Moving Target Classification Based on micro-Doppler Signatures Via Deep Learning

Yonatan D. Dadon*, Shahaf Yamin*, Stefan Feintuch*, Haim H. Permuter†, Igal Bilik†, Joseph Taberkian†
School of ECE, Ben Gurion University of the Negev
*. {*dadonyon, yamins, stefanfe*}*@post.bgu.ac.il* †. {*haimp, bilik, joseph*}*@bgu.ac.il*

***Abstract*—Radar-based classification of ground moving targets relies on Doppler information. Therefore, the classification between humans and animals is a challenging task due to their similar Doppler signatures. This work proposes a Deep Learning-based approach for ground-moving radar targets classification. The proposed algorithm learns the radar targets' micro-Doppler signatures in the 2D fast-time slow-time radar echoes domain. This work shows that the convolutional neural network (CNN) can achieve high classification performance. Also, it shows that efficient data augmentation and regularization significantly improve classification performance and reduce over-fit.**

I. Introduction

Target classification is one of the major radar tasks in a variety of military and security applications. Some of these applications use electro-optical (EO) sensors, such as video surveillance [1]. Radars have some significant benefits compared to EO sensors in terms of immunity to severe weather and poor lighting conditions, low cost, and robustness.

Classification between humans and animals is a challenging task due to their Doppler similarity. However, they are non-rigid bodies, and therefore, different motions of their parts induce additional modulations to the radar echoes [2]. These modulations, denoted as micro-Doppler, were proposed in the literature for radar targets classification [3].

Object classification using feature extraction from grid-like data by convolutional neural networks (CNNs) has been extensively studied in the literature. [4], [5]. It was shown that CNN, trained by visual data can outperform human classification capabilities, under visual distortions [6], [5], [7] [8].

Majority of widely used training data sets contain objects that are characterised by visual features that are apparent for human classifier MNIST [9], ImageNet [10], CIFAR-10 [11]. The main task of the CNN classifier trained with such a dataset is to extract those distinguishing features.

Typically, radar "raw" data differs from these datasets because it does not contain identifiable features. Thus, the MAFAT radar dataset of moving humans and animals [12] was characterized only by an operational radar frequency. The radar measurements were collected from humans and animals moving at a similar radial velocity toward the radar. Therefore, their classification using only spectral data is challenging, even for a human classifier. This work's main idea is that the CNNs can learn some of a data's distinguishable characteristics using sufficiently diverse training data.

This work leverages the CNNs capabilities to extract the micro-Doppler signatures from the collected radar echoes to classify between animals and humans. This work adopts some ingredients from the computer-vision deep learning (DL) for radar target classification: architectures, regularization methods to overcome the over-fitting phenomena, pre-processing methods to emphasize the distinguishing Doppler features, and data augmentations techniques to enrich the data's diversity.

A wide variety of micro-Doppler based target classification methods has been studied in the literature. Ref. [13] showed that computer-vision methods such as the Gabor filter could be used for micro-Doppler feature extraction. Ref. [14] used machine learning approach and micro-Doppler phenomenon to classify various human activities. These activities are distinguishable in Doppler and micro-Doppler signatures, contrary to the data used in this work. The CNN classification performance was evaluated in [15] using different data representations. It was shown that frequency-domain representation could provide a significant classification performance improvement. The maximum-likelihood and majority-voting classifiers were introduced in [3] for a similar classification problem. High classification performance was demonstrated using Gaussian mixture model (GMM). Ref. [16] showed that data augmentation techniques combined with CNN could significantly improve performance on an imbalanced radar echo dataset. This dataset contains longer radar echo signals with known physical characteristics, contrary to MAFAT dataset [12].

This work investigates the impact of the CNN architecture and data augmentation on the radar targets classification performance. Efficient CNN training requires a broad, diverse, and balanced training data set [17]. This work investigates the CNN classification performance as a function of the training data set size and data diversity. The data set in [12] is extremely imbalanced and small. Therefore, straightforward DL methods are prone to over-fitting. This work demonstrates that data augmentation can significantly improve DL-based radar target classification. We show that using the right configuration of known regularization methods can further improve the model performance under the ROC-AUC criteria [18]. Also, we review different depths and blocks of the CNN architectures and their corresponding performance. Furthermore, we demonstrate how effective computer-vision methods can solve the problem above by achieving a single

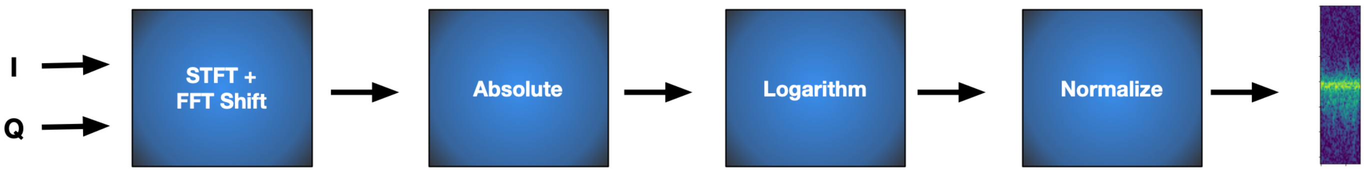

Fig. 1. Pre-process flow.

model with high classification performance of maximal 0.95 and 0.94 on average under the ROC-AUC criteria over MAFAT FULL PUBLIC test set [12]. All results included a hyper-parameter optimization procedure, in which we investigated the effect of each factor on the overall performance.

II. Problem Definition

The radar target classification problem is defined in this Section. Conventionally, radars estimate the target's velocity from the Doppler phenomenon:

$$f_d \approx 2v \cdot \frac{f_t}{c}. \tag{1}$$

where, f_d and f_t are the Doppler and the carrier frequencies, respectively, v is the target radial velocity, and c is the speed of light.

Radar targets classification using Doppler frequency only is impossible when targets have similar velocities. However, the classification between non-rigid targets is possible by exploiting micro-Doppler modulations on the central Doppler shift, f_d. Notice that conventionally, radars operate in the presence of static but large radar clutter that can "mask" slowing moving targets. Fig. 2 shows the Doppler, micro-Doppler, and radar clutter components in the received radar echo.

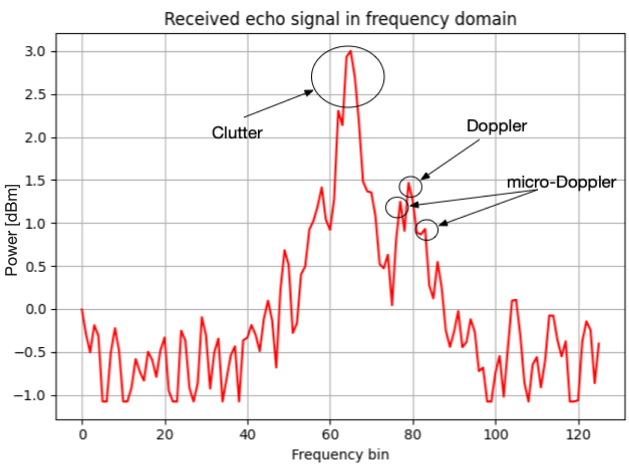

Fig. 2. Human example of a received radar echo.

Consider M received and sampled radar echoes in the time domain. These signals are arranged in a *segment* matrix whose horizontal axis represents the pulse transmission time ("slow time"), and the vertical axis represents the signal samples within each pulse ("fast time"). Let $X \in \mathbb{C}^{128x32}$ be the complex IQ signal matrix in the time domain at the input of the pre-processing, and let $X^F \in \mathbb{R}^{128x32}$ represent a segment at the output of the pre-processing. Let $\widetilde{X}$ be an augmented segment. x_i is the $i-th$ column of X: $X = \begin{bmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_{32} \\ | & | & \cdots & | \end{bmatrix}$. Y is the label of the segment X such that $Y \in \{0,1\}$. In this work, we denote the uniform distribution on a discrete set S as $Unif[S]$, the continuous uniform distribution is $U[S]$, and $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean, μ, and variance σ^2, respectively.

Small movements of the object's parts that differ from the entire object locomotion (for example, the hand movement during a run) induce micro-Doppler signatures [19]. The classification goal is to estimate a model $f_\theta(X)$ with optimal parameters θ^*. We denote the model's output $P^\theta_{Y|X}(\cdot|X)$ and we aim to choose θ^* such that $P^{\theta^*}_{Y|X}(\cdot|X)$ will be equal to the true conditional target distribution $P_{Y|X}(\cdot|X)$.

The performance of the proposed CNN-based classifier was evaluated in this work using the receiver operating characteristic (ROC)-AUC metric [18]. ROC-AUC is the true positive rate (TPR) as a function of the False Positive Rate (FPR) at various decision thresholds. ROC-AUC denotes the area under the curve, and our goal is to maximize it, i.e., achieve ROC-AUC that is as close as possible to 1.

III. Pre-processing

The radar data set contains complex baseband segments in the time domain. Fig. 1 shows the pre-processing of the segment X, transformed to the frequency domain at the processing chain's output. Ref. [15] showed the superiority of the training process in the spectrum domain.

The pre-processing can be formalized as follows:

- STFT: $x_i^F(k) = STFT\left[x_i[n]\right]$ with Hanning window function.
- FFT Shift: $x_i^F(k) \longleftarrow x_i^F\left[((k - \frac{N}{2}))_N\right]$.
- Absolute + Logarithm: $X^F \longleftarrow \log\left(|X^F|\right)$.
- Normalize: $X^F \longleftarrow \frac{X^F - \text{mean}(X^F)}{\text{std}(X^F)}$.
  Then, $X^F[i][j] \longleftarrow \max(X^F[i][j], \text{median}(X^F) - 1)$

The last normalization operation clips the noise floor to a single value per segment, equal to $\text{median}(X^F) - 1)$. This operation emphasizes the difference between the object's valuable information and the noise.

IV. Proposed Solution

Recently, efficiency of the CNN in image-like data classification was shown in [5], [7], [8]. Therefore, this work adopts

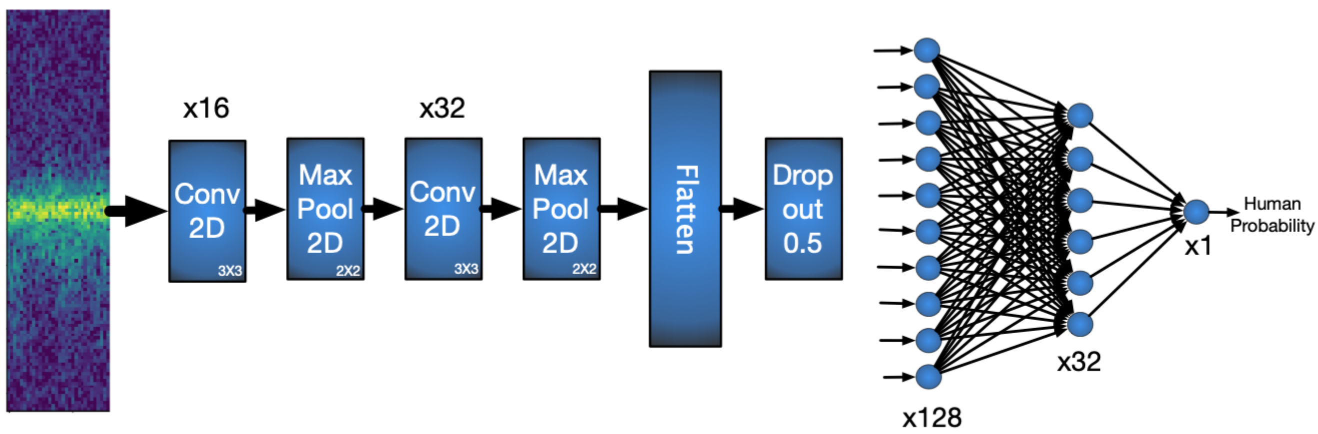

Fig. 3. 2-Layer CNN Architecture.

the CNN architecture for radar targets classification. Fig. 3 shows the proposed model architecture that consists of several DL building blocks: CNN layers, ReLu activation functions, Max Pooling, fully connected network, regularization techniques to reduce over-fitting: Dropout and $L2$ regularization, and learning rate scheduling. Over-fitting occurs when a model adapts too closely to the particular training data while losing its generalization ability. This work adopts regularization, data augmentation, and balancing methods to minimize the over-fitting.

A. *CNN Architecture*

The output from a convolutional layer is a 2D grid, where each grid element has a corresponding number of channels as the number of filters in the layer. The output is obtained by ”sliding” each filter over the layer's input and passing it through a non-linear activation function. Following the ReLu success in image processing [20], this work selected it for the hidden layers activation function.

A Flatten layer follows the convolutional layers. This layer reshapes the multidimensional 2D output into a 1D vector. We chose this vectorization instead of global pooling to empower the network to learn complex connections between the features extracted by the CNN. This vector is passed to a Dropout layer and then to a fully-connected NN with ReLu activation functions [4]. In order to represent the probability function, the activation function of the output layer is sigmoid.

Dropout [21] is a regularization technique of randomly dropping units during training in order to prevent them from co-adapting. Our architecture contains the Dropout right after the Flatten layer. The motivation for this Dropout position is that the following layer is dense with 97.5% of the network's trainable parameters.

B. *Training*

The criteria used for the evaluation of our model performance is the ROC-AUC. However, ROC-AUC is not a differentiable function, and therefore, can not be a loss-function for the network's learning. We selected the Log-Loss function, also denoted as the Binary-Cross Entropy: $L_\theta(X,Y) = -(Y\log(f(X;\theta)) + (1-Y)\log(1-f(X;\theta)))$.

The Binary-Cross Entropy loss contains the *L2 Regularization* term [22], which is commonly used to penalize large weights and biases by addition of a quadratic penalty term to the loss function. The Adam optimizer, an extension of the stochastic gradient descent that considers lower-order gradient moments estimation, was selected for weights adaptation [23].

The learning rate scheduling was used to prevent the model's weights from over-fitting. The learning rate is multiplied by 0.1 when the train set ROC-AUC exceeds the value of 0.99 (over-fitting phenomenon). The learning rate scheduling also contributes to the model's convergence since a lower learning rate localizes the optimizer to a particular area in the weight space.

Randomness is a conventional method to prevent the gradient descent algorithm from a convergence into a local minimum. After each epoch, the entire dataset is shuffled and randomly split into mini-batches. Data augmentation also contains some randomization V.

V. Data Augmentation

Data augmentation is conventionally used in DL network to prevent over-fitting. In addition, it introduces randomness into the dataset and contributes to its balancing. Data augmentation is mandatory in cases where only a small training data set is available. In these cases, CNN tends to over-fit to the available small data set and does not generalize to the unseen data.

This section summarises the data augmentation techniques that were evaluated for the addressed radar-based target classification problem. Some data augmentation methods, such as random frequency shift (RFS), were performed in the original time-domain. Others, such as flipping, noising, and random time shift (RTS), were performed at the pre-processing output, i.e., in the image domain. The following details all the evaluated data augmentation methods, ordered according to their efficiency, from the most efficient to the least efficient.

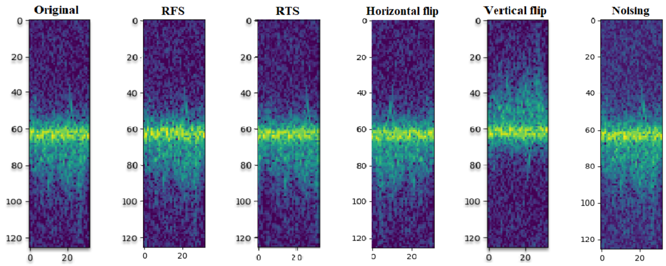

Fig. 4. Data augmentation examples.

*1) **Random Frequency Shift - RFS**:* This technique changes the image for each of the 32 sampled signals. i.e., each segment (in the time-domain, before any processing) is being transformed in the following way:

$$\widetilde{X}[n][m] = X[n][m]e^{j\frac{2\pi\alpha_m n}{N}}, \quad n \in [0,127],\ m \in [0,31],\ n,m \in \mathbb{N}. \tag{2}$$

where $\alpha_m \sim \mathcal{N}(0,\sigma^2)$ is the random shift, n and m are the fast and slow axis indices, respectively. This augmentation results in a random shift of α_m bins in the signal's spectral representation. If α_m is an integer, then each sampled signal at the IQ segment is cyclically shifted. If α_m is not an integer, then the original signal is interpolated in the frequency domain. The kinematic motivation behind this augmentation is to impose minor fluctuations to the image's original micro-Doppler patterns.

*2) **Random Time Shift - RTS**:* This augmentation performs a cyclic shift on the horizontal axis in the image domain:

$$\widetilde{X}^F[n][m] = X^F[n][((m-\alpha))_{32}], \quad n \in [0,127],\ m \in [0,31],\ n,m \in \mathbb{N}. \tag{3}$$

where $\alpha \sim Unif[S]$, $S = \{-L,...,L\} \setminus \{0\}, L \in \mathbb{N}^+$ and $((\cdot))_{32}$ is the 32 modulo operation.

*3) **Noising**:* This data augmentation adding a small random changes to the noised signal in the image domain:

$$\widetilde{X}[n][m] = X[n][m] + v, v \sim \mathcal{N}^c(0,\sigma^2), \quad n \in [0,127],\ m \in [0,31],\ n,m \in \mathbb{N}. \tag{4}$$

*4) **Flipping**:* Two flipping types were considered: vertical and horizontal. This augmentation takes a segment in the image domain and flips it vertically / horizontally. Intuitively, the vertical flip changes the target velocity direction, and the horizontal flip changes the target motion direction. Both augmentations are trivial since they do not affect the micro-Doppler signatures, and they provide only a factor 4 effect on the augmented dataset size.

VI. Target Classification Results

This section summarises the impact of data augmentation and model architecture on CNN learning and shows the classification performance in terms of AUC[1]. All algorithms, data processing, and metric evaluation were implemented with TensorFlow2 [24] library. The used hardware is NVIDIA 1080 Ti.

A. Testbench Dataset

The MAFAT data set [12] published in 2020 contains radar echo signal recordings of both humans and animals labeled objects at different operational environments. Each record contains a complex raw radar signal with its corresponding label, geographic location, signal to noise ratio (SNR) type, track ID, sensor type, etc. The data's physical characteristics, such as sampling frequency, carrier frequency, baseband bandwidth, etc., are unknown. The data set contains four different data groups: 1) train, 2) experiment, 3) synthetic, and 4) background [2].

Only the train data set contains the real-world recorded radar echoes thus, has the best data quality. Unfortunately, the train set is heavily imbalanced in terms of both target types and SNR levels. Therefore, a data balancing method is needed to prevent trained model bias [17].

The goal is to provide the maximum possible diversity in the dataset. Therefore, the dataset was stabilized by merging all segments from the training dataset with the synthetic dataset in a way that ensures the minimal number of segments from the same track. The provided test set does not have any mutual track recording with the train, experiment, and synthetic data sets. Therefore, this "Full Public" data set was used for the performance evaluation.

B. Data Augmentation Impact on Performance

TABLE I
Data Augmentation Ablation Test

Augmentation method	ROC-AUC score
None	0.8862
Vertical Flip	0.9108
Horiz Flip	0.9075
RTS	0.9255
RFS	**0.9270**
Noising	0.9183

The influence of different model-based data augmentation techniques over a 2-layer CNN model in Fig. 3 was evaluated. Table I shows the ROC-AUC for each evaluated data augmentation technique. The results were obtained by averaging the score over several independent training phases. "None" refers to an experiment in which only $15K$ target balanced samples were used for training without data augmentation. Table I shows that the random frequency shift augmentation outperformed the other techniques.

[1]The implemented models, experiments, and augmentations can be found at https://github.com/Shahaf-Yamin/Radar-Moving-Target-Classification-Via-CNN

[2]Further details can be found in [12]

C. Train Data Expansion

Table II shows the balanced-augmented train dataset size's influence on the ROC-AUC score, compared to duplicated dataset performance. The original unbalanced train set size was $6.5K$, and the balanced train set size was $15K$. Note that dataset balancing improves ROC-AUC score by **0.2**.

Data augmentation facilitates the model effort to generalize to unseen data. However, expanding the data with too many augmentations might degrade the performance. This degradation can be explained by the fact that too much augmentation from the same kind may lead the model to "memorize" specific samples. Table II shows that training with $400K$ slightly outperforms the $815K$.

TABLE II
ROC-AUC PERFORMANCE AS A FUNCTION OF BALANCED-AUGMENTED/DUPLICATED DATA SIZE

Size (K)	6.5	15	45	90	135	225	310	400	540	815
Bal. + Aug.	0.69	0.88	0.91	0.92	0.91	0.93	0.92	0.94	0.93	0.93
Dup.	0.69	0.69	0.74	0.71	0.69	0.72	0.71	0.73	0.77	0.75

D. CNN Depth

Fig. 5 shows the evaluation of several CNN architectures with various depths and layer sizes. The 2-layer CNN model outperformed all others, both in terms of ROC-AUC criteria and in over-fitting. Notice the over-fitting phenomenon in Fig. 5 at the 8-layer model. Although regularization techniques can reduce the larger network's over-fitting, good performance is also achievable with lower complexity CNN.

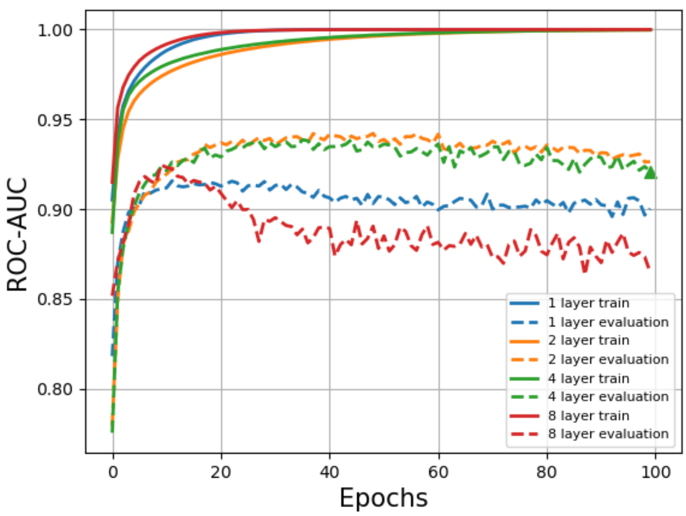

Fig. 5. ROC-AUC as a function of CNN layers.

E. Regularization Methods

The regularization methods and learning rate scheduling impact on the 2-layer CNN model's performance is evaluated. Fig. 6 shows the influence of these methods on over-fitting and Table III summarizes the final ROC-AUC results. Note that for the 2-layer CNN model without augmentation, the gap between validation loss and train loss increases over epochs and creates the over-fitting "fork". This trend means that the model learns to approximate the train set distribution while diverging from the evaluation set distribution. Moreover, inserting L2 regularization significantly reduces over-fitting.

Notice the learning rate scheduling impact without and with scheduling in green and yellow dashed lines, respectively. Both behave similarly until the point where the green line splits to create the over-fitting "fork".

Interestingly, the dropout after the Flatten layer reduces the over-fit even more since most of the network's weights are located in the dropout location (the reduced gap between the validation and train Loss).

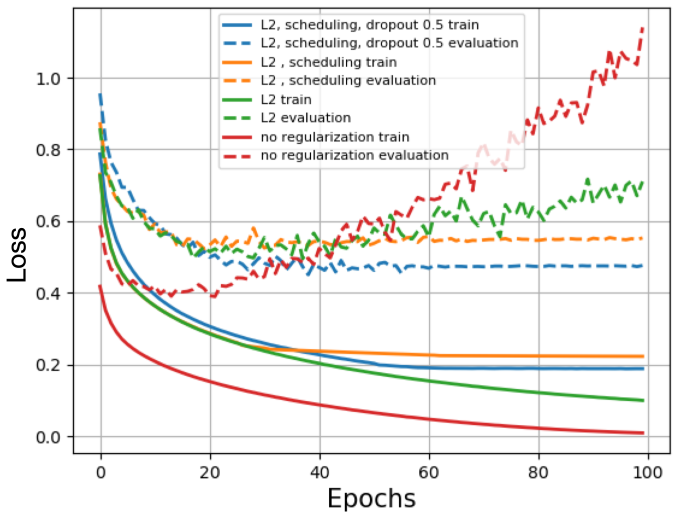

Fig. 6. Regularization methods influence on Loss.

TABLE III
ROC-AUC OF REGULARIZATION METHODS

	L2, Scheduling, Dropout	L2, Scheduling	L2	No Regularization
ROC-AUC	0.9381	0.9362	0.9264	0.9070

VII. CONCLUSIONS

This work proposed the CNN-based approach for the radar-based classification of humans and animals using micro-Doppler signatures. Several model-based data augmentation techniques were analyzed. The importance of the data balancing and data augmentation on the classification performance in the low-quality data scenarios was demonstrated. The importance of regularization techniques for over-fitting minimization was also demonstrated for relatively low-complexity networks with an augmented dataset.

References

[1] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 1993–2016, 2017.

[2] V. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, 2001.

[3] I. Bilik, J. Tabrikian, and A. Cohen, "Gmm-based target classification for ground surveillance doppler radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 267–278, 2006.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[5] A. e. a. Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[6] S. F. Dodge and L. J. Karam, "A study and comparison of human and deep learning recognition performance under visual distortions," *CoRR*, vol. abs/1705.02498, 2017. [Online]. Available: http://arxiv.org/abs/1705.02498

[7] C.-Y. Lee, P. W. Gallagher, and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," 2015.

[8] K. e. a. He, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] L. et al., "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[11] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.

[12] "Mafat competition website," 2020. [Online]. Available: https://competitions.codalab.org/competitions/25389

[13] J. L. et al., "Target classification based on micro-doppler signatures," in *IEEE International Radar Conference, 2005.*, 2005, pp. 179–183.

[14] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE T. Geoscience and Remote Sensing*, vol. 47, pp. 1328–1337, 05 2009.

[15] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18 484–18 501, 2018.

[16] E. Al Hadhrami, M. Al Mufti, B. Taha, and N. Werghi, "Transfer learning with convolutional neural networks for moving target classification with micro-doppler radar spectrograms," in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*. IEEE, 2018, pp. 148–154.

[17] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, p. 20–29, Jun. 2004. [Online]. Available: https://doi.org/10.1145/1007730.1007735

[18] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recogn.*, vol. 30, no. 7, p. 1145–1159, Jul. 1997. [Online]. Available: https://doi.org/10.1016/S0031-3203(96)00142-2

[19] I. Bilik and J. Tabrikian, "Radar target classification using doppler signatures of human locomotion models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1510–1522, 2007.

[20] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[21] N. e. a. Srivastava, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization," in *Advances in neural information processing systems*, 1992, pp. 950–957.

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

[24] M. A. et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/