Indoor Human Walking Path Reconstruction from a FMCW Radar Signal

Federico Matteo[†], Domenico Solimini[‡]

***Abstract*—Surveillance systems call for new solutions. Modern video-cameras, in spite of high resolution performance, are not able to cope with difficult conditions such as not lighted environments and other blindness scenarios. On the contrary, radar devices perform better in this kind of situations and represent a valid alternative for the majority of surveillance applications. In this paper we investigate how the tracking procedure of a person who freely moves into an empty room could be improved considering both computational efficiency and signal quality. We achieve this aim by using a Gaussian Mixture Model in order to estimate the subject position and its shape in the Range Doppler signal. Moreover, we also developed further machine learning models to improve the proposed algorithm and extend its application to a wide range of demands, e.g. surveillance systems based on sensors networks. We also investigate the use of a dimensionality reduction system, i.e. several architectures of autoencoder models, which can compress and decode the signal with varying loss of accuracy. The proposed methods are discussed via experimental analyses and graphical performance results, since they all belong to the unsupervised learning field. We eventually consider our system ready to provide solutions for real-world problems.**

***Index Terms*—Target tracking, feature learning, indoor sensing, micro-Doppler, autoencoders, denoising.**

I. Introduction

The belief that smart sensing can have a crucial property for surveillance systems has motivated a widely research attention on human detection and tracking through years. The majority of the studies on this field have been focused on video cameras and laser scanners, although sensors of this kind present several limitations when collecting information, such as difficulties to deal with darkness and blind conditions caused by weather or concealing clothing. A radar system can be considered a modern and suitable approach to work out these problems. Indeed, radars preserve visual privacy, penetrate walls or obstacles and reveal highly detailed movements. These results are achieved without requiring the target to wear a sensor, enabling radars to be the main alternative to a wide range of applications.

A big challenge for the microwave radar sensor is how to keep the monitoring cost low while providing sufficient range detection and monitoring accuracy for specific applications needs. In this paper we are using measurements obtained from a low-power frequency-modulated continuous-wave (FMCW) radar available in the open dataset IDRad [1] implementing various techniques to handle the noise and providing a good tracking accuracy. We also require computational resources to be as low as possible. Differently from many other studies [5] - [7] our results do not have need of any additional constraint, we simply track a single subject that freely moves inside a room.

The dataset [1] consists of 150 minutes of annotated micro-Doppler data equally spread over five targets who are free to walk and move into two different and unfurnished rooms. In [1] the authors investigate the use of micro-Doppler signatures retrieved from a low-power radar device to identify a set of persons based on their gait characteristics. In this paper we try to improve its preprocessing phase as well as to speed up the subject tracking process and to improve the robustness of the procedure. We use a Gaussian mixture model (GMM) to automatically identify the bump centre and its position through the estimate of the velocity and distance means, then we build an elliptic shape to approximate the bump contours. Once we collect this data for each subject and for each room in the training set, we investigate three different approaches, i.e. a smoothed spline, a self organising map and two deep learning architectures in order to regularise and denoise the trajectory that the target person follows. Owing to the simple structure and fast processing speed, this method is applicable to a variety of radar sensors for ranging and detection.

The main contributions of our project are achieved:

- through the fact that in real world scenarios our method can have a good performance, independently from the radar placement in the room and the movements of the target subject;
- by applying a dynamic thresholding procedure in order to remove noise in tested condition setup;
- by considering automatic and quick subject tracking via Gaussian mixture models, in contrast to a slow feature identification;
- by reconstructing the subject trajectory and its shape in the radar signal via advanced unsupervised learning techniques.

The rest of the paper is organized as follows: Section II briefly lists related works in the area of radar signal processing. Section III and Section IV describe the followed processing pipeline and how the MD signal has been handled, respectively. In Section V, we subsequently explain the proposed approaches. Section VI consists of an extensive description of our methods' performances and Section VII contains conclusions and suggestions for possible applications, as well as some possible directions for future researches.

[†]University of Padua, email: federico.matteo@studenti.unipd.it
[‡]University of Padua, email: domenico.solimini@studenti.unipd.it

II. Related Work

In recent years the main efforts in the literature have been spent on improving the existing state-of-the-art person identification machine learning algorithms. In particular, a large attention have been focused on the transition from supervised learning techniques to the unsupervised learning methods, which in this case relies upon deep learning architectures. Whereas these schemes attained high accuracy, they are based on tracking methods that usually suffer of case sensitivity and high computational requirements. These cons arise from the fact that a significant amount of work has been done in the domain of identifying individual subjects based on video images as input [12] or using wearable devices like accelerometers [11] or smartphone's sensors [10].

The typical analysis of a sensor's study starts from processing the raw data signature to reduce noise and retain only essential information.

In [1] the former stage is performed by studying the distribution of the relative power before and after a subject is identified in the environment. The authors exploit the normal distribution shape of the doppler shifted signal and retrieve an optimal threshold value. We highlight how it is impossible to apply this kind of thresholding whether the empty environment has not been previously recordered from the radar. We notice that when dealing with feature identification, the representation of salient physical movement attributes can be a tedious task involving tuning of parameters and thresholds. Examples of this procedure are [2] and [3]. One way to overcome these issues is to have an approach which learns and captures the detailed and intricate properties of the MD signal. Although this can be achieved via deep learning, this step usually requires a significant computational power and a specific architecture to be trained. In [4] the learning of the input representation is achieved using stacked autoencoders, where output of one autoencoder is input to the next one. In this paper we prefer to maintain the first stage as general as possible by considering a dynamic thresholding procedure and to speed up the second stage by applying unsupervised learning techniques that overcome deep learning architectures velocity.

The last discussed aspect is linked to the increasing interest in lightweight wireless radar networks whose applicability in human tracking has been explored in various studies, as [8]. An important advantage in this field is represented by the ability of compressing the retrieved signal with the minimal information loss. We have explored various solutions to achieve this goal.

III. Processing Pipeline

Our processing framework could be ideally divided into three steps, as shown in Fig. 1.

In the first one, the raw input data coming from the FMCW radar have been processed using a two-dimensional Fourier transform in order to extract a map that for every *time-distance-velocity* coordinate input returns the intensity of the reflected Doppler shifted signal. Those information have been represented through a series of images - one for each input time - and then modified to detect the signal generated from the subject. A dynamic thresholding has been applied to accomplish this last requirement, in order to remove the low intensity signal generated from background noise and subsequently removing the signal generated by the reflection of the radar waves on the static objects in the room. In this way, for every time instant a set of *distance-velocity* points that should have been generated by the moving subject is selected.

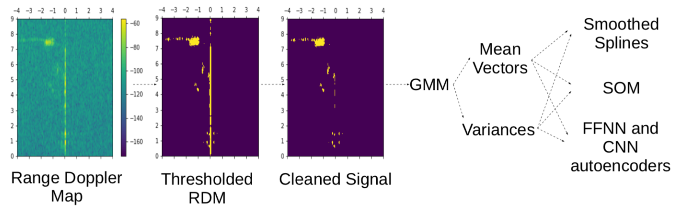

Fig. 1: Schematic representation of the processing pipeline.

The second step is focused on cleaning the signal, removing all the points that are not related with the subject motion and on the extraction of a minimal set of features that can be used for effectively tracking the subject position. Various clustering algorithms have been applied to identify the significative points inside the *distance-velocity-signal intensity* space; the one that achieves the best performance is the Gaussian mixture model (GMM) with two classes. The cluster presenting an estimated centroid with the highest value of intensity is assumed to contain the signal generated by the target movement. The signal shape has shown to be effectively approximated by an ellipse whose axis are estimated multiplying by 3 the variances of the two coordinates of the points belonging to the selected cluster.

In the last step, signals obtained from the time series of the clusters' centres and the ones obtained from the ellipses axis have been regularized by removing sudden variations from the trend, obtaining the estimate of a smoother evolution of those variables. The assumption behind this step is that the real subject trajectory evolves regularly and the irregularity of the estimated evolution patterns are uniquely due to a lack of precision in the parameters estimate.

Several different methods with increasing complexity have been implemented. A *smoothed spline* is the easiest and most straightforward approach for data series regularization. It is very fast and it does not require to manually specify any parameter, so it could be easily applied to a wide range of situations. However, because of its widespread applicability, it is not specifically suited for denoising purposes and its results are easy to overcome.

Self Organizing Map is a standard approach for dimensionality reduction problem. We use it to map the time series onto a small number of points. Interpolating these points with a spline could give an excellent regularization of our data. It suffers the presence of small clusters of noisy points, therefore data have been previously cleaned by applying a DBSCAN clustering and by not considering all the points labeled as noise for the SOM training.

Various *Autoencoders* based on fully-connected feed-forward neural network architectures have been implemented as well as one convolutional neural network autoencoder. These models represent some of the most established approaches for signals denoising and compression, they are discussed in section IV.

IV. Signals and Features

Doppler radars are able to detect objects in relative movement with respect to the radar itself through the emission of electromagnetic waves and by capturing the reflected waves modulated by the target motion. Information about the speed of the target is retrieved from the difference between the emitted and received wave length, distance of the moving object is inferred from the delay between emitted and received signals. In this way, it is possible to compute a *Range Doppler Map* (RDM) that assigns to each *distance-velocity* couple the corresponding ratio of received signal power, measured in decibel. An extensive description of the physics behind this devices, included details of the applications concerning the detection of a walking human are present in [9].

The signals used for this analysis come from the IDRad open dataset that has been collected for a previous study [1] using a $77GHz$ Frequency Modulated Continuous Wave radar. The main drawback of this equipment is due to the low signal-to-noise ratio (SNR) of $8dB$ on average.

Signals used in our work are structured in 123 different folders each containing 179 consecutive frames capturing 12 seconds (15 *fps*) of one person walking in two different rooms without any moving object that could increase sourronding noise. One of the rooms is 6 times 9 meters wide while the other one has dimensions 5 times 8 meters. The radar is located in one of the shorter walls of each room and features a range and a velocity resolution of $10cm$ and $2cm/s$ respectively, allowing the detection of walking gait details. Speed ranges between -3.8 m/s and 3.8 m/s. Five different people - one for each set of measures - are free to walk in any direction in the room, turning, stopping and making casual movements. For these reasons the dataset can be considered as very general and representative of real world condition.

As said, one of the biggest challenge that this dataset presents is due to the low SNR, it is important to clean the signal discriminating low intensity points from the peak generated by the moving subject.

The technique that was chosen to overcome this problem is simple yet effective. As a first step, a threshold is dynamically applied to each RDM, so that only the points presenting an intensity higher than the threshold are considered for further analysis. A suitable threshold value can be obtained for each RDM by considering the distribution of the signal intensity and selecting an intensity value that leaves to its right side a value included between 2 and 3 percent of the distribution. In this case the value leaving the 2.5% to the right tail of the distribution has been chosen. This estimate is robust enough to fit the variety of available person-room combinations.

In the previous analysis of this dataset [1], the detected signal distribution in the empty room was exploited to detect the additional signal generated by the subject, identifying an optimal threshold of $-45dB$. This approach leads to optimal results in terms of signal cleaning but it is sensible to the specific setup and hardly replicable in a different environment. However, despite the lower amount of noise generated by this technique, the subsequent steps yield equivalent results between the two discussed approaches.

Static objects generate a strong intensity reflected signal that is a big source of noise. In spite of this, noisy points are arranged in a line corresponding to very small speed targets, so they can be easily removed. All speed values assumed by a number of points greater or equal than one half of the maximum possible number - i.e. the total distance range divided by the distance resolution - are selected and all the points presenting that speed value are discarded.

In the second step of the feature extraction, selected points are divided into two clusters using a 2 classes GMM in the *range-speed-intensity* space. Other clustering techniques including Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and k-means have been implemented but results have not been fully satisfactory. One cluster is assumed to be the class to which the target subject belongs, so the remaining one is considered to be noise. The algorithm is imposed to automatically identify two gaussians, thus the centre of the gaussian distribution with higher signal intensity can be directly used to estimate the centre of the bump generated by the target.

The shape of the bump, however, is not effectively described by the covariance ellipses generated by the GMM model because of the strong influence that a small number of noisy points can exercise on these quantities. A better result is obtained by selecting the variances of the coordinates of those points belonging to the clusters and approximating the bump with an ellipse having the axis equal to the computed variances multiplied by 3. The interquartile range and the difference between the 90th and the 10th percentiles were also applied, but turned out to be less informative than the variance. This procedure can fit live tracking requirements, since it is capable of retrieving centres and the approximate shapes of the signal peaks through the whole time range of $12s$ in an average time of $16s$, using an Intel Core i3-5010U CPU with a maximum frequency of $2.10GHz$. If we reduce the sampling original frame per second ratio up to a 11 *fps*, our machine is suitable for the live tracking procedure.

V. Learning Framework

Four series of data describing the two components of the centre of the clusters and their extension are retrieved from the previous analysis.

Those series suffer the presence of sharp variations on the top of the regular trajectory generated by the subject's torso movements and by the overriding sinusoidal-like signal generated by legs and arms rocking. Even stronger irregularities are noticeable in the variances sequences since those parameters are not directly linked to a precise physical quantity but are used to estimate the evolution of the shape of the signal generated by the subject. These irregularities are easily noticeable by a human being just looking at the obtained time sequences with a graphical representation. A direct observation of the ellipsoidal approximation of the bumps often does not give such an immediate and unambiguous response. Temporal information, therefore, are extremely valuable and could lead to strongly improved estimates.

The proposed methods - except for the first one - do not limit their relevance to a significant increase in the quality of the temporal series but simultaneously cause a strong reduction in size of data required to reconstruct the subjects' trajectory.

The following methods are discussed in distinct paragraphs:

- Smoothed spline;
- Self Organizing Maps;
- Feed forward fully connected neural network autoencoder;
- Feed forward convolutional neural network autoencoder.

A. Smoothed Spline

For the smoothed spline development we choose the model described in [6] that avoids the knot selection by using a maximal set of knots. The complexity of the fitting procedure is controlled by regularization, as a matter of fact, among all functions f we find the one that minimize the penalized residual sum of squares:

$$\mathrm{RSS}(f,\lambda) = \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int (f''(t))^2 \, dt, \qquad (1)$$

where λ is the smoothing parameter. Values for which $\lambda = 0$ and $\lambda = \infty$ are not interesting, we hope to find a λ in between, i.e. the optimal value establishing the tradeoff between the closeness to the data and the function curvature, i.e. the first and the second term in the Eq. (1). The positive aspects of this method are that the optimal value of the regularizer parameter λ is automatically found via Cross Validation and that this is the fastest procedure that we propose for the statistical quantification purpose. An example is shown in Fig. 2.

B. Self Organizing Maps

The second method is based on the application of the SOM algorithm, which leads to an organized representation of patterns drawn from the *distance-velocity* space. The drawback of this technique is that we initially have to specify the optimal number of neurons. If it is too high, the reduced output dimension follows each point giving a not cleaned output, if it is too low, then it is quite difficult to follow the points belonging to the input domain. In addition, the regularized trajectory suffers the presence of closely distributed outliers. Therefore we use a density based clustering technique, i.e. a DBSCAN

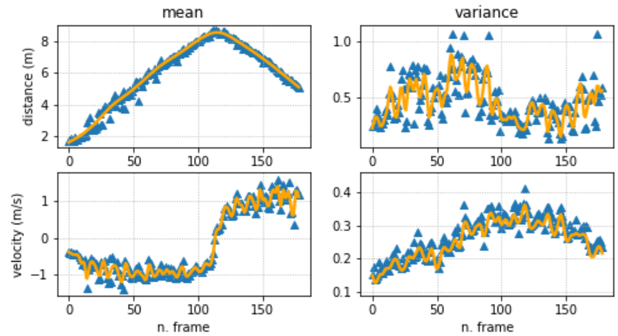

Fig. 2: The smoothed splines regularization for means and variances on distance and velocity.

algorithm in order to remove the noise. This allows the SOM to work well, but by increasing even more the number of parameters that have to be taken into account. As a matter of fact, for the DBSCAN procedure we cannot automatically select the *minpoints* and the *ε-neighbours* values. An estimate for these values can be carried out via experimental results and by a graphical analysis of a k-th nearest neighbour distance distribution.

Good values for *minpoints* and *ε-neighbours* are shown in TABLE 1.

	ε-neighbours	*minpoints*
means	20	10
variances	8	5

TABLE 1

In this way, all the points labeled as noise are removed and a single cluster identified and used to feed the SOM algorithm.

In order to reconstruct the signal, the first measure of the series, the last one and all the neurons are interpolated using a hard spline. With this setting, again using experimental trials, we noticed that a neuron every ten points should be inserted in order to obtain an informative approximation of the signal.

This whole procedure allows to reduce the size of the signal by a factor of ten and - eventually - to easily rebuilt the input signal from the compressed version. Although this procedure can be quite long to retrieve, once the optimal hyperparameters are derived, the whole algorithm outperforms the smoothed spline, results are more extensively discussed in Section VI, while in Fig. 3 an example of the SOM results is shown.

C. Autoencoders Architectures and Training

The encoder-decoder architecture is developed in such a way that the decoder phase is not able to learn a perfect copy of the input signal since a hidden layer of the network (encoded layer) is smaller than the first one, hence the autoencoder is forced to codify a compressed version of the input data. Once the network has been trained, it is possible to split the decoder and the encoder subnets. This aspect makes possible to apply autoencoders as noise filters; the applied

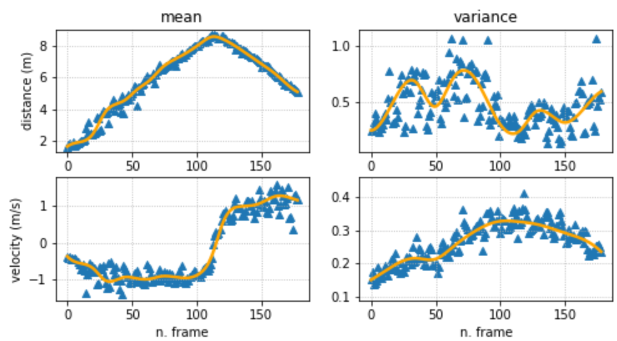

Fig. 3: The SOM regularization for means and variances on distance and velocity.

compression of the signal forces the reconstructed signal to be much more regular than the input one. To reconstruct the signal it is sufficient to pass as input to the decoder the compressed version of the signal generated by the encoder.

In order to quantify the strength of the signal compression, a compression factor has been introduced. It is computed by dividing the number of input neurons of the network by the number of neurons in the encoded layer.

The most important aspect of this architecture is that, provided that the input data present some kind of regularity, the network can adapt to the input distribution returning a cleaned and denoised copy of the input, where outliers are automatically detected and removed. Trough the use of different architectures, different solutions are explored evaluating the compression factors and the ability of the network to interpret the most relevant features of the input series.

Once the main architecture class and the number of hidden layers of the network has been assigned manually, various layer dimensions have been tested, the different architectures are represented in Fig. 4 and in Fig. 7. An exhaustive search of the optimal dimensions goes behind our purposes, however a few architectures showing good performances have been selected. All the models make use of the hyperbolic tangent as activation function and, quite surprisingly, perform better without regularization strategies such as internal dropout layers or L2 regularitazion on the output layer. The reason is related with the complex structure of the pattern that we want to reconstruct: strong regularizing techniques reduce overfitting but tend to ignore fine signal modulations. The loss function, therefore, should be determined uniquely by the closeness of the output to the input signal. A suitable choice turned out to be the mean absolute error function.

Feeding the networks with entire 179 frame long data sequences is unpractical. Data present high variability and the total number of available series is not high enough to effectively train the network.

Even if various data augmentation techniques can be implemented, a simpler solution can be achieved dividing each series in a number of equally spaced windows. In this way the variability of the input data diminishes substantially, as well as the number of neurons needed for an effective encoding of the signal, the time needed to train the network and the needed dimension of the training set decrease too. In addition the number of available data sequences increases.

The drawback of this approach lies in the process of merging the encoded sub-series avoiding sharp transitions. A possible solution can be obtained dividing the original series in a number of partially overlapping windows, interpolating the overlapping decoded sub-series. Nevertheless this approach reduces the compression factor and increases the computational time. In this analysis the number of critical transitions is so low to consider the problem as negligible.

Varying the window size without making any modification to the internal layers of the network, it is possible to observe a variety of behaviours. In Fig. 5 it is represented the loss on the test set as function of the compression factor using the architectures represented in Fig. 4.

For the training procedure one-hundred windows per epoch are specified, each time randomly sampled from the training set. A standard procedure - without random sampling - is applied to the validation set. All the input values have been divided for the maximum range of the respective dimension before being processed by the encoder and the output has been reshaped in the inverse way. Training is stopped once the validation loss did not diminish in ten consecutive epochs. Also a high number of optimizer has been taken into account, we eventually prefer the Adam algorithm.

D. Fully Connected Neural Network Autoencoder

An autoencoder has been built by using a feed-forward fully-connected neural network (FFNN).

We have selected three architectures with an increasing number of hidden layers. The simplest one has only 1 hidden layer with six neurons, while the other two with 3 and 5 hidden layers are respectively described in Fig. 4.

Applying the simplest model, we found a weak capacity to reproduce the signal through a regular shape, but a good capacity of compression. The former is not necessarily a drawback, since it could be good to have a faithful compressed representation of the input. In our case we also seek for regularization, so we consider this aspect a weakness that we would like to prevent.

When complex models are used, it is easily noticed how the signal acquires a more structured form with time evolution, effectively reconstructing the walking gait in the wide majority of the observations as shown in Fig. 6. In this sense the best model is the one presenting an internal structure with $12 \times 6 \times 12$ neurons, as shown in the upper diagram of Fig. 4.

E. Convolutional Neural Network Autoencoder

A convolutional autoencoder ordinarily contains at least two convolution layer along the temporal domain, one pooling layer and one upsampling layer. The input, output and the encoded layer instead are dense. The simplest possible architecture, therefore, is the one represented in Fig. 7. This is also

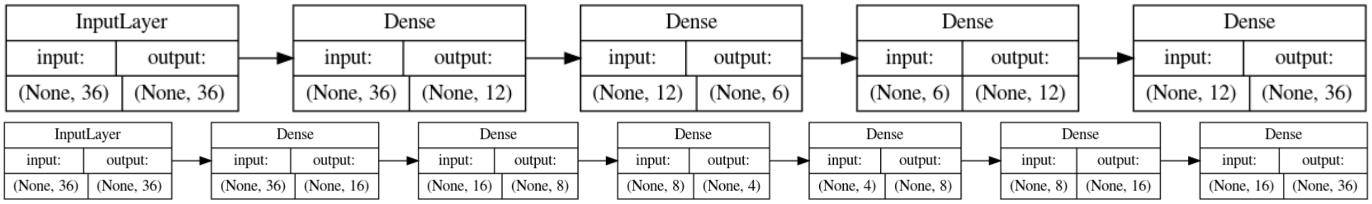

Fig. 4: In the upper diagram the three hidden layers FFNN architecture, in the lower one the five hidden layers FFNN architecture used in the analysis.

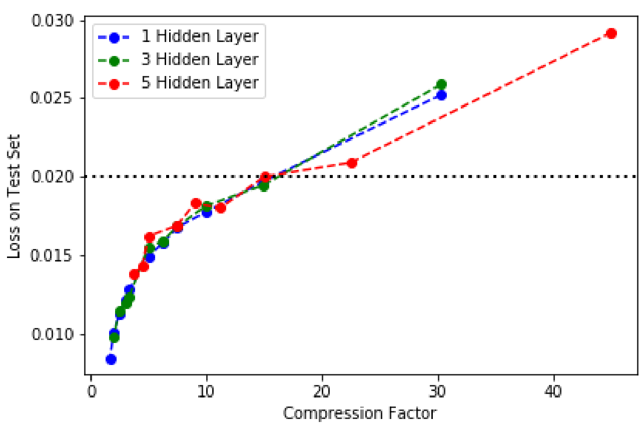

Fig. 5: Loss on the test set vs compression factor using different kind of FFNN architectures. When keeping loss under 0.02, all the selected networks perform the same in terms of compression. Relevant differences can be observed in the signal quality.

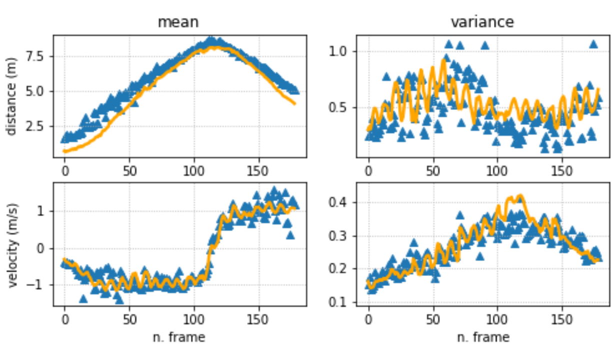

Fig. 6: Plot of the encoded representation of means and variances using the FFNN encoder with 3 hidden layers. It is noticeable how fine signal's patterns in highly noisy scenarios are reconstructed.

the architecture of choice for this analysis since no significant improvement appears increasing the network complexity.

In the convolutional layer a set of filters, or kernels, are multiplied by convolution for the previous layer to obtain highly representative features. Inside the encoder part of the network, the max-pooling layer looks for the maximum within a region of specific width and height. This corresponds to a subsampling which in this case is extremely important for data dimension reduction. The upsampling layer, on the contrary, is a simple layer with no weights that increases the dimensions of input and is used in a generative model when followed by a convolutional layer.

CNN are widely used because of their abilities of extracting relevant features from data at different levels similar to a human brain. This characteristic is extremely interesting for this analysis since the biggest challenge is reconstructing the global trend of the signal plus regular fine modulations generated by the walking gait. On the countrary of the others, this model performs very well showing a highly regular output that closely follows the input sequence without being affected by noisy and irregular points as represented in Fig. 8. This aspect is further discussed in the following section VI.

VI. Results

Each metric and model analysed in this paper relies upon the field of the unsupervised learning. Starting from the pro-processing phase, in which clustering algorithms are applied in order to retrieve the bump location of the FMCW signal, up to the last stage where machine learning models and neural networks systems are tuned for the trajectory regularizarion. Each of this phases is investigated manually; we prooceed by taking into account a single technique, analysing its output and retrieving a benchmark in terms of the behaviour on the testing set by means of a visual analysis of the plots to evaluate its performance.

The first aspect considered for discriminating has been the choice of the best clustering algorithm to use in order to retrieve the target position. We have concluded that the GMM was the best one. In this case, the graphical analysis has been immediate and voluntarly not shown herein. DBSCAN has been difficult to generalize since each frame presented a different density in the background noise. Therefore it would have been impossible to choose a value of the ε-*neighbours* and the *minpoints* parameters that could fit every situation. The decision between k-means and GMM has not been trivial, as a matter of fact for the considered aim they behave quite similarly, although the GMM seems to outperform the k-means results in some critical cases.

After this phase, the most important and more investigated decision has been which learning model choosing to regularize the means' trajectories and the ellipses' shapes. Using a machine learning model we aim to fit a curve with a

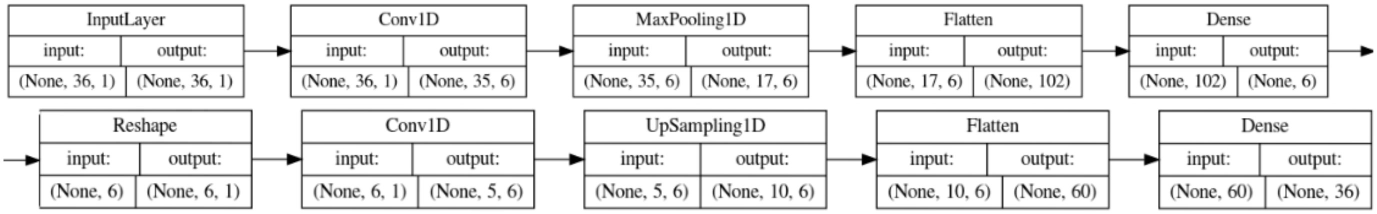

Fig. 7: The CNN architecture used in the analysis.

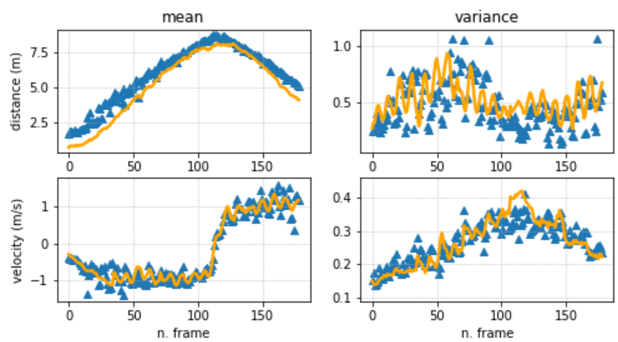

Fig. 8: Plot of the encoded representation of means and variances using the CNN encoder. Performances appear to be quite close to the FFNN architecture.

noise removal capability and at the same time, considering a compression of the whole generated signal.

Fitting the smoothed spline, the self organizing map, the FFNN and the CNN has revealed quite different results for the mean regularization. The same procedure is also applied for the shape regularization development, in Fig. 9, an example of reshaped ellipses thanks to the regularization is shown.

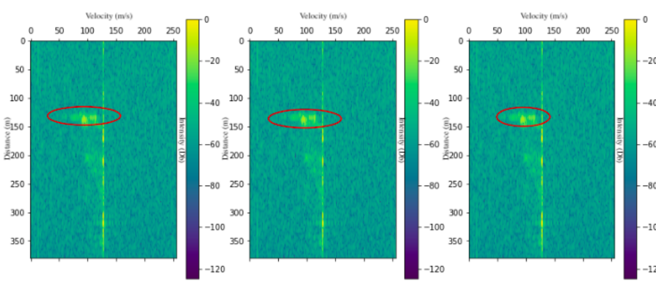

Fig. 9: From the left to right: the initially estimated ellipse, the regularized ellipse via smoothed spline, the regularized ellipse via the CNN.

In this case the presence of an outlier in the variances signal is correctly identified using the CNN but not by the smoothed spline, the output is proposed and compared. It is easy to see how the CNN behaves better than the spline in this case.

The trajectory regularization is an objective task that means and ellipses have in common, however, we noticed that the means suffer also the presence of critical frames where distinguishing the subject bump from the noise shape is challenging also for a human being, as shown in Fig. 10. Both the approaches based on neural networks overcome the simpler regularizing techniques when dealing with this problems.

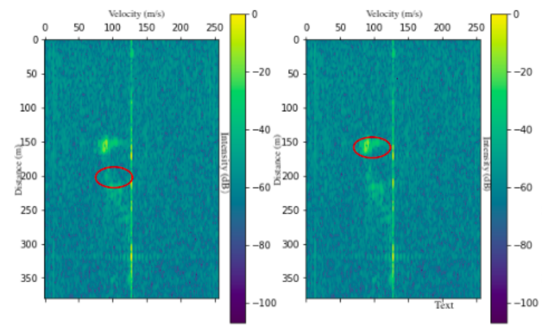

Fig. 10: In the left frame the mean initially estimated, in the right the same observation corrected via the FFNN.

In general, an extensive robustness study of the four proposed technics has been carried out. We point out that several aspects have been considered and that in this sense, when dealing with different applications, one method could appear more appealing than others. For example, the smoothed spline suffers a lot the presence of noisy points, while whether the SOM has a good compression factor reducing input size by ten, it performs a too strong regularization loosing important details.

We highlight how the autoencoders achieve two important feauteres, ie. data compression and walking gait tracking. Similar performances between the two architecture schemes (CNN and FFNN) are obtained.

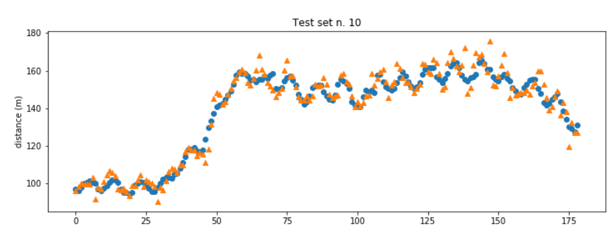

Fig. 11: CNN estimate on the validation set, CNN regularized points are the blue ones, while the originals are the orange.

In Fig. 11 it is easily noticeable a sort of trend which remarks the gait movement of the subject target. This is particularly interesting especially when considering that using the autoencoder it is possible to perfectly reproduce this signal from a dimensional reduced one. This interesting feature could be further investigated for a classification task, in order to identify the identity of the subject walking in the room, in the same manner of what authors do in [1]. Finally, as already said in section IV, live tracking requirements of the person movements can be achieved by using the GMM model without these four machine learning metrics, but discarding the regularization of the values, yielding to a high noise signal.

VII. Concluding Remarks

In this paper we have proposed a gaussian mixture model combined with four machine learning models - smoothed splines, SOM, FFNN, CNN - in order to track a person movements in an empty room by using a FMCW radar signal. We consider our findings interesting because the applied techniques have revealed how it is possible to exploit our methods in order to rebuild the person position and the shape of the bump contours using an ellipse, to automatically live tracking the target person, to reduce the signal dimension through the use of an autoencoder without loss of significance and also to faithfully reproduce the gait motion of the person. Multiple are the applications of each result, for example we highlight the possibility of applying the GMM algorithm via economic simple sensors, indeed we use this algorithm on a low-lever machine, as mentioned in section IV. An important application could be adding a fully connected layer at the end of the decoder using a softmax classification model in order to perform and compare the accuracy results obtained by [1]. While an extension of our work could be considering the same problem setup, but with multiple persons allowed to move into the room.

References

[1] B. Vandersmissen et al., *Indoor Person Identification Using a Low-Power FMCW Radar*, in IEEE Transactions on Geoscience and Remote Sensing, July 2018.

[2] Bijan G. Mobasseri and Moeness G. Amin *A time-frequency classifier for human gait recognition*, in Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI, Orlando, May 2009.

[3] V. C. Chen, *Spatial and temporal independent component analysis of micro-Doppler features*, in IEEE International Radar Conference, Arlington, VA, 2005.

[4] B. Jokanovic, M. Amin and F. Ahmad, *Radar fall motion detection using deep learning*, 2016 in IEEE Radar Conference (RadarConf), Philadelphia, PA, 2016.

[5] G. Garreau, C. M. Andreou, A. G. Andreou, J. Georgiou, S. Dura-Bernal, T. Wennekers, and S. Denham, *Gait-based person and gender recognition using micro-doppler signatures*, in IEEE Biomedical Circuits and Systems Conference (BioCAS), Nov 2011.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, USA, 2001.

[7] D.Tahmoush and J.Silvious,*Radarmicro-dopplerforlongrangefront- view gait recognition*, in IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Sept 2009.

[8] Zhang S., Xiao W. (2012) *Human Tracking for Daily Life Surveillance Based on a Wireless Sensor Network*. In: Wang X., Zheng R., Jing T., Xing K. (eds) Wireless Algorithms, Systems, and Applications. WASA 2012.

[9] V. C. Chen, F. Li, S. - Ho and H. Wechsler, *Micro-Doppler effect in radar: phenomenon, model, and simulation study*, in IEEE Transactions on Aerospace and Electronic Systems, Jan. 2006.

[10] Akram Bayat, Marc Pomplun, Duc A.Tran, *A Study on Human Activity Recognition Using Accelerometer Data from Smartphones*. Procedia Computer Science, 2014.

[11] Huiyu Zhoua, Huosheng Hu, *Human motion tracking for rehabilitation, a survey*, in Biomedical Signal Processing and Control, January 2008.

[12] H. Tsutsui, J. Miura, Y. Shirai, *Optical flow-based person tracking by multiple cameras*, Conference Documentation International Conference on Multisensor Fusion and Integration for Intelligent Systems, Aug. 2001.