Deep Learning for Transport Mode Recognition: A Case Study

Hugues Moreau (CEA LETI, Grenoble, France, hugues.moreau@cea.fr)
Andrea Vassilev (CEA LETI, Grenoble, France, andrea.vassilev@cea.fr)
Liming Chen (LIRIS, Ecole Centrale de Lyon, liming.chen@ec-lyon.fr)

Abstract

In Computer Vision, Deep Convolutional Neural Networks simplified the classification pipeline by replacing the explicit feature computation. With other sensors, can we hope this simplification will occur ? Experiments on inertial sensors (Accelerometer, gyrometer, magnetometer) show some preprocessing steps are still necessary to extract the information.

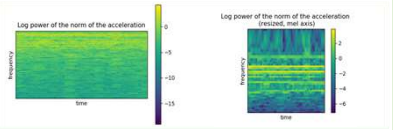
Transport Mode Detection from inertial sensors

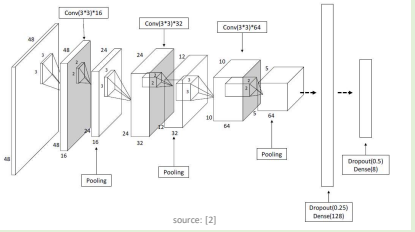
Exemple of applications: automatic carbon footprint tracking, city-scale planning, or heath monitoring

Application to the Sussex-Huawei Locomotion dataset ([1]) which has 7 sensors (both real and virtual) embedded in a smartphone. Despite the organization of a challenge, most methods differ from each other.

Methodology

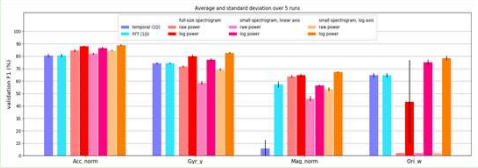
We use a network which architecture comes from the literature [2], on spectrograms which 'frequency' axis is displayed using a logarithmic scale

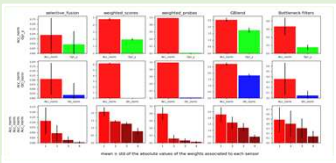




Preprocessing

2D convolutions on spectrograms (time-frequency diagrams) are far better than 1D convolutions on raw (temporal) segments. 1 dimensional convolutions on the norm of the Fourier Transform of the signal ranks between these two methods. Using a log axis for the frequencies allows to compress the information effectively (better than a linear axis).



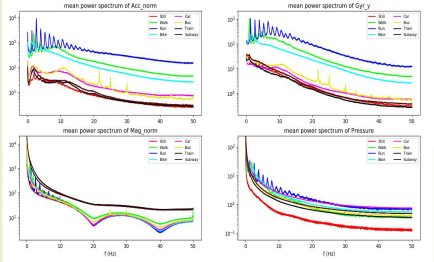


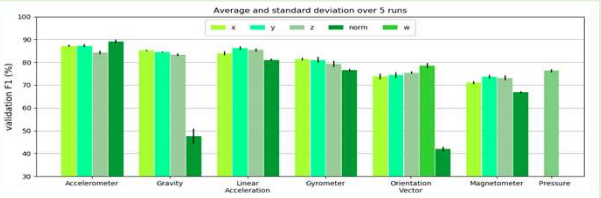
Automatic sensor selection ?

Some fusion methods assign an explicit weight to each sensor. One could think of providing all the sensors to a single network, and letting it choose the best sensor *combination*. However, in most cases, the network uses all sensors available, without restraint.

A pronounced spectral profile

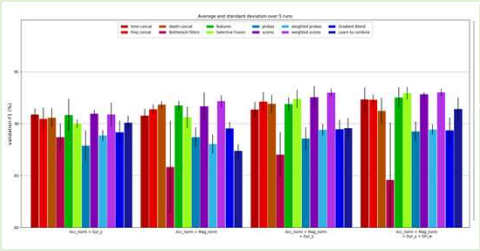
Different signals have specific spectra





Unimodal sensor evaluation

Real sensors (accelerometer, gyrometer, magnetometer) are better than the virtual sensors deriving from them (gravity and linear acceleration for the accelerometer, orientation for the gyrometer). Individually, the best real sensor is the accelerometer, followed by the gyrometer, the barometric pressure, and the magnetometer.



Data fusion methods

Few differences between 'simple' methods (concatenation of the input signals, average of predictions), and the more complex ones coming from the literature.

Conclusion

The choice of an encoding for the signals (1D temporal segments, 2D spectrograms) is paramount, the sensor choice is secondary, and the choice of a fusion method is irrelevant.

What about the other signals ?

Bibliographic study: How do practitionners preprocess temporal data with ? Is it mandatory to compute frequency comonents (spectrograms, Fourier Transform) ?
The answer is not obvious. Few direct comparisons exist, but we can observe some tendencies in the state of the art (STFT spectrograms for ECG/EEG, wavelet spectrograms for accelerometer from roll bearing study, some audio applications use mel-cepstral coefficients while other process raw signals).

Notable exception: [3] experimented with audio signals (tagging of musics), and 2D convolutions on spectrograms are better than convolutions on 1D temporal segments when the training dataset has less than a million songs, both preprocessing methods becoming equal when the dataset reaches this value.

References

[1] Wang, L. *et al.* 2019. « Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition With the Sussex-Huawei Dataset »

[2] Ito, Chihiro *et al.*. 2018. « Application of CNN for Human Activity Recognition with FFT Spectrogram of Acceleration and Gyro Sensors »

[3] Pons *et al.*. 2018. « End-to-end learning for music audio tagging at scale ».