2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)

Ambiguity Function Based Radar Waveform Classification and Unsupervised Adaptation Using Deep CNN Models

Pavel Itkin, and Nadav Levanon

Dept. of Electrical Engineering – Systems, Tel Aviv University, Tel Aviv, 6997801, Israel

Abstract **We present a robust generalized approach to phase and frequency modulated LPI Radar waveform classification and adaptation, inspired by deep convolutional neural architectures. We use a complex Ambiguity Function matrix as a pre-processing step, following which, a waveform classification, or adaptation to unlabeled reference target domains, is performed. We test our method on a wide range of tasks, datasets, and different signal distributions. Our method surpasses the state-of-the-art performance on classification problems on multi-encoding, multi-feature datasets, in diverse and challenging conditions. Our novel approach to an unlabeled Radar waveform adaptation reveals impressive classification improvements to domain shifted unlabeled signals.**

Index Terms **— Ambiguity function, convolutional neural networks, domain adaptation, EW, LPI, waveform classification.**

I. Introduction

Radar emitter detection and recognition, as well as its waveform classification, is a critical issue in military Radar applications. It is also relevant in intercept receivers used for electronic warfare (EW) mission, hostile emitter intelligence collection, and construction of effective jamming responses. Of special importance are low probability of intercept (LPI) waveforms, operating at an extremely low signal to noise ratio (SNR).

Almost all present radar waveform classification algorithms are feature-based methods, incorporating solution-tailored feature extractor, followed by a waveform modulation classifier. Successful waveform classification techniques in the literature are heavily based on signal feature selection and extraction, such as signal statistics, temporal and spectral analysis, signal wavelet transforms, image processing, denoising, and many more. The extracted signal features are then further processed by applying an exhaustive feature search for classification performance improvement. The extracted diverse features are used as a classifier input, which may consist of an ensemble of classical machine learning (ML) or deep learning (DL) models, and in some cases, decision-tree-based classification models. The classifiers are usually trained on a wide range of previously extracted features, and accompanying research of high-level and low-level hyperparameters and architectures is being performed. While most approaches focus on feature extraction and classifier architecture construction, few known methods introduce robust, feature-flexible, and real-time applicable solutions. Popular approaches focus on time-frequency image (TFI) representations, used in speech and communication signal analysis. They use reassigned short-time Fourier transform (RSTFT) [1], Wigner-Ville distributions (WVD) [2] and Choi-Williams distributions (CWD) [3,6]. Other approaches utilize Radar specific analysis tools, such as temporal auto-correlation function (ACF) [4] and ambiguity function (AF) representations [5,10]. Extensive feature extraction-based methods [6] make use of statistical features (such as Pseudo-Zernike [3,6] moments and cumulants), TFI based features (such as CWD time peak location [3,6], Radon-Wigner transform, etc.) PSD related features (e.g. instantaneous frequency and phase) and more.

Further, significant multi-stage pre-processing is performed, e.g., image binarization, noise mitigation, etc. The waveform classifiers (WFCs) are mainly being tailored and adjusted to carefully chosen feature representations. Among the most popular models, there are state vector machines [5,7] (SVM), fully connected (FC) or dense artificial neural networks (ANN) coupled with convolutional neural networks [1,2,8,9] (CNN) and auto-encoder [10] (AE) based methods. The proposed model in [8] is an example of a custom CNN classifier, using a smoothed WVD signal representation. The models in the literature are often coupled with exhaustive parameter and architectural structure search for multi-class modulation classification.

The main drawbacks of known waveform classification solutions include manual feature selection and exhaustive searches for tailor-made classification parameters and rules. Additionally, the variety of known trained classifiers in the open literature do not address the domain shift problem, which essentially means that a well-performing classifier, trained on data generated by synthetic simulations or cooperative emitters, will fail to classify non-cooperative, real-world-scenario emitter recordings, due to different signal feature distributions. One cannot train a classifier using traditional, supervised techniques, without known labels for the intercepted Radar samples (which is usually the case for non-cooperative emitters intercepted by a passive receiver).

To answer the need for an unconstrained, generalized, and an easily adaptable waveform classification, in an emitter-saturated-environment, we present a unified approach, which targets generalization and adaptation driven performance. Robust, yet simple to implement, understand and adapt to, our approach, for the multi-feature waveform classification task, is based on the award-winning Oxford-based Visual Geometry Group's 16-layer CNN architecture (VGG16) [11], along with an incorporation of a surprisingly effective transfer learning optimization technique [12]. The final classification is being done by a generalization-driven, metric K-Nearest-Neighbor (KNN) based methods. Our approach outperforms the current state-of-the-art Radar waveform classifiers while addressing more complex and realistic multi-class configurations, successfully classifying all tested

978-1-5386-9549-4/19/$31.00 ©2019 IEEE

TABLE I
SUMMARY OF USED WAVEFORM MODULATIONS

Modulation Type	Phase Modulation	Frequency Modulation
Rectangular	const.	const.
LFM	$\pi\frac{BW}{T_{PW}}t^2$	$\frac{BW}{T_{PW}}t$
Barker 7	$\pi[1,1,1,0,0,1,0]$	const.
Barker 13	$\pi[1,1,1,1,1,0,0,1,1,0,1,0,1]$	const.
P4 7	$\frac{2\pi}{7}(m-1)\left(\frac{m}{2}-4\right);1\le m\le 7$	const.
P4 13	$\frac{2\pi}{13}(m-1)\left(\frac{m}{2}-7\right);1\le m\le 13$	const.

classes in an extremely harsh noisy environment at very low (-10dB) SNR conditions. Furthermore, we utilize our approach for a generative adversarial network (GAN) [13] based solution to the yet-to-be-addressed hard problem of unlabeled domain shifted Radar signals classification, using an unsupervised adversarial metric Domain Adaptation (DA) technology [14], which proves to be effective on the aforementioned unlabeled datasets. Our novel solution discloses a general, easy to train method for Radar waveform classification, with state-of-the-art performance, and importantly, our method is successfully applied to a diverse range of tasks, while avoiding the need for any feature extraction and classification model parameter searches.

This paper is organized as follows. Section II introduces the intercepted signal model, its AF representation, and an overview of typical Radar modulations. The proposed supervised metric VGG16 deep Radar waveform classifier approach is described in Section III. Section IV covers unlabeled DA solution using unsupervised, metric-based adversarial technology. Finally, Section V presents simulation results for all addressed problems and discusses the observed classification and adaptation performance.

II. INTERCEPTED LPI SIGNAL MODEL

The waveform classification problem requires a passive receiver to be able to successfully classify and recognize the underlying signal characteristics, such as the intentional waveform modulation type, pulse width (PW), spectral bandwidth (BW), etc. Our approach assumes that the input signal to our model has been detected, intercepted, pre-processed, and sampled. The detected signal, of constant amplitude A and instantaneous phase, φ is down-converted to the intermediate frequency by estimating and removing the carrier frequency and then sampled, resulting in a discrete sampled complex envelope of the intercepted signal $y[n]$. Furthermore, we assume that the intercepted signal consists of a single pulse only, of unknown duration τ_{pw}, consisting of a single code period, entirely contained within a sampling interval of duration T_I, which we assume to be larger than the pulse repetition interval (PRI), with a complex noise $m[n]$ added therein. Then, if for example, the signal is transmitted in an additive white Gaussian noise (AWGN) environment, with two-sided power spectral density $N_0/2$, the intercepted discrete-time signal model $y[n]$ is given by:

$$y[n]=u[n]+m[n]=Ae^{j\varphi[n]}+m[n] \tag{1}$$

where $u[n]$ is the complex envelope of the transmitted signal. We define n as the sample index, increasing with every sampling interval T_S, s.t there are $N=T_I/T_S$ samples contained within the intercepted interval T_I, sampled at the sampling frequency of $f_s=1/T_S$. Furthermore, the instantaneous phase $\varphi[n]$ is composed of the instantaneous frequency $f[n]$ and the instantaneous phase offset $\theta[n]$, as follows:

$$\varphi[n]=2\pi f[n]\cdot T_S n+\theta[n] \tag{2}$$

where the temporal dependencies of $f[n]$ and $\theta[n]$ are represented by frequency and phase modulation, respectively.

The Radar signal modulation is a popular signal modification technique that makes each modulated signal a unique waveform. The modulation in the Radar domain is a widely used, almost mandatory, technique. Often used in conjunction to matched (or mismatched) filtering, the waveform modulation is useful for achieving maximum SNR [15], in addition to obtaining the desired signal, ACF and AF properties (such as better delay or Doppler resolution, minimum peak sidelobes, minimum integrated sidelobes, etc.). The Radar modulation types possess the features of an irreplaceable and valuable toolkit for Radar signal engineers. The Radar modulation families essentially contain 3 basic types [15]: (a) Rectangular pulse (no modulation), (b) Frequency modulation (FM) such as Linear Frequency Modulation (LFM), Costas codes etc., and (c) Phase modulation (PM) such as binary phase-shift keying (BPSK) codes (e.g. Barker) and polyphase codes (e.g. P1-P4, Frank code, T1-T4 etc.). The unique benefits each waveform modulation provides can be analyzed by computing the intercepted signal complex envelope's auto-correlation function (ACF) given by:

$$R(\tau)=\int_{-\infty}^{\infty}u(t)u^*(t+\tau)dt \tag{3}$$

where $u(t)$ is the intercepted signal's complex envelope.

The ACF gives us information about the analyzed waveform's temporal properties (e.g., mainlobe width, sidelobe height, recurrent sidelobes, etc.). If we would like to analyze the Doppler shift effect on a given waveform we would look at the doppler shifted complex envelope of $u(t)$, denoted by $u_D(t)$, and plugging it into the ACF, taking the absolute value of the result, which yields the standardized absolute valued AF denoted by $|\chi(\tau,\upsilon)|^2$, as follows[15]:

$$u_D(t)=u(t)\exp(j2\pi\upsilon t) \tag{4}$$

$$|\chi(\tau,\upsilon)|^2=\left|\int_{-\infty}^{\infty}u(t)u^*(t+\tau)\exp(j2\pi\upsilon t)dt\right|^2 \tag{5}$$

where υ expresses the Doppler shift. Since the AF is symmetrical with respect to the origin, usually only one half

of the Doppler plane (positive or negative Doppler shifts) is required. The AF is computed using the code from [16]. For the simulations made in the scope of the current research, we used different practical modulations, including Rectangular, LFM, Barker 7, Barker 13, 7-sized P4, and 13-sized P4 codes. Additionally, we alternated the PWs and the BWs of all pulses, while for LFM modulation we also changed the LFM slope, reflected by the time-bandwidth (TBW) product, which considered one of the most important, performance defining, parameters of the LFM modulation. For each experiment, we used a different combination of modulations and pulse characteristics, defining each class, and making its features distinct. The specific modulations used for our experiments are shown in Table I [15]. It is also worth noting that the AF usually calculated using only a single intercepted pulse, without access to the reference pulse, at the passive intercepting receiver. Our signal representation choice, as an input for the classification and adaptation methods, based on the AF, as follows:

$$T_{AF}=\begin{bmatrix}\Re\mathfrak{e}\left(\chi(\tau,\upsilon)\right)\\ \Im\mathfrak{m}\left(\chi(\tau,\upsilon)\right)\\ \left|\chi(\tau,\upsilon)\right|^2\end{bmatrix}=\begin{bmatrix}\begin{bmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&\ddots&\vdots\\ a_{m1}&\cdots&a_{mn}\end{bmatrix}\\ \begin{bmatrix}b_{11}&\cdots&b_{1n}\\ \vdots&\ddots&\vdots\\ b_{m1}&\cdots&b_{mn}\end{bmatrix}\\ \begin{bmatrix}c_{11}&\cdots&c_{1n}\\ \vdots&\ddots&\vdots\\ c_{m1}&\cdots&c_{mn}\end{bmatrix}\end{bmatrix} \quad (6)$$

More specifically, we use a 3-channel tensor (similarly to RGB channels for images) using the complex AF's real and imaginary matrices shown in (6), where $T_{AF}\in\mathbb{C}^{m\times n\times 3}$ is a 3-channel input tensor. As for the dimensions m,n, we can sample the AF in any desired dimension, using the code in [16]. We chose the code parameters to achieve a sampled positive Doppler AF with $n=99$, $m=99$ resulting in a $99\times 99\times 3$ tensor, as our signal representation choice.

III. Metric-based DNN Waveform Classification

In the last decade, there was an unignorable rise in the popularity of deep neural networks (DNN), proving to be effective and achieving numerous breakthroughs in a wide variety of research areas. In our classification method, we made use of one of the popular, pre-defined, DNN architectures, and performed an effective initialization weights transfer, which contributed greatly to a fast model convergence.

A. VGG16 Based Classifier With Transfer Learning

The CNN [17] is a specific DNN layer architecture category, which has received much attention in computer vision (CV) complicated tasks, especially excelling in image classification and object detection. The convolutional kernel window uses many-to-one transform and reducing the output tensor's dimensions. If we denote the convolution square kernel spatial extent as F, the window stride size as S, the amount of zero padding as P and the number of kernels (or filters) as K, the output of the CNN layer, with regard to its input, will have the following size:

$$\begin{aligned}W_{output}&=\left(W_{input}-F+2P\right)/S+1\\ H_{output}&=\left(H_{input}-F+2P\right)/S+1.\\ D_{output}&=K\end{aligned} \quad (7)$$

There are plenty of popular, pre-defined CNN based architectures specializing in different tasks each. These pre-defined models are the result of long, empirical, and comprehensive researches, which prove to be more than suitable for major CV tasks. Among the most successful models is the 16-layer pre-defined model, proposed by Oxford's Visual Geometry Group [11], referred to as VGG16, best known as an award-winning model in the infamous ImageNet 1000-class classification challenge [18]. For our current research, we chose VGG16 as our single classifier model, for its impressive, multidisciplinary, success record. Due to the lack of available pre-trained Radar classification VGG16 models, we utilized the transfer learning paradigm [12]. Our training process used the pre-trained ImageNet-VGG16 weights as an initialization stage, and despite the significant domain difference, the model converged successfully, learning useful features from our AF signal representation. For further classification improvement and information loss reduction, following ideas from metric-based classification and adaptation techniques [19,20], we replaced the SoftMax layer in the VGG16 CNN with a 256-feature embedding FC layer, forming clustered, space defined classes. We perform the training with optimization of the triplet loss function expressed as follows [14]:

$$\mathcal{L}(\theta_S)=\sum_{a_i,p_i,n_i}\max\left(\left\|f_{\theta_S}(a_i)-f_{\theta_S}(p_i)\right\|^2-\left\|f_{\theta_S}(a_i)-f_{\theta_S}(n_i)\right\|^2+m,\,0\right) \quad (8)$$

where a_i, p_i, n_i represent the arbitrary anchor sample, the positive (anchor shared class) sample and negative (different class) sample, respectively, m is the distance threshold and $f_{\theta_S}(\cdot)$ is the VGG16 model. The triplet loss function optimization may be seen effectively as a way to reassure that every negative sample is drawn away, and every positive, class sharing, embedding sample is pushed towards the anchor point, making same-class samples to gather around their corresponding cluster center. Finally, the actual classification uses the KNN algorithm introduced. Given i^{th} sample denoted x_i, N_C samples associated with class c, a ground-truth label of the i^{th} sample, y_i, and the Euclidian distance operator $d(\cdot)$, the predicted class label is given by:

$$\hat{y}=\underset{c\in(1,\ldots,C)}{\arg\min}\,d\left(x_i,\mu_C\right) \quad (9)$$

where

$$\mu_C = \frac{1}{N_c} \sum_{i,s.t\, y_i=c} x_i \,. \tag{10}$$

with μ_c being the cluster center, representing the spatial extent of class c. The complete architecture of our proposed, VGG16 transfer-learning-based unified model can be seen in Fig. 1.

B. Case Study: AWGN Denoising With Signal Averaging

As seen, our proposed model uses a single pulse with no noise channel assumptions, for achieving robust and generalized classification performance. It is interesting to assess our model's performance while integrating known task-supporting techniques, such as signal denoising. For this case study, we examine the effects of the widely used Signal Averaging (SA) technique on an oversampled signal. The SA technique uses an oversampling factor T, which means sampling the same signal value by T more times, and averaging along the T received samples, as follows:

$$V_{avg} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{T} y_{i,j} \tag{11}$$

where n represents the number of oversampled intervals and $y_{i,j}$ represents the j^{th} oversampled signal sample, in the i^{th} interval. The described SA method practically means exploiting the AWGN zero expectation property, which is expected to mitigate the interleaved noise effectively. For our experiment, we do not train our VGG16 embedding feature space model any further, but test it on an oversampled signal's AFs, only choosing the oversampling factor and performing SA technique.

IV. Unsupervised Adversarial Domain Adaptation

An important and non-trivial problem in passive LPI Radar signal classification is unlabeled and differently distributed datasets in a physical RF-rich environment. One can simulate different disrupted synthetic Radar signals, and easily label them for supervised classification, to achieve acceptable waveform classification performance. However,

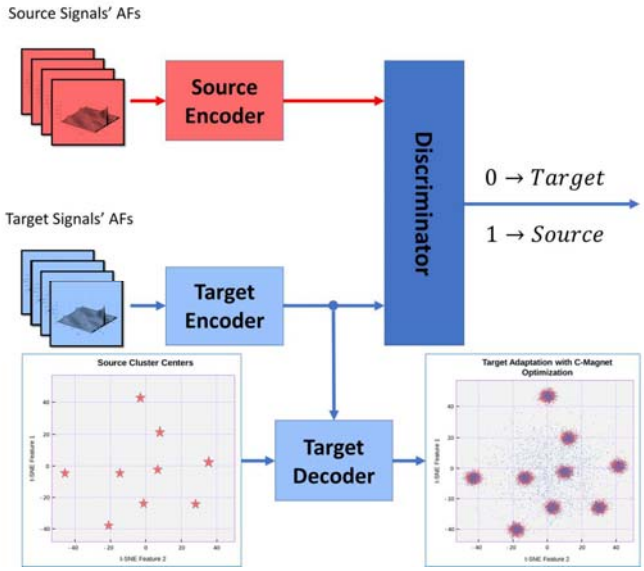

Fig. 2. The model for our proposed unlabeled DA method based on the M-ADDA adversarial model [14]. The source encoder is our metric-based VGG16 model, and it initializes the target encoder's weights.

when immigrating from the synthetic simulation environment in the laboratory to a real-world, typically non-cooperative, scenario, the relatively simple and automatic labeling task becomes nearly impossible. To our knowledge, there are no openly published works to address this difficult problem. The differently distributed, and relatively small recorded datasets that usually do not have associated labels (due to the non-cooperative emitter's nature) are often referred to as the *domain shift* problem in the deep learning literature. The unsupervised generative domain adaptation approaches [22,23] are best known for their impressive and effective results for the domain shift problem. These approaches mainly utilize the GAN technology [13] in order to project the source (i.e., training, labeled dataset) and target (i.e., unlabeled, differently distributed dataset) distributions onto a shared space. We propose an effective DA approach, based on a state-of-the-art DA technique known as M-ADDA [14], while using our aforementioned pre-trained VGG16-based

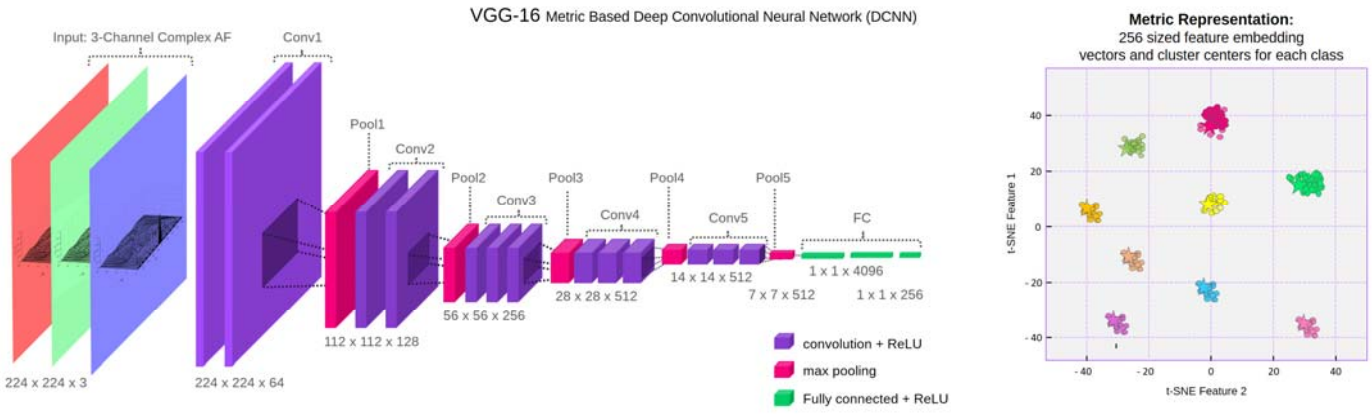

Fig. 1. Full metric based VGG16 model architecture illustration, with 3-Channel AF signal representation and 256-sized embeddings.

metric model's weights and architecture, as a baseline for source and target encoders initialization and training. The proposed model consists of one pre-training source model stage, followed by two alternating target model training stages. The source pre-training is done with the VGG16 using the triplet loss function. The target training starts with a duplication of the trained source encoder's weights, E_{θ_S}, which initialize the similar target VGG16 encoder, $E_{\theta_{T_E}}$. Subsequently, for discriminating source samples from target samples, an alternation between generator training and discriminator D_{θ_D} training (a binary class FC-DNN) is done, using adversarial loss function [13], when our target is to fool the discriminator to achieve shared embedding distribution space projection. The final training stage includes periodic optimization of the center magnet (C-Magnet) loss function [20], using the cluster centers from the source training set embeddings, and pulling the target's unlabeled samples towards the shared classes. The adversarial loss and the C-Magnet loss are presented in eqs. (12) and (13), respectively. C_j is the j^{th} cluster center, and $f_{\theta_T}(x_i)$ represents the target encoder applied to the i^{th} sample. A detailed illustration of our unlabeled Radar DA model appears in Fig. 2.

$$\mathcal{L}\left(\theta_{T_E},\theta_D\right)=\min_{\theta_D}\max_{\theta_{T_E}}-\sum_{i\in S}\log D_{\theta_D}\left(E_{\theta_S}\left(X_{S_i}\right)\right) -\sum_{i\in T}\log\left(1-D_{\theta_D}\left(E_{\theta_{T_E}}\left(X_{T_i}\right)\right)\right) \tag{12}$$

$$\mathcal{L}_{C-Mag}\left(\theta_T\right)=\sum_{i\in T}\min_j\left\|f_{\theta_T}(x_i)-C_j\right\| \tag{13}$$

V. Simulation Results

All simulation results presented are based on our metric VGG16 proposed model. The simulations were done on a GeForce™ GTX 1080Ti single GPU and Intel™ Core i9 7900X CPU. The Radar signal simulation was done in MathWorks™ MATLAB R2019a environment, and the DNN models were implemented using PyTorch™ 1.01 DL framework written in Python 3.7.1 language.

A. Unified-Metric VGG16 Waveform Classification

We train our proposed AF embedding encoder and KNN classifier on a feature-fused diverse dataset, consisting of 9 different classes, taken from Table I. Additionally, we added practically important pulse parameters, such as different PWs and BWs, to different classes with the same modulation. The training used the pre-trained ImageNet VGG16 network weights as the initialization stage and Adam [24] optimization. Firstly, we perform transfer learning from ImageNet to Radar AF represented waveforms, using a modified SoftMax layer (observing only 9 of 1000 output classes) for 150 epochs. Secondly, replacing the SoftMax layer with a randomly initialized (using Xavier [25] initialization method) FC layer that generates the aforementioned 256-dimensional feature embedding representations, which we trained with the triplet loss function.

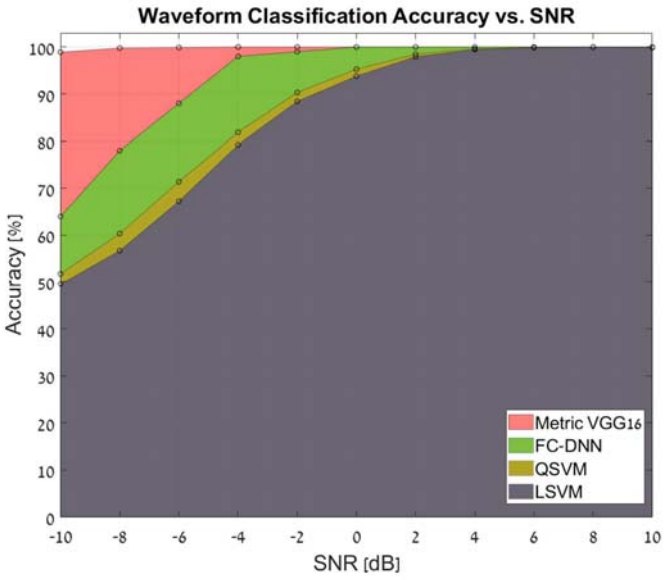

Fig. 3. Waveform classification accuracy results compared to other models' performances. Our proposed AF-based metric VGG16 model achieves 98.9% performance at SNR = -10dB

The training process, with a learning rate decay technique, lasted about 15 hours, while the significant model conversion took less than 2 hours, during which the validation error decreased by 96.2%. The final test results, using an unseen test set, and assessed for each of the tested 11 SNRs, are compared to different, AF based, classification techniques used in previously mentioned works: SVM, quadratic kernel SVM, 8-layered FC-DNN, and our proposed Metric VGG16 models. The comparison results are shown in Fig. 3, and one can easily observe that our approach demonstrates superior performance, for an impressive 98.9% classification accuracy at a very low SNR of -10dB.

B. Case-Study: AWGN With Signal Averaging

For the AWGN special case we used the aforementioned SA technique, with an oversampling factor of 20 (i.e. averaging each 20 sample intervals), using the same test dataset trained VGG16 metric-based model from the waveform classification experiment, with the aim of examining the SA technique on our model's performance at extremely low SNRs (lower than -24dB). The results are shown in Fig. 4, and provide a solid proof of the effectiveness of our method combined with the SA technique in the special, yet practically common, AWGN case.

C. Unlabeled Signals Domain Adaptation

In the current experiment, we will test our proposed novel solution, based on an existing M-ADDA [14] unsupervised domain adaptation solution. The training of the M-ADDA model consists of two main recurrently alternating phases: 1) Source model triplet-based training, and 2) Interleaved adversarial training, followed by C-Magnet-based training instance, for target and discriminator training. The source and target encoders, similarly to our other experiments, are initialized with our metric VGG16 model. We use the pre-

TABLE II
UNLABELED DOMAIN ADAPTATION RESULTS

Waveform Classification Accuracy		
	Source → Target (-12dB)	**Source → Target (-14dB)**
Source Only	81.91%	49.8%
VGG16 + M-ADDA	89.04% (+7.13%)	68.02% (+**18.22%**)

trained source encoder for target training, using -12dB and -14dB SNR unlabeled signal samples as target datasets, which our model performs poorly on. The 60-epoch adversarial training lasted for 5 hours. The waveform classification performance, before and after domain adaptation, is shown in Table II. As demonstrated, we have achieved an impressive maximum of 18.22% improvement on the challenging dataset, proving, our model's efficiency on the domain-shifted unlabeled waveform classification task. This novel approach is the first unsupervised Radar waveform domain adaptation successful solution in the known literature.

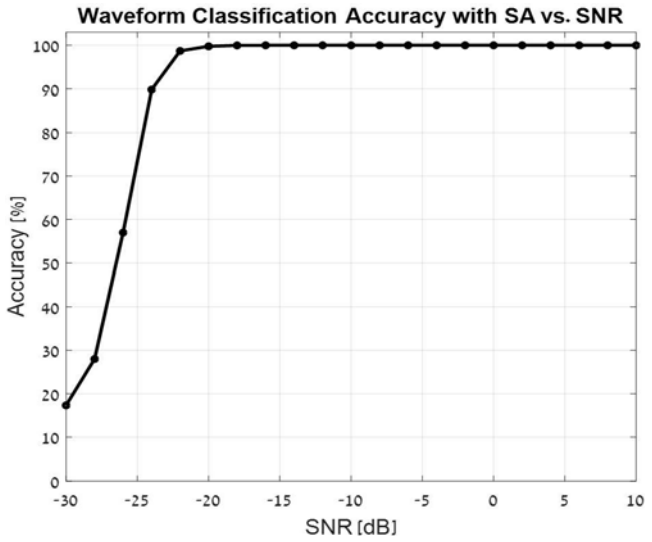

Fig. 4. Classification performance for the Metric-VGG16 model with SA. At SNR = -24dB the classification accuracy is 89.79%.

REFERENCES

[1] F. C. Akyon, Y. K. Alp, G. Gok and O. Arikan, "Classification of Intra-Pulse Modulation of Radar Signals by Feature Fusion Based Convolutional Neural Networks," *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2290-2294, September 2018.

[2] M. Zhang, M. Diao and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, vol. 5, pp. 11074-11082, June 2017.

[3] M. Zhang, L. Liu and M. Diao, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, pp. 1682, October 2016.

[4] B. D. Rigling and C. Roush, "ACF-based classification of phase modulated waveforms," *2010 IEEE Radar Conference*, May 2010.

[5] A. B. Buchenroth, B. Rigling and V. Chakravarthy, "Ambiguity-based classification of phase modulated radar waveforms," *2016 IEEE Radar Conference*, May 2016.

[6] M. Zhang, M. Diao, L. Gao and L. Liu, "Neural networks for radar waveform recognition," *Symmetry*, vol. 9, no. 5, pp. 75, May 2017.

[7] A. M. Pavy and B. D. Rigling, "Phase modulated radar waveform classification using quantile one-class SVMs," *2015 IEEE Radar Conference*, pp. 0745-04750, May 2015.

[8] C. Wang, J. Wang and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2437-2441, March 2017.

[9] X. Liu, D. Yang and A. El Gamal, "Deep neural network architectures for modulation classification," *2017 51st Asilomar Conference on Signals, Systems and Computers, IEEE*, pp. 915-919, November 2017.

[10] A. Dai, H. Zhang and H. Sun, "Automatic modulation classification using stacked sparse auto-encoders," *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 248-252, November 2016.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, April 2015.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, October 2010.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems (NIPS)*, pp. 2672-2680, December 2014.

[14] I. Laradji and R. Babanezhad, "M-ADDA: Unsupervised Domain Adaptation with Deep Metric Learning," *arXiv preprint arXiv:1807.02552*, July 2018.

[15] N. Levanon and E. Mozeson, *Radar Signals*, New York: John Wiley & Sons, 2004.

[16] E. Mozeson and N. Levanon, "MATLAB code for plotting ambiguity functions," *IEEE Transactions on Aerospace and Electronic Systems*, vol 38. No. 3, pp. 1064-1068, July 2002.

[17] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, November 1998.

[18] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255, June 2009.

[19] P. O. Pinheiro, "Unsupervised domain adaptation with similarity learning," *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8004-8013, June 2018.

[20] O. Rippel, M. Paluri, P. Dollar and L. Bourdev, "Metric learning with adaptive density discrimination," *arXiv preprint arXiv:1511.05939*, March 2016.

[21] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, June 2015.

[22] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, pp. 1180-1189, July 2015.

[23] E. Tzeng, J. Hoffman, K. Saenko and T. Darrell, "Adversarial discriminative domain adaptation," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167-7176, July 2017.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, December 2014.

[25] X. Glorot, and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the 13th international conference on artificial intelligence and statistics*, pp. 249-256, May 2010.