Semi-Supervised Autoencoder for Human Activity Recognition

Human Data Analytics

Denis Deronjic

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

15 September 2021

Outline

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

- Problem Description
- Data Pre-Processing
- Conv1D NN
- Inception Style NN
- Semi-Supervised Autoencoder
- Conclusions

Problem Description

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Human Activity Recognition (HAR):

Classification of Human Activity such as Running and Walking from sensors or videos

Sensor type HAR:

- Less difficult than video based (no need for cameras in the env.)
- Wide adoption of devices such as Smartphones and Wearables
- Privacy oriented (as mentioned no need for camera)
- Reduced costs of IMU sensor

Classification Model:

Deep Learning approach using raw data (no hand-crafted features)

Dataset

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Dataset composed by 4.5 hours of measurements in 9 axis from 16 individuals:

- Accelerometer (3 axis)
- Gyroscope (3 axis)
- Magnetometer (3 axis)

Activities Tracked

- Standing
- Sitting
- Lying
- Walking
- Running
- Jumping
- Falling

https://www.xsens.com/products/mtw-awinda

Data Pre-Processing

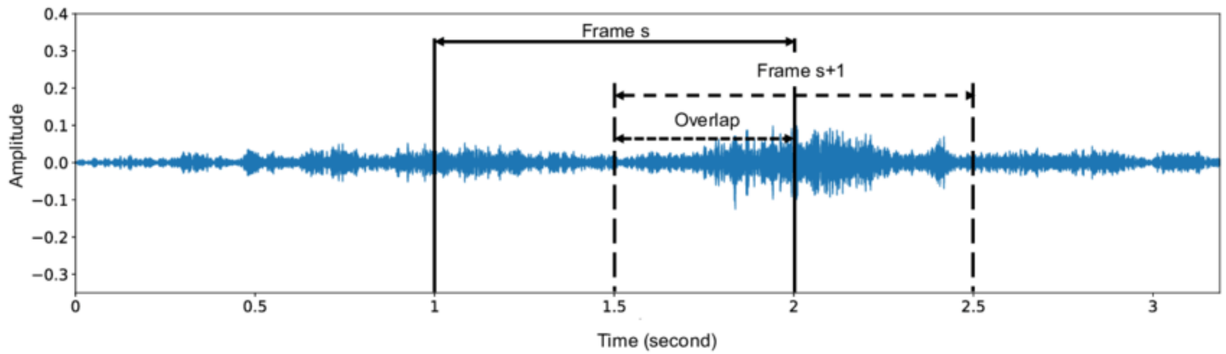
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Difficulties with HAR data:

- Data taken from different individuals
- Every activity has different duration

Solution: Data framing

- Window size of 128
- Step of 8



https://www.researchgate.net/publication/333914934_End-to-End_Environmental_Sound_Classification_using_a_1D_Convolutional_Neural_Network

Data Framing

Magnetometer Drop

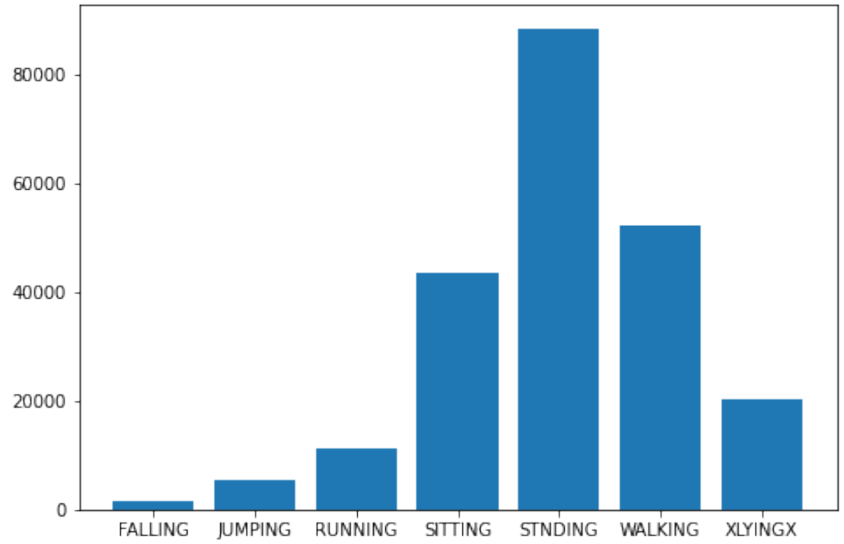
→

Data shape: 128x6

Data Pre-Processing

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Not all labels have enough data



Data augmentation?

Unfortunately data augmentation worsening the results in this case

Data augmentation could add bias to the data

Model learns from a wrong data distribution

Fortunately the following models are able to classify well also the less represented labels

F1-score metric used to test the models

$$\text{F1-score} = 2\frac{Precision \cdot Recall}{Precision + Recall}$$

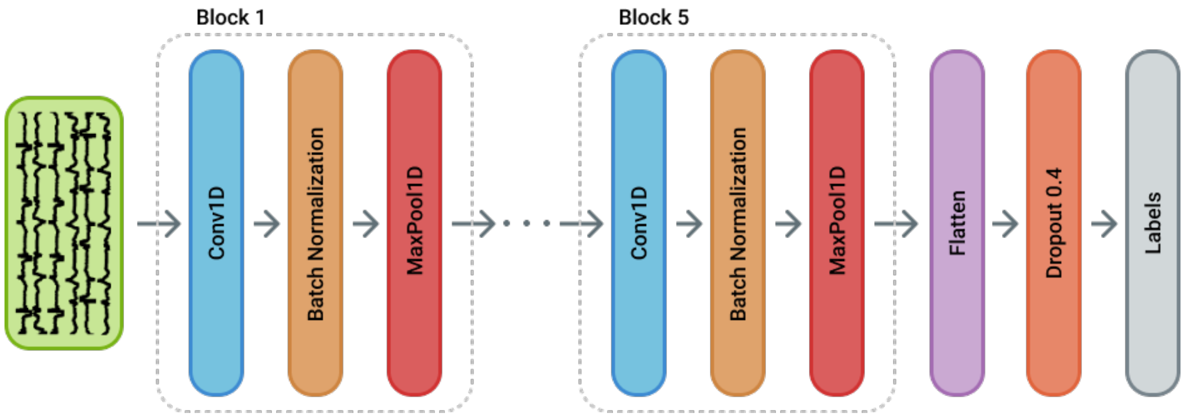
Conv1D NN Model

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

5 Convolutional layers each followed by batch norm and max pooling

Kernel Sizes: 9,7,5,3,3 **Stride:** 1 **Padding:** Same

Num Filters: 16,32,64,128,256

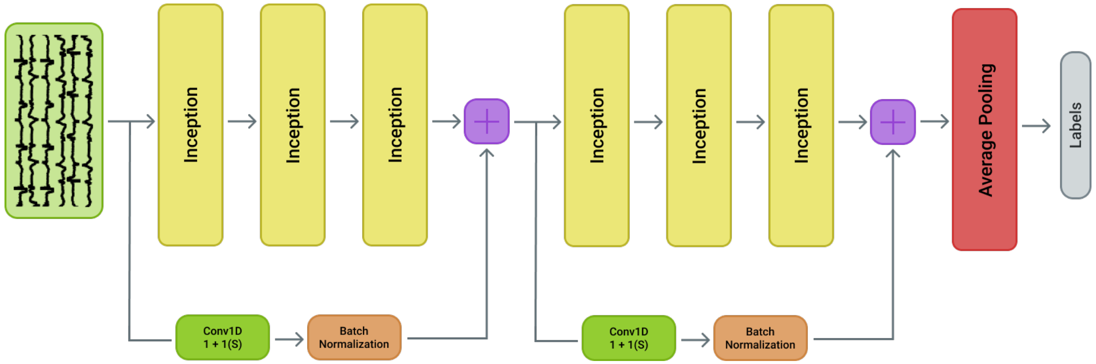


F1-Score	Falling	Jumping	Running	Sitting	Standing	Walking	Lying
Raw	0.92	0.96	0.99	0.98	0.99	0.99	0.98
DCT	0.77	0.87	0.98	0.84	0.92	0.98	0.77
Raw+DCT	0.86	0.94	0.99	0.93	0.97	0.99	0.90

Accuracy
98.83%
91%
96.2%

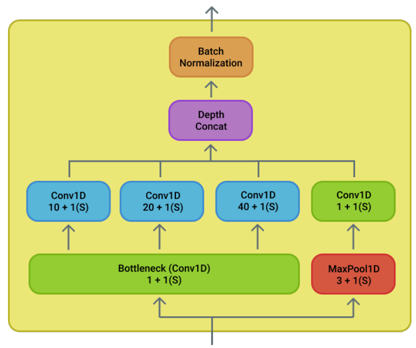
Inception Time NN

- Inspired by Inception Architecture from Computer Vision Domain
- Inception Network designed for Time-Series data
- Residual connection for every three inception blocks



Inception Block

- Four parallel Conv1D branches
- Concatenation
- 1D convolution with kernel size 1 speeds up the training phase
- Each Conv1D has 32 kernels

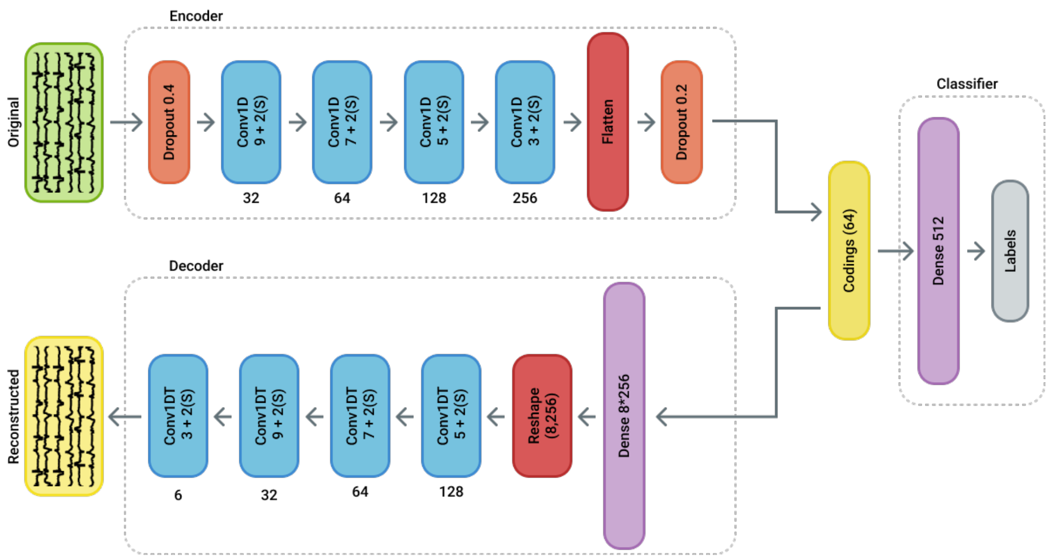


F1-Score	Falling	Jumping	Running	Sitting	Standing	Walking	Lying	Accuracy
Raw	0.9	0.95	0.99	0.95	0.97	0.99	0.93	95.63%
DCT	0.65	0.85	0.97	0.89	0.94	0.98	0.8	92.26%
Raw+DCT	0.83	0.93	0.99	0.94	0.96	0.99	0.91	96.17%

Semi-Supervised AutoEncoder

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

- Composed by encoder, decoder and a classifier
- Learns from reconstruction error and classification error simultaneously
- Training time is similar to a classical AutoEncoder
- De-noising capability thanks to the dropout in the input of the encoder



Loss Function

$$J_{SSAE}(\theta) = \alpha J_{rec}(\theta) + \beta J_{class}(\theta)$$

0.1 1.0

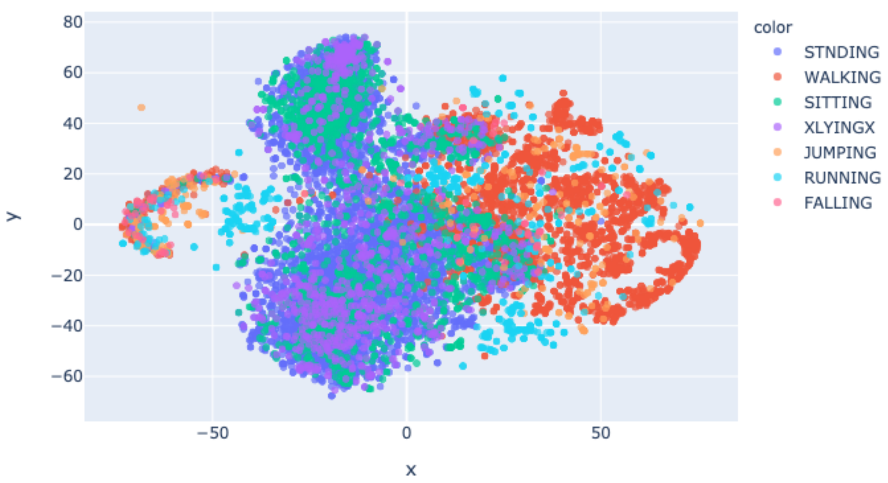
Semi-Supervised AutoEncoder

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

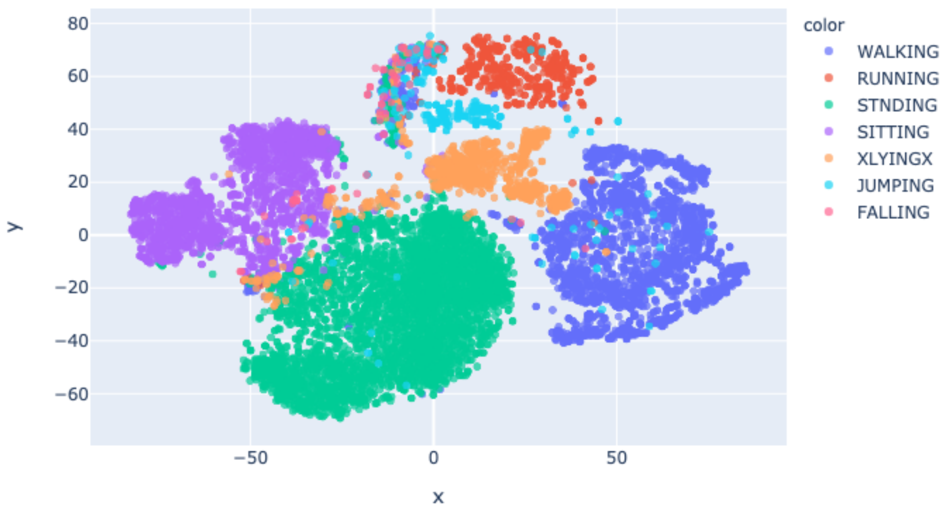
Advantages:

- Classification and Reconstruction in the same network
- Simple architecture but powerful and lightweight
- Good clusters in the latent space

AE latent space



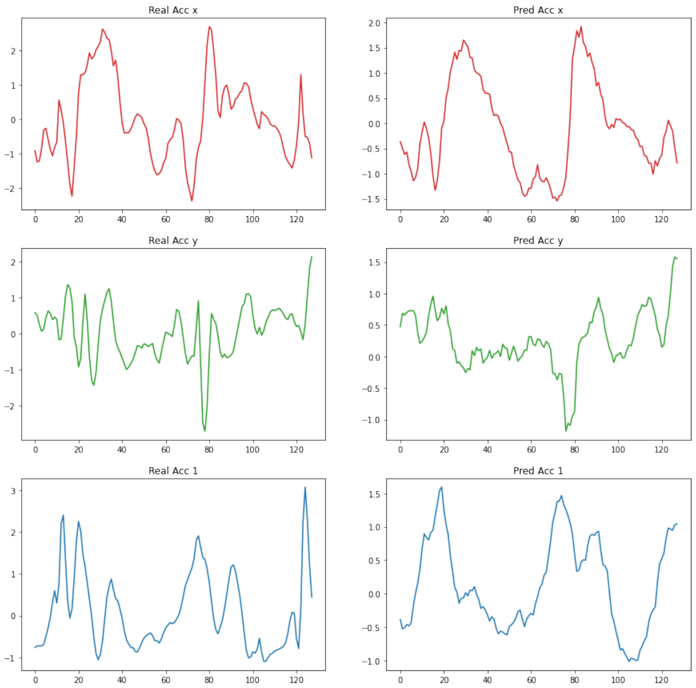
SSAE latent space



SSAE Results

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Reconstruction



Classification

F1-Score	Falling	Jumping	Running	Sitting	Standing	Walking	Lying
Raw	0.79	0.94	0.99	0.97	0.98	0.99	0.95
DCT	0.56	0.68	0.92	0.79	0.88	0.96	0.63
Raw+DCT	0.79	0.93	0.98	0.96	0.98	0.99	0.96

Accuracy
97.42%
85.47%
97.42%

Conclusions

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

- Semi-Supervised AutoEncoder achieves good classification and reconstruction results
- Classification results comparable with Conv1D model (best model)
- Could be used for supervised pre-training (few labeled samples can improve significately the reconstruction error)
- For classification task of Human Activity is sufficient to use raw data

Future works

Implementation of Variational semi-supervised Autoencoders (VSSAE) for generative capability