Self-Supervised Contrastive Learning for Radar-Based Human Activity Recognition

Mohammad Mahbubur Rahman, Sevgi Zubeyde Gurbuz
Electrical and Computer Engineering
The University of Alabama, Tuscaloosa, AL-35487, USA.
mrahman17@crimson.ua.edu, szgurbuz@eng.ua.edu

***Abstract*—Short-range radars are widely used for micro-Doppler-based human activity recognition by using the supervised training paradigm. However, acquiring labeled trained RF data is not always possible. This paper presents a self-supervised contrastive learning (SSCL) framework that utilizes physics-aware augmented radar micro-Doppler signatures for human activity recognition. The SSCL requires two augmented views of the input data samples. For the first augmented view, the Short-time-Fourier-transform properties have been manipulated to generate Multi-Resolution micro-Doppler (MR-mD) signatures. The Second augmented view has been generated through synthesizing micro-Doppler signatures by a Physics-aware Generative adversarial Network (PhGAN). Experimental result shows that the proposed SSCL framework achieved 4% improvement over conventional unsupervised autoencoder pretraining while classifying 14 ambulatory human activities.**

***Index Terms*—Radar micro-Doppler, Activity recognition, Supervised training, Self-supervised learning, Data Augmentation.**

I. Introduction

Human activity recognition (HAR) is an important component of enabling technologies for aging-in-place and remote patient monitoring outside of a hospital setting. An effective HAR system could provide critical medical data that can serve an important role in telemedicine; namely, 1) monitoring of mental and physical health, as well as mobility issues for aging-in-place, 2) diagnoses of neuro-muscular disorders, such as Parkinson's or Cerebral Palsy, and 3)rehabilitation, the objective assessment of establishing a normal gait. Towards this aim, short-range radar could be a game-changer to improve the quality of care and facilitate aging-in-place by enabling continuous ambulatory recordings over an extended duration in real-life settings. On top of that, they are low-cost, non-intrusive, non-contact sensors that can work remotely in an indoor environment without acquiring personal imagery of the environment or user. Eying the promise of this sensor for telemedicine and telehealth applications, plenty of research has been conducted in recent years on radar-based human sensing for daily activity monitoring [1], Fall detection [2], Gait analysis [3], sensing for smart environments [4], and human-computer interfaces [5] via gesture [6] and sign language recognition [7].

Current literature on radar-based activity recognition is predominantly relying on fixed snapshots in time of RF data acquired for a single activity. Collecting activity-wise shorter duration of RF data is helpful in labeling the data to train a supervised deep neural network (DNN). This paradigm of supervised learning has a proven track record for training specialist models that perform extremely well on the task they were trained to do. However, the nature of human motion is continuous, and human motion has nearly unlimited diversity. Practically speaking, it's impossible to collect and label all the combinations of humanly possible motions and used them in training a supervised deep neural network. Moreover, a massive amount of data is also required to attain state-of-the-art recognition performance. Even though, someone takes the burden of collecting a massive amount of continuous RF data, labeling those collected data is a daunting task because such signals are not human interpretable. Therefore, Supervised learning is a bottleneck for building more intelligent generalist models that can do multiple tasks and acquire new skills without massive amounts of labeled data. If DNNs can garner a deeper, more nuanced understanding of reality beyond the specified training data set, they will be more useful and ultimately bring AI closer to human-level intelligence [8].

The above limitations of supervised learning motivate the need for learning from unlabelled RF data. Self-supervised learning (SSL) enables DNNs to learn good representations of high-dimensional observations from large amounts of unlabelled data [8]. In SSL, a common approach is to learn a joint embedding of similar observations or views such that their representation is close [9]–[12]. These methods usually rely on data augmentations techniques to generate additional views to preserve the semantic characteristics of observation, while changing other "nuisance" aspects [13]. As most unsupervised/ self-supervised representation learning methods are designed for computer vision, therefore, these methods rely on RGB-specific augmentation techniques such as random cropping & resizing, flipping, rotating, color distortion, Gaussian blur, etc. However, those RGB augmentation technique does not directly translate for Radar data as they might change the underlying physics of the radar perception for human motion sensing. Therefore, a lot of effort has been made in the literature to do physics-aware RF data augmentation which can be broadly categories into model-based synthesis and direct synthesis. Model-based synthesis generates the time-frequency transform of the expected radar return computed from a skeletal model comprised of point targets animated using motion capture (MOCAP) data [14]. Whereas, direct synthesis generates the micro-Doppler signature using generative adversarial networks (GANs) trained on a small number of measured signatures.

Specifically, in [15], authors proposed a physics-aware Multi-Branch GAN (PhGAN), to generate micro-Doppler signatures for ambulatory human activities with high kinematic fidelity.

State-of-the-art SSL method like contrastive learning [9]–[12], [16] requires two distinct augmented views for the same instance of the unlabeled data so that the augmentation of an instance can be accurately retrieved from a pool of instances through learning joint embedding. In RF domain, these two aforementioned physics-based micro-Doppler data augmentation techniques could play a pivotal role to apply SSL in radar data. However, the unavailability of MOCAP data could make the model-based synthesis infeasible. Therefore, we propose a new augmentation technique named "Multi-Resolution micro-Doppler (MR-mD)" view that can generate multiple micro-Doppler signatures for the same RF data instance at different spatial and temporal resolution levels. As a result, MR-mD and PhGAN-generated micro-Doppler are the two augmented views that can be leveraged in SSL training. So, the contribution of this work is proposing a self-supervised training framework that utilizes novel physics-aware data augmentation techniques for radar-based human activity recognition.

The paper is organized as follows: Section II describes state-of-the-art self-supervised methods and introduces SSL contrastive learning framework. In section III, the physics-aware data augmentation techniques and SSL for Radar data have been presented. Finally, section IV describes the experimental results for 14 ambulatory human motion recognition using SSL.

II. Self-Supervised Learning

In recent years, a lot of new approaches have been proposed for self-supervised learning, in particular for computer vision, that have resulted in great improvements over supervised models when few labels are available. Early work on unsupervised representation learning has focused on designing pretext tasks and training the network to predict their pseudo labels [17], [18]. Later on, unsupervised learning based on predictive models including auto-regressive (AR) and auto-encoding (AE) models [19] shows less dependence on the labled data. The family of autoencoders such as denoising auto-encoders [20], and variational auto-encoders [21] provide a popular framework for unsupervised representation learning using a predictive loss. Recently, contrastive learning has become widely popular for learning effective representations. The core idea of contrastive learning is to keep features from positive sample pairs close, and features from negative samples far from each other. DIfferent contrastive learning frameworks namely SimCLR [9], the momentum-contrastive approach (MoCo) [10], ContrastiveMultiview-Coding [22], and BYOL [16] are available in literature.

In the Radar literature, the concept of SSL has appeared in the context of radar image enhancement [23], super-resolved radar beamforming [24], unsupervised human sensing [25], and human activity recognition [26]. In [23], the SSL network is trained to learn the mapping function between the original radar image and the high-quality radar image. In [24], authors used SSL to super-resolve a radar array, by training an auto-encoder with a diluted radar array and used to reconstruct the amplitude and phase of missing receiving channels. In [26], authors consider semi-supervised training of an attention-augmented convolutional autoencoder (AA-CAE) for human activity recognition using radar micro-Doppler signatures. The AA-CAE learns global information along with spatially localized features to help the classifier overcome the limited receptive field of a CAE. Khatabi et.al. [25] proposed a trajectory-guided CAE-based unsupervised learning framework for human pose and human activity recognition. The authors used Range-Azimuth and Range-elevation heatmaps from customized FMCW hardware. As the heatmaps are sparse, therefore, authors applied an ROI selection in each frame and tried to reconstruct the ROI. Once the CAE learns to reconstruct the ROI, the decoder part is taken off and downstream task-specific modules are integrated.

This work presents a self-supervised contrastive learning framework [9] for Radar micro-Doppler based human activity recognition. Contrastive learning first uses unlabeled training data to generate two augmented views of the same data samples. These augmented views are then passed through an encoder network for latent feature representation learning. Afterward, the non-linear projections of the latent representation are computed using a fully-connected network (i.e., MLP). The projection head highlights the invariant features from the latent representation. The generic representations are learned by simultaneously maximizing agreement between differently augmented views of the input image and minimizing agreement between transformed views of different images, following a method called contrastive learning. This simple self-supervised contrastive learning framework is called SimCLR.

III. Self-Supervised Contrastive Learning in Radar-based activity recognition

The heart of the SimCLR is the strong data augmentation techniques (e.g., resize & random crop, random color distortion, and random Gaussian blur) that help the learned representation to be more robust under different transformations, which can help the model to learn transformation invariant representations. However, these RGB-specific augmentation techniques are not directly applicable to RF Data. RF data is not an image, but rather a time-frequency representation of the received I/Q data. These I/Q data are obtained by analyzing the back-scattered signal power reflected from different objects located in space. There is no color information in RF signals, and the signal is not invariant to rotation transformation, making most of the existing unsupervised learning methods not directly applicable to RF signals. Therefore, designing an effective data augmentation technique for RF data that does not change the underlying physics of radar perception of human motion, is a novel research question, which we tend to answer in this work.

A. RF Micro-Doppler Data Augmentation

The radar received signals are raw complex I/Q Data. These I/Q data go through a signal processing pipeline before

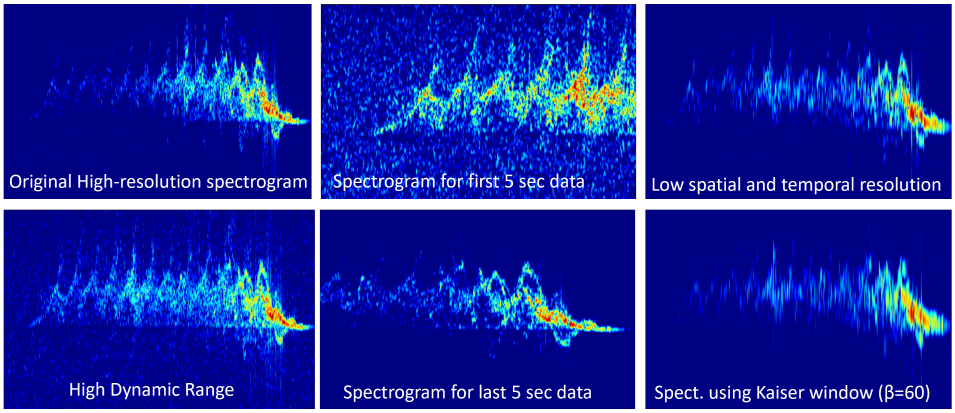

Fig. 1. Multi-Resolution micro-Doppler (MR-mD) Augmented View

generating time-frequency representation. The raw complex RF data are first reshaped in a 2D matrix based on the fast-time and slow-time samples. A Fast Fourier Transform (FFT) is applied in the fast-time dimension to generate the range profile. Afterward, a fourth-order butter-worth high pass filter is applied to remove static clutter. Finally, the micro-Doppler signature is computed by finding the spectrogram, $S(t,\omega)$, which is the square modulus of the short-time Fourier transform (STFT), of the signal $x(t)$, which spans the appropriate range bins. For a window function $w(t)$, the spectrogram is found as

$$S(t,\omega) = \left| \int_{-\infty}^{\infty} w(t-u)x(u)du \right|^2. \quad (1)$$

The application of the SSL framework requires two distinct augmented views of a data sample. We propose the following two techniques for micro-Doppler data augmentation:

1) ***Multi-Resolution Micro-Doppler (MR-mD) Augmented View:*** This augmentation technique takes advantage of STFT properties to generate multi-resolution micro-Doppler signatures from the same RF data. Specifically, varying the window size, window function, overlapping length, and dynamic range generates micro-Doppler signatures of different spatial and temporal resolutions.

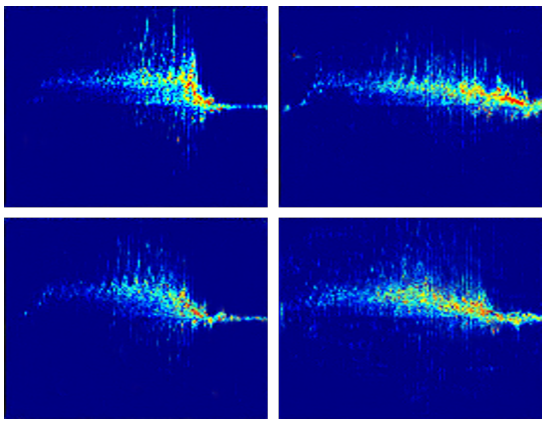

Fig. 2. PhGAN Augmented View for the activity class 'walking towards the radar'.

- While applying STFT, the choice of window size and overlapping length dictates the temporal resolution in micro-Doppler patterns. Therefore, a fixed window length of 256, and a small overlapping length of 90 generate a low temporal resolution micro-Doppler signature. On the contrary, a high-temporal resolution signature can be generated by choosing an overlapping length of 200.
- Varying the time duration in time-frequency transform helps in generating multiple spectrograms from long-duration RF data. For example, from 10 seconds of collected data, two distinct micro-Doppler signatures can be generated by separating the received pulses for the first and last 5 seconds. This resembles cropping the data.
- Changing both the temporal and spatial resolution of the micro-Doppler signature is another way of augmenting the data. This is achieved by lowering the temporal resolution by a small overlapping window length, followed by a lower number of FFT points. usually choosing the number of FFT points the same as the window length lowers the frequency resolution.
- Another way of data augmentation is to use a different window function. Instead of using a hanning window, choosing a kaiser window [27] with a high shape factor β changes the resolution of the micro-Doppler signature. Increasing β widens the main-lobe and decreases the amplitude of the side-lobes.

$$w(n) = \frac{I_0(\beta\sqrt{1-(\frac{n-N/2}{N/2})^2})}{I_0(\beta)}; 0 \le n \le N, \quad (2)$$

  where I_0 is the zeroth-order modified Bessel function of the first kind. The length $L = N + 1$.
- Finally, choosing a high dynamic range allows the weakest reflected signals to appear in the micro-Doppler signature, this in turn allows some background noises in the spectrograms.

Therefore, by diversifying the STFT properties, the RF micro-Doppler signatures can be augmented without violating the physics of radar perception for human motion.

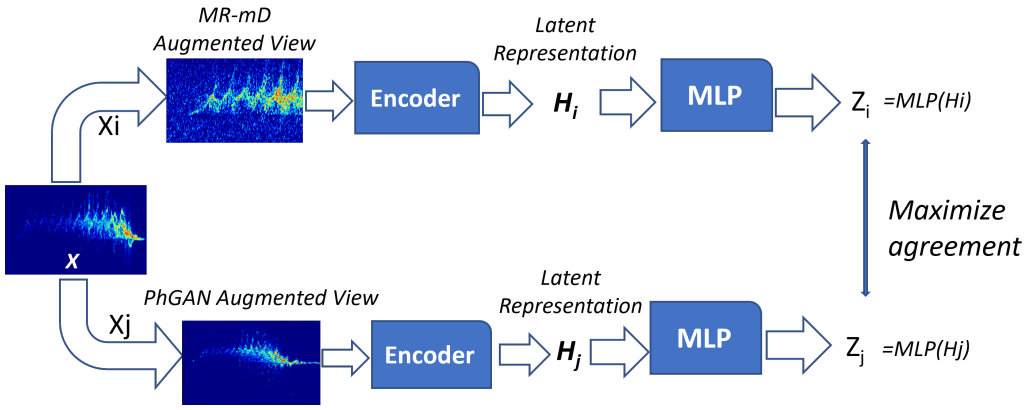

Fig. 3. Self-supervised contrastive learning framework for RF Micro-Doppler feature extraction.

Figure 1 depicts the MR-mD augmented signatures for the activity class 'walking toward the radar'.

2) ***Physics-aware GAN (PhGAN) Augmented View:*** Generative adversarial network (GAN) has been shown effective in generating micro-Doppler signatures [15], [28]–[30] in recent years. Generative models are able to map from simple latent distributions to arbitrarily complex data distributions. The latent space of such generators captures semantic variation in the data distribution. However, conventional GANs generate a lot of kinematically inconsistent signatures, while synthesizing micor-Doppler signatures. Using those erroneous data degrades the classification performance [31]. Therefore, the newly proposed physics-aware GAN (PhGAN) [15] has been employed to synthesize RF micro-Doppler signatures with enhanced kinematic fidelity. The PhGAN architecture integrates the envelopes of the signatures as the auxiliary branches into the discriminator and adds a physics-based constraint into the discriminator loss function. This modification in architecture and in loss function provides the most accurate synthetic signatures with distinct torso response, accurate Doppler shift, and consistent periodicity. Therefore, the second augmented view of the proposed SSL framework has been generated through PhGAN. Figure 2 shows the PhGAN augmented views used in SSL training.

B. The Self-Supervised Framework

The contrastive learning framework for self-supervised Radar micro-Doppler feature extraction has been illustrated in figure 3. specifically, given a randomly sampled mini-batch of activity micro-Doppler spectrogram, two augmented views *Xi* and *Xj* are generated through MR-mD and PhGAN augmentation techniques. The two images are encoded via an encoder network, namely ResNet [32] to generate representations *Hi* and *Hj*. The representations are transformed again with a non-linear transformation network, namely a MLP projection head, yielding *Zi* and *Zj* that are used for the contrastive loss. The projection head amplifies the invariant features and maximizes the ability of the network to identify different augmented views of the same image. For a minibatch of N samples, a total of 2N augmented examples are considered, the contrastive loss between a pair of augmented image is given as follows:

$$l_{i,j} = -log\frac{exp(sim(zi,zj)/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq i]} exp(sim(zi,zj)/\tau)} \quad (3)$$

where $1[k \neq i]$ is an indicator function evaluating to 1 iff $[k \neq i]$ and τ denotes a temperature parameter. Updating the parameters of the neural network using this contrastive objective causes representations of corresponding views to "attract" each other, while representations of non-corresponding views "repel" each other.

IV. Experimental Results

A. RF Data acquisition and Pre-Processing:

The data for this study is collected with a TI AWR1642 single-chip 76-GHz to 81-GHz automotive radar in an indoor laboratory environment. The radar has been placed on a table 1.5 meters up from the ground facing a walkway that was 6 meters long and 3 meters wide. The Data are acquired for 14 different ambulatory activities, which are articulated while moving toward the radar. A total of 100 samples (10 sec. duration for each sample) per activity have been recorded from 10 participants. The data have been split into 80/20 % train-test set. For the first augmented view, the 80% raw training data

List of Activities	
Vacuuming the Floor	Forward Marching
Normal Walking	Limping with Knee Brace
Short Step Walking	Walking W/Loads on Both Hands
Walking with a Cane	Walking W/Loads on One Hand
Dragging Furniture	Walking on toes
Walking with Crutches	Scissor Gait Walking
Skipping	Putting Books on the Bookshelf

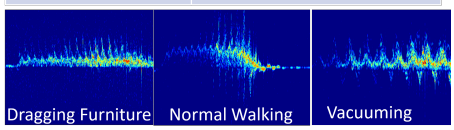

Fig. 4. List of activities and example spectrograms.

TABLE I
CLASSIFICATION PERFORMANCE FOR 14 HUMAN DAILY MOTION RECOGNITION.

Training Data	Pre-training Network	Fine-tuning data	Accuracy (%)
Real samples: 1120 (80% training Data)	Pre-training using CAE	No Fine-tuning	80.36
7000 (MR-mD Augmented Data)	Pre-training using CAE	Fine-tuned with 1120 real samples	84.64
7000 (PhGAN Augmented Data)	Pre-training using CAE	Fine-tuned with 1120 real samples	89.00
14000 (MR-mD & PhGAN Augmented Data)	Pre-training using CAE	Fine-tuned with 1120 real samples	87.14
7000 pairs (MR-mD & PhGAN Augmented Data)	Pre-training using SimCLR	Fine-tuned with 224 real samples (20%)	88.00
		Fine-tuned with 448 real samples (40%)	90.00
		Fine-tuned with 672 real samples (60%)	92.00
		Fine-tuned with 1120 real samples (100%)	93.00

have been passed through the MR-mD augmentation module to generate a total of 7000 augmented views for 14 classes.

For the PhGAN augmented view, the raw training data have been converted into micro-Doppler spectrograms by applying STFT with a window size of 256, an overlapping length of 200, a Hanning window function, and 2^{12} FFT points. After generating the micro-Doppler signatures, these data has been utilized in training the PhGAN to generate 7000 PhGAN augmented views for all the classes (500 per class).

B. Training and Evaluation

The Augmented views are paired in a way, so that, each pair consists of a MR-mD augmented view and a GAN-augmented view. A total of 7000 pairs have been formed and fed into the self-supervised framework without any labels. A resnet-18 architecture (Up to the average pooling layer) has been utilized as the encoder network, which is followed by a non-linear MLP projection head. The projection head consists of a fully connected layer followed by a ReLu activation and a Dropout of 0.3, which is finally followed by another fully connected layer. While training, a batch size of 128, temperature value of 2 (used in contrastive loss function), and trained for 500 epochs. The decreasing training loss shown in figure 5, indicates the model is learning similar features from the corresponding views in a self-supervised setting.

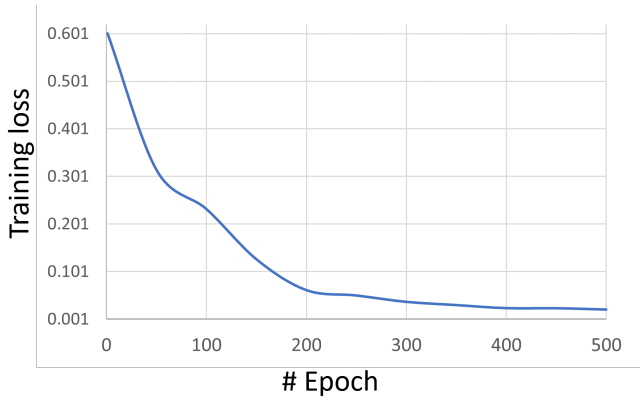

Fig. 5. Contrastive learning training loss.

To Evaluate the efficacy of the self-supervised contrastive learning, a 14-class activity recognition downstream task has been conducted. After training is completed, the projection head is replaced with a linear classification layer with 14 output neurons. The entire resnet encoder has been frozen, except the final convolution layer. Freezing the final convolutional layer provides sub-optimal recognition performance. The network is now fine-tuned with labeled data and compared with a couple of baselines performance as shown in table I. First of all, unsupervised pertaining a Convolutional AutoEncoder (CAE) [33] with the 80% (1120 samples) training data achieved an accuracy of 80.36% while tested on 20% (280 samples) test set. Pretraining the CAE separately with the two augmented views and finetuning with all the real training samples acquire an accuracy of 84.64% and 89.00% respectively for the MR-mD and PhGAN augmented data. While pretraining the CAE with the combined data from both augmented views achieved an accuracy of 87.14% after finetuning. This shows that combining both views in training actually degrades the classification performance.

Finally, while pre-training the proposed SSL framework with the proposed paired augmented views, an accuracy of

88% is achieved while fine-tuning with just 20% of labeled data. This performance is comparable with the classification accuracy achieved while pretraining with PhGAN augmented data and finetuned with all the real samples. With the increasing fine-tuning data, the SimCLR provides increasing classification performance and a maximum of 93% accuracy is obtained while using all the available labeled real data for fine-tuning.

V. CONCLUSION

This paper presents a Self-supervised contrastive learning (SimCLR) framework for radar-based human activity recognition. The proposed framework uses multi-resolution micro-Doppler augmentation and Physics-aware GAN augmentation to learn the joint embedding between the corresponding views. The contrastive loss maximize the agreement between the paired corresponding view and minimize the agreement between the dissimilar view. The experimental result shows that the proposed self-supervised framework achieved a 4% higher accuracy compared to the conventional unsupervised CAE-based pre-training while classifying 14 ambulatory human activities.

VI. ACKNOWLEDGEMENTS

This work was funded in part by AFOSR Award #FA9550-22-1-0384 and National Science Foundation (NSF) Cyber-Physical Systems (CPS) Program Award #1932547. Human studies research was conducted under UA Institutional Review Board (IRB) Protocol #18-06-1271.

REFERENCES

[1] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[2] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[3] M. M. Rahman, D. Martelli, and S. Z. Gurbuz, "Gait variability analysis with multi-channel fmcw radar for fall risk assessment," in *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2022, pp. 345–349.

[4] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, 2021.

[5] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-doppler based human-robot classification using ensemble and deep learning approaches," *CoRR*, vol. abs/1711.09177, 2017.

[6] H. Liang, X. Wang, M. S. Greco, and F. Gini, "Enhanced hand gesture recognition using continuous wave interferometric radar," in *2020 IEEE International Radar Conference (RADAR)*, 2020.

[7] M. M. Rahman, E. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for asl recognition with rf sensors," *IEEE Transactions on Aerospace and Electronic Systems*, 2022.

[8] Y. LeCun and I. Misra, "Self-supervised learning: The dark matter of intelligence," *Meta AI*, vol. 23, 2021.

[9] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[10] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[11] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020.

[12] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. E. Hinton, "Big self-supervised models are strong semi-supervised learners," *Advances in neural information processing systems*, vol. 33, pp. 22 243–22 255, 2020.

[13] J. Von Kügelgen, Y. Sharma, L. Gresele, W. Brendel, B. Schölkopf, M. Besserve, and F. Locatello, "Self-supervised learning with data augmentations provably isolates content from style," *Advances in neural information processing systems*, vol. 34, pp. 16 451–16 467, 2021.

[14] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, and M. G. Amin, "Dnn transfer learning from diversified micro-doppler for motion classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, 2019.

[15] M. Rahman, S. Gurbuz, and M. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–15, 2022.

[16] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.

[17] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," *arXiv preprint arXiv:1803.07728*, 2018.

[18] M. Noroozi and P. Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *European conference on computer vision*, 2016, pp. 69–84.

[19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[20] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1096–1103.

[21] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," *Advances in neural information processing systems*, vol. 29, 2016.

[22] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *European conference on computer vision*. Springer, 2020, pp. 776–794.

[23] S. Huang, "Self-supervised enhanced radar imaging based on deep-learning-assisted compressed sensing," *arXiv preprint arXiv:2208.03911*, 2022.

[24] I. Orr, M. Cohen, H. Damari, M. Halachmi, M. Raifel, and Z. Zalevsky, "Coherent, super-resolved radar beamforming using self-supervised learning," *Science Robotics*, vol. 6, no. 61, p. eabk0431, 2021.

[25] T. Li, L. Fan, Y. Yuan, and D. Katabi, "Unsupervised learning for human sensing using radio signals," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 3288–3297.

[26] C. Campbell and F. Ahmad, "Attention-augmented convolutional autoencoder for radar-based human activity recognition," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 990–995.

[27] J. F. Kaiser, "Nonrecursive digital filter design using the i_0-sinh window function," in *Proc. 1974 IEEE International Symposium on Circuits & Systems, San Francisco DA, April*, 1974, pp. 20–23.

[28] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware design of multi-branch gan for human rf micro-doppler signature synthesis," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

[29] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 4, pp. 3197–3213, 2020.

[30] H. Lee, J. Kim, E. K. Kim, and S. Kim, "Wasserstein generative adversarial networks based data augmentation for radar data analysis," *Applied Sciences*, vol. 10, no. 4, p. 1449, 2020.

[31] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures," *IEEE Trans on AES*, vol. 56, no. 4, pp. 3197–3213, 2020.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[33] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems*.