HDA COURSE PROJECTS

Instructor

Michele Rossi - michele.rossi@unipd.it

Lab. classes

Francesca Meneghello - francesca.meneghello@dei.unipd.it

Silvia Zampato

MIME
Master's degree ICT Internet Multimedia Engineering

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Exam dates and submission deadlines

- Exam: February 1-2, 2023
  - report+code submission deadline: January 30, 2023
- Exam: February 15-16, 2023
  - report+code submission deadline: February 13, 2023
- Exam: July 5, 2023
  - report+code submission deadline: July 3, 2023
- Exam: July 19, 2023
  - report+code submission deadline: July 17, 2023
- Exam: September 13, 2023
  - report+code submission deadline: September 11, 2023

Guidelines

- Max 2 people per group
  - and both have to contribute to the work
- For the final evaluation you need to
  - submit a report (see the template on Moodle)
  - submit the code of the implementation
  - prepare a presentation through slides (20 minutes strict, possibly including a demo)
- The evaluation will consider different aspects
  - about the report, the presentation …
  - see the details in the Latex template for the final report (Moodle)

Guidelines

- Be creative!
  - We provided you with some ideas for possible project developments, but original works are always welcome!
  - You can use the neural network architectures seen during the labs but…experiment with new approaches!
  - Pre-processing techniques may be useful
  - We provide you with some references but try to explore a little bit other contributions in the literature that may be helpful
  - Implement your own neural network architecture…DO NOT simply use pre-trained models from Keras: the objective of the project is that you put into practice the things you learned during the theoretical lessons, not to improve your qualities of reusing other networks/code

Guidelines

- Prepare the project and the report considering the grid we use for the exam evaluation (see the LaTex template in Moodle)
  - pay attention to the pre-processing phase
  - create an original neural network architecture
  - evaluate the performance of the algorithms in terms of running time and complexity (memory occupation)
  - compare the performance of different approaches

Guidelines

- Some projects require you to collect **custom datasets**
  - available sensors:
    - 3 pulse oximeter
    - 8 heart monitors
    - 2 respiration belts
  - long story short… **reserve your spot!**
  - start soon to acquire your dataset!

Proposed Projects

PART A – ON BODY AND ENVIRONMENTAL SENSORS

1) A1: Activity recognition with four accelerometers
2) A2: Pathological gait recognition
3) A3: Sleep posture monitoring
4) A4: Subject identification from ECG, PPG, and respiratory signal

collect your own dataset

PART B – AUDIO SIGNALS

1) B1: Speech command recognition (keyword spotting)
2) B2: Environmental sound classification

PART C – IMAGES

1) C2: Lymphoma subtype classification
2) C3: Bone age prediction from hand radiographs
3) C4: Lung disease prediction from X-ray images

PART D – WIRELESS SIGNALS

1) D1: Activity recognition through Wi-Fi devices
2) D2: Person identification through Wi-Fi devices
3) D3: People counting through Wi-Fi devices

Master's degree ICT Internet Multimedia Engineering

PART A: ON BODY AND ENVIRONMENTAL SENSORS

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

1222·2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

Master's degree ICT Internet Multimedia Engineering

PROJECT A1

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

1222·2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. A1 “activity recognition with four accelerometers”

Reference papers

[Fadel19] Fadel, W. F., Urbanek, J. K., Albertson, S. R., Li, X., Chomistek, A. K., & Harezlak, J., Differentiating Between Walking and Stair Climbing Using Raw Accelerometry Data, in Statistics in Biosciences, 11(2), 334–354, 2019.

[Karas19] Karas, M., Bai, J., Strączkiewicz, M., Harezlak, J., Glynn, N. W., Harris, T., … Urbanek, J. K., Accelerometry Data in Health Research: Challenges and Opportunities. Review and Examples, in Statistics in Biosciences, 11, 210–23, 2019.

Dataset (760.8 MB uncompressed)

https://physionet.org/content/accelerometry-walk-climb-drive/1.0.0/

Why is activity recognition important?

- Navigation systems
  - adapt to user movement
  - e.g., predict direction and only use that portion of the map(s)
  - put the system into power saving mode when there is no mobility
- First responders
  - security personnel, firefighters
  - e.g., who has to be assisted first
- Assisted living
  - react to reduced activity levels
  - unusual mobility patterns
  - user motion-aware services and/or environments
- Rehabilitation
  - measure recovery of motor functions
  - measure effectiveness of rehabilitation
- In most of these cases the use of cameras is not possible
- The system has to be: unobtrusive, lightweight, portable, …

Dataset description

- Four 3-axial ActiGraph GT3X+ wearable accelerometers at
  - left ankle
  - right ankle
  - left hip
  - left wrist
- Dataset: collected from 13 males & 19 females subjects aged between 23 and 52
- Six activities considered
  - 1=walking
  - 2=descending stairs
  - 3=ascending stairs
  - 4=driving
  - 77=clapping
  - 99=non-study activity

Dataset description

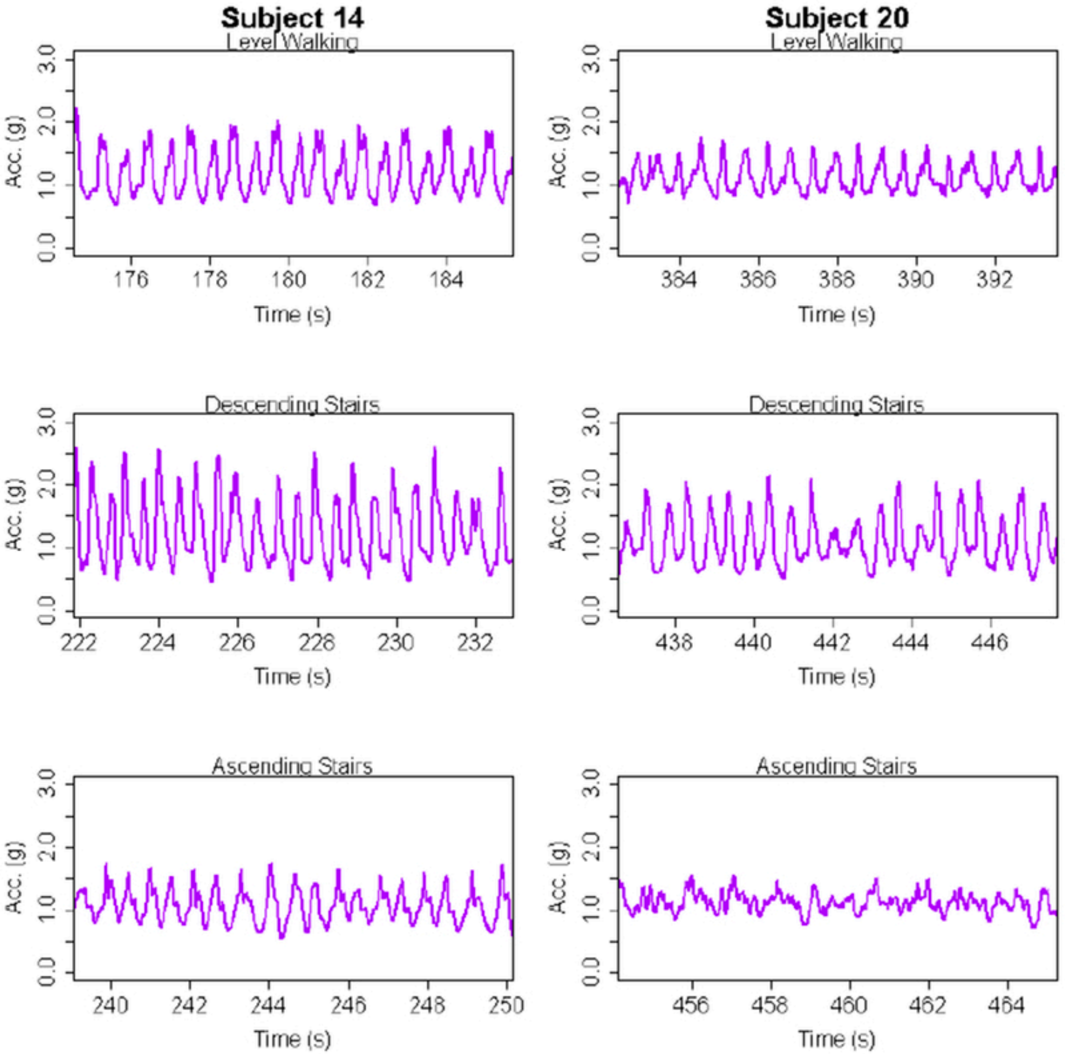
- Dataset format → Text file: 14 variables
  - activity: Type of activity
  - time_s: Time from device initiation (seconds [s])
  - lw_x: Left wrist x-axis measurement (gravitation acceleration [g])
  - lw_y: Left wrist y-axis measurement (gravitation acceleration [g])
  - lw_z: Left wrist z-axis measurement (gravitation acceleration [g])
  - lh_x: Left hip x-axis measurement (gravitation acceleration [g])
  - lh_y: Left hip y-axis measurement (gravitation acceleration [g])
  - lh_z: Left hip z-axis measurement (gravitation acceleration [g])
  - la_x: Left ankle x-axis measurement (gravitation acceleration [g])
  - la_y: Left ankle y-axis measurement (gravitation acceleration [g])
  - la_z: Left ankle z-axis measurement (gravitation acceleration [g])
  - ra_x: Right ankle x-axis measurement (gravitation acceleration [g])
  - ra_y: Right ankle y-axis measurement (gravitation acceleration [g])
  - ra_z: Right ankle z-axis measurement (gravitation acceleration [g])

Dataset description

- No information is provided to convert the data into the global frame

- In [Fadel19] authors consider the vector magnitude to remove the effects of the sensor orientation

$$vm(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2}.$$



Possible project developments

- Walking VS stairs climbing
  - 3 classes classification problem
- Features
  - consider the first three activities as in [Fadel19]
  - use the magnitude or try to use the information on the 3 axes
  - use the features defined in [Fadel19], or
  - compute features with FFT/DCT in an audio-like fashion, or
  - use raw signals with automatic feature extraction
- Classification architecture
  - CNN or RNN…

PROJECT A2

Project no. A2 “pathological gait recognition”

Reference papers

[Jun20_1] K. Jun, Y. Lee, S. Lee, D.-W. Lee, and M. S. Kim, Pathological Gait Classification Using Kinect v2 and Gated Recurrent Neural Networks, IEEE Access, vol. 8, pp. 139881-139891, 2020.

[Jun20_2] K. Jun, D. W. Lee, K. Lee, S. Lee, and M. S. Kim, Feature Extraction Using an RNN Autoencoder for Skeleton-based Abnormal Gait Recognition, IEEE Access, vol. 8, pp. 19196-19207, 2020.

[Lee19] D. W. Lee, K. Jun, S. Lee, J. K. Ko, and M. S. Kim, Abnormal gait recognition using 3D joint information of multiple Kinects system and RNN-LSTM, in Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019.

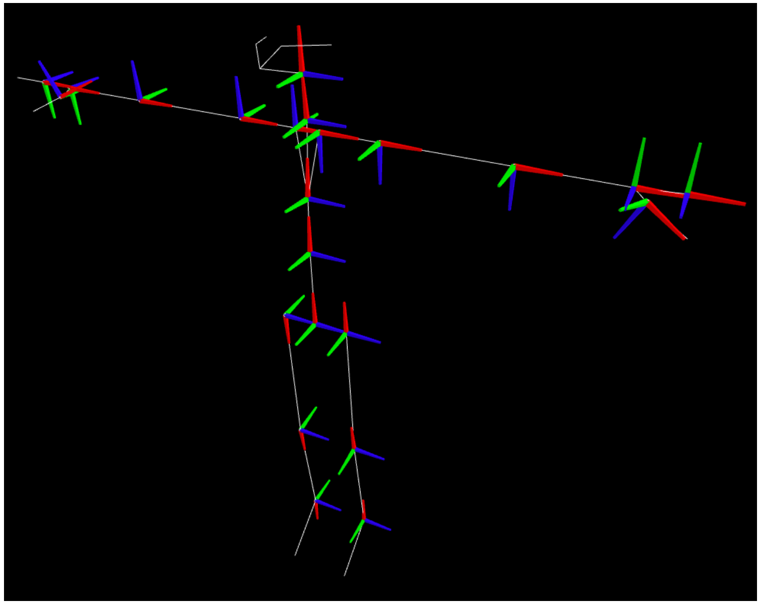
Dataset (111.22 MB uncompressed)

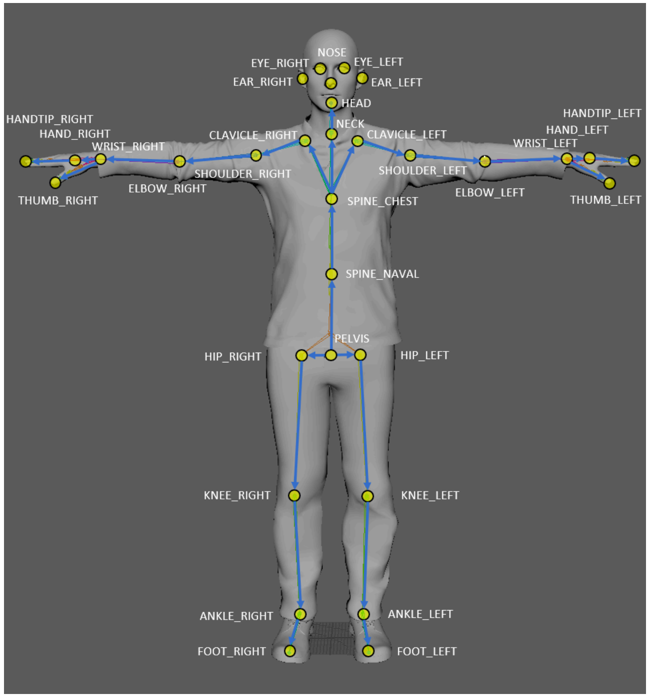
https://drive.google.com/file/d/1BLUXyk_59agThysXwamfVyZahSI-sScl/view?usp=sharing

https://ieee-dataport.org/documents/azure-kinect-3d-skeleton-and-foot-pressure-data-pathological-gaits

Dataset description

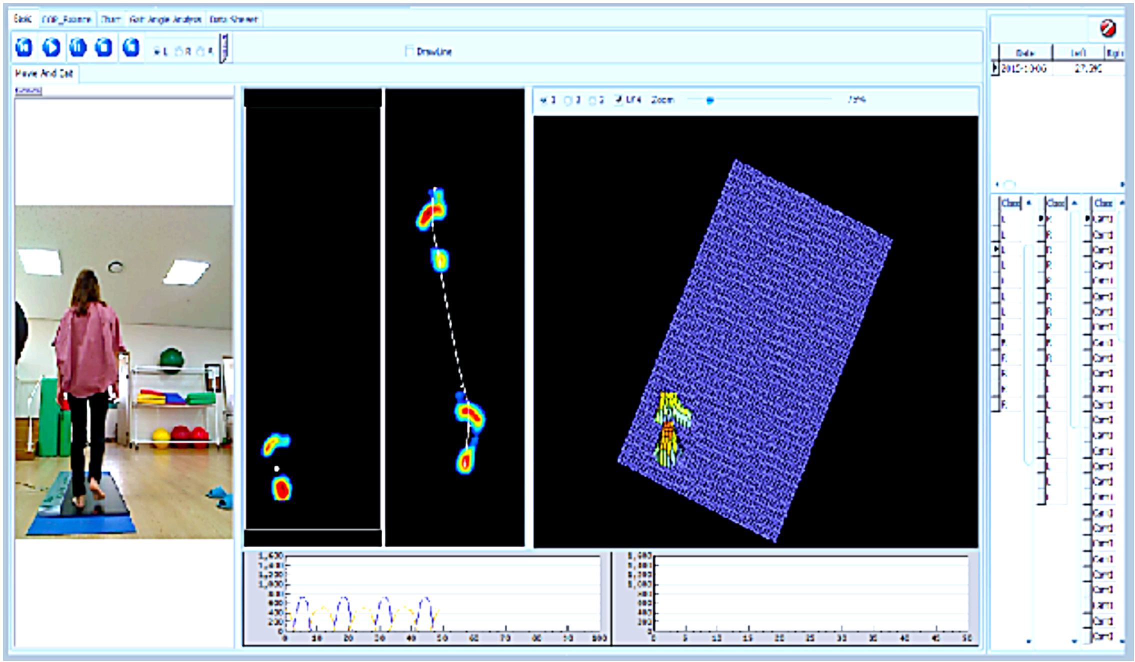
- Sequential skeleton data
  - Azure Kinect (Microsoft Corp. Redmond, WA, USA)
  - https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints





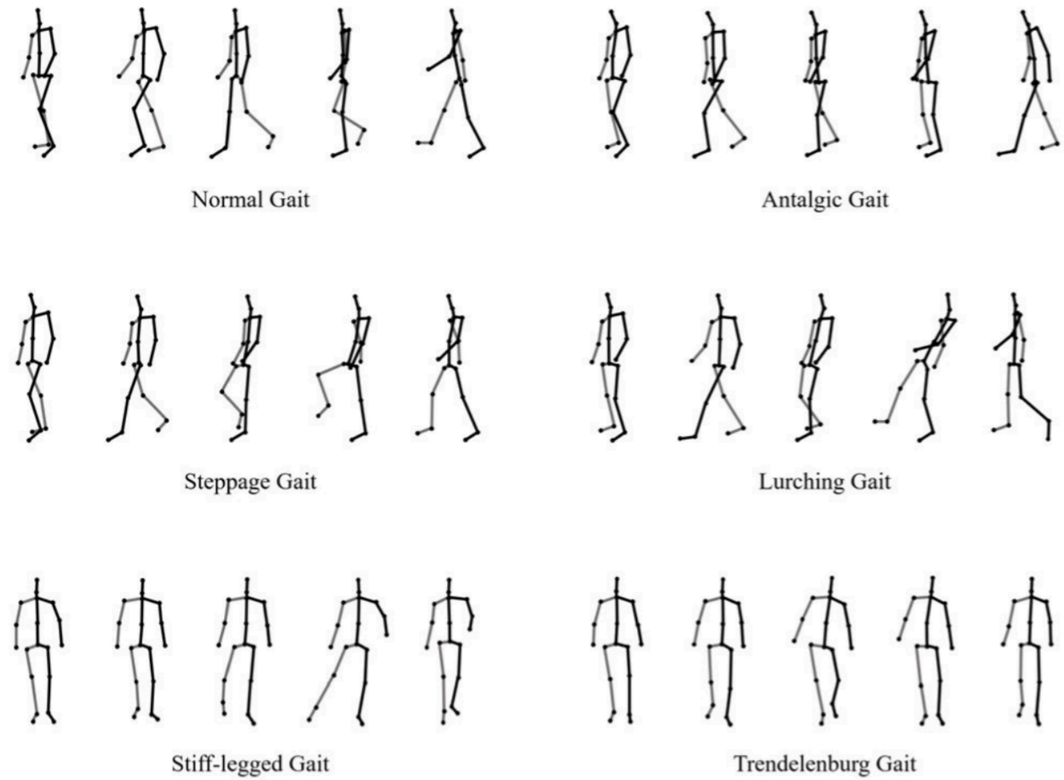
Dataset description

- Average foot pressure data
  - GW1100 (GHIWell, Korea)
  - size = 48 x 128
  - GW1100 is a 1080mm x 480mm sized pressure plate and contains 6,144 high-voltage matrix sensors with maximum pressure 100 N/cm^2



Dataset description

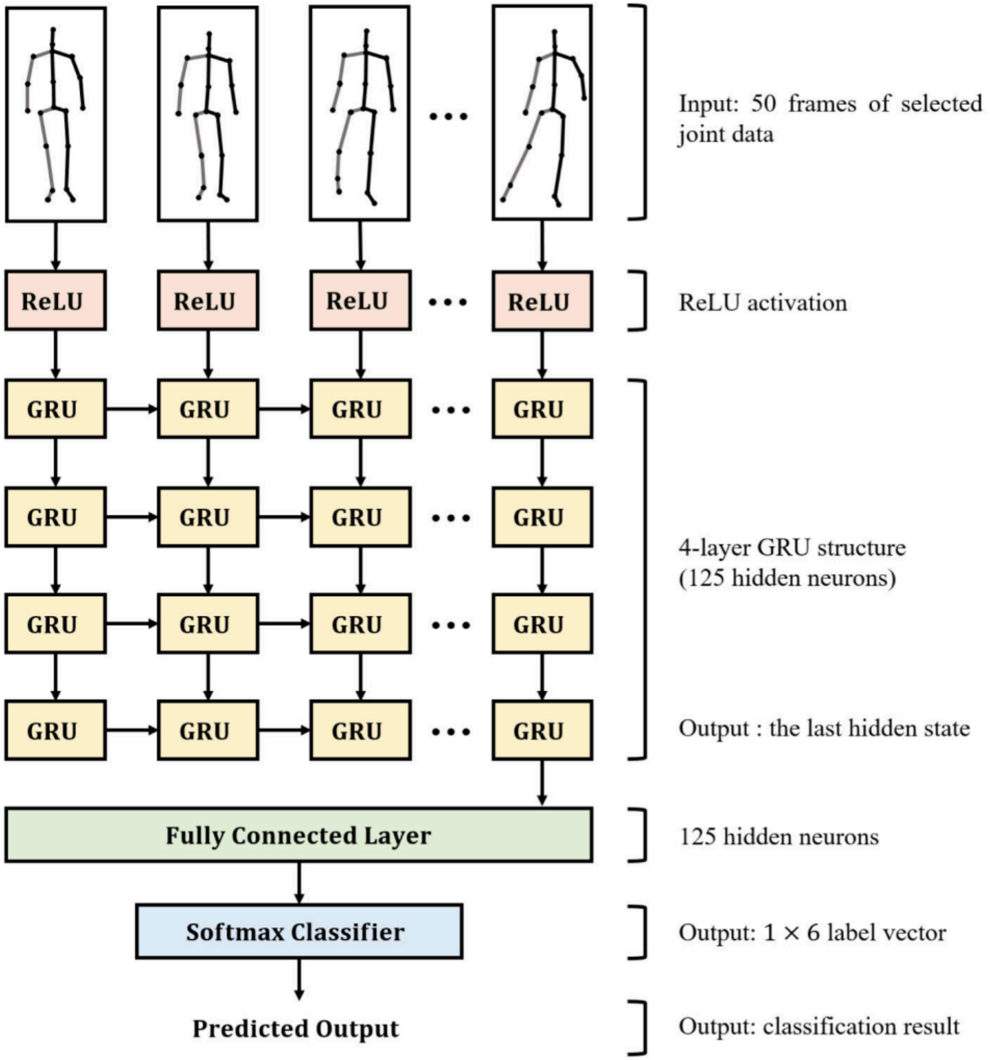
- Simultaneously collected data from the two sensors for normal and five pathological gaits
  - antalgic
  - lurching
  - steppage
  - stiff-legged
  - Trendelenburg



- 1,440 data instances (12 people x 6 gait types x 20 walkings)

Approach in [Jun20_1]

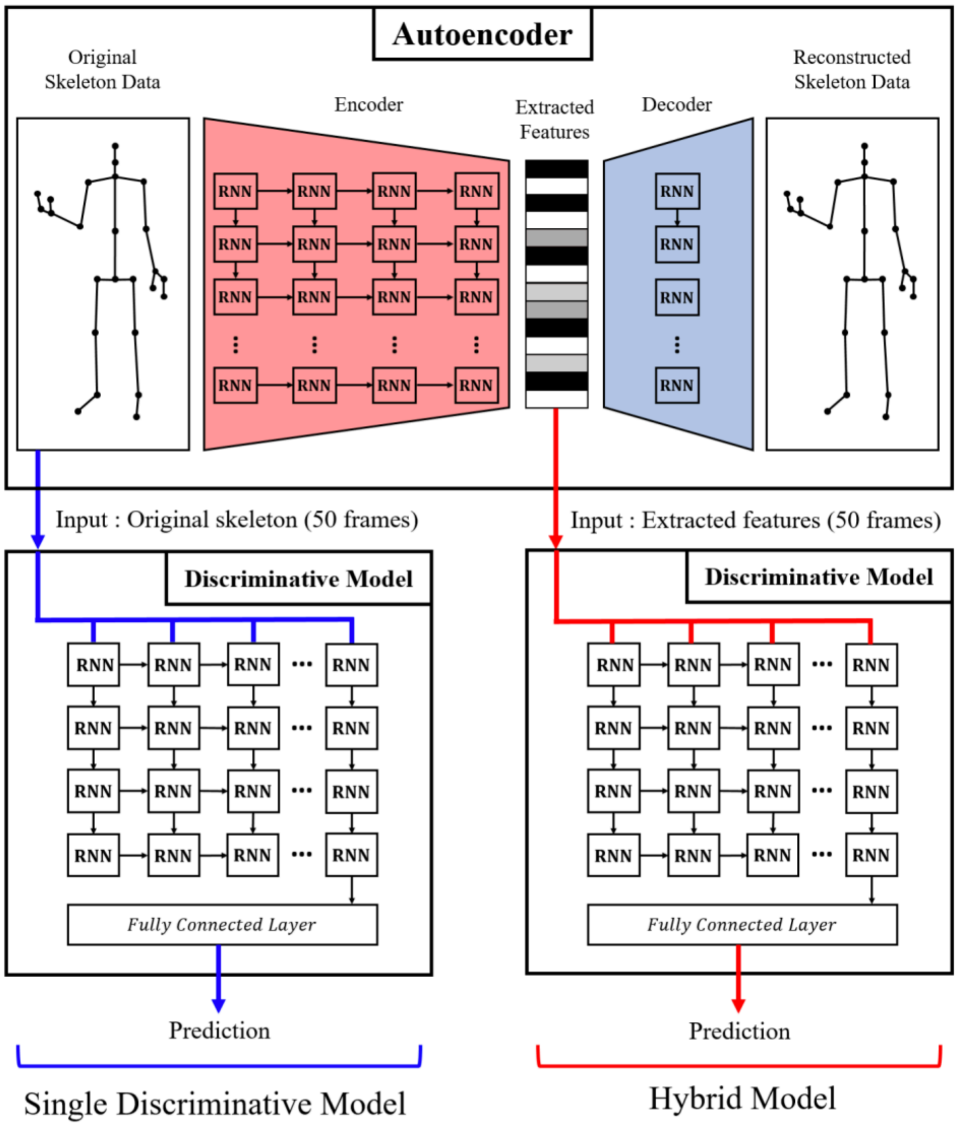
- RNN with GRU or LSTM cells
- About 90% of accuracy
- Performance are evaluated considering sub-groups of joints



Approach in [Jun20_2]

- RNN with GRU or LSTM cells combined with an RNN autoencoder for feature extraction

- Improvement of about 5% with respect to the previous approach



Possible project developments

- Consider the Sequential skeleton data
- Learning architectures
  - use raw signals with automatic feature extraction or
  - manually extract features or
  - combine the classification network with a preliminary automatic feature extraction network (e.g., autoencoder)
- Average foot pressure data never used by the authors…maybe they can be useful as a side information?

MiME
Master's degree ICT Internet Multimedia Engineering

PROJECT A3

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. A3 “sleep posture monitoring”

Reference paper

[Pouyan17] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani and S. Ostadabbas, A pressure map dataset for posture and subject analytics, in Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, FL, 2017.

PmatData dataset (102.3 MB uncompressed)

https://physionet.org/content/pmd/1.0.0/

Contains in-bed posture pressure data

- multiple adult participants
- two different types of pressure sensing mats

High level description of the dataset

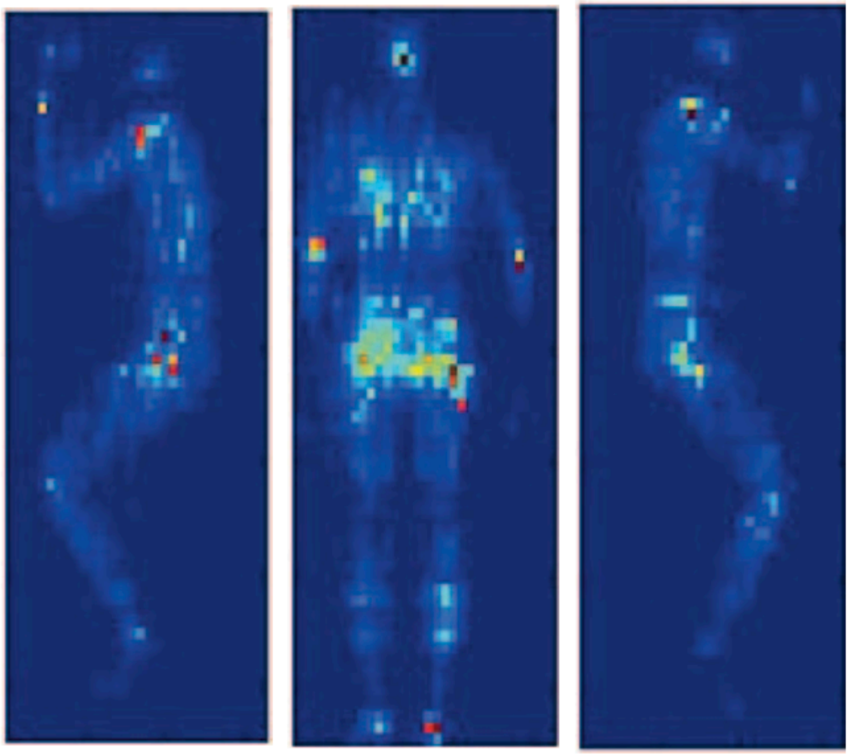
- Pressure data from two separate experiments

- Experiment 1: 13 participants
  - Size of pressure mat is 32*64: 2048 values per acquisition
  - Sampling rate: 1Hz
  - Sample values range: [0-1000]
  - 17 files for each subject (8 standard postures and 9 additional states)
  - Each file includes around 2 minutes of acquisitions

Index	Posture	Bed Inclination (degree)	Body-roll (degree)	Symbol	Duration	Spec's of mat
1	Supine	0	0		2 mins	Vista
2	Right	0	0		2 mins	Vista
3	Left	0	0		2 mins	Vista
4	Right	0	30 (1 wedge)		2 mins	Vista
5	Right	0	60 (2 wedges)		2 mins	Vista
6	Left	0	30 (1 wedge)		2 mins	Vista
7	Left	0	60 (2 wedges)		2 mins	Vista
8	Supine	0	0		2 mins	Vista
9	Supine	0	0		2 mins	Vista
10	Supine	0	0		2 mins	Vista
11	Supine	0	0		2 mins	Vista
12	Supine	0	0		2 mins	Vista
13	Right Fetus	0	0		2 mins	Vista
14	Left Fetus	0	0		2 mins	Vista
15	Supine	30	0		2 mins	Vista
16	Supine	45	0		2 mins	Vista
17	Supine	60	0		2 mins	Vista

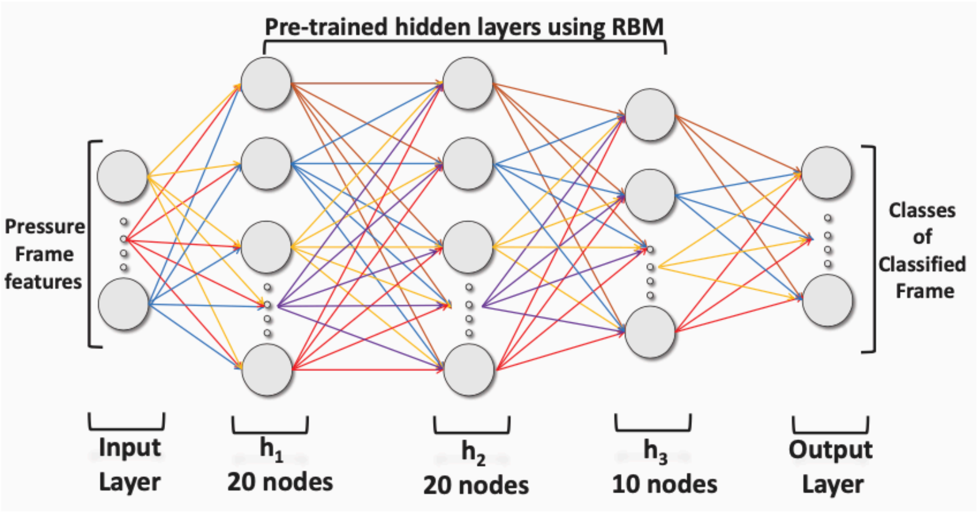
High level description of the dataset

- Experiment 2: 8 participants
  - The data is collected for both sponge and air mattresses
  - Size of pressure mat is 27*64: 1728 values per acquisition
  - Each file contains the average of around 20 acquisitions
  - 29 different states of 3 standard postures
  - Sample values range: [0-500]
  - Sampling rate: 1Hz



In [Pouyan17]

- Subject identification in three standard postures:
  - right side
  - supine
  - left side
- Main idea: each subject has a personalized sleeping pattern in each posture
- Architecture: one FFNN for each posture
- Manual feature extraction
  - 18 statistical features



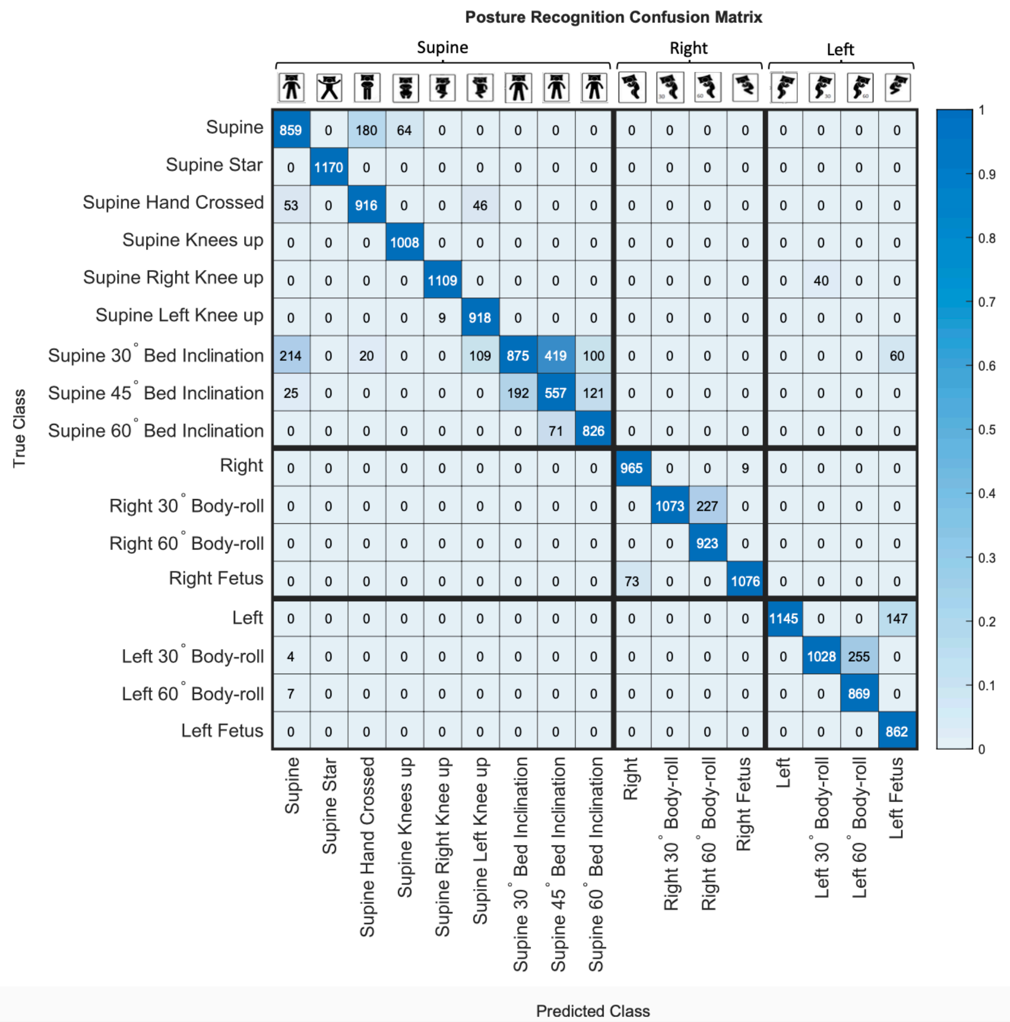
In [Pouyan17]

- Results:

Posture	Predicted/Actual	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Supine	Recall	100	90	90	88.8	80	45.4	90	100	90	90	90	80	81.8
	Specificity	100	100	98.3	99.1	99.1	97.6	96.0	96.7	100	99.1	99.1	99.1	100
	Precision	100	100	81.8	88.8	88.8	62.5	64.2	71.4	100	90	80	88.8	100
	Accuracy						85.5							
Right Side	Recall	70	60	80	70	90	90	70	90	100	100	70	60	90
	Specificity	100	98.4	98.3	97.6	95.2	97.5	99.1	100	97.5	98.3	97.6	100	100
	Precision	100	75	72.7	70	60	75	87.5	100	76.9	8.3	70	100	100
	Accuracy						80.4							
Left Side	Recall	33.3	70	100	100	90	81.8	70	100	100	90	80	80	70
	Specificity	100	96.0	96.7	97.5	99.1	98.3	97.5	99.1	98.3	99.1	98.3	98.3	100
	Precision	100	58.3	71.4	76.9	90	75	87.5	90.9	83.3	90	80	80	100
	Accuracy						82.3							
Participants' Details	Age	19	23	23	24	24	26	27	27	30	30	30	33	34
	Height (cm)	175	183	183	177	172	169	179	186	174	174	176	170	174
	Weight (kg)	87	85	100	70	66	83	96	63	74	79	91	78	74

Other reference [Davoodnia19]

- Subject identification and posture recognition using the same data of [Pouyan17]
- Architecture: CNN
- Automatic feature extraction



[Davoodnia19] V. Davoodnia and A. Etemad, Identity and Posture Recognition in Smart Beds with Deep Multitask Learning, in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019.

Possible project developments

- Classification tasks
  - Subject identification
  - Posture recognition
  - Joint subject identification and posture recognition
- Datasets
  - use one of both the available datasets
  - different mattresses
- Features
  - manual feature extraction or raw data
- Architecture
  - CNN, RNN, combinations…

MIME
Master's degree ICT Internet Multimedia Engineering

PROJECT A4

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. A4 "subject identification from ECG, PPG and respiratory signal"

Collect your own dataset

- Pulse oxymeter with our application
  - Signals:
    - PPG

- Heart monitor with our application
  - Signals:
    - ECG
- Respiration belt
  - Signals:
    - respiratory signal

Reference papers

[Wieclaw17] L. Wieclaw, Y. Khoma, P. Fałat, D. Sabodashko and V. Herasymenko, Biometrie identification from raw ECG signal using deep learning techniques, 2017, 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems

[Yaacoubi20] C. Yaacoubi, R. Besrour and Z. Lachiri, A multimodal biometric identification system based on ECG and PPG signals, 2020 In Proceedings of the 2nd International Conference on Digital Tools & Uses Congress

[Biswas19] D. Biswas, CorNET: Deep Learning Framework for PPG-Based Heart Rate Estimation and Biometric Identification in Ambulant Environment, IEEE Transactions on Biomedical Circuits and Systems, 2019

Possible project developments

- Classification tasks
  - Subject identification
- Signals
  - All the signals indicated or a subset of them (e.g., only ECG)
- Features
  - manual feature extraction or raw data
- Architecture
  - CNN, RNN, combinations…

Master's degree ICT Internet Multimedia Engineering

PART B
AUDIO SIGNALS

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO
MATEMATICA

1222·2022
800 ANNI

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

PROJECT B1

MiME

Master's degree ICT Internet Multimedia Engineering

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. B1 “speech recognition”

Reference papers

[Sainath15] Tara N. Sainath, Carolina Parada, Convolutional Neural Networks for Small-footprint Keyword Spotting, INTERSPEECH, Dresden, Germany, September 2015.

[Warden18] Pete Warden, Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition, arXiv:1804.03209, April 2018.

https://arxiv.org/abs/1804.03209

- The authors are from Google Inc.
- Reference dataset recently released by Google [Warden18]

Dataset description

- Reference dataset for small-footprint keyword spotting (KWS)
  - Released in August 2017
  - 65,000 one-second-long utterances of 30 words
  - by thousands of different people
  - released under creative commons 4.0 license
  - collected by AYI

Google blog

https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html

Speech dataset (2.11 GB uncompressed)

http://download.tensorflow.org/data/speech_commands_v0.02.tar.gz

Approaches for implementing a KWS engine

- **LVCSR based KWS** - This approach uses a two-stage process. In the first stage, the transcription of the speech into words is done using a Large Vocabulary Continuous Speech Recognition (LVCSR) engine, outputting formatted text. In the second stage, a textual search for the key-words within the text is performed. Using this approach, results from LVCSR and the text search are combined to spot the key-words
- **Phoneme Recognition based KWS** - This approach also uses a two-stage process. In the first stage, the speech is transformed to a sequence of phonemes. In the second stage, the application searches for phonetically transcribed key-words in the phoneme sequence obtained from the first stage
- **Word Recognition based KWS** [Sainath15] - This approach searches for the key-words in a one stage operation. The recognition is phoneme-based and the KWS engine looks for the keyword in the speech stream based on a target sequence of phonemes representing the key-word

CNN model from [Sainath15]

- Features are obtained from raw audio data
- 40-dimensional log Mel filterbanks coefficients
  - audio frame length 25 ms
  - with a 10 ms time shift
- At every new audio frame
  - Feature vector is obtained
  - And stacked with 23 frames to the left and 8 to the right (32 frames total)
  - This returns 32 frames at a time, spanning over 31 x 10 ms + 25 ms = 0.335 s
- A Convolutional Neural Network (CNN) is used to detect words
- Input to the CNN is a matrix of size t x n = 32 x 40 = 1,280 elements
  - t represents the number of elements in time (number of audio frames)
  - n represents the number of elements in the frequency domain (Mel features)

CNN model from [Sainath15]

- 27-44% improvement for KWS with respect to traditional neural networks

- Paper focus is on
  - Devise CNN architectures with small memory footprint
  - Playing with CNN parameters (number of kernels, strides, pooling, etc.)

Possible project developments

- Experiment with different audio features (+)
  - Type of coefficients (e.g., discrete Wavelet transform)
  - Design of Mel filterbanks, e.g.,
  http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.1426&rep=rep1&type=pdf
- Play with a standard/deep CNN using (++)
  - dropout, regularization
- Investigate recent/new ANN architectures (+++)
  - Autoencoder-based (CNN/RNN autoencoder + following SVM)
  - Attention mechanism and/or inception-based CNN networks
  - Comparison of different architectures: memory *vs* accuracy

Useful resources

Recent developments

[Chorowski15] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, Y. Bengio, Attention-Based Models for Speech Recognition, Conference on Neural Information and Processing Systems (NIPS), Montréal, Canada, 2015.

[Tang18] R. Tang and J. Lin, Deep residual learning for small-footprint keyword spotting, in IEEE ICASSP, Calgary, Alberta, Canada, 2018.

[Andrade18] D. C. de Andrade, S. Leo, M. L. D. S. Viana, and C. Bernkopf, A neural attention model for speech command recognition, arXiv:1808.08929, 2018. https://arxiv.org/pdf/1808.08929.pdf

White Paper: “Key-Word Spotting - The Base Technology for Speech Analytics”

https://pdfs.semanticscholar.org/e736/bc0a0cf1f2d867283343faf63211aef8a10c.pdf

Example code:

https://github.com/tensorflow/tensorflow/tree/master/tensorflow/examples/speech_commands/

MIME
Master's degree ICT Internet Multimedia Engineering

PROJECT B2

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. B2 “environmental sound classification”

Reference papers

[Piczak15] K.J. Piczak, ESC: Dataset for Environmental Sound Classification, in Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015.

[Piczak15-1] K. J. Piczak, Environmental sound classification with convolutional neural networks, in Proceedings of the IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, 2015.

ECS-50 dataset (884 MB uncompressed)

- https://github.com/karolpiczak/ESC-50
- Annotated collection of 2000 short clips comprising 50 classes of various common sound events

High level description of the dataset

- 5-second-long clips, 44.1 kHz, single channel
- Arranged into 5 uniformly sized cross-validation folds, ensuring that clips originating from the same initial source file are always contained in a single fold

High level description of the dataset

- 50 classes in the dataset

Animals	Natural soundscapes & water sounds	Human, non-speech sounds	Interior/domestic sounds	Exterior/urban noises
Dog	Rain	Crying baby	Door knock	Helicopter
Rooster	Sea waves	Sneezing	Mouse click	Chainsaw
Pig	Crackling fire	Clapping	Keyboard typing	Siren
Cow	Crickets	Breathing	Door, wood creaks	Car horn
Frog	Chirping birds	Coughing	Can opening	Engine
Cat	Water drops	Footsteps	Washing machine	Train
Hen	Wind	Laughing	Vacuum cleaner	Church bells
Insects (flying)	Pouring water	Brushing teeth	Clock alarm	Airplane
Sheep	Toilet flush	Snoring	Clock tick	Fireworks
Crow	Thunderstorm	Drinking, sipping	Glass breaking	Hand saw

High level description of the dataset

- ESC-10: selection of 10 classes from the bigger dataset
  - The differences between classes are much more pronounced, with limited ambiguity
  - Classes: *sneezing, dog barking, clock ticking, crying baby, crowing rooster, rain, sea waves, fire crackling, helicopter, chainsaw*

- meta/esc50.csv data description, the "esc10" column indicates if a given file belongs to the *ESC-10* subset

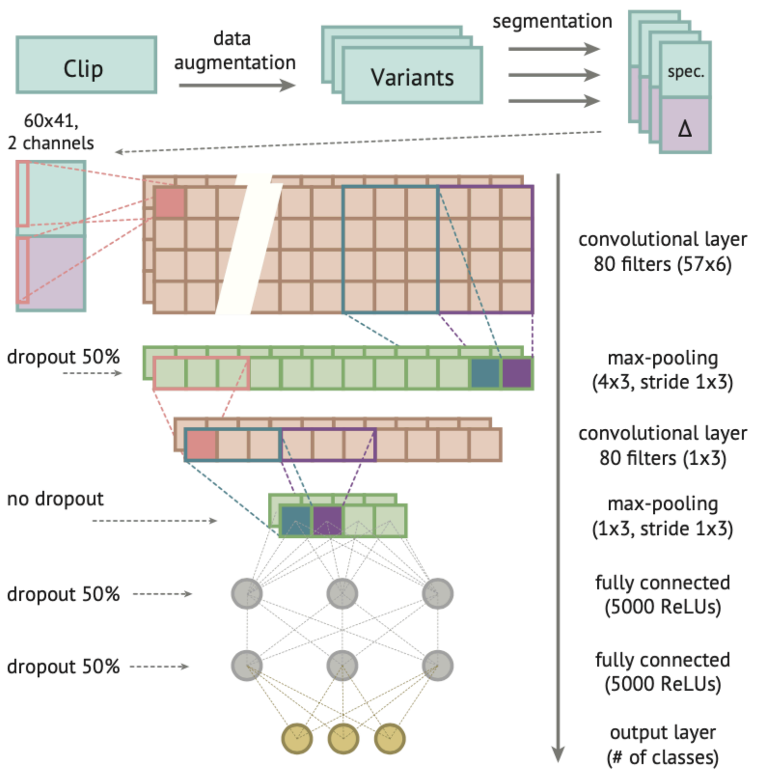
- meta/esc50-human.xlsx contains the human classification accuracy

Approach in [Piczak15-1]

- Data augmentation: apply random time delays to the original recordings

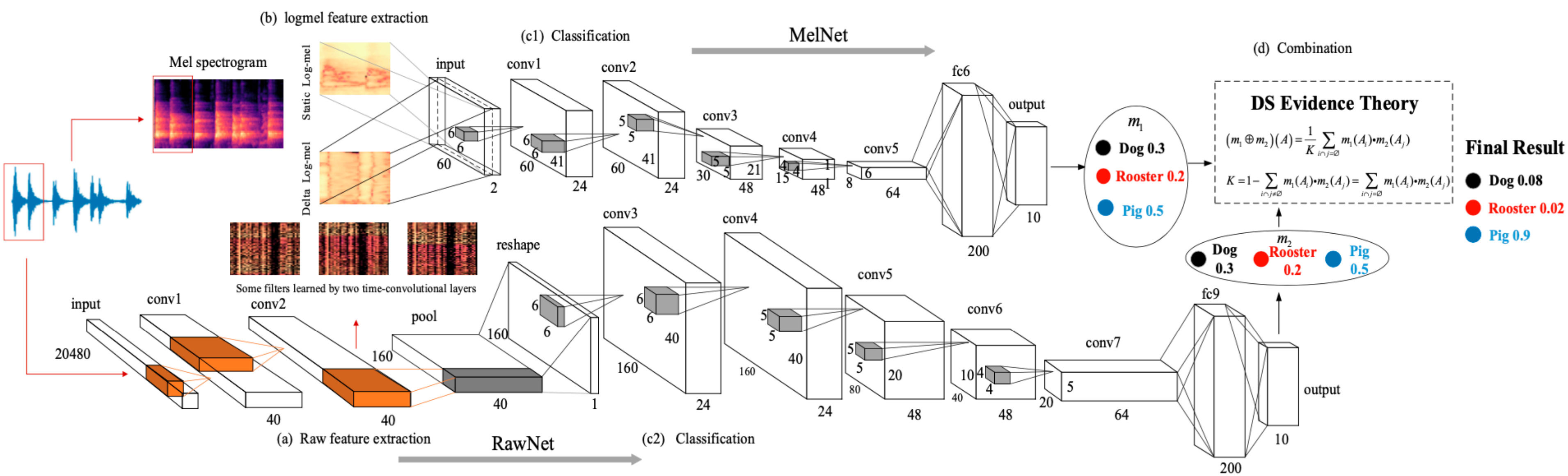
- Feature extraction: log-scaled mel-spectrograms with 60 mel-bands
  - resampled to 22,050 Hz
  - windows size 1024
  - hop length 512

- Learning architecture: CNN



Other reference [Li18]

- Combines mel-spectrogram features and raw audio waveform



[Li18] S. Li, Y. Yao, J. Hu, G. Liu, X. Yao and J. Hu, An Ensemble Stacked Convolutional Neural Network Model for Environmental Event Sound Recognition, Applied Science, vol. 8, no. 1152, July 2018.

Useful links

- Some useful functions https://nbviewer.jupyter.org/github/karoldvl/paper-2015-esc-dataset/blob/master/Notebook/ESC-Dataset-for-Environmental-Sound-Classification.ipynb

Possible project developments

- Classification tasks
  - on the entire ESC-50 dataset
  - on the restricted ESC-10 dataset
  - on each of the 5 groups of sounds:
    - animals
    - natural soundscapes & water sounds
    - human, non-speech sounds
    - interior/domestic sounds
    - exterior/urban noises
- Features: try with different approaches: mel-spectrogram, other manual-extracted features, raw data, combinations
- Architectures
  - different possibilities: CNN, RNN, …

Master's degree ICT Internet Multimedia Engineering

PART C
VISUAL DATASETS

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

1222·2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

MIME
Master's degree ICT Internet Multimedia Engineering

PROJECT C1

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. C1 “lymphoma subtype classification”

Reference papers

[Orlov10] N. V. Orlov, W. W. Chen, D. M. Eckley, T. J. Macura, L. Shamir, E. S. Jaffe, and I. G. Goldberg, Automatic classification of lymphoma images with transform-based global features, IEEE transactions on information technology in biomedicine, vol. 14, no. 4, pp. 1003–1013, 2010.

[Janowczyk16] A. Janowczyk and A. Madabhushi, Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, Journal of Pathology Informatics, vol. 7, no. 1, pp. 29, July 2016.

‘Lymphoma Subtype Classification Use Case’ p.14

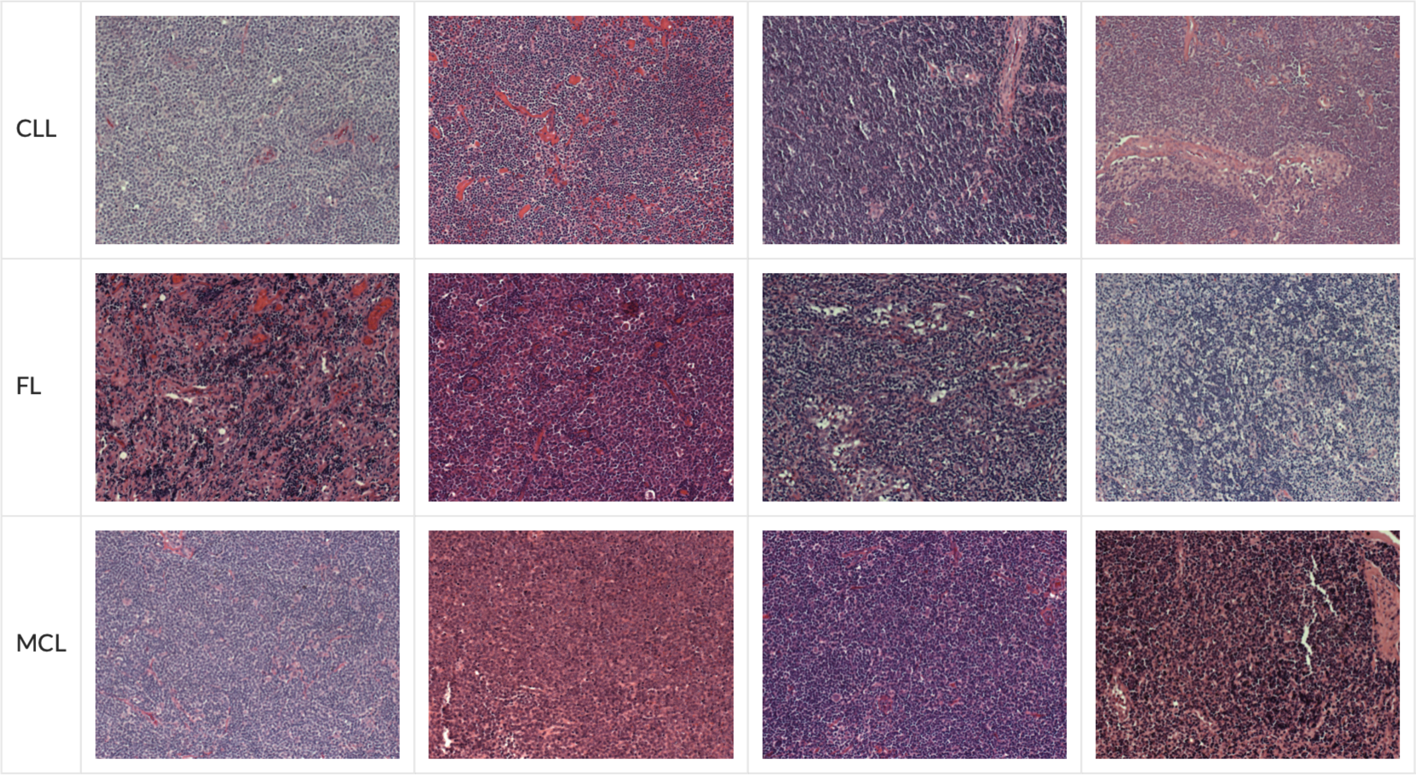
[Tambe19] R. Tambe, S. Mahajan, U. Shah, M. Agrawal, and B. Garware, Towards Designing an Automated Classification of Lymphoma subtypes using Deep Neural Networks, in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD ’19), Association for Computing Machinery, New York, NY, USA, 2019.

Dataset (1.62 GB uncompressed)

https://drive.google.com/file/d/1VRRptGC_1OSRaA_WWI28X0M_Fhtfzjby/view?usp=sharing

High level description of the dataset

- 374 images of size 1388 x 1040:
  - 113 for the Chronic Lymphocytic Leukemia (CLL) class
  - 139 for the Follicular Lymphoma (FL) class
  - 122 for the Mantle Cell Lymphoma (MCL) class



Approach in [Orlov10]

- Feature fusion: two-stage approach
  - compute spectral planes: simple (Fourier, Chebyshev, and wavelets) and compound transforms (Chebyshev of Fourier and wavelets of Fourier)
  - compute features for each pixel plane (raw data and spectral planes) as in [Orlov08]
- Classification
  - weighted neighbor distance (WND)
  - naïve Bayes network (BBN)
  - radial basis functions (RBF)
- Several color spaces were used: RGB, gray, Lab, H&E

[Orlov08] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, I. G. Goldberg, WND-CHARM: Multi-purpose image classification using compound image transforms, Pattern Recognition Letters, vol. 29, pp. 1684–1693, 2008.

Approach in [Janowczyk16]

- Approach
  - split the images in sub-patches
  - collect classification output for each patch
  - winner-take-all: the class with the highest number of votes became the designated class for the entire image
- Architecture
  - AlexNet ([Krizhevsky17])
- Accuracy: 96.58%

Author blog: http://www.andrewjanowczyk.com/use-case-7-lymphoma-sub-type-classification/

[Krizhevsky17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Communications of the ACM, vol. 60, no. 6, pp. 84–90, May 2017.

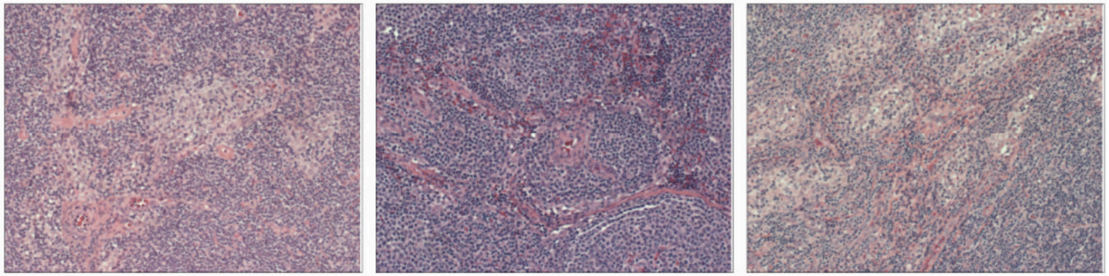
Approach in [Tambe19]

- Inception V3 network ([Szegedy16])
  - several branches used to determine the appropriate type of convolution to be made at each layer

[Szegedy16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

Possible project developments

- 3-classes classification task
  - CLL
  - FL
  - MCL



a) b) c)

- Different color spaces can be used
  - RGB
  - gray scale …
- Architectures
  - CNN, RNN, …
  - solve the classification on small patches from the image and then apply decision fusion mechanisms to classify the image based on the combination of the patches classification

PROJECT C2

Master's degree ICT Internet Multimedia Engineering
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DIPARTIMENTO MATEMATICA
1222·2022 800 ANNI
UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. C2 “bone age prediction from hand radiographs”

Reference papers

[Larson18] D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence, C. P. Langlotz, Performance of a Deep-learning neural network Model in assessing skeletal Maturity on Pediatric hand radiographs, Radiology, vol. 287, no. 1, pp. 313-322, April 2018.

[Halabi19] S. S. Halabi *et al.*, The RSNA Pediatric Bone Age Machine Learning Challenge, Radiology, vol. 290, pp. 498-503, 2019.

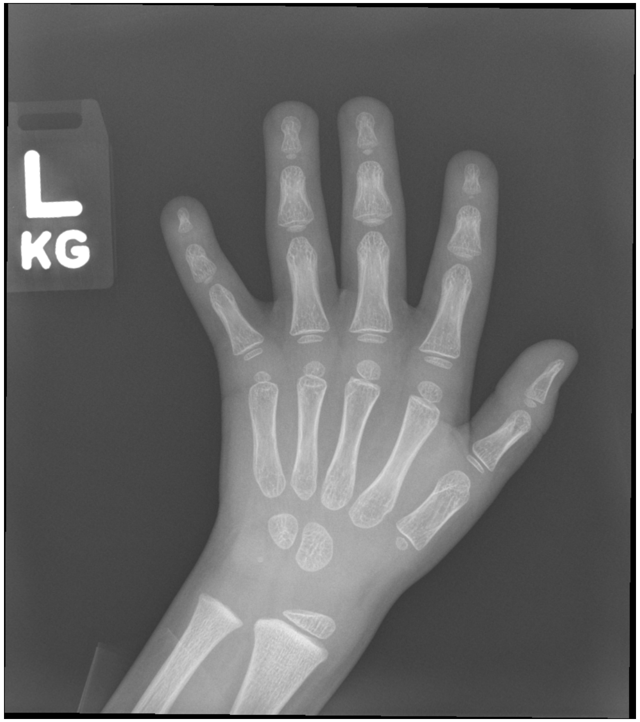
https://www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge/RSNA-Pediatric-Bone-Age-Challenge-2017

Dataset (10.3 GB uncompressed)

https://stanfordmedicine.app.box.com/s/4r1zwio6z6lrzk7zw3fro7ql5mnoupcv/folder/42459416739

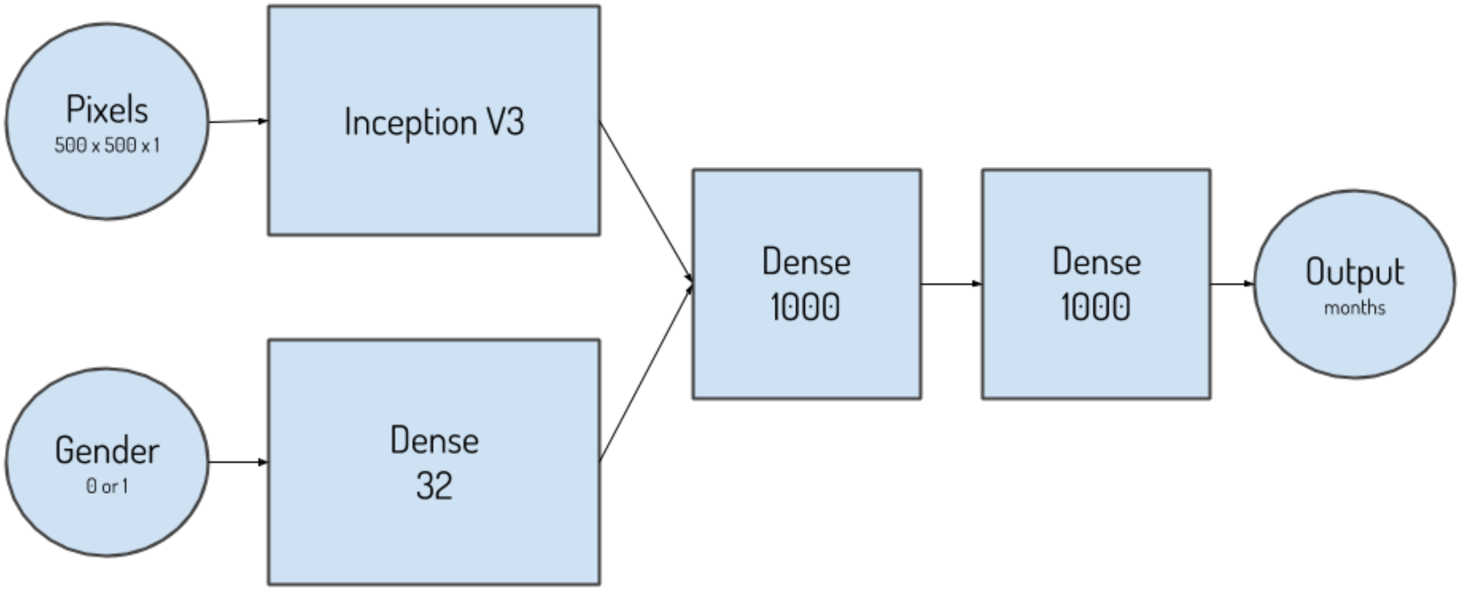
Dataset description

- 12,612 training hands' X-ray images (digital and scanned) from two U.S. hospitals
- CSV file containing the age (to be predicted) and the gender (useful additional information)



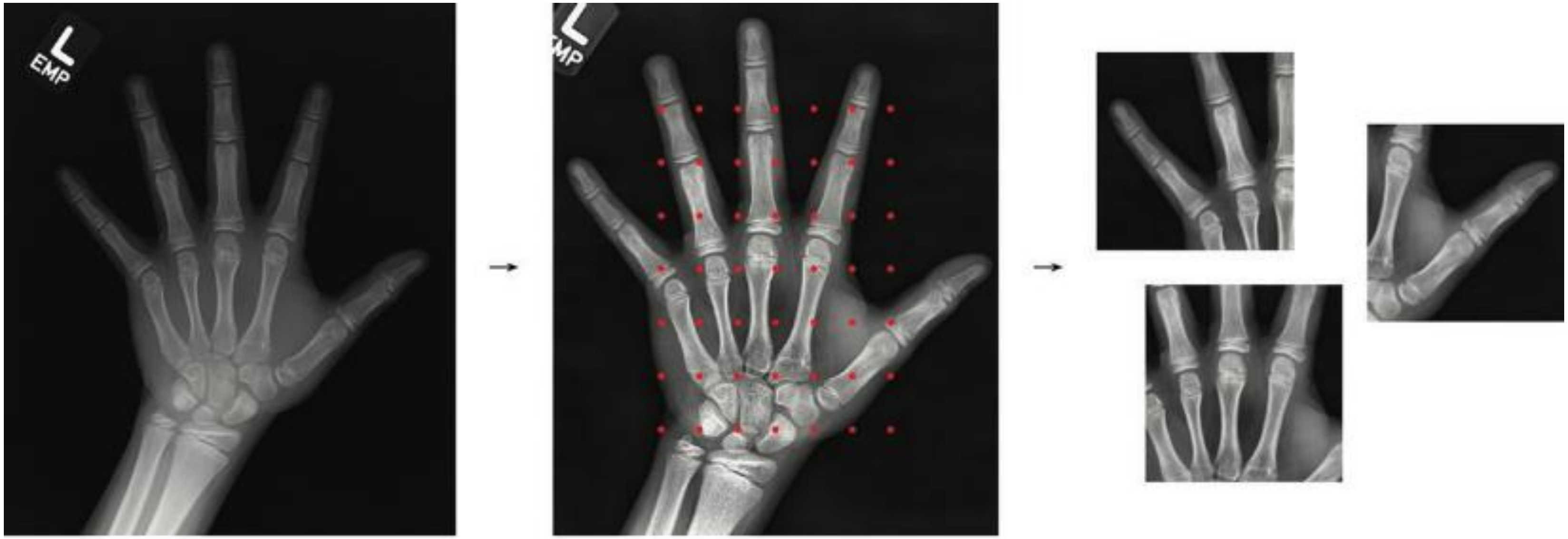
Winner model from [Halabi19]

- https://www.16bit.ai/blog/ml-and-future-of-radiology
- The age is predicted with an accuracy of 4 months



Second-place model from [Halabi19]

- Gender-specific models
- Each image was divided into 49 overlapping patches
- Use ResNet-50



Original Raw Image

Cropped + Resized + CLAHE

Each red point represents the center of an extracted patch.

Each patch is used as a training example for the CNN.

Possible project developments

- Solve the bone age prediction as a regression or classification task
  - as input: use row images or extract features
  - as output: use the age value (for regression) or classes of ages (for classification)
  - assess the importance of the gender information into the classification
  - possible idea: classify the entire image or use subpatches and then apply a decision fusion mechanism
- Architectures
  - different possibilities: CNN, RNN, attention, …

MIME
Master's degree ICT Internet Multimedia Engineering

PROJECT C3

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. C3 "lung disease prediction from X-ray images"

Reference papers

[Cohen2020] Cohen, Joseph Paul, et al. Covid-19 image data collection: Prospective predictions are the future, *arXiv preprint arXiv:2006.11988* (2020).

[Cohen2020_1] Cohen JP, Dao L, Roth K, et al. Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning, *Cureus*. 2020;12(7):e9448, Jul. 2020.

[Wang2020] Wang, L., et al. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images, *Sci Rep* **10,** 19549 (2020).

Dataset (1 GB uncompressed)

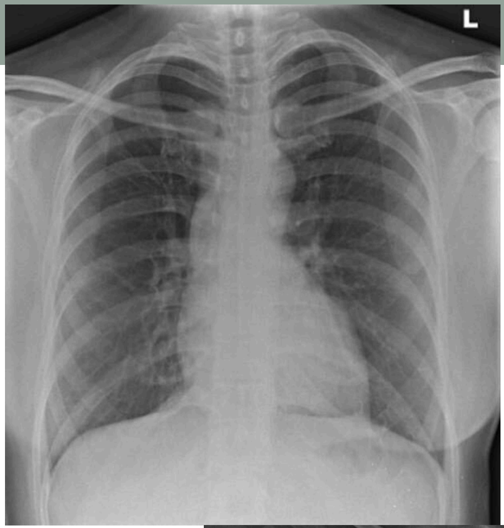
https://data.mendeley.com/datasets/jctsfj2sfn/1

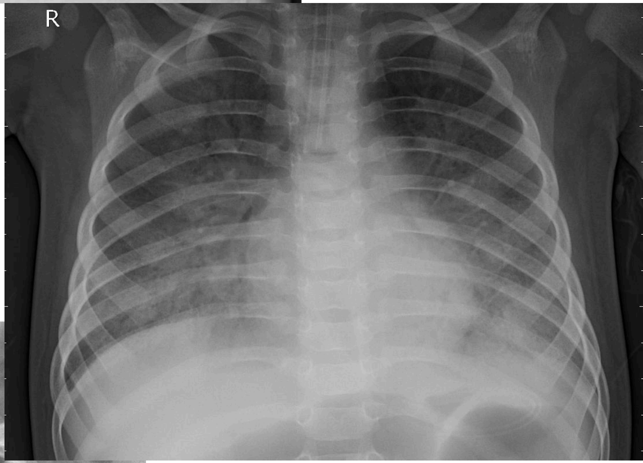
Dataset description

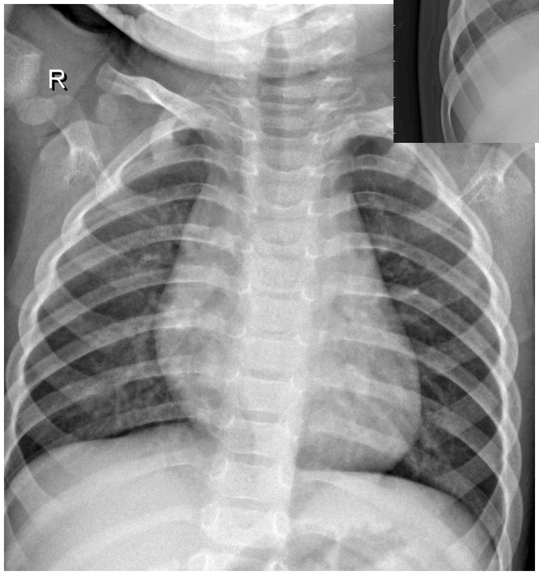
- The dataset collects the images from several institutes

- Three classes
  - COVID19
  - pneumonia
  - normal chest

- 1525 X-ray images each

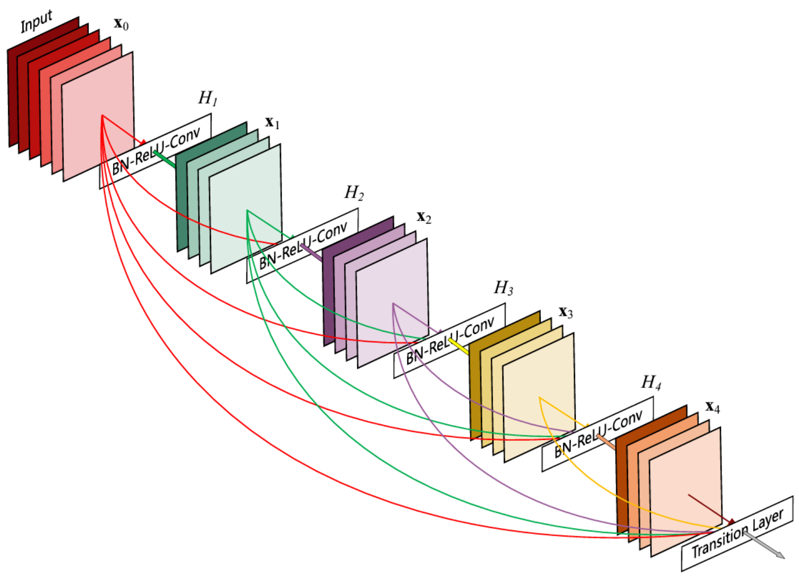






Approach in [Cohen2020]

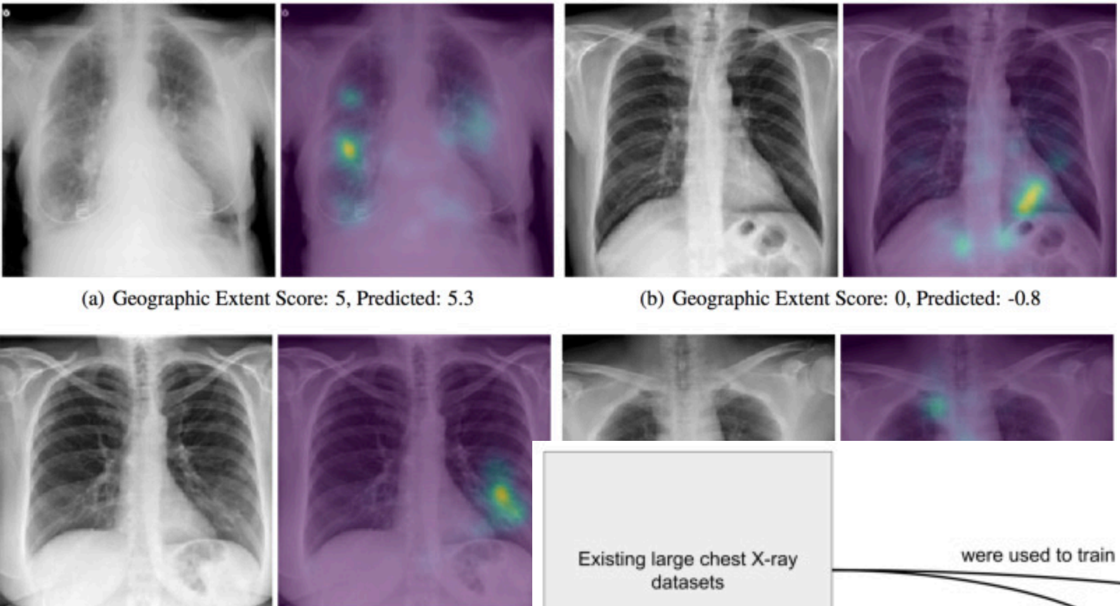
- use DenseNet - CNN where each layer is connected to every other layer in a feed-forward fashion



- [Huang2017] Huang, Gao, et al. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Approach in [Cohen2020]

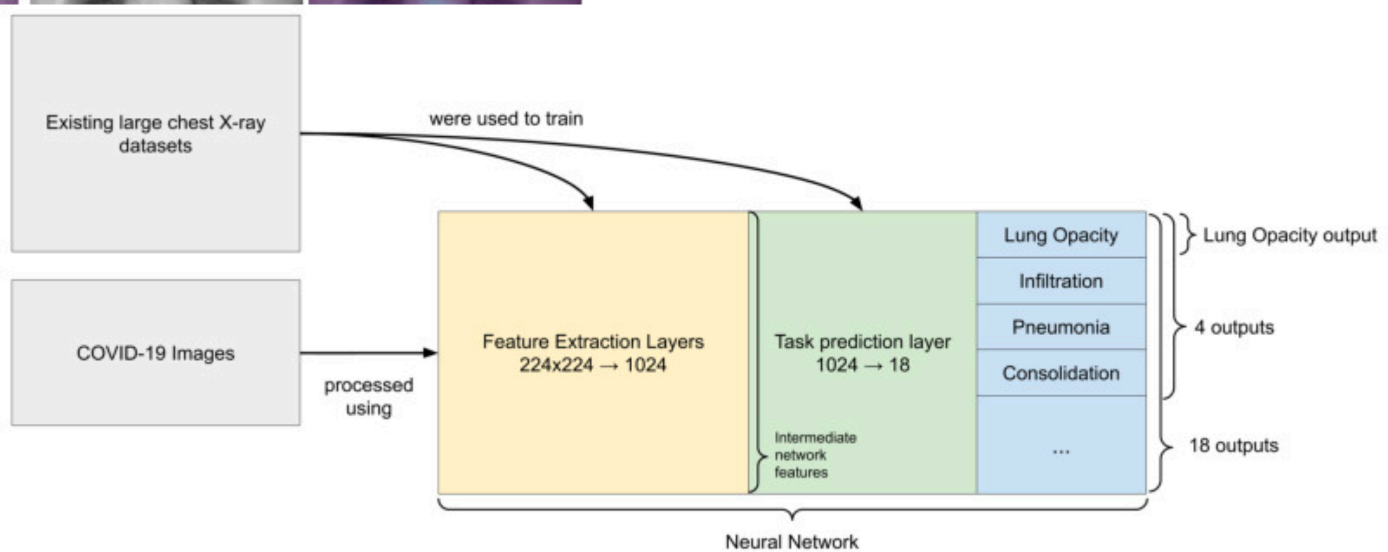
- The objective was to identify the severity of the disease but their ideas can be useful also for classification task on the three classes (normal – COVID-19, pneumonia)



(a) Geographic Extent Score: 5, Predicted: 5.3

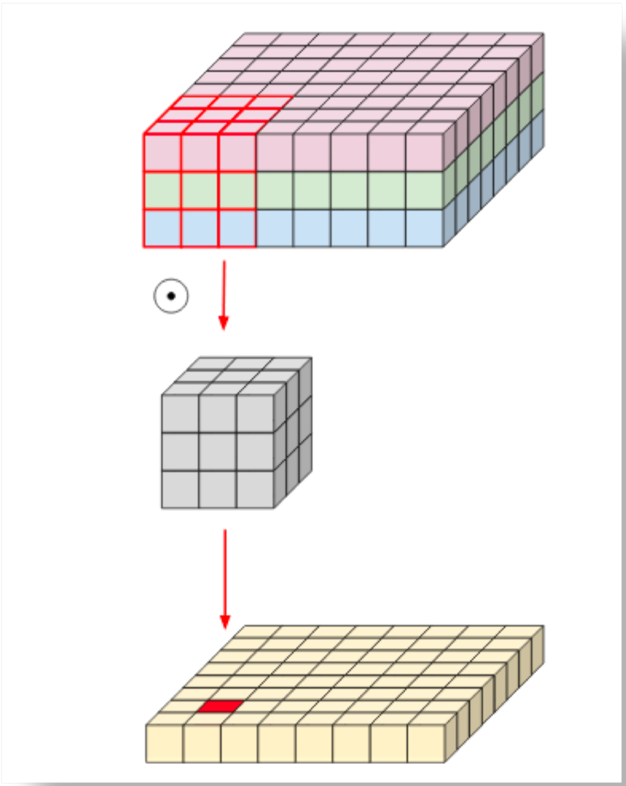
(b) Geographic Extent Score: 0, Predicted: -0.8

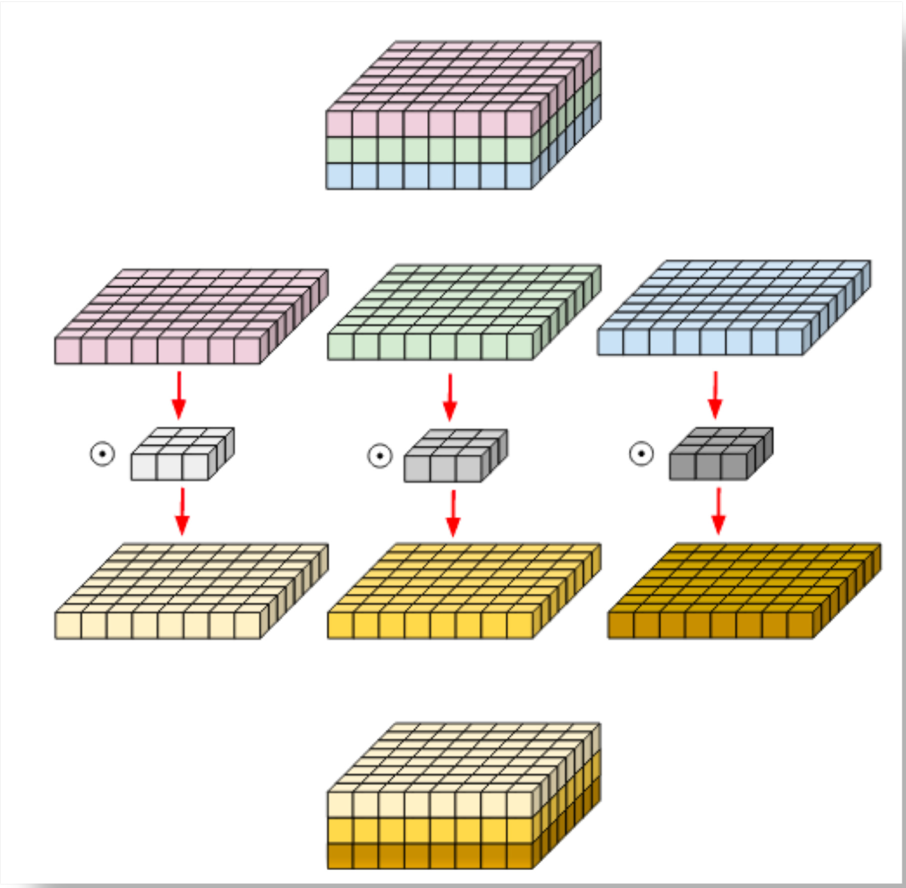
(c) Geographic Extent Score: 2, Predicted: 0.62



Approach in [Wang2020]

- convolutional and depth-wise convolutional layers

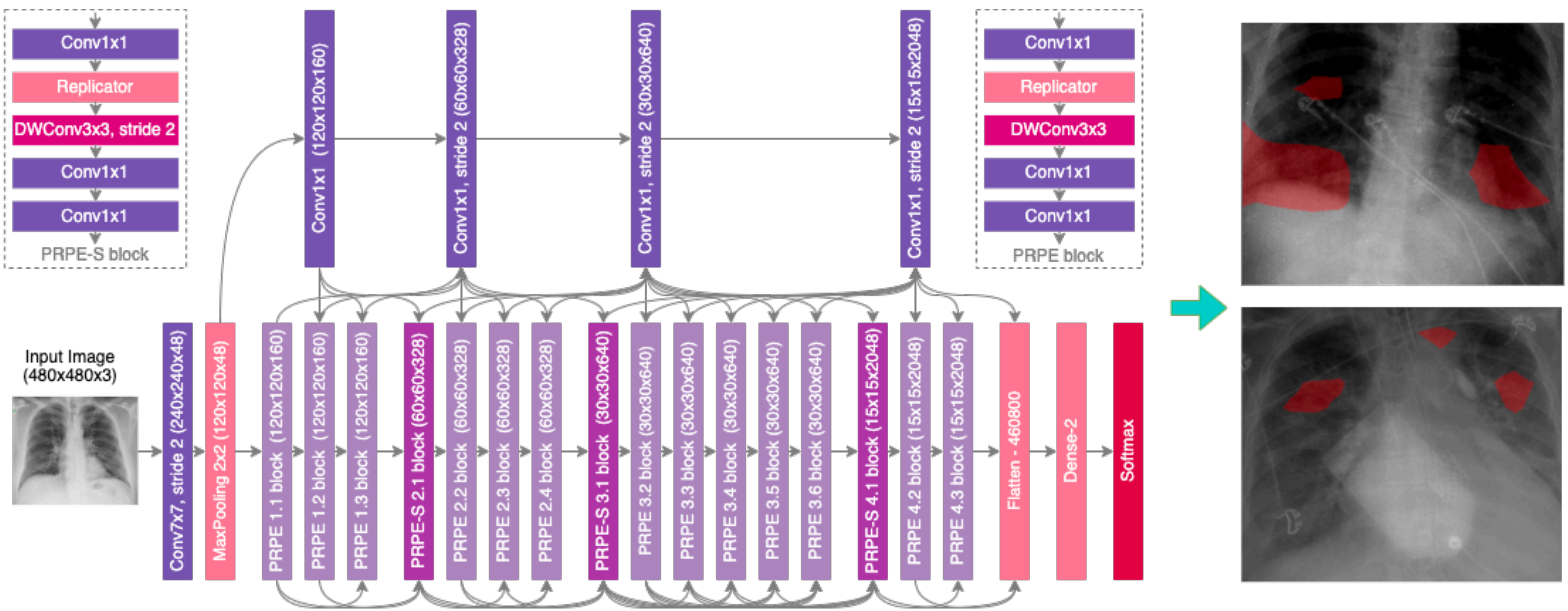




https://github.com/lindawangg/COVID-Net

Approach in [Wang2020]

- classification task on the three classes



https://github.com/lindawangg/COVID-Net

Possible project developments

- Classification task
  - use row images or extract features
  - possible approaches: classify the entire image or use subpatches and then apply a decision fusion mechanism
  - use attention mechanisms
- Architectures
  - CNN, RNN, attention, …

PART D
WIRELESS SIGNALS

MIME
Master's degree ICT Internet Multimedia Engineering

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO
MATEMATICA

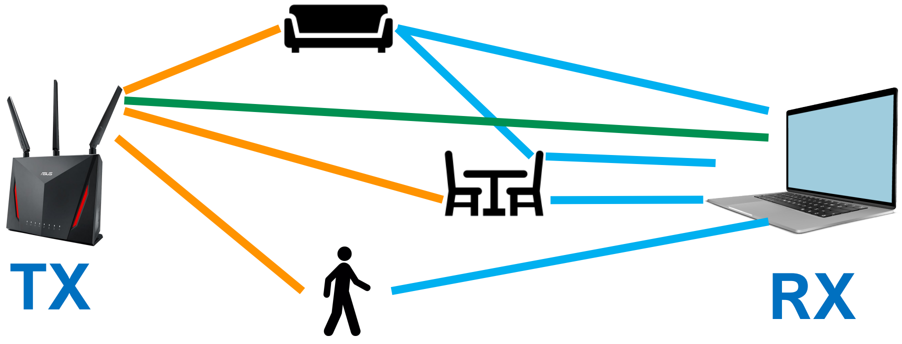
800 ANNI 1222·2022

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

General idea

- **Important disclaimer. The D1, D2 and D3 projects require some preliminary knowledge about wireless communications**

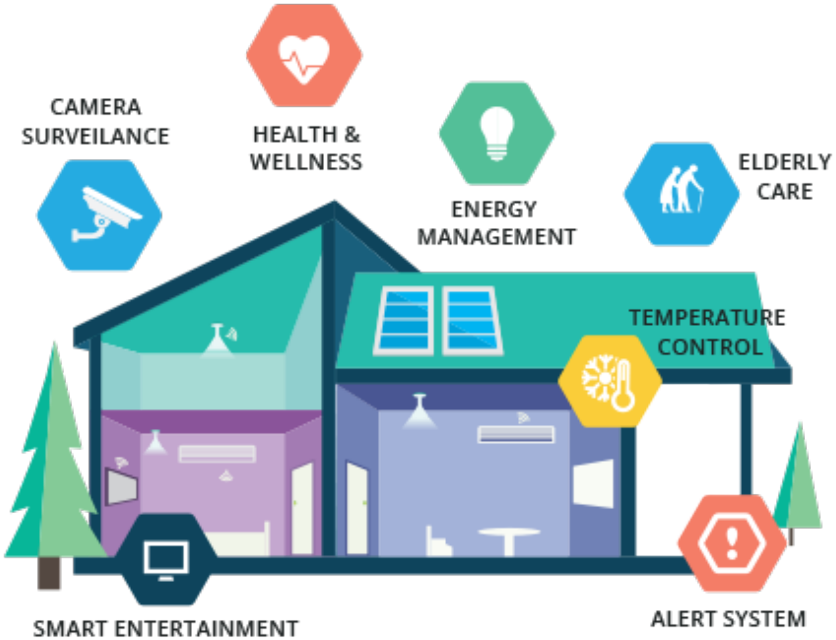
- **Main idea.** The presence and the movement of objects in the environment affect the Wi-Fi signal (multi-path) propagation
- These modifications can be
  - estimated via dedicated signal processing on the Wi-Fi channel frequency response (CFR) and
  - used as a proxy for human sensing applications



General idea

- Why radio waves for sensing?
  - more user friendly than approaches based on wearable devices: the user is not required to wear anything
  - privacy preserving: no images of the subjects are captured
  - insensitive to light conditions and presence of dust, smoke
  - see through walls/obstacles (low frequencies)

- Applications
  - Human detection
  - Fall detection
  - Human activity recognition
  - Human vital signs monitoring
  - People tracking
  - Gesture recognition



General idea

- In recent years, spurred by the pervasiveness of Wi-Fi-enabled devices, Wi-Fi sensing have been widely investigated
- September 2020: the IEEE 802.11bf working group was established to empower Wi-Fi devices with sensing capabilities [Restuccia2021]
- The goal is to allow Wi-Fi routers to perform a dual role
  - communication access points (AP)
  - monitoring devices, leveraging ongoing Wi-Fi traffic as well as ad-hoc packets to deliver the sensing service

[Restuccia2021] F. Restuccia, IEEE 802.11bf: Toward Ubiquitous Wi-Fi Sensing. *arXiv preprint arXiv:2103.14918*, 2021.

MIME
Master's degree ICT Internet Multimedia Engineering

PROJECT D1

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO
MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Project no. D1 “activity recognition through Wi-Fi channel frequency response”

Reference paper

[Meneghello2022] F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello and M. Rossi, SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points. *IEEE Transactions on Mobile Computing*. 2022.

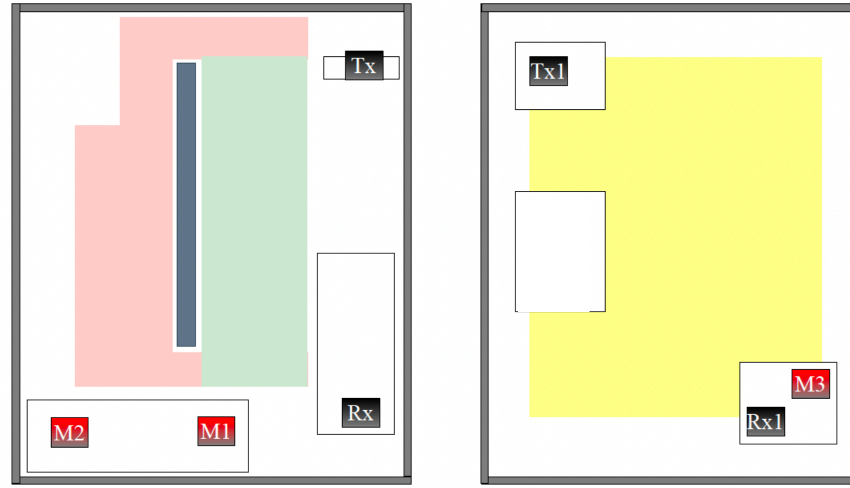
Dataset (23 GB uncompressed) – **AR-** sub-folders

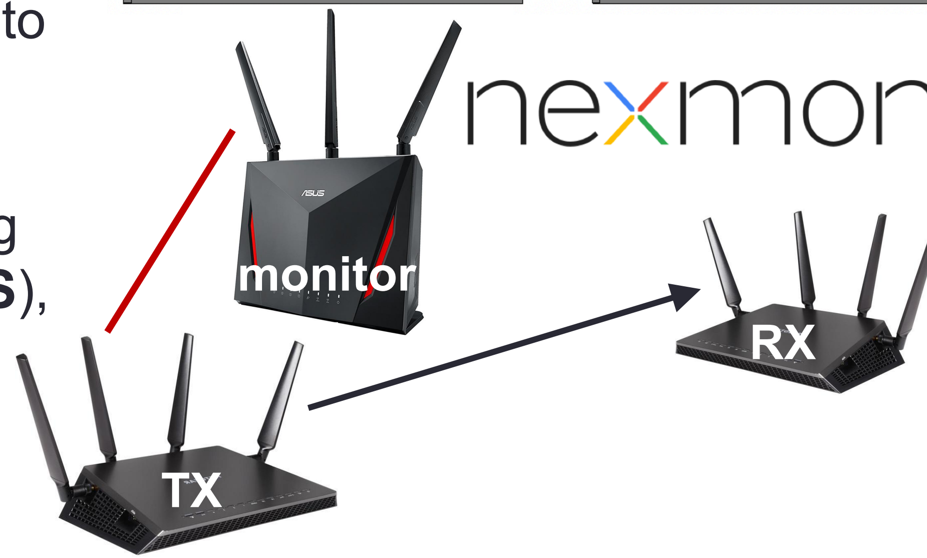
https://drive.google.com/file/d/1t9wrxCyk1_AqXj3j_NQ81tNzwOdhH1j1/view?usp=sharing

https://ieee-dataport.org/documents/csi-dataset-wireless-human-sensing-80-mhz-wi-fi-channels

Dataset description

- More than 6 hours of Wi-Fi channel readings acquired while a volunteer (4 in total) performs up to six different activities in different indoor environments
- Walking (**W**), running (**R**), jumping (**J**), sitting still (**L**), standing still (**S**), sitting down/standing up (**C**) and doing arm exercises (**G**)





set	campaigns	environment	w × l × h [m]	obstructed path	devices pos.	Tx hardware	Rx hardware	person, Pi
AR-1	a-b-c-d-e	bedroom	5 × 6 × 4	-	M1-Tx-Rx	Netgear	Netgear	P1
AR-2	a	bedroom	5 × 6 × 4	-	M1-Tx-Rx	Netgear	Netgear	P2
AR-3	a-b	bedroom	5 × 6 × 4	✓	M2-Tx-Rx	Netgear	Netgear	P1
AR-4	a	bedroom	5 × 6 × 4	✓	M2-Tx-Rx	Netgear	Netgear	P2
AR-5	a-b	living room	5 × 6 × 4	-	M3-Tx1-Rx1	Netgear	Netgear	P1
AR-6	a	kitchen	3.5 × 3 × 3.2	-	M3-Tx1-Rx1	Netgear	Netgear	P1
AR-7	a	laboratory	7.5 × 3.5 × 2.9	-	M3-Tx1-Rx1	Netgear	Netgear	P3
AR-8	a-b	office	4 × 6 × 3	-	M3-Tx1-Rx1	Asus	Asus	P4
AR-9	a-b-c	semi-anechoic	9 × 7 × 3.4	-	M3-Tx1-Rx1	Asus	Asus	P4

Dataset description (also for D2, D3)

- For each pair of transmit and receive antennas, the CFR consists of a vector of complex numbers specifying the attenuation and the phase shift experienced by the signal over each OFDM sub-channel
- Each .mat file collects the M CFR vectors (referred to as a CFR trace) acquired during the transmission time
- The CFR trace is saved as a $(N * Nant) \text{ X } M$ dimensional complex matrix, where each row is a CFR vector
- The CFR vectors estimated on different monitor antennas are stored as subsequent rows in the CFR trace

Dataset description (also for D2, D3)

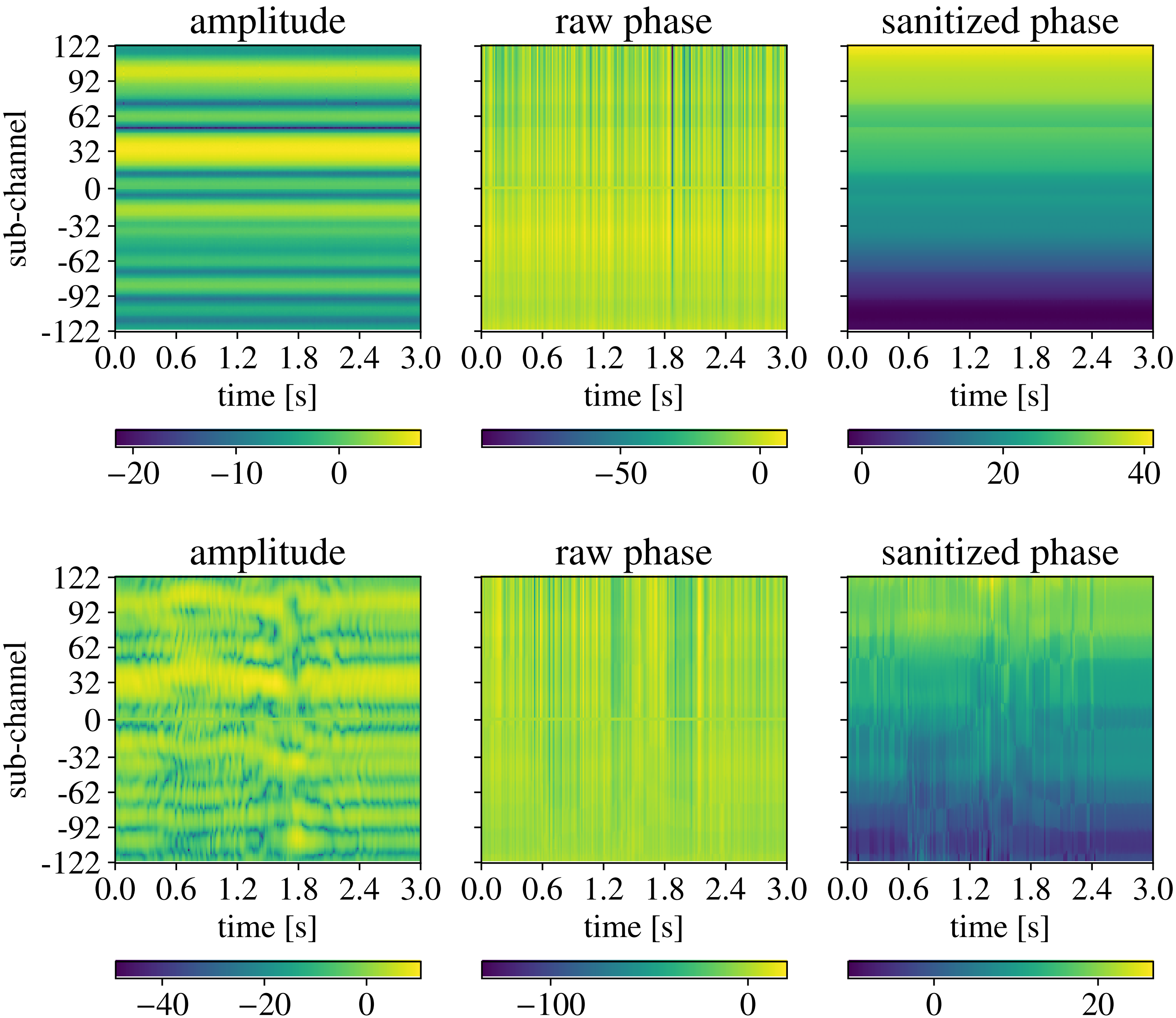
- A Python script is available at https://github.com/signetlabdei/SHARP to process the data and obtain an M X N X $Nant$-dimensional matrix → more convenient

- The repository also contains Python scripts to
  - invert the sign on sub-channels from −63 to 122 (artifact introduced by the Nexmon tool)
  - sanitize the CFR phase to remove the phase offsets introduced in the CFR recordings due to hardware artifacts

- The CFR can also be sanitized by considering one antenna as a reference and multiplying the CFR on the other antennas by the complex conjugate CFR of the reference antenna → ok if you use the raw amplitude and phase, but not so good if you want to leverage Doppler

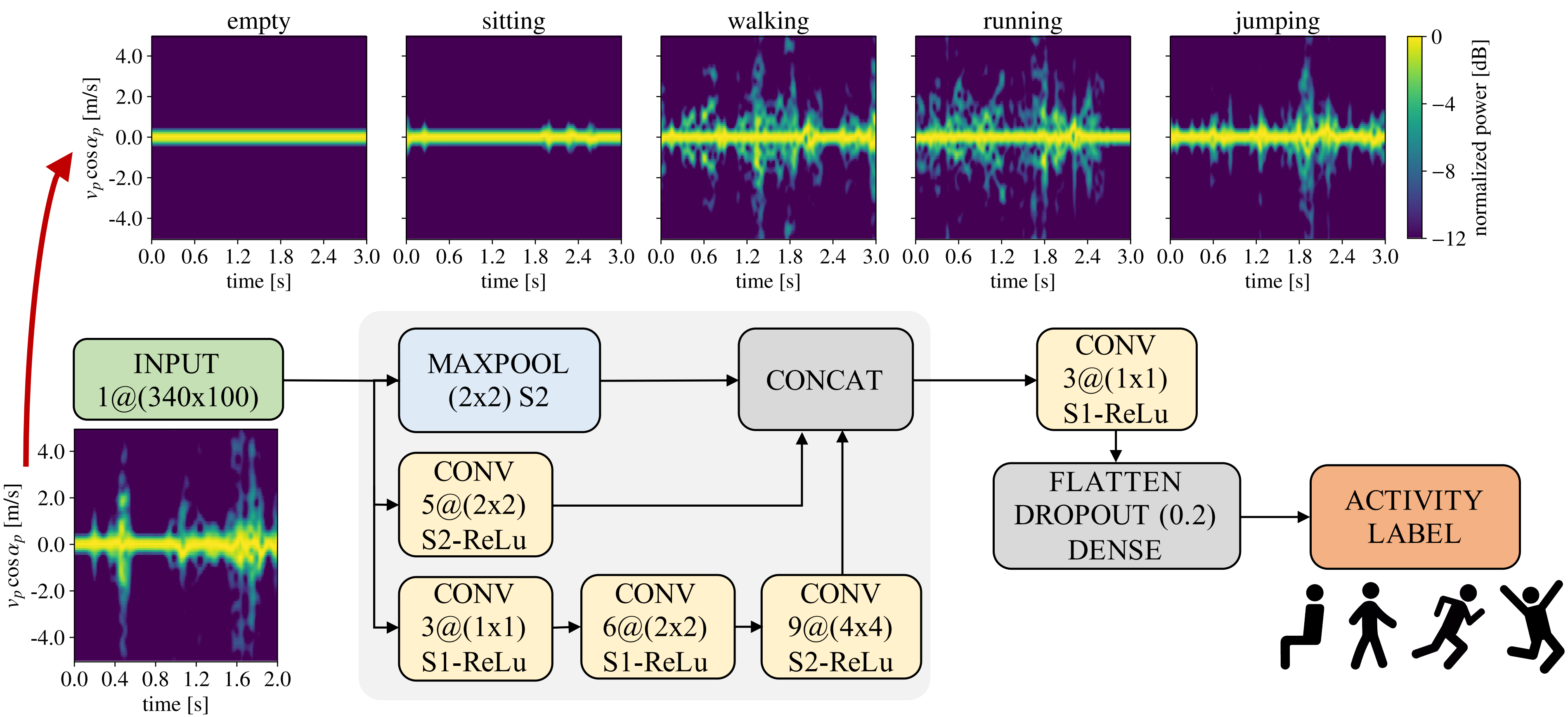
empty room

running person



Approach in [Meneghello2022]

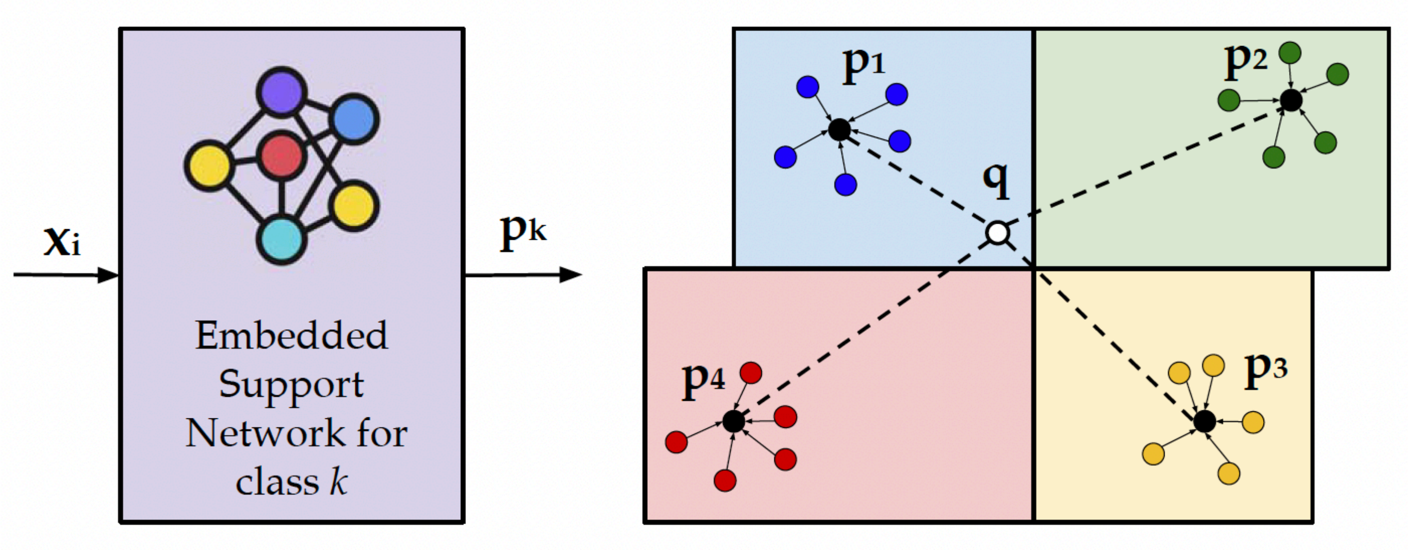
- Estimate the Doppler shift from the channel frequency response
- Use a CNN-based algorithm to classify the activities



Python implementation available at https://github.com/signetlabdei/SHARP

Approach in [Bahadori2022]

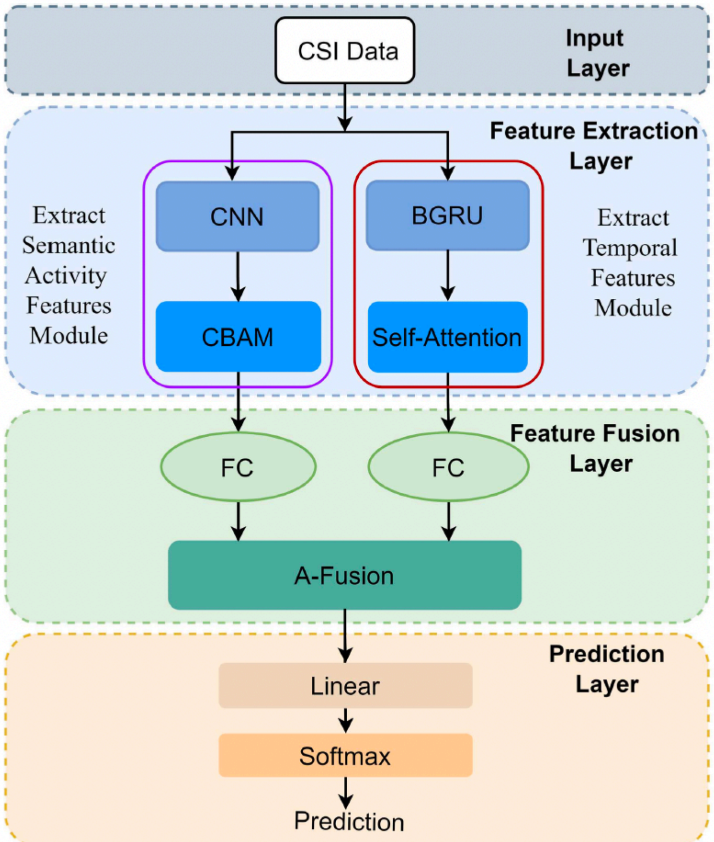
- Embedded prototype network - few-shot learning
  - can generalize to new environments by leveraging only a few new samples



Implementation available at https://github.com/niloobah/ReWiS

[Bahadori2022] N. Bahadori, J. Ashdown, and F. Restuccia, ReWiS: Reliable Wi-Fi Sensing Through Few-Shot Multi-Antenna Multi-Receiver CSI Learning. *in Proc. of IEEE WoWMoM*, 2022

Approach in [Zhang2022]



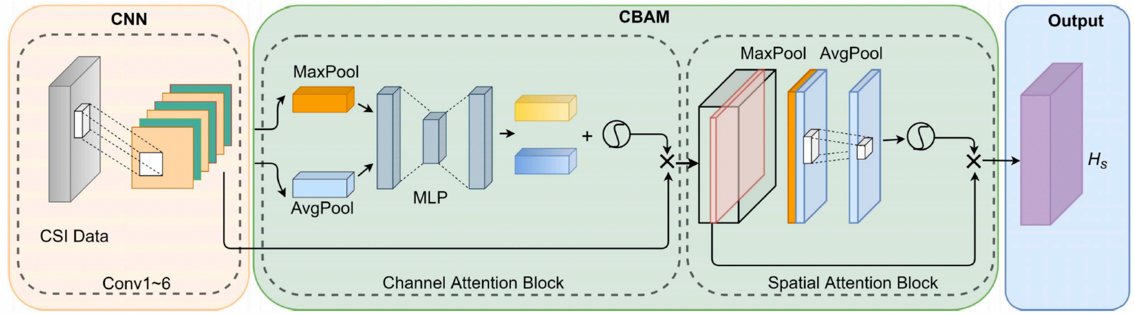


Fig. 6. Structure of semantic activity features extraction module.

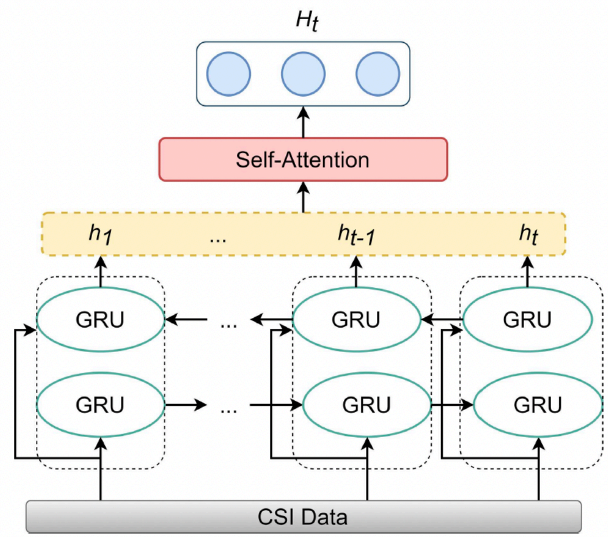
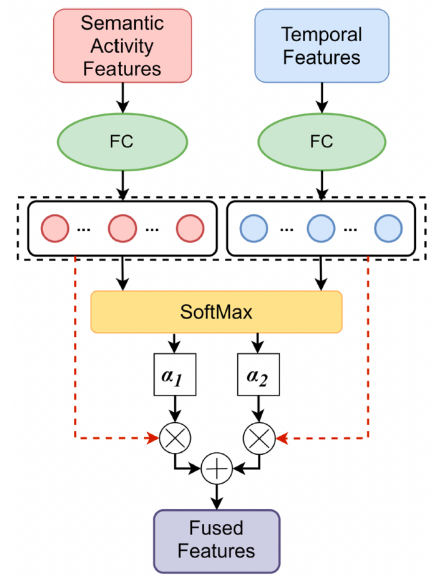


Fig. 7. Structure of temporal features extraction module.



Structure of A-Fusion.

[Zhang2022] Y. Zhang, Q. Liu, Y. Wang and G. Yu, CSI-Based Location-Independent Human Activity Recognition Using Feature Fusion. *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-12, 2022

Possible project developments

- Classification task
  - try different approaches to obtain robust algorithms, i.e., that work well when tested on different environments, people and hardware with respect to the ones considered during training
  - use different features as input: raw data (CFR amplitude and/or phase), processed features (Doppler shift or others), combinations of them…
  - try networks with memory cells (recurrent) to capture the correlation in time
  - try including attention mechanisms to capture relevant characteristics of different movements
  - compare the approaches in [Meneghello2022] and [Bahadori2022] on the same dataset/datasets (you can also train in one of the datasets and test on the other)

PROJECT D2

Project no. D2 “person identification through Wi-Fi channel frequency response”

Reference paper

[Meneghello2022] F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello and M. Rossi, SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points. *IEEE Transactions on Mobile Computing*. 2022.

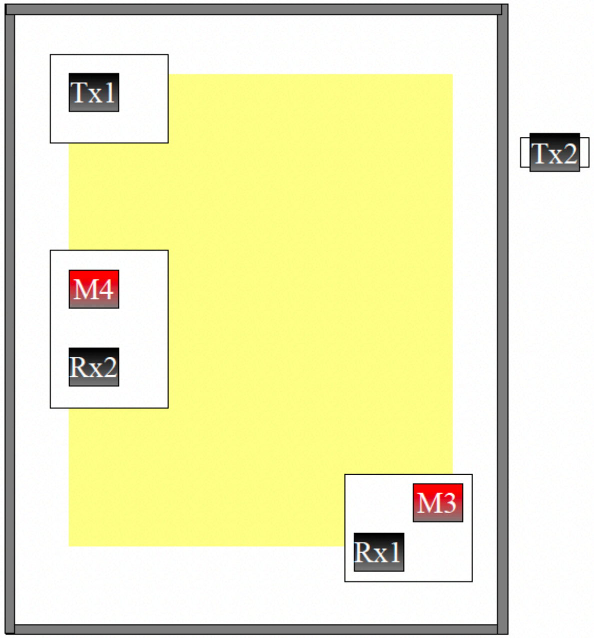
Dataset (23 GB uncompressed) – **PI-** sub-folders

https://drive.google.com/file/d/1t9wrxCyk1_AqXj3j_NQ81tNzwOdhH1j1/view?usp=sharing

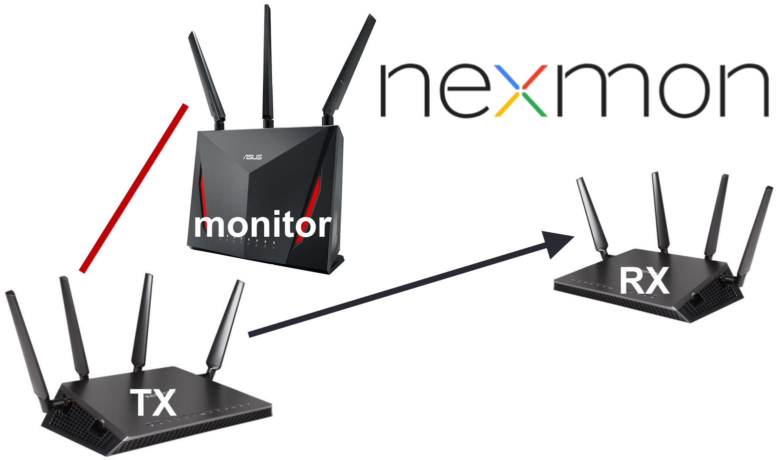
https://ieee-dataport.org/documents/csi-dataset-wireless-human-sensing-80-mhz-wi-fi-channels

Dataset description

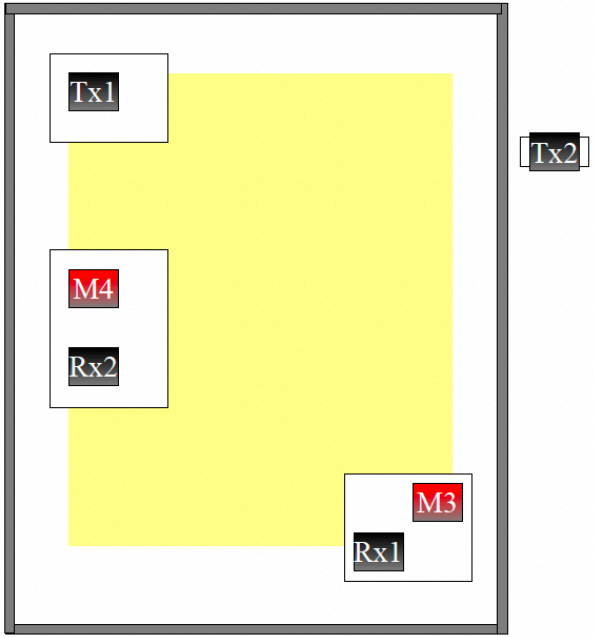
- The dataset collects more than 3 hours of Wi-Fi channel readings collected while a person moves freely in the environment
  - 10 volunteers



set	campaigns	environment	w × l × h [m]	obstructed path	devices pos.	Tx hardware	Rx hardware	person, Pi
PI-1	a	meeting room	7 × 7.5 × 3.5	-	M3-Tx1-Rx1	Netgear	Netgear	P3, P5-P13
PI-2	a	meeting room	7 × 7.5 × 3.5	✓	M3-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13
PI-3	a	meeting room	7 × 7.5 × 3.5	-	M4-Tx1-Rx1	Netgear	Netgear	P3, P5-P13
PI-4	a	meeting room	7 × 7.5 × 3.5	✓	M4-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13



Dataset description



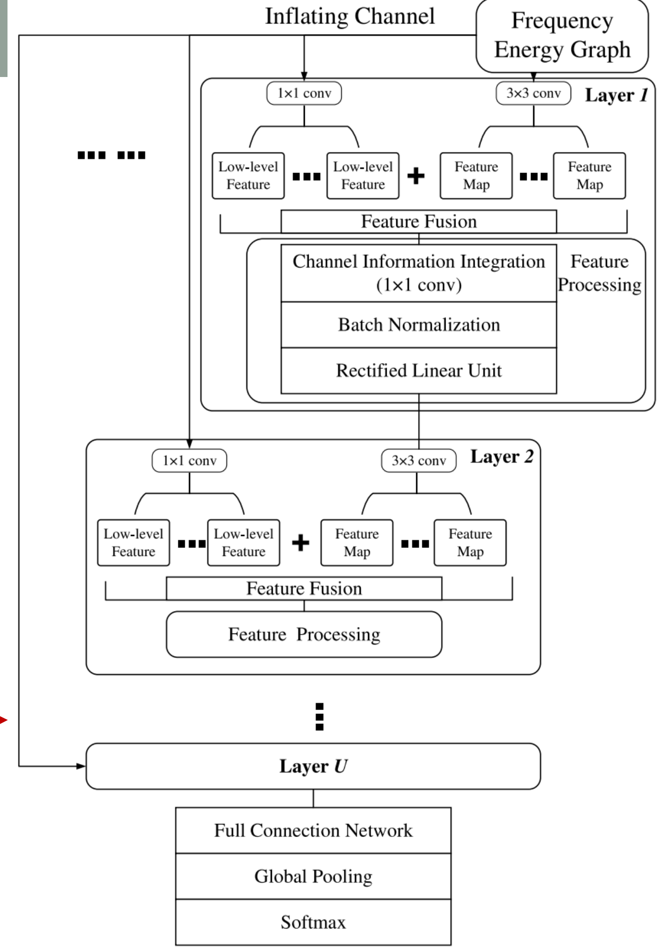
- The data was concurrently collected from two Wi-Fi networks, entailing **Tx1-Rx**1 and **Tx2-Rx2** respectively, through two monitor devices (**M3** and **M4**)

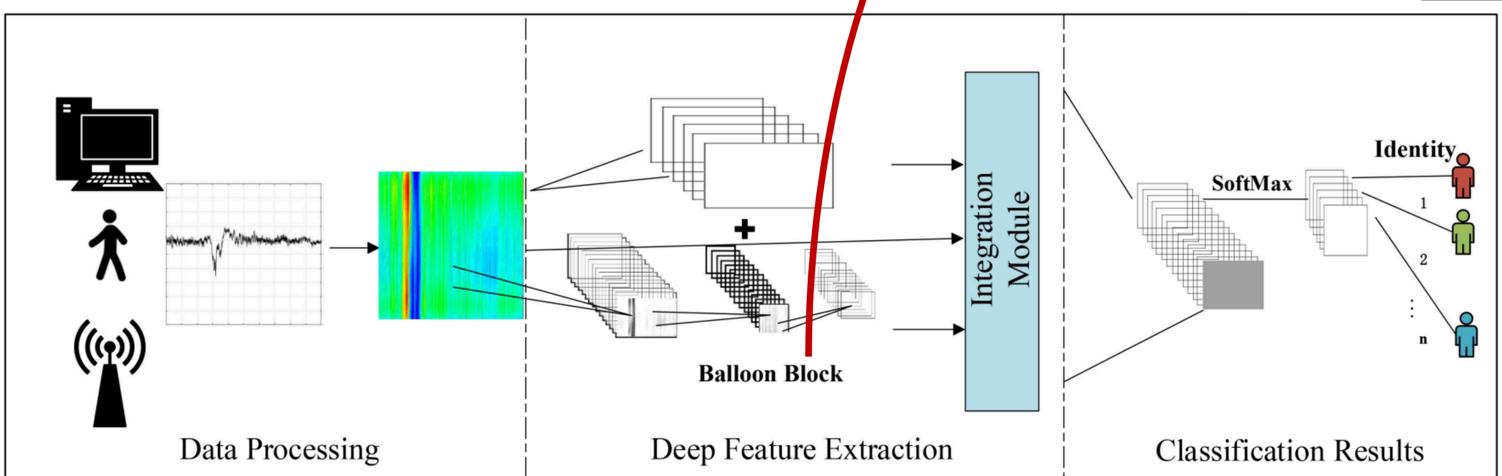
- This configuration generated four simultaneous acquisitions for each experiment:
  - the files with the same suffix in the associated sub-folders – e.g., “PI1a_p06.mat”, “PI2a_p06.mat”, “PI3a_p06.mat”, “PI4a_p06.mat” – refer to simultaneous collections

The rest of the description of the dataset is as in project D1

Approach in [Cao2021]

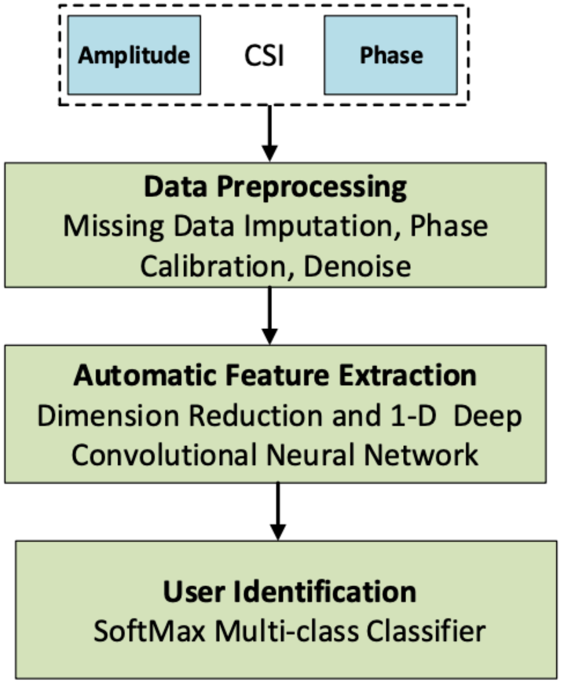
[Cao2021] Y. Cao, Z. Zhou, C. Zhu, P. Duan, X. Chen and J. Li, A Lightweight Deep Learning Algorithm for WiFi-Based Identity Recognition. *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17449-17459, 2021

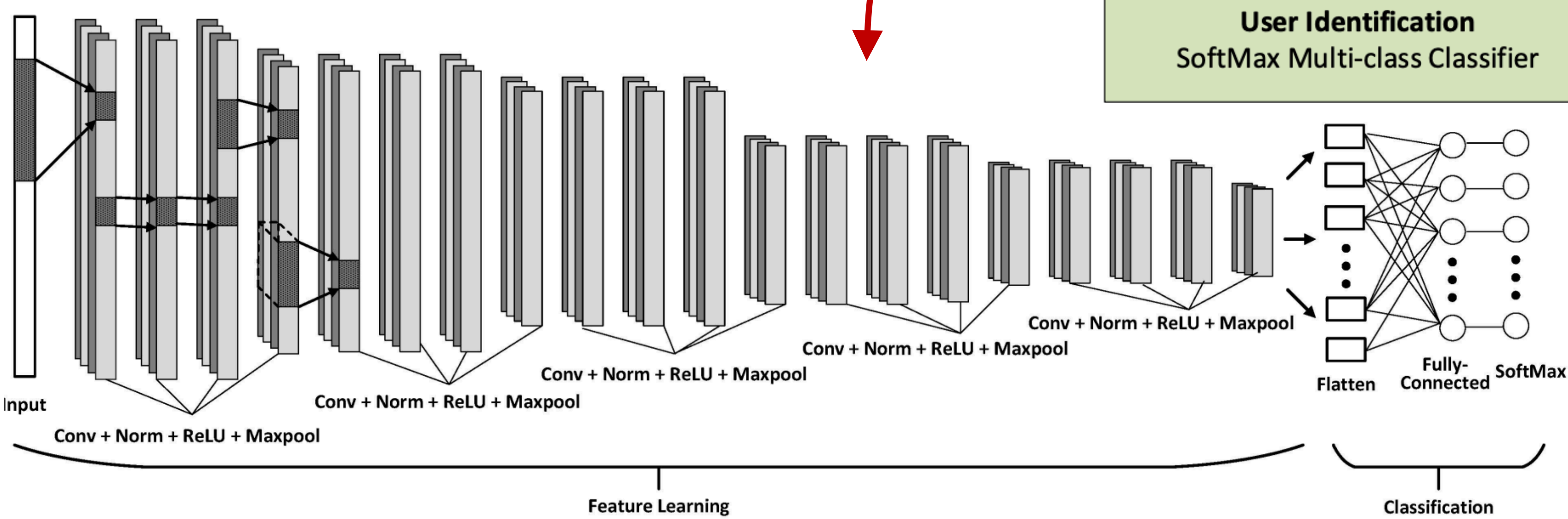




Approach in [Pokkunuru2018]

[Pokkunuru2018] A. Pokkunuru, K. Jakkala, A. Bhuyan, P. Wang and Z. Sun, NeuralWave: Gait-Based User Identification Through Commodity WiFi and Deep Learning. *IECON*, 2018





Possible project developments

- Classification task
  - train a classifier able to discriminate the person moving in the environment
  - use raw data (CFR amplitude and/or phase), or manually compute other features (average, variance, …), or process the raw signals (e.g., Doppler as shown for project D1)
  - compare the data and the results of the four different acquisitions (different points of view)

- Architectures
  - CNN, RNN, attention, …

PROJECT D3

MIME
Master's degree ICT Internet Multimedia Engineering

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO MATEMATICA

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Project no. D3 “people counting through Wi-Fi channel frequency response”

Reference paper

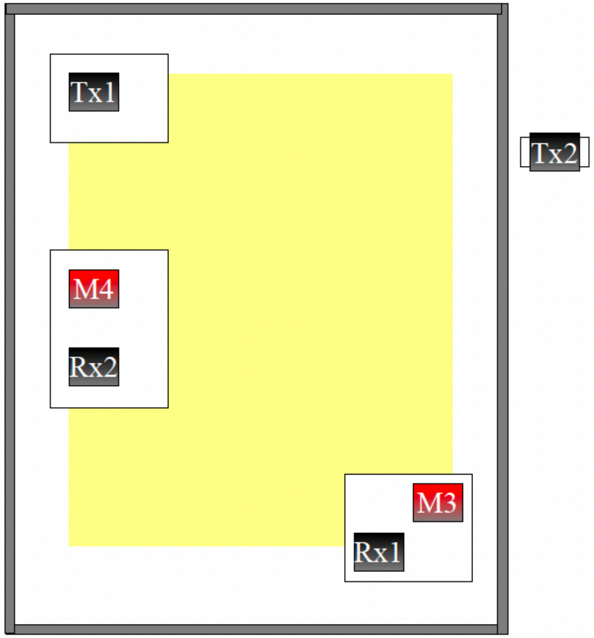
[Meneghello2022] F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello and M. Rossi, SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points. *IEEE Transactions on Mobile Computing*. 2022.

Dataset (23 GB uncompressed) – **PC-** sub-folders

https://drive.google.com/file/d/1t9wrxCyk1_AqXj3j_NQ81tNzwOdhH1j1/view?usp=sharing

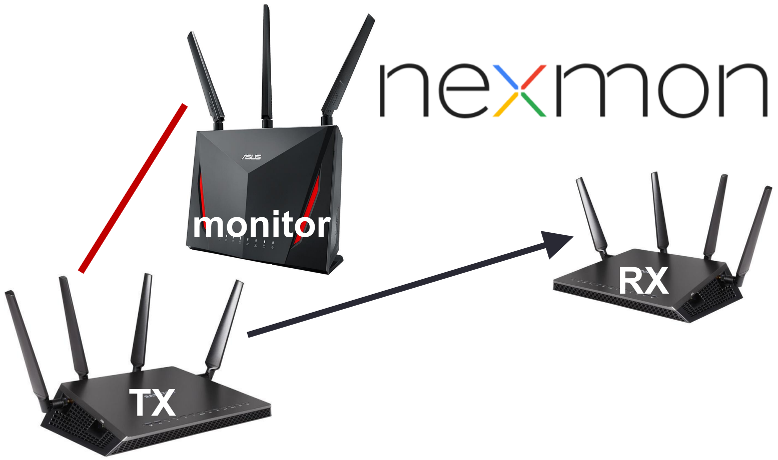
https://ieee-dataport.org/documents/csi-dataset-wireless-human-sensing-80-mhz-wi-fi-channels

Dataset description



- The dataset collects more than 3 hours of Wi-Fi channel readings collected while a different number of people move freely in the environment
  - from 1 to 10

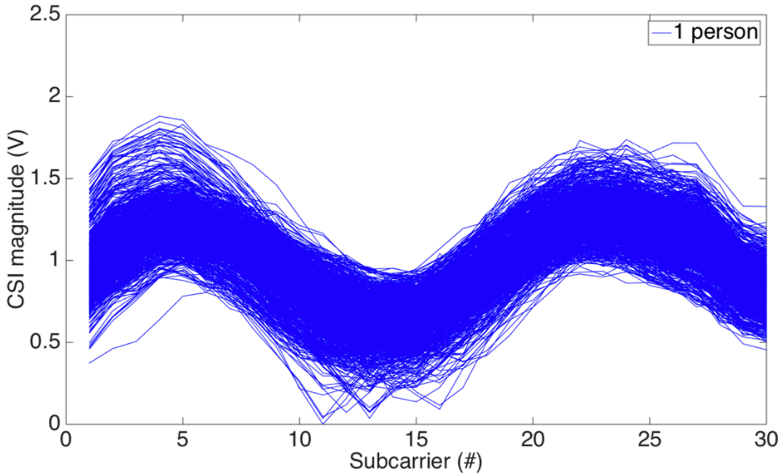
set	campaigns	environment	w × l × h [m]	obstructed path	devices pos.	Tx hardware	Rx hardware	person, Pi
PC-1	a	meeting room	7 × 7.5 × 3.5	-	M3-Tx1-Rx1	Netgear	Netgear	P3, P5-P13
PC-2	a	meeting room	7 × 7.5 × 3.5	✓	M3-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13
PC-3	a	meeting room	7 × 7.5 × 3.5	-	M4-Tx1-Rx1	Netgear	Netgear	P3, P5-P13
PC-4	a	meeting room	7 × 7.5 × 3.5	✓	M4-Tx2-Rx2	Netgear	TP-Link	P3, P5-P13

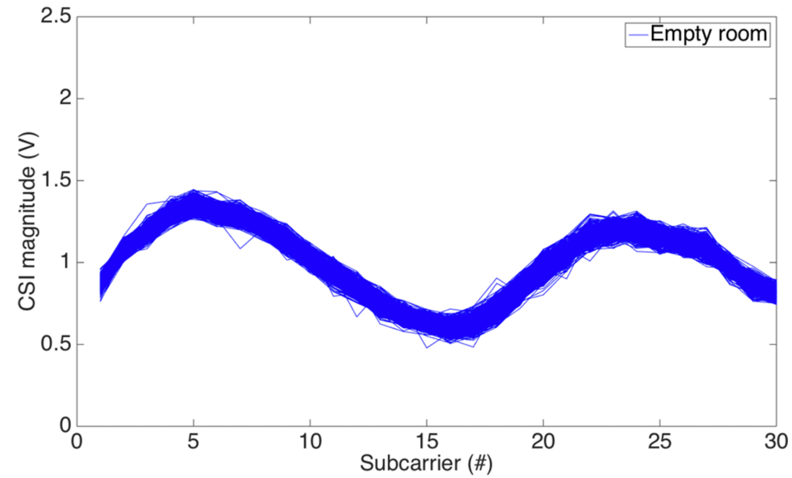


The rest of the description of the dataset is as in projects D1 and D2

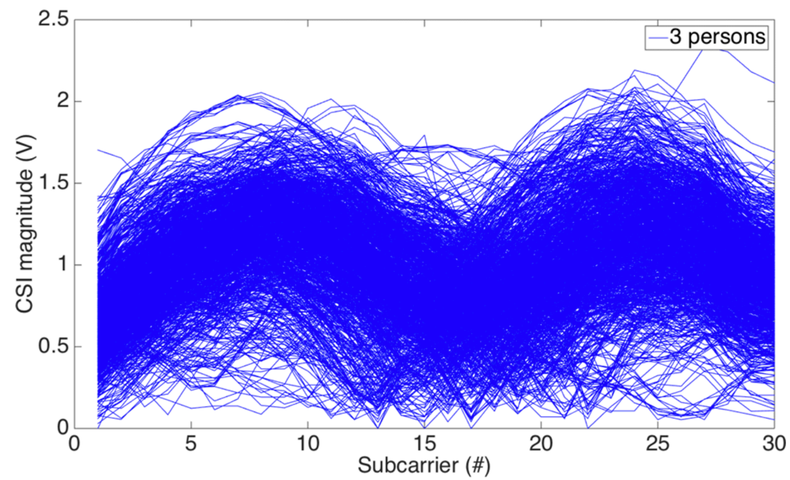
Approach in [DiDomenico2016]

- Manual feature extraction and selection
- Linear discriminant classifier



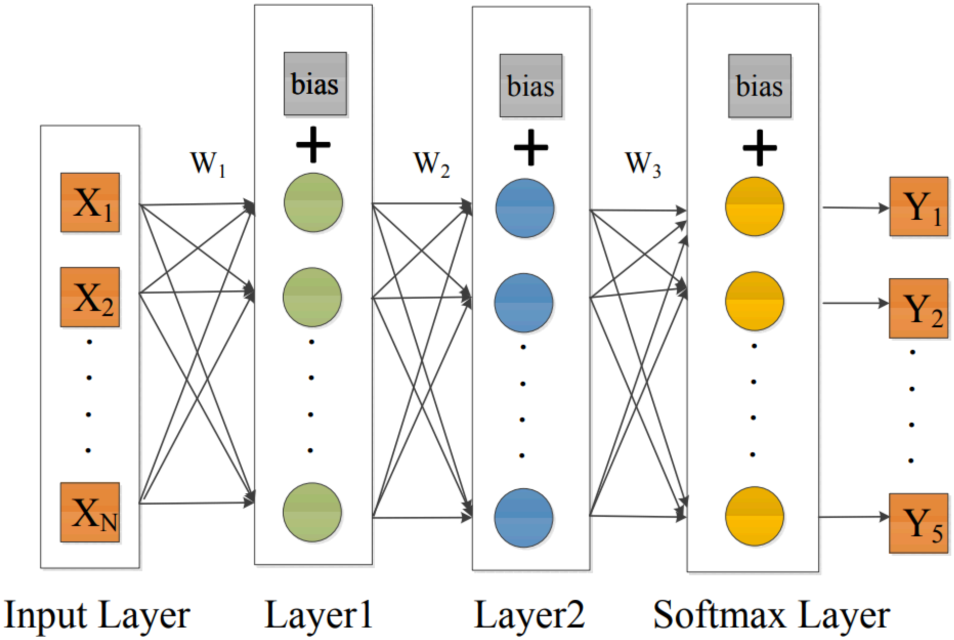


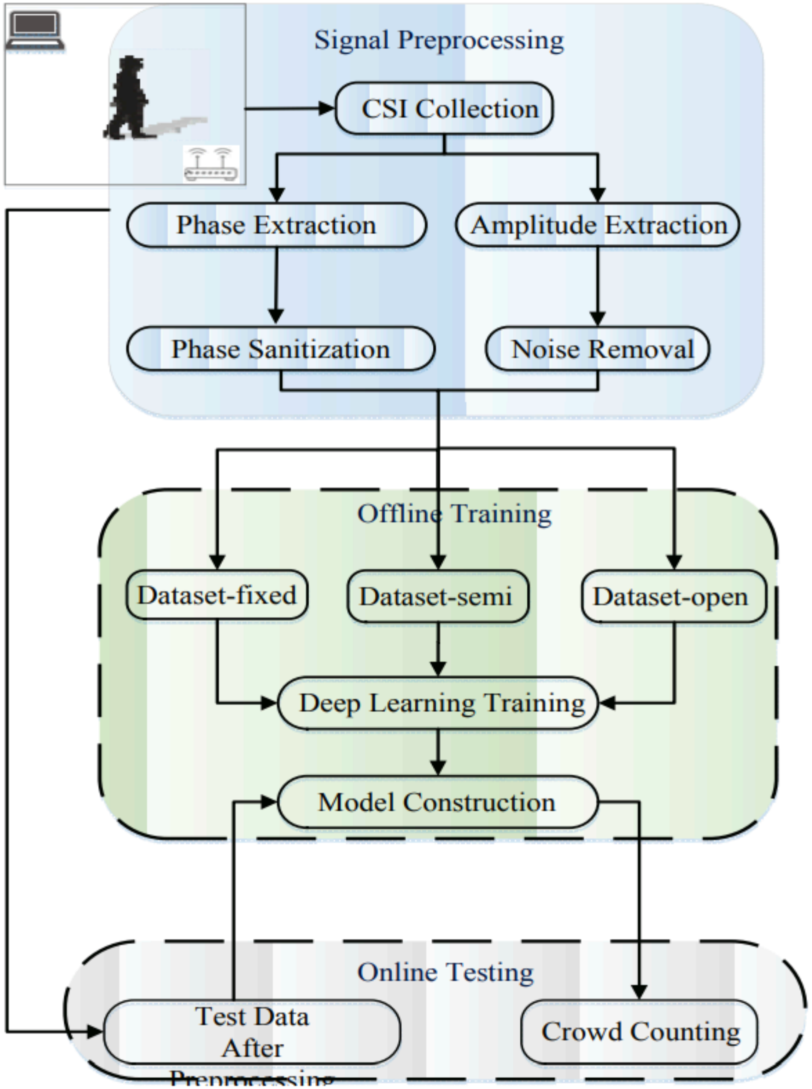
[DiDomenico2016] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, A Trainedonce Crowd Counting Method using Differential WiFi Channel State Information. Proc. of ACM WPA, 2016.



Approach in [Liu2017]

[Liu2017] S. Liu, Y. Zhao and B. Chen, WiCount: A Deep Learning Approach for Crowd Counting Using WiFi Signals. *Proc. of IEEE ISPA/IUCC*, 2017





Possible project developments

- Classification tasks
  - train a classifier able to discriminate the exact number of people in the room
  - use bigger classes for classification (e.g., from 1 to 3 people, from 4 to 6 …)
  - use raw data (CFR amplitude and/or phase), or manually compute other features (average, variance, …), or process the raw signals (e.g., Doppler as shown for project D1)
  - compare the data and the results of the four different acquisitions (different points of view)

- Architectures
  - CNN, RNN, attention, …

HDA COURSE PROJECTS

Instructor

Michele Rossi - michele.rossi@unipd.it

Lab. classes

Francesca Meneghello - francesca.meneghello@dei.unipd.it

Silvia Zampato