An Equivariant Neural Network with Hyperbolic Embedding for Robust Doppler Signal Classification

Pierre-Yves Lagrave*, **Yann Cabanes**,*****, **Frédéric Barbaresco****

*Thales Research and Technology
Palaiseau, FRANCE
email: pierre-yves.lagrave@thalesgroup.com

**Thales Land and Air Systems
Limours, FRANCE
email: yann.cabanes@gmail.com; frederic.barbaresco@thalesgroup.com

***Institut Mathématiques de Bordeaux
Bordeaux, FRANCE

Abstract: *This paper focuses on the robustness aspects of Doppler signal processing tasks with Machine Learning algorithms within the context of pathological radar clutter classification. More precisely, group-based convolution operators are combined with the hyperbolic embedding technique to build an equivariant neural network operating on the Doppler signal represented as complex covariance matrices. Our numerical testing performed on simulated data has shown the superiority of our approach when compared to conventional neural networks, from both accuracy and robustness standpoints.*

1. Introduction

The deployment of a radar on a new geographical site is long and costly, and this process could be alleviated by an automated recognition of pathological clutter. As in [1], we aim at developing a Machine Learning (ML) algorithm allowing to identify specific clutter characteristics from their Doppler spectrum fluctuation. It has been shown in [6, 7, 8, 9] that ML techniques can be used to obtain very good results for the classification of radar micro-Doppler data. Although very promising from an accuracy standpoint, these works do not focus on the robustness of the considered ML algorithms, leaving for instance aside the impact of an increase of the sensor thermal noise on the inference results.

ML algorithms can be made more robust to a given set of transforms of their inputs by using training data augmentation strategies, as demonstrated within the image processing area [3]. With this approach, the robustness is learnt directly from the data by spending some useful capacity of the algorithm. There is also no theoretical robustness guarantee for the trained algorithm, although attempts have been made to theorize the data augmentation process and its effects [4, 5].

In this paper, we introduce an approach combining hyperbolic embedding and equivariant inner representation of the input Doppler signal through a group-based convolutional neural network. The proposed architecture in particular ensures that local robustness to a specific group action is enforced through the algorithmic design.

978-3-944976-30-3

2. From Covariance Matrices to Cosets

We follow [1] and represent each cell by its covariance matrix, which is Toeplitz Hermitian Positive Definite (THPD). We explain in the following how a given THDP matrix can actually be seen as a manifold valued function for which there exists a natural group action.

To do so, we denote by $\mathbb{D}$ the Poincaré unit disk $\mathbb{D} = \{z = x + iy \in \mathbb{C}/\, |z| < 1\}$ and by $\mathbb{T}$ the unit circle $\mathbb{T} = \{z = x + iy \in \mathbb{C}/\, |z| = 1\}$. We then consider the following Lie Groups:

$$\mathrm{SU}(1,1) = \left\{ g_{\alpha,\beta} = \begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix} /\, |\alpha|^2 - |\beta|^2 = 1, \alpha, \beta \in \mathbb{C} \right\} \tag{1}$$

$$\mathrm{U}(1) = \left\{ \begin{bmatrix} \alpha/\,|\alpha| & 0 \\ 0 & \bar{\alpha}/\,|\alpha| \end{bmatrix}, \alpha \in \mathbb{C} \right\} \tag{2}$$

We can endow $\mathbb{D}$ with a transitive group action $\circ$ of $\mathrm{SU}(1,1)$ defined as it follows

$$\forall g_{\alpha,\beta} \in \mathrm{SU}(1,1),\ \forall z \in \mathbb{D},\ g_{\alpha,\beta} \circ z = \frac{\alpha z + \beta}{\bar{\beta} z + \bar{\alpha}} \tag{3}$$

The Cartan decomposition associated with $\mathrm{SU}(1,1)$ is the following

$$\begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix} = |\alpha| \begin{bmatrix} 1 & z \\ \bar{z} & 1 \end{bmatrix} \begin{bmatrix} \alpha/\,|\alpha| & 0 \\ 0 & \bar{\alpha}/\,|\alpha| \end{bmatrix} \tag{4}$$

with $z = \frac{\beta}{\bar{\alpha}}$ and $|\alpha| = \frac{1}{\left(1-|z|^2\right)^{1/2}}$. This allows building the following diffeomorphism between $\mathrm{SU}(1,1)$ (seen as a topological space) and $\mathbb{T} \times \mathbb{D}$

$$\phi : \begin{cases} \mathrm{SU}(1,1) & \to \mathbb{T} \times \mathbb{D} \\ g_{\alpha,\beta} & \mapsto \left(e^{i\theta_\alpha}, \frac{\beta}{\bar{\alpha}}\right) \end{cases} \tag{5}$$

where $\theta_\alpha = \arg(\alpha) \mod 2\pi$. As $\mathbb{T} \simeq \mathrm{U}(1)$ as a group, we can therefore associate $z \in \mathbb{D}$ with an element of the quotient space $\mathrm{SU}(1,1)/\,\mathrm{U}(1)$, which can then further be lifted to a group element $g_{\alpha,\beta} \in \mathrm{SU}(1,1)$ by choosing

$$\begin{cases} \alpha & = \frac{e^{i\theta_z/2}}{\left(1-|z|^2\right)^{1/2}} \\ \beta & = \frac{|z| e^{i\theta_z/2}}{\left(1-|z|^2\right)^{1/2}} \end{cases} \tag{6}$$

Let's then denote by $\mathcal{T}_n^+$ the set of Toeplitz Hermitian Positive Definite (THPD) matrices of size n. As described in [1], the regularized Burg algorithm allows transforming a given matrix $\Gamma \in \mathcal{T}_n^+$ into a power factor in $\mathbb{R}_+^*$ and reflection coefficients in a $(n-1)$ dimensional product of Poincaré spaces $\mathbb{D}^{n-1} = \mathbb{D} \times ... \times \mathbb{D}$ via the bijective map

$$\psi_n : \begin{cases} \mathcal{T}_n^+ & \to \mathbb{R}_+^* \times \mathbb{D}^{n-1} \\ \Gamma & \mapsto (p_0, \mu_1, ..., \mu_{n-1}) \end{cases} \tag{7}$$

Combining all of the above, we can therefore see an element $\Gamma \in \mathcal{T}_n^+$ as a power factor in $\mathbb{R}_+^*$ and $n-1$ cosets elements in $\mathrm{SU}(1,1)/\,\mathrm{U}(1)$. In the following, we will focus on radar clutter classification and we will consider as in [1] rescaled THPD matrices which can be represented by the reflection coefficients only.

3. Related Work and Contribution

Classifying Doppler data with deep learning methods has been investigated from several angles by leveraging on the multiple representations of the signal, including in particular the use of off-the-shelf algorithms such as Convolutional Neural Networks (CNN) [7] or Long-Short Term Memory (LSTM) modeling [6]. Other approaches leverage on the natural representation of the Doppler signal as Symmetric Positive Define (SPD) matrices [9] or HPD matrices [8] when the complex phase term is taken into account.

Classifying SPD, HPD and THPD matrices is a specific case of the problem of classifying manifold valued data, for which generic algorithms have been introduced. For instance, the usual convolution operator on $\mathbb{R}^2$ is generalized to manifold valued images in [10] and provides equivariance to the isometry group admitted by the underlying manifold.

Also, Group-Convolutional Neural Networks (G-CNN) are becoming more and more popular as they are shown to be efficient for a wide range of operational applications [12, 13]. Following several pieces of work such as [14, 15] generalizing usual CNN [11] by introducing group-based convolution kernels, [13] has recently introduced a very generic approach providing equivariance to any Lie Group with a surjective exponential map and which is applicable to any input data representable by a function defined on a smooth manifold and valued in a vectorial space.

By leveraging on [13], we introduce in the present paper a G-CNN equivariant to $\mathrm{SU}(1,1)$ and applicable to the grid of points of $\mathbb{D}$ obtained from the hyperbolic embedding of the covariance matrices. We have implemented our approach in TensorFlow in a differentiable way to allow end-to-end training of the proposed architecture.

Numerical experiments have been performed by working with simulated data and we provide in the following some evidence for the robustness of the approach to perturbations arising with the increase of the thermal noise of the sensor. To further appreciate the benefits of using equivariant layers operating in the hyperbolic space, we also show the superiority of our approach by providing a comparison with the use of conventional fully connected neural networks.

4. Group-Equivariant Neural Network with Hyperbolic Embedding

We propose building a $\mathrm{SU}(1,1)$-equivariant neural network by combining hyperbolic embedding and group-based convolution feature maps as depicted on Fig. 1.

4.1. Group-based Convolution Operator

For $G = \mathrm{SU}(1,1)$, we will build feature maps $\psi : G \to \mathbb{C}$ by leveraging on the convolution operators on groups. More precisely, for a kernel function $k_\theta : G \to \mathbb{C}$ parameterized by $\theta \in \mathbb{R}^\ell$, an input feature map $f : G \to \mathbb{C}$ and a group element $g \in G$, we will consider the following operator ψ^θ

$$\psi^\theta(g) = (k_\theta \star_G f)(g) = \int_G k_\theta\left(h^{-1}g\right) f(h)\, d\mu^G(h) \tag{8}$$

as long as the right handside integral is well defined, i.e for $k_\theta, f \in L^1\left(G, d\mu^G\right)$ where $L^1(X, d\mu)$ refers to the set of functions from X to $\mathbb{C}$ which are integrable with respect to the measure μ. In equation (8), μ^G refers to the Haar measure of G that is normalized according to

$$\int_G F(g \circ 0_{\mathbb{D}})\, d\mu^G(g) = \int_{\mathbb{D}} F(z)\, dm(z) \tag{9}$$

for all $F \in L^1(\mathbb{D}, dm)$, with $0_{\mathbb{D}}$ the center of $\mathbb{D}$, and where the measure dm is given, for $z = z_1 + iz_2$, by

$$dm(z) = \frac{dz_1 dz_2}{\left(1 - |z|^2\right)^2} \tag{10}$$

We emphasize here that the convolution operator ψ^θ is equivariant with respect to the action of G so that $L_g\psi^\theta = k_\theta \star_G (L_g f)$, for all $g \in G$ and where L_g is the left shift operator such that $\forall h \in G$ and $\forall f : G \to \mathbb{C}$, $L_g f(h) = f\left(g^{-1}h\right)$. Said differently, it means that the convolution feature map is transformed consistently with the input data, as illustrated on Fig. 2.

As we don't have much prior knowledge about the shape of the kernel $k_\theta : G \to \mathbb{C}$ in practice, we generically model it as a simple Fully Connected Neural Network (FCNN) as in [13]. After adequate localization of the kernel function, the integral (8) can be computed by a Monte-Carlo approach or by using Helason-Fourier analysis [17] on symmetric spaces following the ideas underlying to the approach [15] introduced for compact groups.

4.2. Lifting Operator

In order to process the input $\Gamma \in \mathcal{T}_n^+$ corresponding to the reflexion coefficients μ_1,...,μ_{n-1} with the group-based convolution layers as previously introduced, a lifting step is required to represent the data as a function $f_\Gamma : \mathrm{SU}(1,1) \to \mathbb{C}$. Assuming that the ordering of the reflection coefficients is not material for our considered classification task, we will represent a given element $\Gamma \in \mathcal{T}_n^+$ as it follows:

$$f_\Gamma : \begin{cases} \mathrm{SU}(1,1) & \to \mathbb{C} \\ g & \mapsto \sum_{i=1}^{n-1} \mu_g^{\mathbb{C}} \mathbf{1}_{\{\mu_g = \mu_i\}} \end{cases} \tag{11}$$

where $\mu \to \mu^{\mathbb{C}}$ is the canonical embedding from $\mathrm{SU}(1,1)/\,\mathrm{U}(1) \simeq \mathbb{D}$ to $\mathbb{C}$, and $\mu_g \in \mathrm{SU}(1,1)/\,\mathrm{U}(1)$ refers to the class of $g \in \mathrm{SU}(1,1)$ in the quotient space $\mathrm{SU}(1,1)/\,\mathrm{U}(1)$.

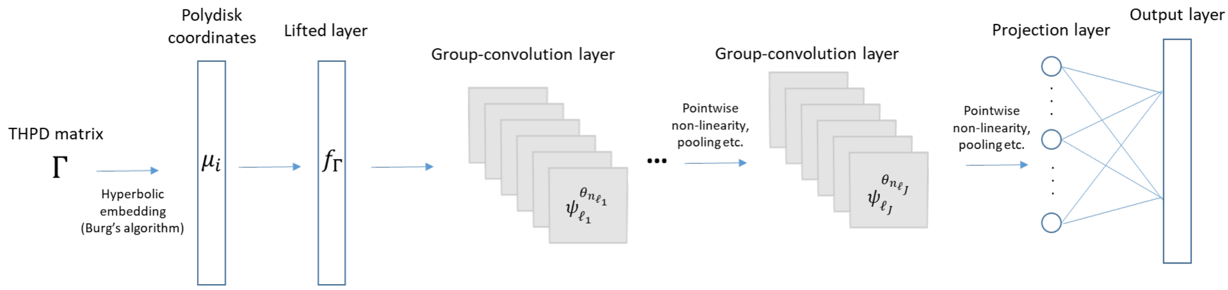

Figure 1: $\mathrm{SU}(1,1)$-equivariant Neural Network with hyperbolic embedding for THPD matrix classification. The initial THPD matrix Γ is embedded into $\mathbb{D}^{n-1}$ through the computation of the reflexion coefficients μ_i, which will be themselves lifted to a function $f_\Gamma : \mathrm{SU}(1,1) \to \mathbb{C}$. The lifted representation is then processed with J group-based convolution layers of ℓ_i feature maps each and denoted $\psi_{\ell_j}^{\theta_{n_{\ell_j}}}$.

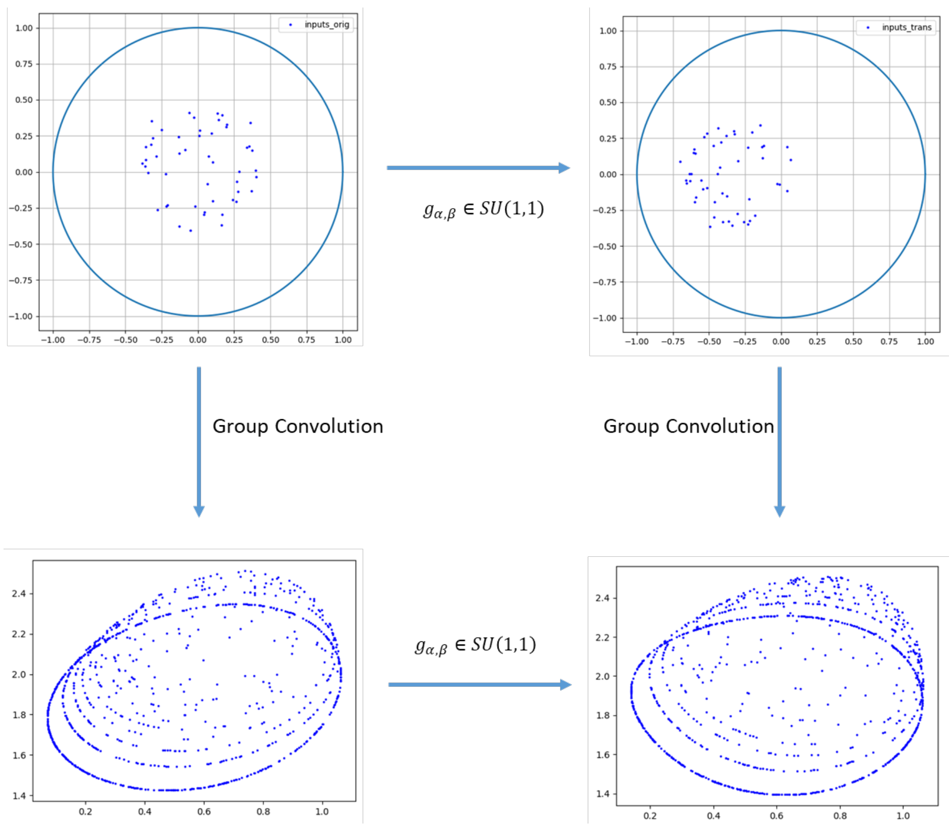

Figure 2: Example of an equivariant gaussian convolution feature map and corresponding inputs in dimension $n = 50$. The input data points are represented in $\mathbb{D}$ (after folding from $\mathbb{D}^{n-1}$) while the features maps discretized over 1000 samples are shown in the complex plane. The initial points $\mu_1, ..., \mu_{n-1} \in \mathbb{D}$ (top-left) are subject to the action of an element $g_{\alpha,\beta} \in \mathrm{SU}(1,1)$, leading to transformed points (top-right). The group-convolution operator is applied to the corresponding lifted features maps (bottom) and the action of $g_{\alpha,\beta} \in \mathrm{SU}(1,1)$ transforms the initial convoluted data (bottom left) to the convolution of the transformed data (bottom right), making the overall diagram commutative.

5. Numerical Experiments

We have implemented our approach in TensorFlow and we provide in the following the results of experiments for pathological radar clutter classification. All tests have been run using a single workstation equiped with a NVIDIA GeForce RTX 2080 GPU card.

5.1. Classification Problem

We have considered the problem of pathological radar clutter classification as formulated in [1]. More precisely, we represent each cell by its THPD auto-correlation matrix Γ, our goal being to predict the corresponding clutter $c \in \{1, ..., n_c\}$ from the observation of Γ. The input data have been obtained by simulating the auto-correlation matrix of a given cell according to

$$Z = \sqrt{\tau} R^{1/2} x + b_{radar} \tag{12}$$

where τ is a positive random variable corresponding to the clutter texture, R a THPD matrix associated with a given clutter, $x \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma^x\right)$ and $b_{radar} \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma\right)$, with $\mathcal{N}_{\mathbb{C}}\left(0, t\right)$ referring to the complex Gaussian distribution with mean 0 and standard deviation t. In the following, b_{radar} will be considered as a source of thermal noise inherent to the sensor.

5.2. Testing Set-Up

We will be interested in the following in evaluating the accuracy and the robustness of our approach by considering a neural network constituted of one $\mathrm{SU}(1,1)$ convolutional layer with one feature map, followed by one fully connected output layer. The kernel function is modeled as a neural network with one layer of 32 neurons with swish activation functions.

To appreciate the improvement provided by our approach, we will compare the obtained results with those corresponding to the use of a conventional neural network operating on the complex reflections coefficients. By using one hidden layer of 40 neurons, both approaches rely roughly on the same numbers of trainable parameters.

In the following, we will denote $\mathcal{N}_\sigma^G$ (resp. $\mathcal{N}_\sigma^{FC}$) the neural network with $\mathrm{SU}(1,1)$ equivariant convolutional (resp. fully connected) layers and trained on 400 THPD matrices of dimension 10 corresponding to 4 different classes (100 samples in each class) which have been simulated according to (12) with a thermal noise standard deviation σ.

5.3. Results and Discussion

In order to evaluate the algorithms $\mathcal{N}_\sigma^G$ and $\mathcal{N}_\sigma^{FC}$, we have considered several testing sets T_σ consisting in 2000 THPD matrices of dimension 10 (500 samples in each of the 4 classes) simulated according to (12) with a thermal noise standard deviation σ.

T_σ	$\mathcal{N}_1^G$	$\mathcal{N}_1^{FC}$	$\mathcal{N}_{250}^G$	$\mathcal{N}_{250}^{FC}$
T_1	0.9873 (0.0015)	0.9735 (0.0037)	0.9891 (0.0021)	0.9846 (0.0026)
T_{50}	0.9606 (0.0039)	0.9301 (0.0054)	0.9653 (0.0034)	0.9621 (0.0047)
T_{100}	0.9327 (0.0074)	0.8811 (0.0041)	0.9386 (0.0054)	0.9401 (0.0039)
T_{150}	0.9051 (0.0054)	0.8419 (0.0044)	0.9148 (0.0032)	0.9191 (0.0056)
T_{200}	0.8805 (0.0056)	0.8042 (0.0058)	0.8955 (0.0068)	0.8988 (0.0043)

Table 1: Accuracy results of the algorithms $\mathcal{N}_1^G$, $\mathcal{N}_1^{FC}$, $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ on the testing sets T_σ, averaged over 10 realizations and together with the corresponding standard deviation in brackets.

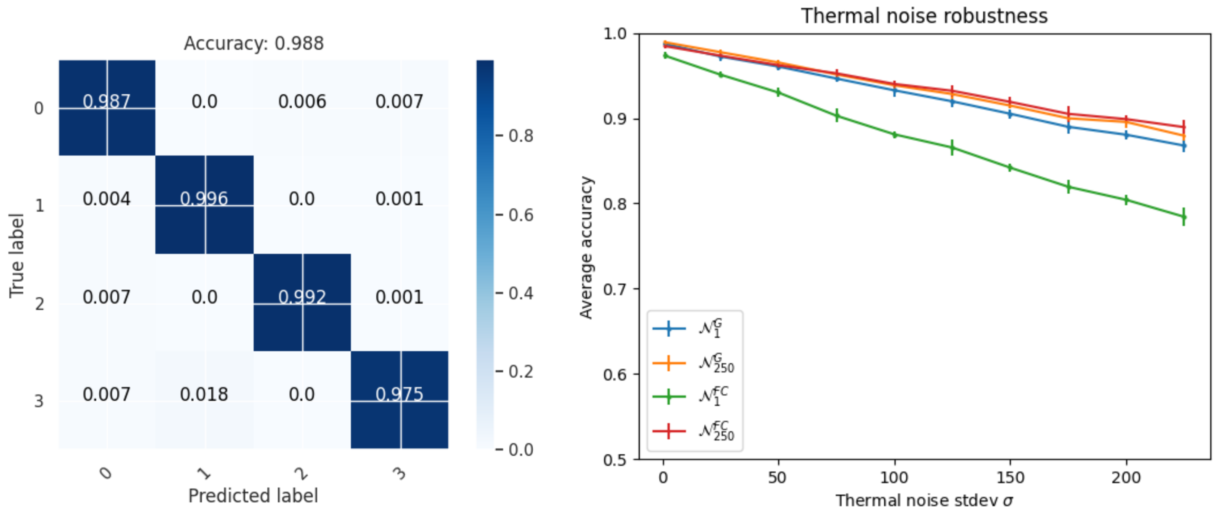

Figure 3: The confusion matrix on the left handside corresponds to the evaluation of $\mathcal{N}_1^G$ on the testing set T_1, averaged over 10 realizations. The robustness of the approach is illustrated on the right handside where average accuracy results of the algorithms $\mathcal{N}_1^G$, $\mathcal{N}_1^{FC}$, $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ on the testing sets T_σ are shown as a function of σ, together with the corresponding standard deviation as error bars.

Our results are shown in Table 1 and on Figure 3. To interpret the numbers, we first emphasize that the algorithms $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ could be seen as $\mathcal{N}_1^G$ and $\mathcal{N}_1^{FC}$ trained on augmented data with respect to the thermal noise. The fact that $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ achieve similar performances is therefore expected from a theoretical standpoint and gives us some comfort with respect to the correctness of the implementation of our approach.

We then observe that $\mathcal{N}_1^G$ reaches similar performances as $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ as σ increases, meaning that our approach allows to achieve some degree of robustness with respect to the variation of σ through the use of equivariant layers and pooling steps. The small discrepancies ($\approx 1\%$ for $\sigma = 200$) between the results obtained with $\mathcal{N}_1^G$ and the performances achieved by $\mathcal{N}_{250}^G$ and $\mathcal{N}_{250}^{FC}$ can be attributed to discretization effects in the implementation arising in particular with the computation of the $\mathrm{SU}(1,1)$ convolution maps. This behavior has to be compared with that of $\mathcal{N}_1^{FC}$ which consistently achieves lower accuracy results than $\mathcal{N}_1^G$, hence demonstrating the superiority of our approach from both accuracy and robustness standpoints.

6. Conclusions and Further Work

By leveraging on the hyperbolic embedding approach to represent THPD matrices, we have introduced a $\mathrm{SU}(1,1)$ equivariant neural network to perform supervised classification taks with application to pathological radar clutter classification.

Our approach has been implemented in TensorFlow and we have demonstrated its superiority from both accuracy and robustness standpoints by working on simulated data and by focusing on thermal noise related perturbations.

Further work will include extending the approach to perform Positive Definite Block-Topelitz matrices classification by leveraging on their embedding into Siegel spaces [2] and by building $\mathrm{SU}(n,n)$ equivariant neural networks. This would in particular allow considering the spatial dimension of the signal in the learning algorithm.

References

[1] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon and Jérémie Bigot, Non-supervised Machine Learning Algorithms for Radar Clutter High-Resolution Doppler Segmentation and Pathological Clutter Analysis, IRS 2019, Ulm, Germany, 26-28 June 2019

[2] Barbaresco, F.: Radar micro-doppler signal encoding in siegel unit poly-disk for machine learning in fisher metric space. In: 2018 19th International Radar Sympo- sium (IRS). pp. 1-10 (2018). https://doi.org/10.23919/IRS.2018.8448021

[3] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[4] Lyle, Clare & Wilk, Mark & Kwiatkowska, Marta & Gal, Yarin & Bloem-Reddy, Benjamin. (2020). On the Benefits of Invariance in Neural Networks.

[5] Shuxiao Chen and Edgar Dobriban and Jane H Lee (2019) A Group-Theoretic Framework for Data Augmentation

[6] P. Molchanov, R. I. Harmanny, J. J. de Wit, K. Egiazarian, and J. Astola. Classification of small UAVs and birds by micro-Doppler signatures. International Journal of Microwave and Wireless Technologies, 6(3-4):435–444, June 2014.

[7] R. P. Trommel, R. I. A. Harmanny, L. Cifola, and J. N. Driessen. Multitarget human gait classification using deep convolutional neural networks on micro-doppler spectrograms. In 2016 European Radar Conference (EuRAD), pages 81–84, Oct. 2016.

[8] Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, Matthieu Cord. A Hermitian Positive Definite neural network for micro-Doppler complex covariance processing. International Radar Conference, Sep 2019, Toulon, France.

[9] Z. Huang and L. J. Van Gool. A Riemannian Network for SPD Matrix Learning. In AAAI, volume 1, page 3, 2017.

[10] JJ Bouza, CH Yang, BC Vemuri MVC-Net: A Convolutional Neural Network Architecture for Manifold-Valued Images With Applications - arXiv preprint arXiv:2003.01234, 2020 - arxiv.org

[11] LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11): 2278–2324.

[12] Bekkers, E., Lafarge, M., Veta, M., Eppenhof, K., Pluim, J., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis. Accepted at MICCAI 2018, arXiv preprint arXiv:1804.03393 (2018)

[13] Finzi, M.; Stanton, S.; Izmailov, P.; and Wilson, A. G. 2020. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. arXiv preprintarXiv:2002.12880

[14] Cohen, T., Welling, M.: Group equivariant convolutional networks. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research, PMLR, New York, 20–22 June 2016, vol. 48, pp. 2990–2999 (2016)

[15] Kondor, R., Trivedi, S.: On the generalization of equivariance and convolution in neural networks to the action of compact groups. In: Dy, J.G., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmassan, Stockholm, Sweden, 10–15 July 2018, Proceedings of Machine Learning Research, vol. 80, pp. 2752–2760. PMLR (2018).

[16] R Chakraborty, M Banerjee, BC Vemuri A CNN for homogneous Riemannian manifolds with applications to Neuroimaging - arXiv preprint arXiv:1805.05487, 2018 - arxiv.org

[17] S. Helgason, Groups and geometric analy-sis, Academic Press, Orlando, 1984.