Détection d'événements à partir de capteurs sols – application au suivi de personnes fragiles

Soutenance de thèse

Ludovic Minvielle

Directeur: Nicolas Vayatis, Centre Borelli, ENS Paris-Saclay
Co-encadrante: Mathilde Mougeot, Centre Borelli, ENS Paris-Saclay

Thèse industrielle entre l'ENS Paris-Saclay et Tarkett

Mercredi 15 Juillet 2020

Table of contents

1. Introduction

2. A tour of monitoring systems
    - Systems
    - Processing

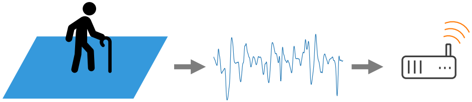
3. Step proposal network
    - Region proposal network
    - Main architecture
    - Signal embedding
    - Step proposal network
    - Results

4. Conclusion

Introduction

Context

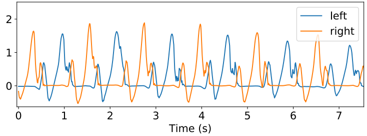
- Elderly population is growing
- Higher levels of frailty globally
- Increasing demand for reliable monitoring devices
- Tarkett, French company with 12,500 employees, 13 industrial sites, sells 1.3 millions m^2 of flooring every day
- *Floor in Motion*: a floor-based sensor for elderly care
- **Objective**: providing tools for elderly monitoring in nursing homes
    - First aimed application: fall detection



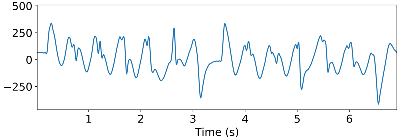
Introduction

Motivation

- Processing and understanding time series
  - Proliferation of sensor-based systems
  - Redundancy, interpretability, external pertubations
- Real world application
  - Real-time processing in a limited system
  - Convenient hypotheses not granted



(a) Foot-attached accelerometer



(b) Tarkett's floor sensor

A tour of monitoring systems

Systems

What makes a good monitoring system ?

- coverage and occlusion
- intrusiveness
- signal quality / information
- robustness
- ease of installation / use
- scalability

Criteria	RGB cam	Depth cam	Wearable	Acoustic	Radar / Wi-Fi	Vibration	Floor
Coverage/Occlusion							
Intrusiveness							
Signal quality / info							
Robustness			★★★	★☆☆	★☆☆	★☆☆	★★☆
Ease of instal. / use			★★☆	★★☆	★★☆	★★★	★☆☆
Scalability			★★★	★★☆	★☆☆	★★☆	★★★

A tour of monitoring systems

Systems

What makes a good monitoring system ?

- coverage and occlusion
- intrusiveness
- signal quality / information
- robustness
- ease of installation / use
- scalability

Signal acquisition

Vision

RGB-Camera
Omni-Camera
Depth-Camera

Criteria	RGB cam	Depth cam	Wearable	Acoustic	Radar / Wi-Fi	Vibration	Floor
Coverage/Occlusion	★☆☆	★☆☆					
Intrusiveness	★☆☆	★☆☆					
Signal quality / info	★★★	★★★					
Robustness	★★☆	★★★	★★★	★☆☆	★☆☆	★☆☆	★★☆
Ease of instal. / use	★☆☆	★☆☆	★★☆	★★☆	★★☆	★★★	★☆☆
Scalability	★☆☆	★☆☆	★★★	★★☆	★☆☆	★★☆	★★★

A tour of monitoring systems

Systems

What makes a good monitoring system ?

- coverage and occlusion
- intrusiveness
- signal quality / information
- robustness
- ease of installation / use
- scalability

Signal acquisition

Vision	Wearable
RGB-Camera Omni-Camera Depth-Camera	Accelerometer Gyroscope Barometric pressure

Criteria	RGB cam	Depth cam	Wearable	Acoustic	Radar / Wi-Fi	Vibration	Floor
Coverage/Occlusion	★☆☆	★☆☆	★★★				
Intrusiveness	★☆☆	★☆☆	★★☆				
Signal quality / info	★★★	★★★	★★☆				
Robustness	★★☆	★★★	★★★	★☆☆	★☆☆	★☆☆	★★☆
Ease of instal. / use	★☆☆	★☆☆	★★☆	★★☆	★★☆	★★★	★☆☆
Scalability	★☆☆	★☆☆	★★★	★★☆	★☆☆	★★☆	★★★

A tour of monitoring systems

Systems

What makes a good monitoring system ?

- coverage and occlusion
- intrusiveness
- signal quality / information
- robustness
- ease of installation / use
- scalability

Signal acquisition

Vision	Wearable	Ambient
RGB-Camera	Accelerometer	Microphone
Omni-Camera	Gyroscope	Radar
Depth-Camera	Barometric pressure	Wi-Fi
		Vibrational
		Pressure

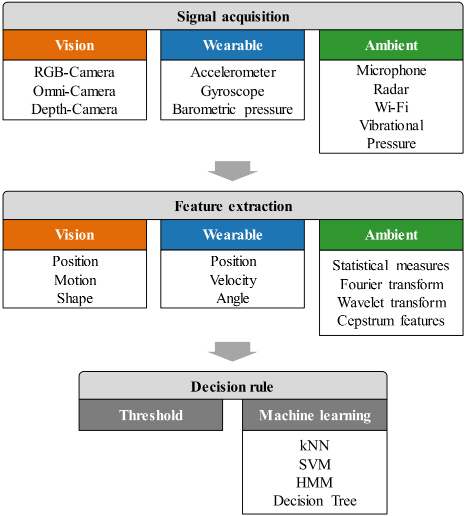
Criteria	RGB cam	Depth cam	Wearable	Acoustic	Radar / Wi-Fi	Vibration	Floor
Coverage/Occlusion	★☆☆	★☆☆	★★★	★★☆	★☆☆	★★★	★★★
Intrusiveness	★☆☆	★☆☆	★★☆	★☆☆	★★☆	★★★	★★★
Signal quality / info	★★★	★★★	★★☆	★★☆	★☆☆	★★☆	★★☆
Robustness	★★☆	★★★	★★★	★☆☆	★☆☆	★☆☆	★★☆
Ease of instal. / use	★☆☆	★☆☆	★★☆	★★☆	★★☆	★★★	★☆☆
Scalability	★☆☆	★☆☆	★★★	★★☆	★☆☆	★★☆	★★★

A tour of monitoring systems

Processing

How to process the inputs ?

-

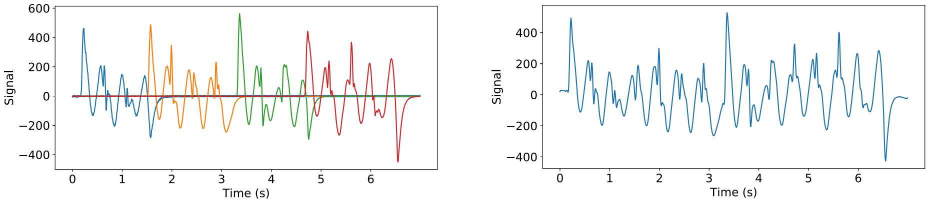


Introduction

- **Issue**: one-dimensional signals for large areas
- Goal: Classify elderly from other individuals
  - Most signals are made of walks of staff individuals
- **Subtask**: Bring the model's attention over step-related signals
- A model to recognize steps ?

Introduction

- **Issue**: one-dimensional signals for large areas
- Goal: Classify elderly from other individuals
  - Most signals are made of walks of staff individuals
- **Subtask**: Bring the model's attention over step-related signals
- A model to recognize steps ?



(a) Raw signal (b) Preprocessed signal

Figure: Healthy individual walking on the sensor.

- Signals are complex
- How to **localize** steps ?
- **This presentation:** A step detector using convolutional neural network: Step Proposal Network

1. Introduction

2. A tour of monitoring systems
   - Systems
   - Processing

3. Step proposal network
   - Region proposal network
   - Main architecture
   - Signal embedding
   - Step proposal network
   - Results

4. Conclusion

Region proposal network

Object detection

- Classification: What is the image class ?

Figure: Classification vs Object detection. Source: [Girshick et al., 2014], [Ren et al., 2015]

Region proposal network

Object detection

- Classification: What is the image class ?
- Object detection: Where are the objects and what are they classes ?

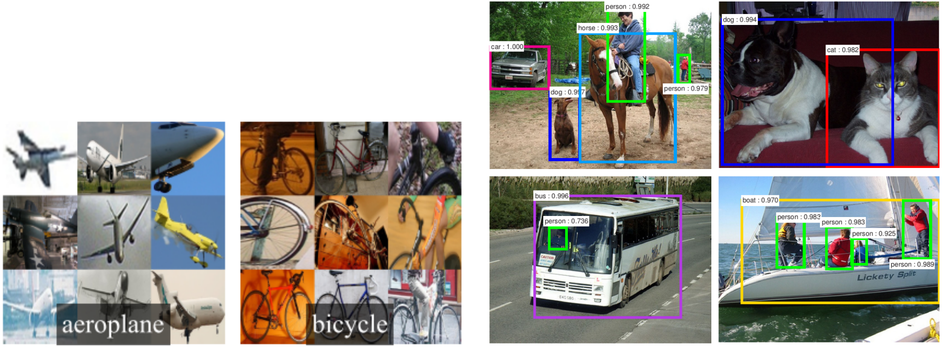


Figure: Classification vs Object detection. Source: [Girshick et al., 2014], [Ren et al., 2015]

Region proposal network

Object detection

- Classification: What is the image class ?
- Object detection: Where are the objects and what are they classes ?
- How to efficiently localize objects ?
- Proposal models [Hosang et al., 2016]
- Faster R-CNN [Ren et al., 2015]

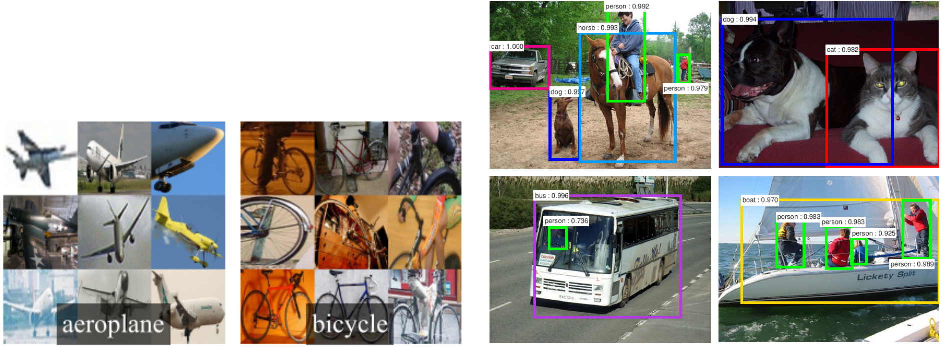


Figure: Classification vs Object detection. Source: [Girshick et al., 2014], [Ren et al., 2015]

Region proposal network

Faster R-CNN

- Main idea: proposals are generated by a CNN called Region Proposal Network

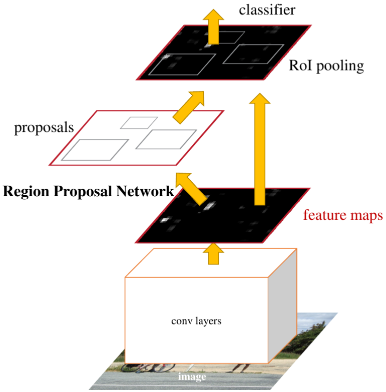


Figure: Region proposal network. Source: [Ren et al., 2015]

Region proposal network

Faster R-CNN

- Main idea: proposals are generated by a CNN called Region Proposal Network
- A sliding window is passed: multiple *anchors* over each location (various sizes and scales)
- Two layers: Classification (Object / Not Object) and Regression (anchor coordinates)

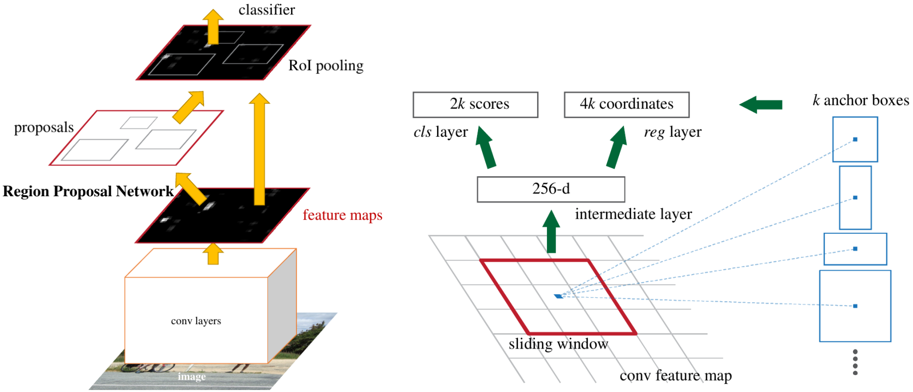


Figure: Region proposal network. Source: [Ren et al., 2015]

1. Introduction

2. A tour of monitoring systems
 Systems
 Processing

3. Step proposal network
 Region proposal network
 Main architecture
 Signal embedding
 Step proposal network
 Results

4. Conclusion

Step proposal network

Main architecture

- Directly inspired from RPN
- Simple architecture with three hidden layers, all **convolutional**
- Output: probability of having a step at a specific window location and size
  - Here 3 sizes and all discrete locations are considered

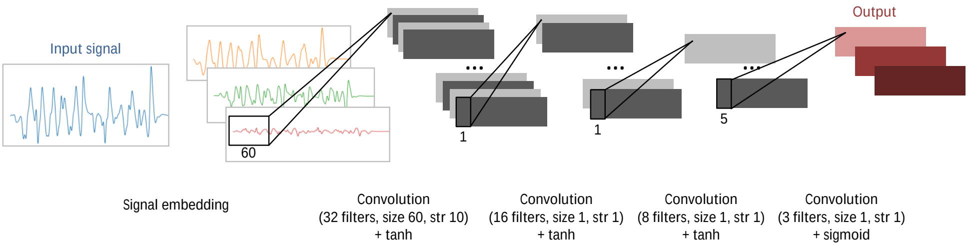


Figure: Architecture of SPN .

Step proposal network

Main architecture

- Directly inspired from RPN
- Simple architecture with three hidden layers, all **convolutional**
- Output: probability of having a step at a specific window location and size
  - Here 3 sizes and all discrete locations are considered

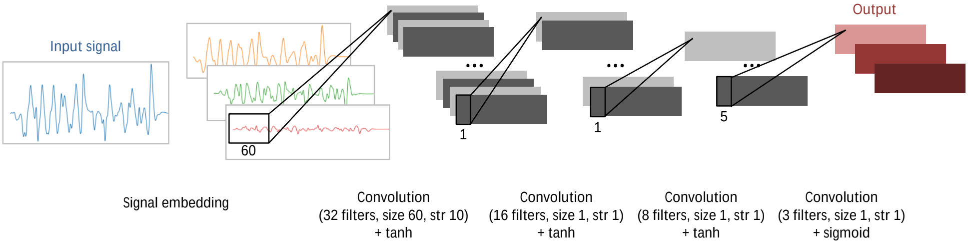


Figure: Architecture of SPN .

- Use the convolutional representation to "boost" training
- First layer (Signal embedding) of SPN is trained **separately** using convolutional dictionary learning

1. Introduction

2. A tour of monitoring systems
   - Systems
   - Processing

3. Step proposal network
   - Region proposal network
   - Main architecture
   - Signal embedding
   - Step proposal network
   - Results

4. Conclusion

Signal embedding

Convolutional dictionary learning

- $\mathbf{s}$: data to be represented
- Objective : find M atoms $\mathbf{d}_m$ and activation signals $\mathbf{x}_m$ such that

$$\mathbf{s} \approx \sum_{m=1}^{M} \mathbf{x}_m * \mathbf{d}_m$$

- $*$: convolution

Signal embedding

Convolutional dictionary learning

- $\mathbf{s}$: data to be represented
- Objective : find M atoms $\mathbf{d}_m$ and activation signals $\mathbf{x}_m$ such that

$$\mathbf{s} \approx \sum_{m=1}^{M} \mathbf{x}_m * \mathbf{d}_m$$

- $*$: convolution

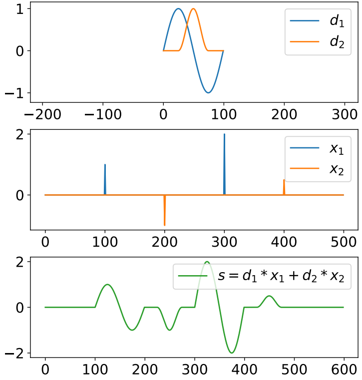


Figure: Convolutional dictionary learning.

Signal embedding

Convolutional dictionary learning

- $\mathbf{s}$: data to be represented
- Objective : find M atoms $\mathbf{d}_m$ and activation signals $\mathbf{x}_m$ such that

$$\mathbf{s} \approx \sum_{m=1}^{M} \mathbf{x}_m * \mathbf{d}_m$$

- $*$: convolution

CDL general problem:

$$\underset{\mathbf{x}_m, \mathbf{d}_m}{\arg\min} \frac{1}{2} \left\| \sum_{m=1}^{M} \mathbf{x}_m * \mathbf{d}_m - \mathbf{s} \right\|_2^2 + \lambda \sum_{m=1}^{M} \|\mathbf{x}_m\|_1$$
$$\text{s.t.} \quad \|\mathbf{d}_m\|_2 \leq 1 \quad \forall m \,.$$

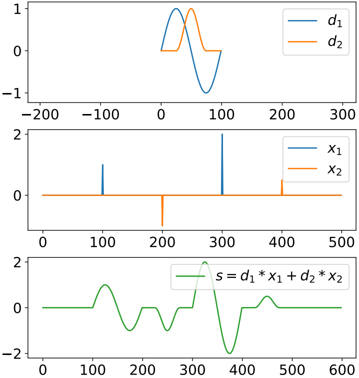


Figure: Convolutional dictionary learning.

Signal embedding

Learning step atoms

- Learning with Alternating Direction Method of Multipliers (ADMM) [Bristow et al., 2013]
- 3 atoms of length 0.7 second

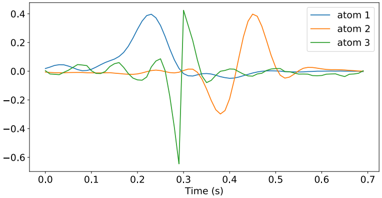


Figure: Dictionary.

Signal embedding

Learning step atoms

- Learning with Alternating Direction Method of Multipliers (ADMM) [Bristow et al., 2013]
- 3 atoms of length 0.7 second

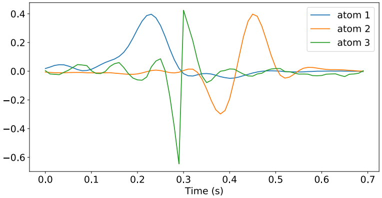
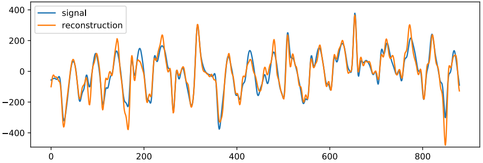
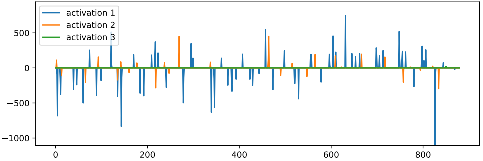


Figure: Dictionary.



(a) Original signal and its reconstruction



(b) Signal activations

Signal embedding

Learning step atoms

- Learning with Alternating Direction Method of Multipliers (ADMM) [Bristow et al., 2013]
- 3 atoms of length 0.7 second
- Use the following embedding:

$$\mathbf{S} \doteq \left(\mathbf{s} * \mathbf{d}_m\right)_{1 \leq m \leq 3}$$

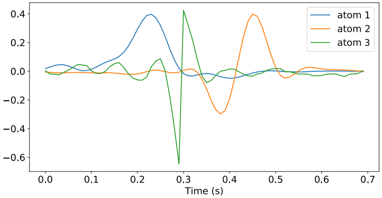
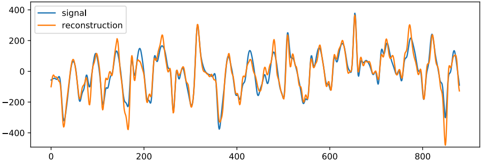
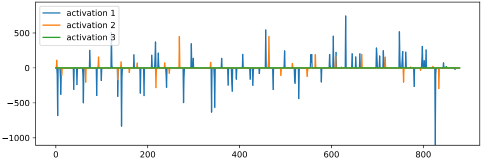


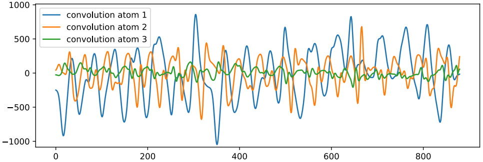
Figure: Dictionary.



(a) Original signal and its reconstruction



(b) Signal activations



Signal embedding

Learning step atoms

- Learning with Alternating Direction Method of Multipliers (ADMM) [Bristow et al., 2013]
- 3 atoms of length 0.7 second
- Use the following embedding:

$$\mathbf{S} \doteq \left(\mathbf{s} * \mathbf{d}_m\right)_{1 \leq m \leq 3}$$

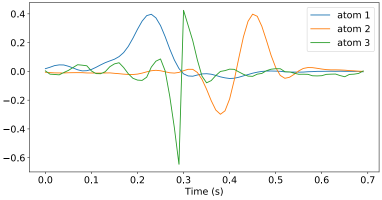
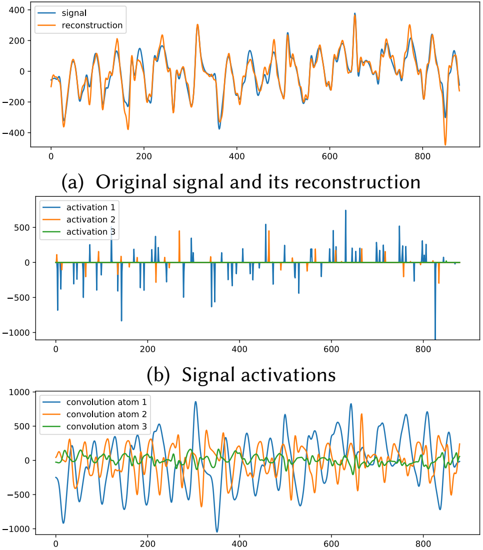


Figure: Dictionary.



(a) Original signal and its reconstruction

(b) Signal activations

1. Introduction

2. A tour of monitoring systems
   - Systems
   - Processing

3. Step proposal network
   - Region proposal network
   - Main architecture
   - Signal embedding
   - Step proposal network
   - Results

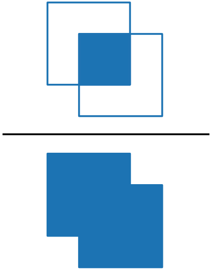
4. Conclusion

Step proposal network

Principle

- Objective of SPN : output boxes with largest Intersection over Union (IoU)
- IoU: $\mathbf{b_j}$ are labelled boxes, $\hat{b}$ is an estimated box:

$$\mathrm{IoU}(\hat{b}) \doteq \max_j \frac{|\mathbf{b_j} \cap \hat{b}|}{|\mathbf{b_j} \cup \hat{b}|}$$

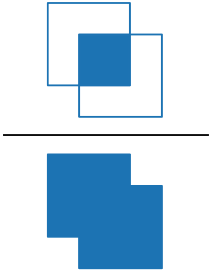


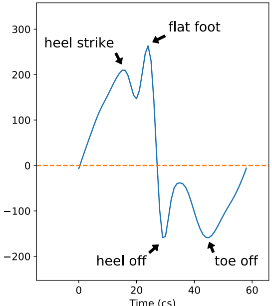
Step proposal network

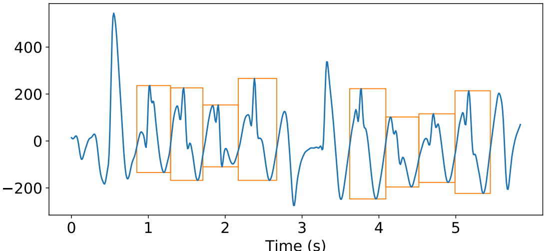
Principle

- Objective of SPN : output boxes with largest Intersection over Union (IoU)
- IoU: $\mathbf{b_j}$ are labelled boxes, $\hat{b}$ is an estimated box:

$$\mathrm{IoU}(\hat{b}) \doteq \max_j \frac{|\mathbf{b_j} \cap \hat{b}|}{|\mathbf{b_j} \cup \hat{b}|}$$







Step proposal network

Training

- Output: a matrix $\mathbf{W} \in \mathbb{R}^{T \times K}$
  - T: signal length
  - K: number of different box sizes
- $\mathbf{W}_{t,k}$: probability that the box b_t^k starting at time t and of size 0.4s, 0.5s, or 0.6s (for respectively $k = 1, 2,$ or 3) has a large IoU score
- Positive boxes: $\mathrm{IoU}(b_t^k) > \sqrt{0.7}$
- Negative boxes: $\mathrm{IoU}(b_t^k) < \sqrt{0.3}$
- Other are not used for training

The loss function $\mathcal{L}$ over a signal $\mathbf{s}$ is defined as:

$$\mathcal{L}(\mathbf{s}, \mathbf{W}) = \sum_t \sum_{k \in [1,2,3]} \mathbb{1}_{\mathrm{IoU}(b_t^k) > \sqrt{0.7}} \log(\mathbf{W}_{t,k}) + \mathbb{1}_{\mathrm{IoU}(b_t^k) < \sqrt{0.3}} \log(1 - \mathbf{W}_{t,k}).$$

1. Introduction

2. A tour of monitoring systems
   - Systems
   - Processing

3. Step proposal network
   - Region proposal network
   - Main architecture
   - Signal embedding
   - Step proposal network
   - Results

4. Conclusion

Results

Data

- 43 signals recorded in a nursing home
- Manually labeled steps

Training

- SPN is trained using classical gradient descent
- Training time: < 5 minutes
- Inference (detection over a 10s signal): < 1 second
- Optimization details
  - learning rate of 10^{-3}
  - learning rate decay ($\times 0.9$ every 10 epochs)
  - Nesterov momentum

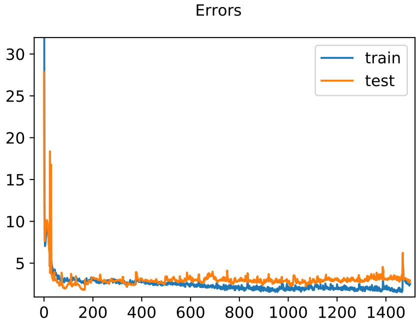


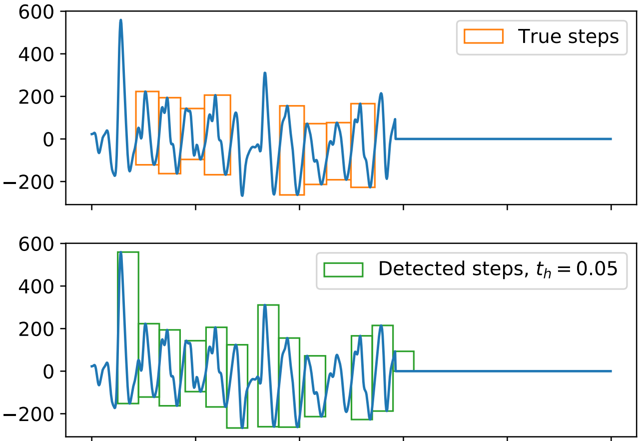
Figure: SPN training and testing errors.

Results

- Object detection use the mean Average Precision (mAP): area under the Precision-Recall curve
- **Without** embedding, mAP = 72,5%
- **With** embedding, mAP = 78,6%

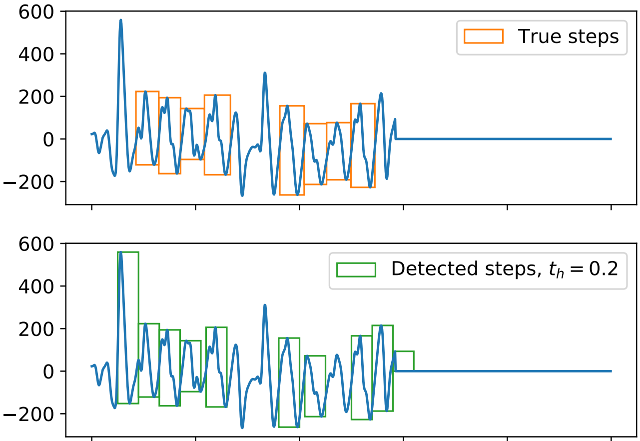
Results

- Object detection use the mean Average Precision (mAP): area under the Precision-Recall curve
- **Without** embedding, mAP = 72,5%
- **With** embedding, mAP = 78,6%



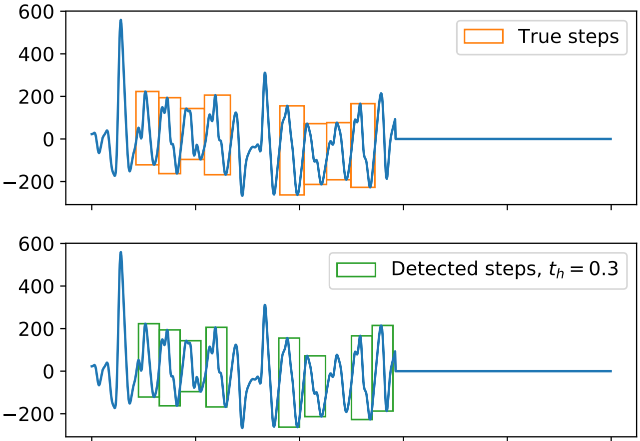
Results

- Object detection use the mean Average Precision (mAP): area under the Precision-Recall curve
- **Without** embedding, mAP = 72,5%
- **With** embedding, mAP = 78,6%



Results

- Object detection use the mean Average Precision (mAP): area under the Precision-Recall curve
- **Without** embedding, mAP = 72,5%
- **With** embedding, mAP = 78,6%



Conclusion

Conclusion

- SPN uses the convolutional representation to detect steps
- Allows to located steps in complex signals
- Training and inference are fast

Future work

- Add a regression layer on the step proposals
- Tests on step detection benchmarks data sets

Conclusion

Conclusion

- SPN uses the convolutional representation to detect steps
- Allows to located steps in complex signals
- Training and inference are fast

Future work

- Add a regression layer on the step proposals
- Tests on step detection benchmarks data sets

Thanks

Contact: minvielle@cmla.ens-cachan.fr

Reference: L. Minvielle and J. Audiffren. Nursenet : Monitoring elderly levels of activity with a piezoelectric floor. Sensors, 19(18), 2019 link

References

Bristow, H., Eriksson, A., and Lucey, S. (2013).
Fast convolutional sparse coding.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014).
Rich feature hierarchies for accurate object detection and semantic segmentation.
In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587.

Hosang, J., Benenson, R., Dollár, P., and Schiele, B. (2016).
What makes for effective detection proposals?
IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(4):814–830.

Ren, S., He, K., Girshick, R., and Sun, J. (2015).
Faster r-cnn: Towards real-time object detection with region proposal networks.
In *Advances in neural information processing systems*, pages 91–99.