Adaptive Target Enhancer: Bridging the Gap between Synthetic and Measured SAR Images for Automatic Target Recognition

Alexandre B. Campos*, Ricardo D. Molin Jr.*, Lucas P. Ramos†, Renato Machado†, Viet T. Vu‡, and Mats I. Pettersson‡

*Microwaves and Radar Institute, German Aerospace Center (DLR), Wessling, Germany.
Email: {alexandre.campos, ricardo.dinizdalmolinjunior}@dlr.de
†Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos, Brazil.
Email: {lucaspr, rmachado}@ita.br
‡Department of Mathematics and Natural Sciences, Blekinge Institute of Technology (BTH), Karlskrona, Sweden.
Email: {viet.thuy.vu, mats.pettersson}@bth.se

***Abstract*—Automatic target recognition (ATR) algorithms have been successfully used for vehicle classification in synthetic aperture radar (SAR) images over the past few decades. For this application, however, the scarcity of labeled data is often a limiting factor for supervised approaches. While the advent of computer-simulated images may result in additional data for ATR, there is still a substantial gap between synthetic and measured images. In this paper, we propose the so-called adaptive target enhancer (ATE), a tool designed to automatically delimit and weight the region of an image that contains or is affected by the presence of a target. Results for the publicly released Synthetic and Measured Paired and Labeled Experiment (SAMPLE) dataset show that, by defining regions of interest and suppressing the background, we can increase the classification accuracy from 68% to 84% while only using artificially generated images for training.**

***Index Terms*—Adaptive filtering, automatic target recognition (ATR), MSTAR, SAMPLE, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) technology has witnessed a remarkable growth in the past few decades [1]. SAR systems have been successfully employed in a wide range of applications, such as target detection and classification in conflict zones or monitoring illegal activities along borders [2]. In such scenarios, automatic target recognition (ATR) algorithms are often used to assist the decision-making process.

While a massive amount of SAR data is continuously produced, the potentially classified nature of the radar signatures from military vehicles arises as a challenge since the most successful supervised classifiers depend on large and varied databases, as well as reliable labels. To mitigate these problems, SAR image simulators can be used for expanding or even replacing small or outdated datasets [3]–[6].

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES-Brazil)—Finance Code 001 (Pró-Defesa IV), and by Brazilian National Council for Scientific and Technological Development (CNPq-Brazil). The authors also thank the Brazilian Institute of Data Science (BI0S), grant 2020/09838-0, São Paulo Research Foundation (FAPESP).

Several databases have been artificially expanded with computer-aided drawing (CAD) modeling, which has been shown to improve classification performance by adding variability to the training. The well-known Moving and Stationary Target Acquisition and Recognition (MSTAR) database [7] was partially recreated through the recently released Synthetic and Measured Paired and Labeled Experiment (SAMPLE) database [8], which introduces the challenge of maintaining a satisfactory classification performance even when more significant fractions of computer-simulated (synthetic) data are used for training. In [8], a baseline ATR algorithm based on convolutional neural networks (CNNs) was considered to assess the proposed challenge without any pre-processing, and it has been verified that the classification performance decreases drastically as synthetic samples are used for training instead of the original data. Following up this experiment, an approach that pre-processes the images based on variance-based joint sparsity (VBJS) despeckling, quantization, and clutter transfer is used in [9] to tackle the measured and synthetic domain mismatch problem. Another solution has been proposed in [10] to map a common domain between real and synthetic samples through a small number of matching images from both sets. Though these techniques have been shown to significantly increase the classifier's performance, they must rely on *a priori* information of the targets, i.e., measured SAR images with known and well-defined labels.

In this paper, we propose the so-called adaptive target enhancer (ATE) for pre-processing synthetic and measured SAR images for ATR. The method aims to detect and enhance targets in these images through a combination of adaptive thresholding, morphological operations, pixel weighting, and a median filter. After both the synthetic and measured images are transformed, a single classifier can be trained with only the synthetic data to predict the outcome class of measured SAR images. Results show that the classification accuracy significantly increases when the proposed pre-processing is applied, endorsing the use of synthetic images for training.

The remainder of this paper is organized as follows. Section II introduces the ATE technique. In Section III, the SAMPLE challenge and related works are presented. The performance of the proposed method is then assessed and discussed in Section IV. Finally, Section V summarizes the main conclusions and presents some final remarks.

II. Proposed model

A. Problem statement

Consider two sets of SAR images, $\mathbf{D}_m$ and $\mathbf{D}_s$. The former is composed of measured SAR images acquired in a real-world scenario. The latter is composed of artificially generated SAR images that share common characteristics with the measured images but were created without a measurement campaign. We can define a classifier C_m that predicts the outcome class of each observation within $\mathbf{D}_m$ in a way that $C_m : \mathbf{D}_m \rightarrow \mathbf{y}$, where $\mathbf{y}$ is the vector of classes associated with each observation. Analogously, for the set $\mathbf{D}_s$, we can define $C_s : \mathbf{D}_s \rightarrow \mathbf{y}$. However, the availability of measured SAR images is often an issue. Hence, we want to estimate a classifier $\hat{C}$ that is able to predict the vector of outcome classes $\mathbf{y}$ for a set of measured SAR images while training only on a set of synthetic SAR images.

We propose to pre-process the images of each set in an unsupervised way by employing the ATE technique, as shown in Figure 1. The same ATE process runs independently for the two image sets, $\mathbf{D}_s$ and $\mathbf{D}_m$, and aims to transform them into the sets $\mathbf{D}_s^*$ and $\mathbf{D}_m^*$, respectively, to mitigate the gap between the synthetic and measured images. Hence, a robust and unique classifier $\hat{C}$ can be achieved by training only on the transformed synthetic data $\mathbf{D}_s^*$.

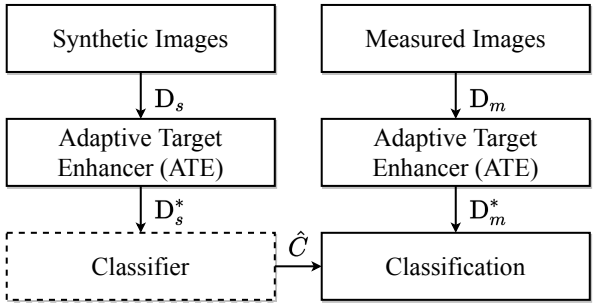


Fig. 1. Proposed scheme for ATR. The ATE step creates two new sets, $\mathbf{D}_s^*$ and $\mathbf{D}_m^*$. The former is used for training a classifier, which is represented by the dash-lined block. When the training is complete, the resulting classifier $\hat{C}$ is used for evaluating the set $\mathbf{D}_m^*$.

B. Adaptive Target Enhancer (ATE)

The ATE can be described as a pre-processing technique composed of an adaptive thresholding step, morphological operations, pixel weighting, and a median filter. The complete ATE flowchart can be seen in Figure 2.

The thresholding step can be described as follows: i) vectorize the SAR image; ii) standardize (subtract the mean and divide by the standard deviation) the resulting vector; iii) create a new vector that contains only the pixels whose amplitude is lower than a threshold constant λ; iv) repeat step (ii) and (iii) for each new vector created from the original data until all pixels are within the range $(0,\ \lambda)$; v) reconstruct the image and assign each pixel that surpassed the threshold λ at any given iteration as '1', setting the remaining ones as '0'.

After thresholding, several clusters of connected pixels are typically scattered through the image. Two morphological operations are then used to define a single target structure. First, an operation of closing (a dilation followed by an erosion) is applied with a fixed structuring element (SE) to ensure that pixels close to each other can form larger objects. Second, an operation of opening (an erosion followed by a dilation) is used to eliminate noisy pixels. In this case, since we are interested in a single target structure, the SE is defined for each image according to the maximum number of locally connected pixels. The proposed morphological operations are based on the assumption that a target is reported and will be the most prominent object in the image. Please note that this assumption is valid for any given target shape at any given position within the image.

Once the target is well defined, regions surrounding it are considered for gradually weighting the image pixels. Pixels closer to the target structure are weighted higher than those far from the target's influence zone. Additionally, the background is not entirely excluded as it can still provide insightful information for the feature extraction process. In total, we define four possible regions:

1) Target region (S_1): Embraces the pixels detected by the threshold and morphological operations steps, and aims to report the pixels that most likely belong to a target;
2) Border region (S_2): Includes the immediate surroundings of S_1 and captures information about pixels that may have a partial contribution from the target, such as the vehicle's edges;
3) Neighboring region (S_3): Comprises the pixels external to $S_2$and aims to capture any remaining useful information, such as the vehicle's shadows;
4) Background region (S_4): The most external zone, composed of the remaining pixels of the image.

For every pixel of an image $\mathbf{I}$, a function $W(\sigma_p)$ will be used to weight its amplitude σ_p in regards of its region:

$$W\left(\sigma_p\right) = \begin{cases} \alpha_1 \times \sigma_p, & \text{if the pixel is within } S_1; \\ \alpha_2 \times \sigma_p, & \text{if the pixel is within } S_2; \\ \alpha_3 \times \sigma_p, & \text{if the pixel is within } S_3; \\ \alpha_4 \times \sigma_p, & \text{if the pixel is within } S_4. \end{cases} \tag{1}$$

Please note that to enhance the pixels closest to the target structure, we need to ensure that $\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4$. Figure 3 illustrates an example of a mask generated by the pixel weighting step. Finally, a median filter is applied to the weighted image to further reduce the influence of noise within the enhanced regions.

III. The SAMPLE Challenge

A. Image database

The experiments presented in this paper have been conducted on high-resolution SAR images originally acquired by

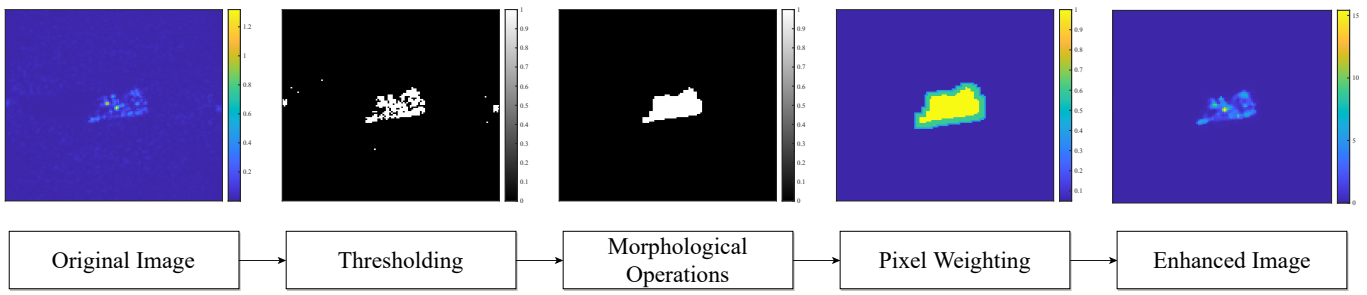


Fig. 2. The proposed ATE method. An image, either measured or synthetic, is thresholded in an adaptive way to capture the target structure, but other bright pixels are also detected. Morphological operations are employed to preserve only the dominant structure, i.e., the target. Pixel weighting is applied to ensure that pixels close or within the target location are considered to a greater extent than background pixels. Finally, the mask is applied to the original image, and the target is enhanced after a median filter is used.

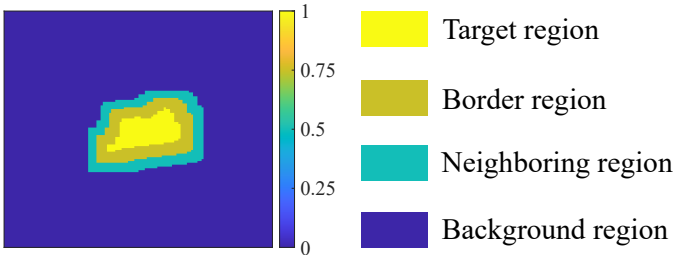


Fig. 3. Illustration of the mask generated by the pixel weighting step. Each of these non-overlapping regions is associated with a different multiplier.

the Air Force Research Laboratory (AFRL) and Defense Advanced Research Projects Agency (DARPA) between 1995 and 1997. This project resulted in the MSTAR dataset, available in [7], and widely used for testing ATR techniques. However, since this database is relatively small and the images share several characteristics – like radar operating parameters and target articulation – ATR classifiers may fail to generalize. Aiming to overcome this issue, a portion of the data has been artificially recreated as part of the SAMPLE [8] database, which is based on CAD models.

The SAMPLE database consists of 1366 paired images from 10 vehicle classes, where each pair is composed of a measured X-band SAR image from the MSTAR dataset and its corresponding computer-simulated one. Each image has 128×128 pixels with an element spacing of 0.2×0.2 m and a spatial resolution of 0.3×0.3 m. The data have been sampled in azimuth from $10°$ to $80°$ and in elevation from $14°$ to $17°$. The vehicle classes are described in detail according to the number of available images in Table I, while a visual comparison between measured and synthetic data can be seen in Figure 4.

Among the several challenges that have arisen with the SAMPLE dataset, a critical one is replacing measured SAR images with their synthetic counterparts with minimal performance loss. The experiments are thus conducted for different proportions of real data used for training. The metric

$$k = \frac{\text{Number of measured images}}{\text{Total number of images}}, \tag{2}$$

is employed to express the percentage of real data used in the experiments.

TABLE I
DESCRIPTION OF VEHICLE CLASSES CONSIDERED IN THIS PROBLEM.

Class	Measured	Synthetic	Type	Traction
2S1	177	177	Tank	Tracked
BMP2	108	108	Tank	Tracked
BTR70	96	96	Tank	Wheeled
M1	131	131	Tank	Tracked
M2	129	129	Tank	Tracked
M35	131	131	Truck	Wheeled
M548	129	129	Truck	Tracked
M60	178	178	Tank	Tracked
T72	110	110	Tank	Tracked
ZSU23	177	177	Tank	Tracked

B. Related works

Along with the SAMPLE database, a baseline classifier based on CNNs – which will be further discussed in Section IV – is also proposed in [8]. Since it has been shown through their experiments that the classification accuracy is severely affected as k decreases, a series of investigations has been devoted to address this issue.

The benefits of adapting the data for increasing the classification performance when k decreases were shown in [9], where the synthetic images are pre-processed in three steps. First, a VBJS technique is used to remove speckle noise. Second, a quantization technique is applied to the target region. Finally, the background of the synthetic images is replaced by the background of the equivalent measured images. Additionally, they considered a modified DenseNet classifier instead of the baseline CNN from [8]. An accuracy of over 90% was reported even for smaller values of k, achieving an accuracy of around 84% for the specific case when $k = 0$, i.e., when all of the images used for training are synthetic.

Another approach to handle the inherent differences between simulated and measured data was proposed in [10]. The so-called matching component analysis (MCA) aims to identify a common domain between training and test sets, here represented by synthetic and measured images, respectively. To that end, they define a subset of labeled samples from the test domain and match them with points of the original training set. After the MCA is run to find affine linear mappings to a common domain, a k-nearest neighbor (kNN) classifier is used by training with synthetic images and testing on measured

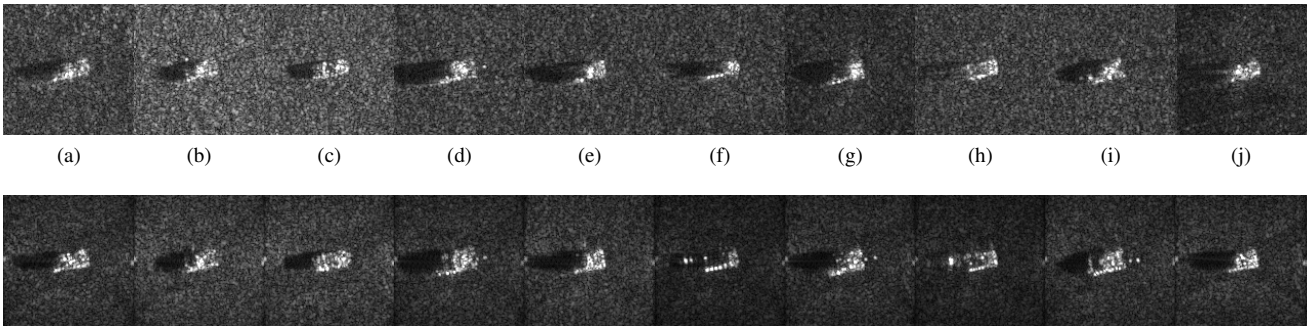


Fig. 4. Ten possible vehicle classes within the MSTAR dataset. (a)-(j): 2S1, BMP2, BTR70, M1, M2, M35, M60, M548, T72, and ZSU23 vehicles represented by measured images, respectively. (k)-(t): Corresponding synthetic images of the vehicles.

ones. With a matching set of 100 samples – which can be seen as an effective k of approximately 0.1 – an accuracy of 87% was achieved.

It is important to highlight some drawbacks of both approaches. In [9], the use of clutter information from the measured images implies the necessity of having access to real data for classification, even when k is low. This also suggests that their performance is heavily dependent on the terrain in which the target is placed. Furthermore, the complexity of the approach is increased, given the VBJS computational cost and the DenseNet classifier. In [10], although a simpler classifier is used, their feature extraction process relies on selecting a number of matching samples composed both of the target signature and the background clutter. This implies that access to real data is still required. Therefore, the challenge of successfully classifying measured images of the SAMPLE dataset when only using information of synthetic images for training remains open.

IV. Experimental Results

A. Experimental setup

To demonstrate the effectiveness of the ATE technique, we implemented the same classifier as proposed in [8], a CNN with four convolutional layers and four fully connected layers. The details of this architecture are outlined in Table II. The CNN was implemented in MATLAB, using the Adam optimization algorithm with a learning rate equal to 1^{-3}. We train the algorithm using batch sizes of 16 images for 60 epochs, considering only samples with elevation angles between $14°$ and $16°$, leaving data collected at a $17°$ elevation angle for testing. The images were evenly cropped on the sides to a size of 64×64 pixels each, in accordance with [8]. During training, 15% of the images were used for validation.

For the baseline, we implemented the network of Table II without pre-processing any of the images. The parameters used for the ATE results are disposed in Table III. To define the target region S_1, the thresholding constant λ should be chosen carefully. If it is set too high, potential target pixels can be missed, and the morphological operations will not be able to fully reconstruct the target. On the other hand, if it is set too low, the closing operation can impact the target's shape. Values in the range of $\lambda = 2$ to $\lambda = 6$ were tested, as well as several SEs for the closing operation. As for the extension of the non-overlapping surrounding regions S_2 and S_3, we defined their coverage radius as an expansion of six and four pixels, respectively. Both the weight values and the extent of the regions were defined empirically. For the specific case of the SAMPLE dataset, we found that the most critical parameter was the α_4 weight. This was expected given the major discrepancy between the backgrounds of the measured

TABLE II
Description of the implemented CNN.

Layer	Dimension	Activation
Input	$64 \times 64 \times 1$	
Convolutional Max-Pooling	$64 \times 64 \times 16$ $32 \times 32 \times 16$	ReLU
Convolutional Max-Pooling	$32 \times 32 \times 32$ $16 \times 16 \times 32$	ReLU
Convolutional Max-Pooling	$16 \times 16 \times 64$ $8 \times 8 \times 64$	ReLU
Convolutional Max-Pooling	$8 \times 8 \times 128$ $4 \times 4 \times 128$	ReLU
Flatten	2048	
Fully connected	1000	ReLU
Fully connected	500	ReLU
Fully connected	250	ReLU
Fully connected	10	ReLU

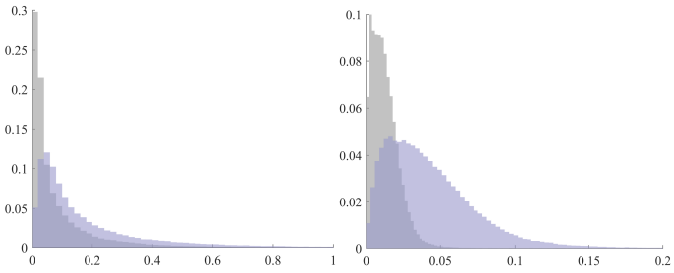
TABLE III
Parameters used for the experiments.

Processing step	Parameters	Adopted value
Thresholding	λ	5
Morphological operations	Closing SE Opening SE	5×5 Adaptive
Pixel weighting	α_1 α_2 α_3 α_4	1.00 0.60 0.40 0.05
Median filter	Window size	3×3

and synthetic images, reported in [9] and further discussed in Section IV-B. Regarding the coverage radius, narrower surrounding regions were also better than larger ones.

B. Data analysis

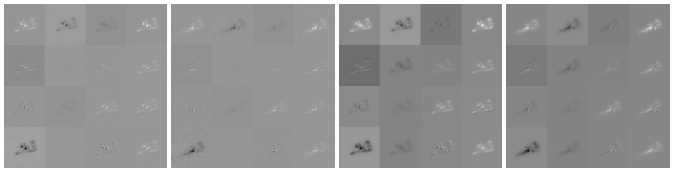
We first look into the discrepancy between the background of the synthetic and measured images of the SAMPLE dataset, shown in Figure 5. In this analysis, we extracted the pixels associated with a vehicle in every single image of both synthetic and measured domains. The remaining pixels are considered to be part of the image background. While Fig. 5(a) shows that similar statistical behavior is expected for both the synthetic and measured images' pixels within the target region, Fig. 5(b) suggests that the data distribution is very different when comparing the background pixels of both domains. This result reinforces the need to adapt or suppress the image background to increase the similarity between the domains.



(a) (b)

Fig. 5. Histogram of the set of pixels associated with the (a) target and (b) background region of the images. The data from the synthetic and measured images are represented in gray and blue, respectively.

In Figure 6, we show the activation of the feature maps of the network's first convolutional layer with and without considering the ATE technique. The network is trained only with synthetic data in both cases. The ATE pre-processing enhances the target structure within the image while suppressing the image background, which directly affects the feature extraction process. With the proposed pre-processing, edges and basic structures are now more visible in the filters.



(a) (b) (c) (d)

Fig. 6. Feature maps generated on the first convolutional layer: (a) synthetic and (b) measured images without using the proposed ATE method, (c) synthetic and (d) measured images with the ATE method.

C. Classification performance

We first compare the classification results when employing the CNN of Table II with and without using the proposed ATE step. In this paper, we focus solely on the case when $k = 0$, i.e., when all of the images used for training are artificially generated. We used the following metrics:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{F-measure} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \tag{5}$$

where TP, FN, and FP stand for true positive, false negative, and false positive, respectively.

The normalized prediction outcomes are shown in Figure 7 as confusion matrices for both cases. Please note that the main diagonal of each confusion matrix denotes the recall per class. The overall results for the baseline (Fig. 7(a)) and when using the proposed ATE (Fig. 7(b)) are summarized in Table IV.

TABLE IV
CLASSIFICATION PERFORMANCE.

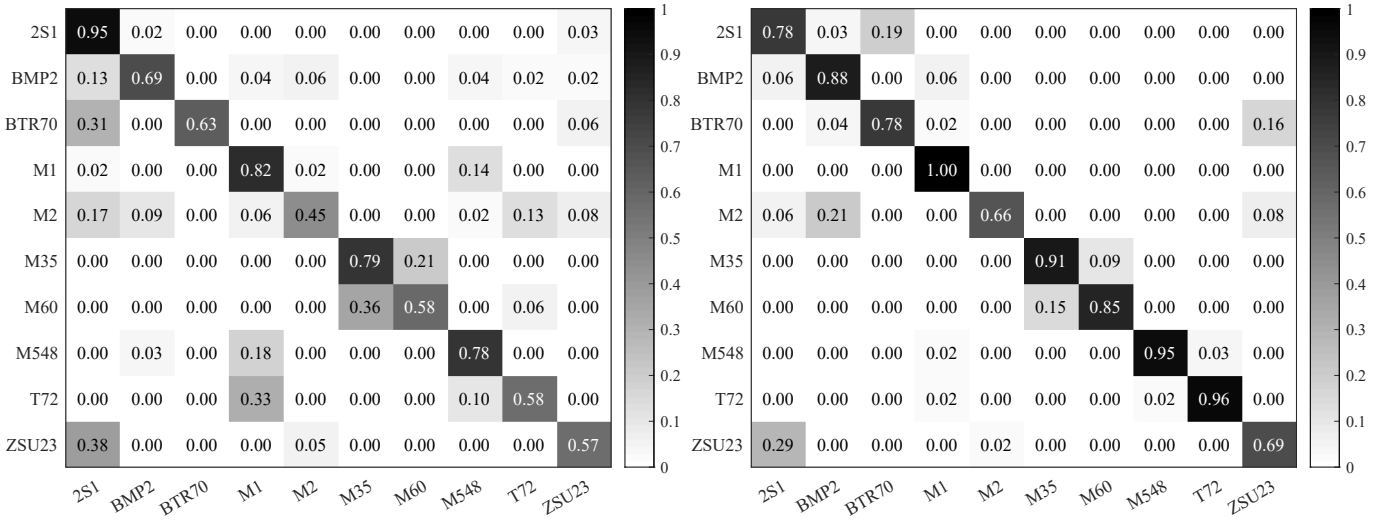
Method	Recall	Precision	F-measure
CNN (baseline)	68.6%	73.0%	70.7%
ATE + CNN (proposed)	84.5%	85.3%	84.9%

V. CONCLUSIONS

This paper proposes an easy-to-implement preprocessing technique to bridge the gap between synthetic and measured SAR images for ATR algorithms. The method delimits a target structure through adaptive thresholding and morphological operations steps, then defines regions surrounding its zone of influence. Each region is multiplied by a weight, and the findings presented in this paper suggest that the classification performance can be improved if the background region is suppressed. A possible explanation for this outcome is the mismatch between the environmental conditions of the synthetic and measured images, which has been reported in some related works and further investigated in this paper. The method provides a considerable gain in accuracy compared to the baseline classifier and performs similarly to the approaches that used measured data. Future works should address the possible implications of limiting the feature extraction process to only the target's structure, investigating ways of automatically tuning the weights for extracting the most useful information for other datasets.

ACKNOWLEDGMENTS

This work was supported in part by the Brazilian Agencies National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES) – Finance Code 001 (Pró-Defesa IV), and the BI0S - Brazilian Institute of Data Science, grant 2020/09838-0, São Paulo Research Foundation (FAPESP).



(a) (b)

Fig. 7. Confusion matrix, normalized by the number of images per class: (a) without using the proposed ATE and (b) using the ATE method.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A Tutorial on Synthetic Aperture Radar," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[2] W. L. Melvin and J. A. Scheer, *Principles of modern radar: radar aplications. Vol. 3*. SciTech Publishing, 2014.

[3] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic Aperture Radar Image Synthesis by Using Generative Adversarial Nets," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1111–1115, 2017.

[4] C. Mao, L. Huang, Y. Xiao, F. He, and Y. Liu, "Target recognition of SAR image based on CN-GAN and CNN in complex environment," *IEEE Access*, vol. 9, pp. 39 608–39 617, 2021.

[5] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 3, pp. 56–81, 2021.

[6] J. Qin, Z. Liu, L. Ran, R. Xie, J. Tang, and Z. Guo, "A Target SAR Image Expansion Method Based on Conditional Wasserstein Deep Convolutional GAN for Automatic Target Recognition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7153–7170, 2022.

[7] AFRL, "MSTAR overview," https://www.sdms.afrl.af.mil/index.php?collecti on=mstar, 1995.

[8] B. Lewis, T. Scarnati, E. Sudkamp, J. Nehrbass, S. Rosencrantz, and E. Zelnio, "A SAR dataset for ATR Development: the Synthetic and Measured Paired Labeled Experiment (SAMPLE)," in *Algorithms for Synthetic Aperture Radar Imagery XXVI*, vol. 10987. International Society for Optics and Photonics, 2019, p. 109870H.

[9] T. Scarnati and B. Lewis, "A deep learning approach to the Synthetic and Measured Paired and Labeled Experiment (SAMPLE) challenge problem," in *Algorithms for Synthetic Aperture Radar Imagery XXVI*, vol. 10987. International Society for Optics and Photonics, 2019, p. 109870G.

[10] C. Clum, D. G. Mixon, and T. Scarnati, "Matching Component Analysis for Transfer Learning," *SIAM Journal on Mathematics of Data Science*, vol. 2, no. 2, pp. 309–334, 2020.