Generating Micro Doppler Signatures with Conditional Generative Models: Challenges and Future Prospects

Noemi Manara†, Pietro Miglioranza†, Gaetano Ricucci†

***Abstract*—This study examines the potential of deep learning-based models for generating Micro-Doppler signatures that can be used to classify human activities. Micro-Doppler signatures refer to the unique spectral and temporal characteristics of a target's motion that are used to differentiate different human activities. The study focuses on four deep learning models, including three variations of Conditional Generative Adversarial Networks (cGANs) and one Conditional Variational Autoencoder (cVAE), to generate synthetic micro-Doppler signatures for human activity classification. The results indicate that while the generated micro-Doppler signatures bear a visual resemblance to real ones, they do not perform as well in classification tasks as real data does. It is concluded that further research is needed to enhance the suitability of generated micro-Doppler signatures for human activity classification. The limitations of the current models are discussed, and suggestions for future improvements are provided.**

***Index Terms*—Conditional Variational Autoencoders, Conditional Generative Adversarial Networks, micro-Doppler signatures, human activity recognition**

I. Introduction

Healthcare is facing a major challenge in monitoring and surveillance of patients, as a result of a shortage of healthcare personnel. One solution to this problem is the use of video systems for surveillance, but this approach is not without its limitations. Video cameras can struggle to capture subjects' movements in low light conditions or if they are obscured by objects, and video surveillance can be invasive in situations where privacy is of high importance.

Radio frequency (RF) sensors are devices that detect and respond to radio frequency signals. In recent years, they have gained popularity in various fields, including smart homes and human-computer interaction. RF sensors can be an effective alternative to video cameras, as they are more cost-efficient, less energy-intensive, and, most importantly, non-invasive.

An RF device operates by transmitting an electromagnetic radio signal in its line of sight. This signal is reflected by targets and objects, and the reflected signal is then captured by the receiver after a set time delay. The received signal is used to determine the range and angle of the target, as well as its speed if it is moving. In situations with multiple moving objects, the superimposed reflected signals can be represented as a Micro-Doppler (MD) signature [1] [2]. In the case of human data collection, the MD signature con be associated to a specific motion or activity.

In recent years, there have been numerous advancements in Machine Learning (ML), which have enabled the use of algorithms to recognize human movements based on both RF and video signals. DNNs have been found to be particularly effective in these types of tasks, as they can achieve good accuracy in motion recognition and classification. [3] [4]

The challenge with Deep Neural Networks (DNNs) in the healthcare industry is obtaining sufficient training data due to privacy concerns. To overcome this, data augmentation is often used to increase data, however, this method can negatively impact RF signals. RF signals contain unique patterns that are linked to the electromagnetic scattering and motion of the target being monitored. Traditional data augmentation techniques, such as scaling and rotation, can disrupt these patterns and decrease the performance of the DNN.

In this work, we present machine learning approaches for generating synthetic micro-doppler (MD) signatures using data obtained from a Frequency Modulation Continuous Wave (FMCW) radar at 77 GHz, as well as a video camera, which is utilized solely for validation and annotation purposes. The proposed methods include the use of Conditional Generative Adversarial Networks (cGANs) and Conditional Variational Autoencoders (cVAEs) to generate multiple MD signature images.

The main contributions of our study are as follows:

1) **Dataset Pre-processing:** The PARrad dataset, which consists of radar data collected from 24 subjects in various conditions, is a huge dataset that nearly 1 TB in size. We pre-process the dataset by extracting only the relevant MD signature images and class information for human motion, resulting in a final dataset size of approximately 260 Mb. The full process is described in Chapter IV.
2) **Conditional Learning Architectures:** We propose and compare two different architectures, a cGAN and a

†Department of Physics and Astronomy, University of Padova,
email: {noemi.manara}@studenti.unipd.it,
email: {pietro.miglioranza}@studenti.unipd.it
email: {gaetano.ricucci}@studenti.unipd.it

cVAE, which are described in detail in Chapter O. Both of them show some limitations for the problem at stake. Our results demonstrate that GANs can be challenging to train, even with a complex architecture.

This report is structured as follows. In Section II we describe the state of the art, while the processing pipeline and dataset characteristics are presented respectively in Section III and Section IV. The proposed modelling techniques are detailed in Section V and their performance evaluation is carried out in Section VI. Concluding remarks are provided in Section VII.

II. Related Work

Two main approaches are present in literature for synthetic data generation: Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs).

Variational Autoencoders are a type of neural network that build upon traditional autoencoders by learning a distribution in the latent (feature) space for each training example. [5] The VAE creates new data by reconstructing the sampled latent space with the decoder network. To ensure that the latent space is continuous and maps similar data close together, a regularization term is added to the loss function. This term uses the Kullback-Leibler divergence to compare the latent distribution to a standard Gaussian. [5] As a result of this regularization, the latent space can be easily manipulated to create continual variations on the data.

A Conditional Variational Autoencoder is a type of generative model that extends the traditional Variational Autoencoder (VAE) by conditioning the generative process on additional information. In other words, the model generates new data samples given specific conditions or inputs, such as class labels or attributes.

CVAEs are useful for many tasks in computer vision, natural language processing, and other domains where it is desirable to generate new data samples with specific characteristics. For example, a cVAE can be used to generate images of a specific object category (e.g. "generate an image of a cat") or to generate speech samples with a specific accent or emotion. [6]

However, due to their characteristic smoothing of the output, other approaches are favoured for detailed reconstructions and find broader treatment in literature.

Generative Adversarial Networks instead have been extensively researched for synthesizing realistic images in a wide range of applications, including Synthetic Aperture Radar (SAR) imaging [7]. In 2018, a study was conducted to explore the use of GANs for generating synthetic micro-Doppler data [8]. In this study, a Deep Convolutional GAN (DCGAN) was used to generate synthetic data that emulated the Boulic walking model, which consists of mathematically defined trajectories. However, this study only considered well-defined, systematic data and did not reflect the complexity of actual measured micro-Doppler signatures.

In another study [9], GAN-generated spectrograms were used to classify a test set of 50 simulated signatures comprising of three activity classes (running, walking, and jumping). This study only considered a small number of samples and easily identifiable classes, making it inadequate for evaluating the validity of using GANs for simulating micro-Doppler signatures.

The first study exploring the use of GANs for classifying real micro-Doppler data was published in [10], where the authors evaluated the effectiveness of adversarial learning in addressing the open-set problem (where the training dataset does not include all classes in the test dataset). Further studies in [11] and [12] utilized GANs to mitigate the problem of low sample support and reported the classification accuracy of Deep Neural Network models trained with GAN-generated synthetic data for human activity recognition.

However, despite their ability to synthesize visually similar radar micro-Doppler signatures, the use of GANs for this purpose is challenged by differences between radar phenomenology and optics. The pixel values in micro-Doppler signatures are not representative of physical shapes, but instead, they relate to human kinematics. This leads to the possibility of GANs generating synthetic samples that are visually similar but incompatible with the kinematics of human motion.

In this study, we compare and contrast the results of utilizing a cVAE and a cGAN to generate synthetic Micro-Doppler signature images for the purpose of human motion classification. Our use of a cVAE represents a novel approach to the field and serves as an important contribution to the current state of the art. Through this comparison, we aim to demonstrate the potential and limitations of each method in synthesizing these images and further advancing the classification of human motion.

III. Processing Pipeline

In this paper, we present two deep learning-based approaches, Conditional Variational Autoencoders (cVAEs) and Conditional Generative Adversarial Networks (cGANs), for the task of synthetic micro-Doppler signatures generation. Both approaches have been implemented and evaluated on a large dataset of real micro-Doppler signatures.

The processing pipeline for both approaches consists of several stages, which are summarized in fig. 1 and outlined below:

- **Data preparation**: The real micro-Doppler signatures were pre-processed to remove some noise and prepare them for input into the deep learning models. The data was also divided into training, validation, and test sets to facilitate the evaluation of the models.
- **Model training**: The cVAE and cGAN models were trained on the training set using the Adam optimization algorithm. The models were trained for more than 1000 epochs, and the performance was monitored using the validation set in the case of the cVAE.
- **Model evaluation**: The evaluation phase included several steps: the inspection of the loss functions to see if

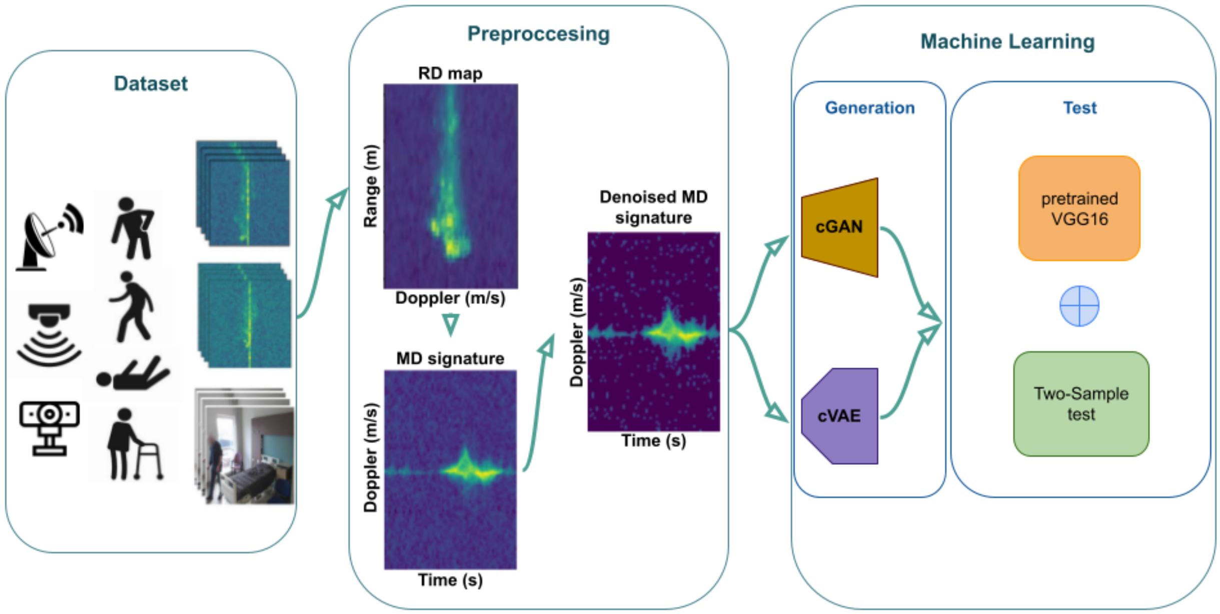

Fig. 1: Processing Pipeline

convergence was reached, visual quality of synthetic data (and also reconstructed data from the test set in the case of the cVAE) and their diversity.

- **Generation of synthetic signatures**: a large dataset of cVAE synthetic signatures was compared with real-world signatures to evaluate its representativeness and quality through the accuracy metrics of Classifier Two-Sample Test (CTST) and Train on Generated/Test on Real.

The processing pipeline was designed to ensure that both approaches were implemented and evaluated in a consistent and fair manner, and to facilitate the comparison of their performance. The results of the evaluations were used to determine the best approach for the problem at stake.

IV. Dataset

The Patient Activity Recognition with Radar sensors (PARrad) dataset [13] used in this paper is a subset of a larger dataset that was collected using two Frequency Modulated Continuous Wave (FMCW) radars with center frequencies of 77GHz and 60GHz. The original data was collected in two different environments: a synthetic hospital room (Homelab) and a real-life hospital room (Hospital). The PARrad data set focuses specifically on the real-life hospital room data, which was recorded in a local hospital.

The data set contains 21569 activities performed by elderly and adult individuals with ages ranging from 24 to 90 years, and heights and weights ranging from 155 to 185 cm and 47 to 95 kg, respectively. The Texas Instruments Millimeter Wave FMCW radars were used in Multiple Input Multiple Output mode to collect the data. These radars have the advantage of being both high power efficient and low-cost, but their high power efficiency comes at the cost of lower Signal to Noise Ratios that can pose challenges in data analysis.

The FMCW radar works by continuously emitting electromagnetic signals from a transmitting antenna and capturing the reflections from the target using a receiving antenna array. Essential information about the target, such as range, angle, and speed, is extracted from the reflections based on the time delay or phase shift (the Doppler effect). The data collected by the radar is processed using 2D Fourier transforms to generate Range-Doppler (RD) maps, which summarize the range and velocity information, and generate a Motion Detector (MD) signature by concatenating over the time dimension.

In addition to the radar sensors, video sensors were used for validation and annotation purposes. The data set consists of 22 hours, sampled with 11.1 fps, of annotated activity data distributed over 14 original classes.

A. Pre-processing

The PARrad dataset is a collection of data that includes 23 subjects performing 14 different types of human activities under various conditions. The data consists of two components: the MD signature maps and the video registration of the activities. The MD signature map is obtained by applying a 2D Fourier transform to the raw radar data and then converting the absolute values of the RD signal to decibels (dB). The resulting RD map is three-dimensional, containing range, Doppler, and time units with 93 range bins and 128 Doppler bins.

For each subject, the MD map is a rectangular 2D map with velocity bins on the x-axis and time bins on the y-axis. The data was collected using two different radars, but for this study, only the data collected by the 77Ghz radar was considered, corresponding to the radar placed near the ceiling. The three middle Doppler bins, which represent objects with zero velocity, were removed without losing any information. [14]

To extract the MD signature of a single activity, the MD map was sliced along the time axis according to the start and finish times of the activity, which were declared in another file within the dataset. Only 16 subjects were used for the study because the duration of some activities for the remaining subjects was systematically over double the average, or the radar signal was not synchronized with the corresponding activity information. In particular, subjects 0, 9, 10, 11, 16, 17, 18 and 23 were excluded.

The MD signature maps are adjusted to a size of 64x96. The 96 velocity bins are determined empirically by dropping the first 14 bins on one side and the last 15 bins on the other. The 64 time bins are selected as the average of the duration of all selected activities, which is of 5.8s. The MD map is resized by repeating the data using the wrap method on both sides if it is smaller than the target size. If the map is larger, it is centered trimmed to fit the desired size.

B. Denoising

The obtained micro-Doppler (MD) maps are quite noisy, so an algorithm is needed to denoise them and obtain better images for the generator algorithms. The following simple approach was chosen.

First, the minimum value of all the MD maps is found. Then, by scanning the first row of all the MD matrices of all subjects, the maximum value of the first row of all the matrices is found. An empirical threshold is then set to this value minus 400, as the signals should be around the center of the image, and if there is a high value at the top of the image, it should be considered noise.

Next, all the entries in the matrices that have a value lower than this threshold are set to the minimum value of all the matrices. This is preferred over setting them to 0, as it maintains a reasonable gradient between the signal and the ground.

An example of noised and denoised images are shown in fig. 1.

A simple pseudo-code of the process is as follow:

Algorithm 1 Denoising algorithm for MD maps

Require: $mDoppler \in \mathbb{R}^{X \times Y \times Z}$
Ensure: $mDoppler_{denoised}$

$$minimum \leftarrow \min_{i=1}^{X} \min_{j=1}^{Y} \min_{k=1}^{Z} mDoppler_{i,j,k}$$

$$maxFirstRow \leftarrow \max_{i=1}^{Z} \max_{j=1}^{Y} mDoppler_{1,j,i}$$

$$threshold \leftarrow \max_{i=1}^{Z} maxFirstRow - 400$$

for $i = 1$ to X **do**
  for $j = 1$ to Y **do**
    for $k = 1$ to Z **do**
      if $mDoppler_{i,j,k} < threshold$ **then**
        $mDoppler_{i,j,k} \leftarrow minimum$
      end if
    end for
  end for
end for

$$mDoppler_{denoised} \leftarrow mDoppler$$

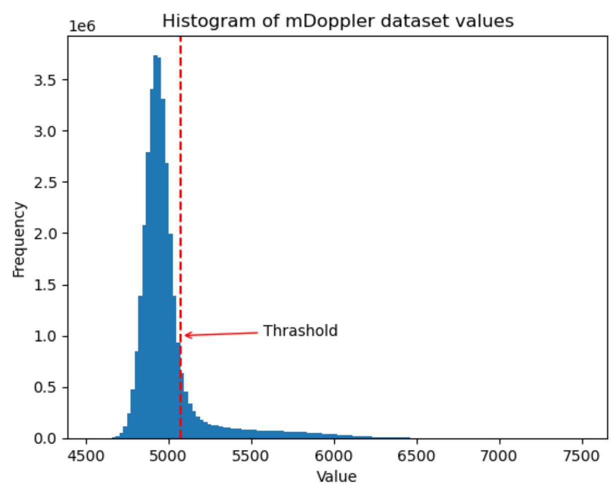

Fig. 2: Histogram of the values of all the micro-Doppler matrices showing that most values (noise) are below the selected threshold.

V. Learning Framework

First and foremost, two different approaches are evaluated in parallel for the generation of synthetic micro Doppler signatures. As they present different characteristics and training procedures, they are illustrated in two distinct sections.

Since both of them are conditioned to the class labels, they include one or more embedding blocks in order for the network to learn an informative representation of the one hot encoded labels, rather than just memorizing the training data. This can also help prevent the network from collapsing to a single mode, where it only generates samples from a single class, and instead encourages it to generate diverse and representative samples for each class.

A. Conditioned Variational Autoencoder

Conditional Variational Autoencoders are powerful tools for generating synthetic data in various domains. They are appealing because, after training on a small dataset, the model can learn the underlying distributions by building a latent space representation of the data and generate new diverse synthetic samples that are not only similar to the real-world ones, but show the continuity property. The variational inference framework allows for a more flexible and expressive model compared to traditional autoencoders, as well as acting as an intrinsic regularization factor for the model. The use of cVAEs for our task allows for control over the class of the target motion, by conditioning the generation process on these attributes. In this way, cVAEs offer a flexible and efficient solution for generating micro-Doppler signatures for use in various research and development applications.

The structure of the **cVAE** implemented is represented in fig. 3 . The reparametrization trick (sampling from the latent space) is tacit for space reasons, but is carried out in the standard way: a random noise variable ϵ is sampled from a standard normal distribution and it is transformed into a sample z from the approximate posterior as follows:

$$z = \mu + \sigma \odot \epsilon$$

The Convolutional Encoder acts as a feature extractor exploiting spatial information and short range correlations, while the Decoder is primarily a Fully Connected architecture to allow for fine-grained reconstruction, smoothed by the last Convolutional layer. Many attempts with Decoder architectures symmetric to the Encoder have been carried out, but each of them resulted in excessively smoothed reconstructions even with long training.

The networks' weights are initialized through the Xavier Uniform algorithm, apart for Batch Normalization, for which the weights are initialized through a normal distribution with mean 1.0 and standard deviation 0.2 while the biases are initialized to 0.

In order to perform a thorough training, the whole dataset is randomly split in: 80% training set, 10% validation set for monitoring that the network does not overfit and 10% test set, used to check the ability of the network to reconstruct novel inputs.

The training is performed for 1000 epochs, using the Adam optimizer with learning rate 0.0001 and batch size of 32.

B. Conditioned Generative Adversarial Networks

Another widely used generative approach are Generative Adversarial Networks. Their architecture consists of two primary components: the generator and the discriminator. The generator produces fake data samples, while the discriminator attempts to determine if each sample is real or fake. The two components engage in a minimax game, where the generator tries to generate samples that deceive the discriminator, and the discriminator tries to correctly distinguish between real and fake samples. This process ultimately results in a generator that can produce high-quality synthetic data.

One of the main benefits of GANs is their ability to create diverse and realistic data samples in all fields, especially images. Additionally, GANs are capable of learning complex relationships within a dataset, making them useful for a wide range of tasks beyond data generation.

Conditional Generative Adversarial Networks are an extension of GANs that involve conditioning both the generator and discriminator on extra information, such as labels. By providing additional information to the model, cGANs improve the quality of the generated images and allow for greater control over the attributes of the generated data.

In this work, we present and elaborate on three architectures, all similar to each other, to show how little or how much simple modifications impact on the performances of the network.

The first attempt, **cGAN-simple** is analogous to what shown in fig. 5, with the exception that no MiniBatchDiscrimination module is present nor the second embedding addition in the Generator.
The second attempt, **cGAN-refresh** is represented in fig. 5, here a second embedding block is added, which has the output of the same size as the output, and is element-wise added to the output of the layer before the last. In this way the information on the class to be generated is propagated through the network and refreshed at the end, in a skip-connection fashion.
The architecture of the final cGAN attempt, **cGAN-minibatch**, is displayed in fig. 4. To solve the problem of mode collapse we attempted to implement a final MiniBatchDiscrimination layer at the end and feeding the label information just once, near the end of the Generator, but two blocks before the output.

In this case, due to the more difficult he training is performed for at least 2000 epochs, using the Adam optimizer with learning rate 0.0001 and batch size of 32.

C. Metrics

In order to evaluate the performances of the two models, we implement two different tests used for synthetic data quality determination, found in literature. [15]

The first, Classifier Two-Sample Test (CTST), consists in the discrimination between real and fake data with no class information by a simple neural network. The second, Pre-Trained VGG16 Test, on training a classifier on generated samples and testing it on real ones.

The fake dataset to test is created by the cVAE through independent normal sampling of 10000 points of the latent space along with randomly generated labels.

D. Classifier Two-Sample Test (CTST)

The Classifier Two-Sample Test is a statistical test that measures whether two sets of data come from the same underlying distribution. It determine whether a classifier's predictions are significantly different from random chance,

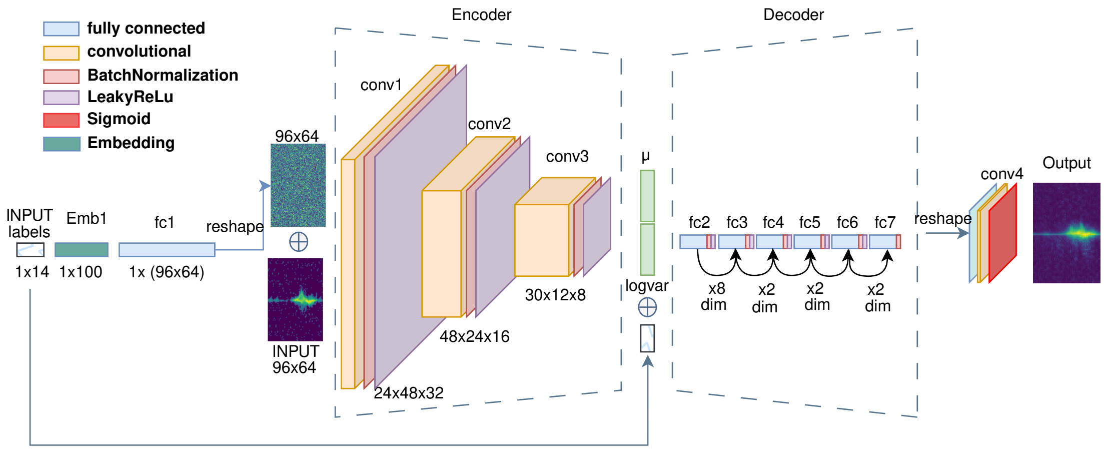

Fig. 3: The architecture of the cVAE model. The class label is one-hot encoded, then undergoes dimensionality upscaling through an Embedding layer and then a Eully connected layer, with LeakyReLU activation. Then it is reshaped and appended to the original image. Each Convolutional block of the encoder doubles the feature maps while halving the dimensions, through the combination of the parameters $\{$kernel = 4, stride = 2, padding = 1$\}$, and has a dropout of 0.1. The output of the Encoder is then flattened and fed to the Linear layers of mean and logvar. The latent space has 32 dimensions. For the decoding, the one-hot encoded label is appended to the end of the extracted latent variable and then they undergo upscaling through 6 Fully connected blocks, with dropout of 0.1 each. The output is then reshaped to the original dimension and fed to a final Convolutional layer that keeps the same dimensions $\{$kernel = 3, stride = 1, padding = 1$\}$. At the end, a Sigmoid activation function is applied.

or if they are indicative of a real relationship in the data. The test works by training a classifier on one set of data and then evaluating its performance on a separate set of data. The accuracy of the classifier is compared to a baseline accuracy that would be expected by random chance. If the classifier's accuracy is significantly higher than random chance, it is concluded that there is a real relationship between the features and the class labels in the data. For this step a simple neural network for binary classification is exploited. It consists of one hidden layer with 20 units and a ReLU activation function followed by a sigmoid activation [16]. The procedure is as follows: balanced training, validation and test sets are built by mixing true and fake data, dropping the information about the class, and finally labeling according to whether the sample is real (1) or fake (0). The sets consists of: 10707 samples in training, 1338 samples in validation and test.

After 100 epochs and Adam optimizer with learning rate of 0.001 the performances were evaluated on the test set in terms of accuracy.

E. *Pre-Trained VGG16 Test*

To further assess the quality of the data generated by the cVAE, a comparison of the performance of a fine-tuned VGG16 is made. The model has all its layers blocked except for the last two linear layers, the latter of which is adjusted to have 14 output features. The model is first trained on a real data training set, and then on a cVAE-generated data training set. The performance of the trained models are then compared on a real test data set, which was not used for cVAE training. If the two model's performances on the test data are comparable, we can conclude that the conditional Variational Autoencoder generates good quality data, useful to perform labeled data augmentation.

VI. Results

In this section, we present the results of our study on the generation of synthetic micro-Doppler signatures. Our goal was to develop a deep generative model that could effectively capture the complex patterns and structures present in micro-Doppler signatures and use it to generate synthetic data that are representative of real-world scenarios. A summary comparison of generated data samples with real data samples is shown in fig. 8.

A. *Preliminary cVAE results*

Experimental results show that the cVAE model was able to learn the underlying patterns and structures in the micro-Doppler signatures and generate high-quality synthetic samples that are similar to real signatures. This is deducible from the visual quality of the generated samples, that is satisfactory both in the case of training samples, fig. 6, and test samples, fig. 7. Our results showed that the cVAE was able to generate visually plausible synthetic signatures, and produce a diverse range of signatures for each class.

B. *Preliminary cGAN results*

In the case of the cGAN the task was proven to be extremely hard. In none of our trials we managed to find a balance between Generator and Discriminator so that the losses would converge.

Among all the implementations tested, one had a significant impact on the quality of the generated signatures: as the generator each time collapsed in the generation of a single sample,

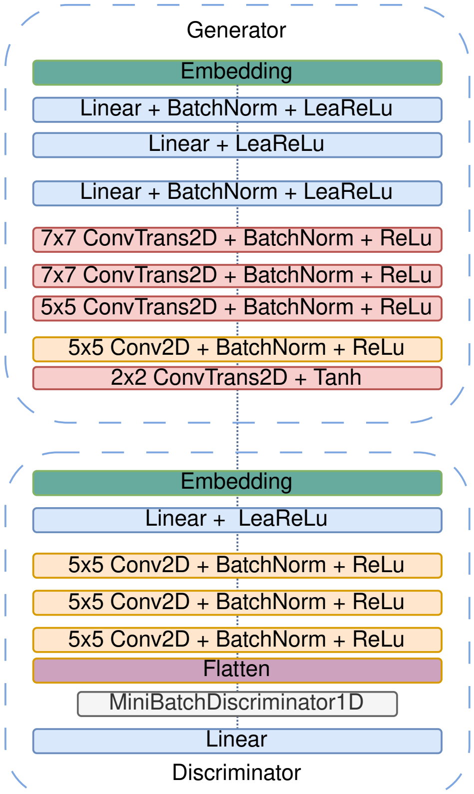

Fig. 4: The architecture of the **cGAN-minibatch** model. **Generator**: the input noise is fed to a Linear layer with 12288 neurons, then reshaped in a 12x8 image with 128 channels and passed to a sequence of three Convolutional Transpose (ConvT) layers, with dropout 0.15 each. Then the class label is one-hot encoded and upscaling through an Embedding layer of dimension 14 and a couple of Linear layers with respectively 1536 and 6144 neurons. It is then reshaped and element-wise added to the output of the ConvT block. Finally, a couple of ConvT and Conv layers are applied, which do not change the dimension but simply bring the channels back to 1. **Discriminator**: the class label is directly upscaled to the sample dimension through an Embedding and a Linear layer and appended as a new feature channel, all the Convolutional layers have dropout of 0.25 and the last layer has no activation function.

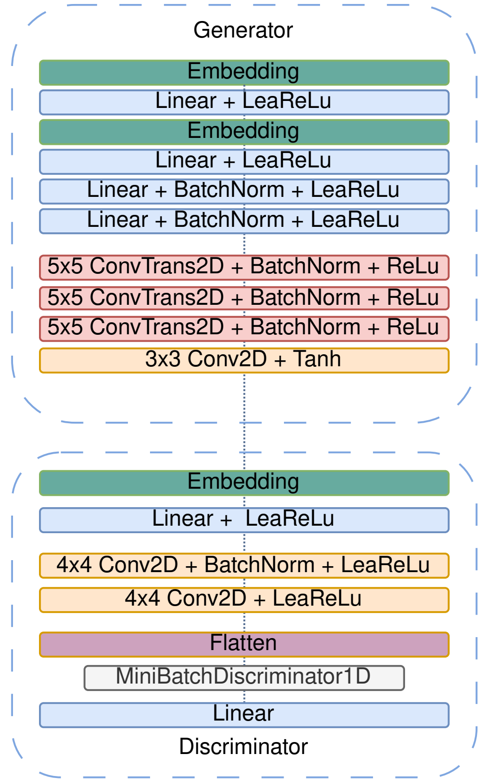

Fig. 5: The architecture of the **cGAN-refresh** model. **Generator**: the class label is one-hot encoded, then undergoes dimensionality upscaling through an Embedding layer of dimension 14 and a Linear layer with 384 neurons. The result is reshaped in a 24x16 matrix and appended as an ulterior feature map channel to the reshaped output of the two-layers Linear block (light blue in the figure) with hidden dimensions of 7168 and 114688. Then two ConvT 2D layers are applied, which bring the number of channels to 1 and double the image dimensions. At this point is added element-wise the second Embedding (analogous to the first, but with an output size of 96x64). Linear layers of 14, 1536 and 6144 neurons respectively. Finally, after the last two blocks the tanh activation function is applied. Dropout of 0.15 is added after each ConvT block. **Discriminator**: here the class label is directly upscaled to the sample dimension and appended as a new feature channel, the first Convolutional layer has dropout of 0.2 and the last layer has no activation function.

we tried implementing a *label refresher* skip connection-inspired layer to propagate the label information across the architecture. The *label refresher* works as follows: along the initial embedding of the one hot encoded label, a second higher dimensional embedding is learnt by the Generator, which is added to the output of one of the last layers of the Generator.
Even though the quality of the so generated data is satisfactory, another open issue arises: this information propagation is too preponderant and draws the generator to mode collapse on the 14 classes as shown in the third row of fig. 8. To tackle this issue the third architecture with only one label encoding was developed, but mode collapse persisted.
We made the decision to not present all the pictures generated by the third architecture, since it would not add any value to showcase multiple repetitions of the same micro-Doppler map. Since conducting the aforementioned tests over a generated dataset composed of 14 micro-Doppler maps repeated many times is meaningless we decided to only test the quality of the images produced by the cVAE.

C. *Quality of generated images*

Even though visually the fake signatures are similar to the real ones, in terms of classification accuracy they do not seem to display the same similarity when inspected by a neural network. As reported in tab. 1, the accuracy of the classification is far from random as the classifier is perfectly

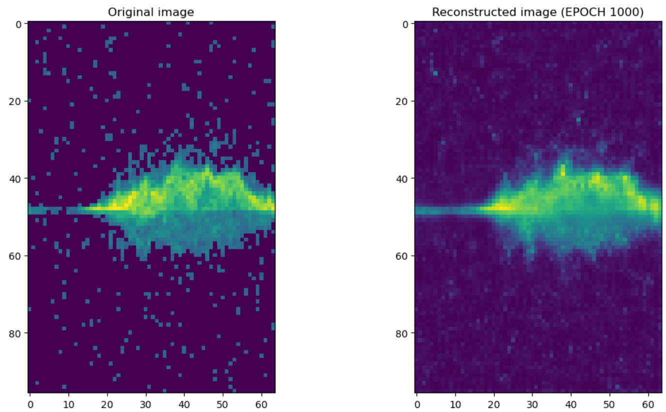

Fig. 6: The **cVAE** reconstruction of a sample at epoch 1000.

able to distinguish between real and fake samples. A possible reason for this behaviour is that the reconstructed images present a different background noise with respect to the pre-processed data, which is easy to spot by a neural network.

	Binary Classifier	Random
cVAE	0.99	0.50

TABLE 1: Results of the **CTST** test

The test conducted through **VGG16** also yielded negative results: as indicated by tab. 2, the accuracy of the classifier trained on the generated data was comparable to random chance, while it performed significantly better when trained on real data, showing that generated data cannot be used to perform data augmentation, as it seems that they do not display the same class-specific features as the real dataset. Probably this is due to a ineffective representation in the latent space, that draws the cVAE to the generation of samples irrespective to the label information.

	real data	generated data
accuracy	0.49	0.09
precision	0.5	0.05
recall	0.5	0.1

TABLE 2: **VGG16** test results

VII. Concluding Remarks

Overall, our results show that cVAEs are a promising technique for generating synthetic micro-Doppler signatures. Future work to perfect this method include some more exploration on the noise reconstruction and the implementation of the *label-refresher* module or other ways to increase the importance of the labels in the latent space mapping.
We also raised the bar for cGAN synthetic micro-Doppler signature generation with some visually appealing results. In this regard, mode collapse remains an open problem that shall be tackled by future research. Our work lays the foundation for future developments in the field of micro-Doppler signature generation.

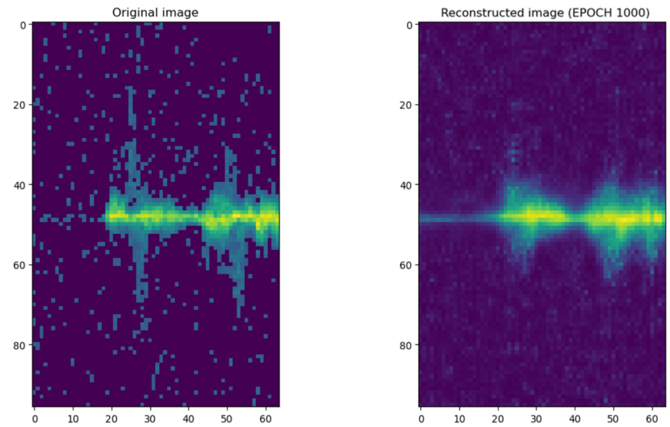

Fig. 7: The **cVAE** reconstruction of a sample not included in the training.

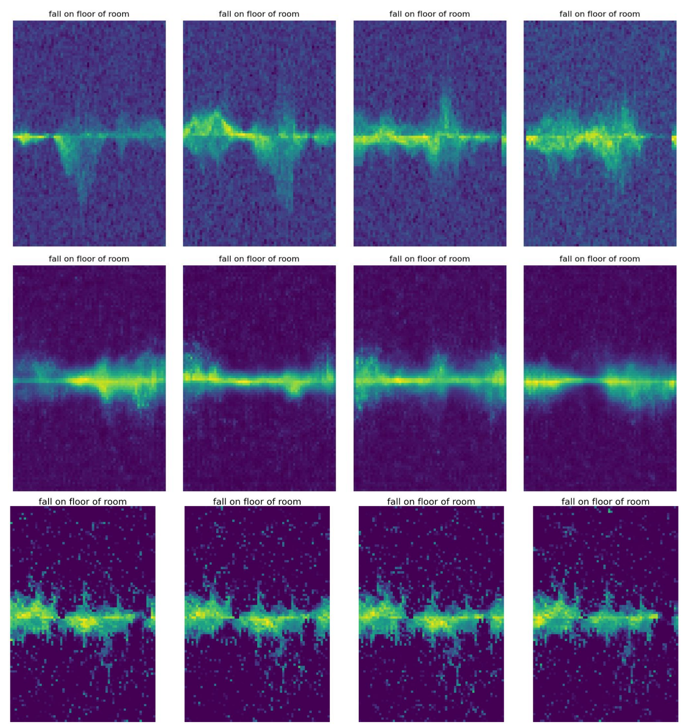

Fig. 8: Comparison of real MD Signature maps describing the 'Fall on Floor' motion with generated maps via **cVAE** (second row) and **cGAN** (third row).

With this project we had the canche to implement from scratch for the first time cutting-the-edge conditional generative architectures and face all the difficulties that come with that: we learnt that finding a combination of Generator-Discriminator that can play the minimax game is a very hard task, as well as that VAEs' latent spaces are not as trivial as one would expect (no clusters of labels were found).

Regarding the work subdivision: Noemi Manara mainly took care of the cVAE, Pietro Miglioranza of the cGAN and Gaetano Ricucci of the preprocessing and final tests, but as each of these were challenging tasks we worked together on most of them: each of us had good ideas or suggestions also for the work of the others.

References

[1] G. M. Brooker *et al.*, "Understanding millimetre wave fmcw radars," in *1st international Conference on Sensing Technology*, vol. 1, 2005.

[2] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and electronic systems*, vol. 42, no. 1, pp. 2–21, 2006.

[3] H. Arab, I. Ghaffari, L. Chioukh, S. O. Tatu, and S. Dufour, "A convolutional neural network for human motion recognition and classification using a millimeter-wave doppler radar," *IEEE Sensors Journal*, vol. 22, no. 5, pp. 4494–4502, 2022.

[4] F. J. Abdu, Y. Zhang, M. Fu, Y. Li, and Z. Deng, "Application of deep learning on millimeter-wave radar signals: A review," *Sensors*, vol. 21, no. 6, 2021.

[5] J. Cha, K. S. Kim, and S. Lee, "On the transformation of latent space in autoencoders," 2019.

[6] Y. ANTOLÍ N, "Counting subgraphs in fftp graphs with symmetry," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 170, pp. 327–353, nov 2019.

[7] B. P. Lewis, J. Y. Liu, and A. Wong, "Generative adversarial networks for sar image realism," in *Defense + Security*, 2018.

[8] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, 2022.

[9] Y. Mi, X. Jing, J. Mu, X. Li, and Y. He, "Dcgan-based scheme for radar spectrogram augmentation in human activity classification," *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, pp. 1973–1974, 2018.

[10] Y. Yang, C. Hou, Y. Lang, D. Guan, D. Huang, and J. Xu, "Open-set human activity recognition based on micro-doppler signatures," *Pattern Recognition*, vol. 85, pp. 60–69, 2019.

[11] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Gan-based synthetic radar micro-doppler augmentations for improved human activity recognition," in *2019 IEEE Radar Conference (RadarConf)*, pp. 1–5, 2019.

[12] I. Alnujaim, D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-doppler signatures of human activity," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 396–400, 2020.

[13] G. Bhavanasi, L. Werthen-Brabants, T. Dhaene, and I. Couckuyt, "Patient activity recognition using radar sensors and machine learning," *Neural Computing and Applications*, pp. 1–16, 2022.

[14] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[15] M. Ehrhart, B. Resch, C. Havas, and D. Niederseer, "A conditional gan for generating time series data for stress detection in wearable physiological sensor data," *Sensors*, vol. 22, no. 16, 2022.

[16] D. Lopez-Paz and M. Oquab, "Revisiting classifier two-sample tests," 2016.