	95.1	95.0	78.5	93.7	93.0	93.2	93.0	91.3	92.8	93.4
Toes	97.4	97.4	97.8	94.0	94.0	73.1	95.1	96.8	93.5	92.7	93.7	93.7	93.3
Total	95.5	95.5	95.8	94.2	94.2	74.9	92.7	92.1	91.2	92.5	92.9	92.6	

# Pretraining evolution - corrected

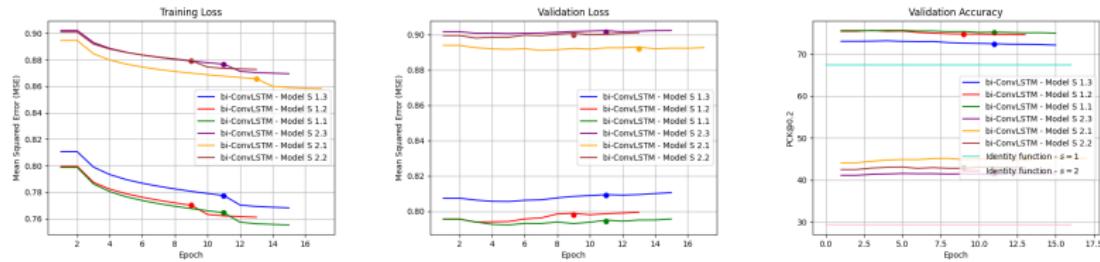


(a) Pretraining results of 3DConv

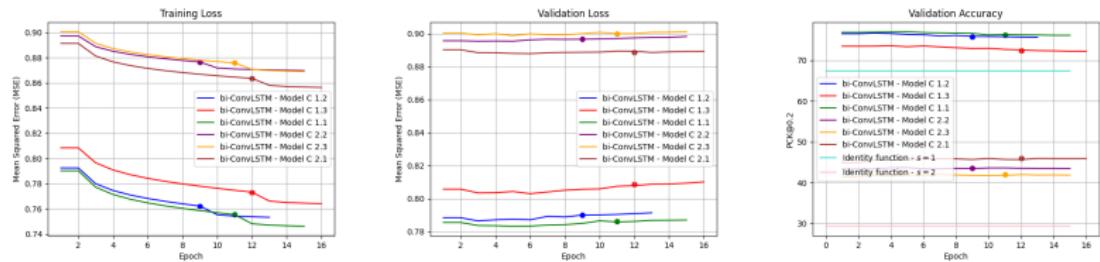


(b) Pretraining results of DeciWatch

# Pretraining evolution - corrected

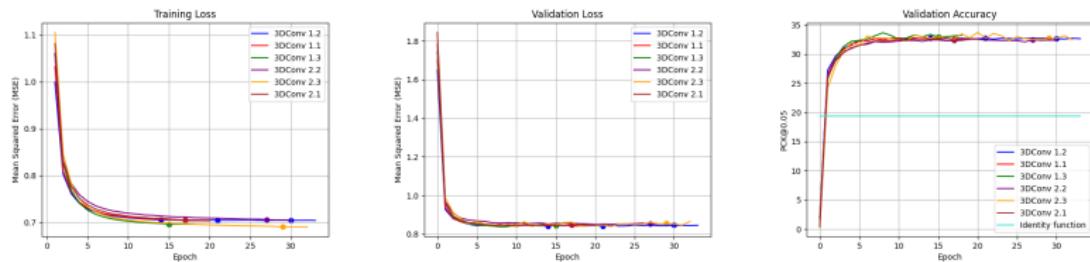


(a) Pretraining results of bi-Conv LSTM Model S.

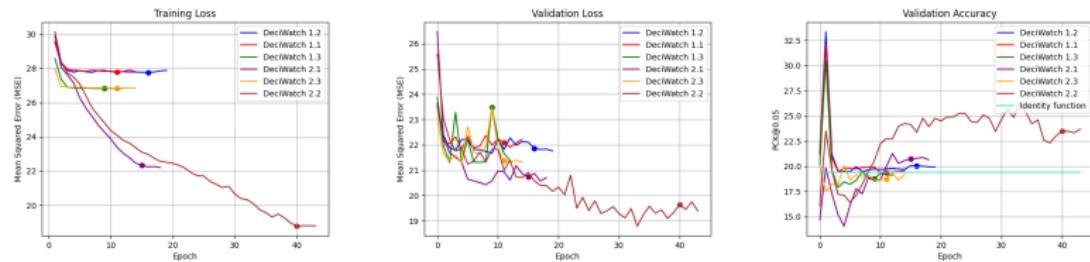


(b) Pretraining results of bi-Conv LSTM Model C.

# Finetuning evolution - corrected

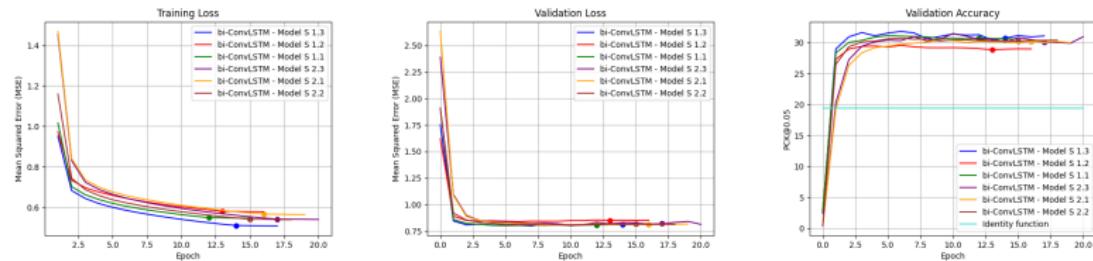


(a) Finetuning results of 3DConv

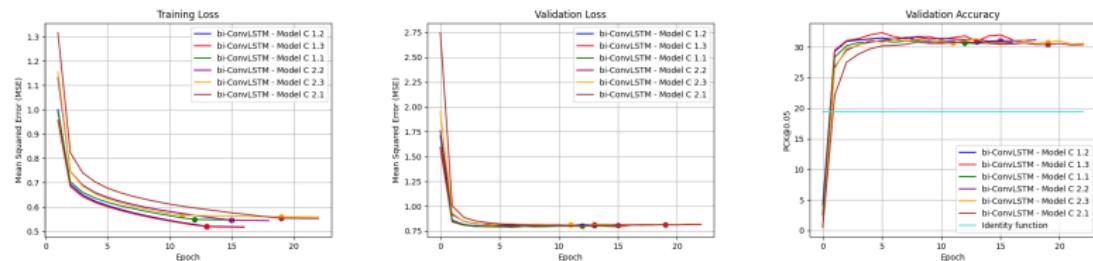


(b) Finetuning results of DeciWatch

# Finetuning evolution - corrected



(a) Finetuning results of bi-Conv LSTM Model S.



(b) Finetuning results of bi-Conv LSTM Model C.

# Finetuning keypoint testing results - corrected

Experiment	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	1.1	1.2	1.3	1.1	1.2	1.3	1.1	1.2	1.3	1.1	1.2	1.3	
Nose	56.6	56.6	56.6	50.7	50.3	48.7	57.5	56.1	54.6	54.9	56.6	56.1	54.6
Ear	73.9	73.8	75.3	68.2	67.7	65.0	74.4	74.6	73.6	77.0	73.8	74.8	72.7
Shoulder	74.7	75.1	75.0	70.2	70.5	66.1	74.6	75.5	74.6	74.3	75.2	74.9	73.4
Elbow	71.8	72.2	71.6	64.2	63.1	60.7	70.0	68.3	70.4	71.4	68.2	69.0	68.4
Wrist	71.4	71.4	71.5	65.8	66.3	61.3	69.9	70.6	68.3	71.2	71.2	70.1	69.1
Pinky	51.4	51.4	52.3	51.4	52.5	48.7	51.8	52.6	54.0	55.1	54.7	54.7	52.6
Index finger	59.9	59.9	60.5	57.1	56.6	51.1	56.4	58.9	55.2	57.4	60.3	59.0	57.7
Thumb	55.5	54.8	57.8	54.0	54.6	50.2	51.4	55.1	55.4	54.6	58.0	53.3	54.6
Hip	76.4	76.4	77.2	72.1	71.8	67.8	75.7	75.5	75.6	75.7	75.7	74.9	74.6
Knee	82.0	82.0	81.0	77.3	77.0	72.0	80.4	81.8	81.2	81.5	81.3	80.3	79.8
Ankle	92.0	91.5	91.6	86.9	87.7	79.8	91.3	92.1	92.3	92.2	91.9	91.1	90.0
Heel	90.5	90.5	90.6	84.0	84.7	76.9	90.4	45.1	90.7	89.8	90.0	89.7	84.4
Toes	77.9	77.8	78.0	73.1	73.7	70.6	77.5	76.0	77.1	76.8	77.6	77.5	76.1
Total	72.5	72.4	73.1	68.0	68.1	62.7	71.5	68.3	71.3	72.2	72.2	71.4	

# Finetuning keypoint testing results - corrected

Experiment	3DConv			DeciWatch			bi-ConvLSTM Model S			bi-ConvLSTM Model C			Total
	2.1	2.2	2.3	2.1	2.2	2.3	2.1	2.2	2.3	2.1	2.2	2.3	
Nose	56.5	56.5	55.5	49.1	50.5	50.9	55.7	58.7	52.9	54.6	57.5	56.9	54.6
Ear	74.1	71.7	74.3	55.2	66.1	66.8	73.9	73.2	73.7	75.5	75.9	74.6	71.3
Shoulder	75.3	74.5	74.1	61.9	69.1	67.7	72.9	74.5	72.1	73.4	73.7	74.5	72.0
Elbow	71.8	72.3	71.2	51.9	56.0	61.7	68.3	69.5	68.9	69.5	69.9	69.4	66.7
Wrist	71.3	71.4	70.8	56.0	56.6	36.3	70.2	70.8	69.5	68.7	70.6	70.2	65.2
Pinky	51.6	51.3	51.6	40.4	40.0	9.24	53.4	51.2	52.5	52.7	52.3	54.7	46.7
Index finger	59.8	59.9	59.6	48.7	46.1	32.5	56.3	57.5	56.4	58.4	58.1	56.3	54.1
Thumb	54.8	54.4	55.6	43.2	44.2	26.0	54.5	52.5	54.8	54.3	53.7	55.3	50.3
Hip	76.2	76.2	76.6	57.5	66.7	67.0	74.9	75.7	74.0	74.6	75.2	73.8	72.4
Knee	81.7	81.6	80.3	68.8	74.5	70.8	81.4	80.8	81.3	82.1	81.0	80.9	78.8
Ankle	91.5	90.4	91.1	42.9	65.4	44.5	91.7	91.6	90.4	91.7	92.0	91.5	81.3
Heel	90.6	90.7	89.4	41.4	62.5	35.1	90.1	90.2	89.5	89.5	89.6	90.3	79.1
Toes	77.6	77.1	77.6	49.1	62.3	57.6	75.3	75.1	75.6	76.4	76.3	77.1	71.4
Total	72.3	72.1	72.6	51.3	58.7	48.1	71.2	71.6	70.7	71.8	72.0	71.9	