Explainability of Deep SAR ATR Through Feature Analysis

Carole Belloni*, Nabil Aouf‡, Alessio Balleri*, Jean-Marc Le Caillec†, and Thomas Merlet§. *Centre for Electronic Warfare, Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA. †IMT Atlantique, Brest, France. ‡City University of London, United Kingdom. §Thales LAS France SAS , Elancourt, France.

Abstract—Understanding the decision-making process of deep learning networks is a key challenge which has rarely been investigated for Synthetic Aperture Radar (SAR) images. In this paper, a set of new analytical tools is proposed and applied to a Convolutional Neural Network (CNN) handling Automatic Target Recognition (ATR) on two SAR datasets containing military targets.

Firstly, an analysis of the respective influence of target, shadow and background areas on classification performance is carried out. The shadow appears to be the least used portion of the image affecting the decision process, compared to the target and clutter, respectively.

Secondly, the location of the most influential features is determined with classification maps obtained by systematically hiding specific target parts and registering the associated classification rate (CR) relative to the images to be classified. The location of the image areas without which classification fails is target type and orientation specific. Nonetheless, a strong contribution of specific parts of the target, such as the target top and the areas facing the radar, is noticed.

Lastly, results show that features are increasingly activated along the CNN depth according to the target type and its orientation, even though target orientation is absent from the loss function.

Index Terms—Deep Learning, SAR, ATR, Explainability, Features.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) data presents multiple advantages over electro-optical data. Meaningful SAR data can be collected from long distances under a wider range of weather conditions and at night. However, the interpretation of SAR images is challenging and specialised operators are required to extract the correct information and provide target classification. To reduce this workload and to speed up the recognition process, Automatic Target Recognition (ATR) can be applied to SAR data. ATR is indeed key for current and future military requirements [1].

ATR methods based on robust and well-known image descriptive features have been developed. Such features could consist of Krawtchouk Moments, Pseudo-Zernike moments or Scale-Invariant Feature Transform (SIFT) [2]-[5]. They tend, however, to provide lower classification performances than deep learning methods.

Recent breakthroughs in deep learning techniques applied to several technical domains make deep learning a valuable choice for ATR application. Artificial intelligence algorithms developed have been used and outperform classical feature methods in the SAR domain for several applications, such as classification, matching or change detection [6]-[8]. The current state of the art on SAR ATR methods evaluated on the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset [9] rely heavily on Convolutional Neural Networks (CNNs). Some neural networks have been specifically developed for SAR applications and are shallower than those used in the visual domain [10]-[12]. Deep learning networks have been also used to simulate realistic SAR synthetic images with Generative Adversarial Network (GANs) [13]-[15].

Unlike feature based models, which often extract man-made target features, features used by neural networks are created using artificial intelligence concepts. Earlier work addressing neural networks for ATR on SAR amongst other data exists, dating back to the 1990s and became more popular around the 2010s particularly with the introduction of CNNs [10], [16]. Deep learning features are quite complex, being the result of stacked convolutions and activations. This complexity makes it difficult to understand which information triggers a CNN classification decision.

Unlike classical features, deep learning features cannot be easily improved or even understood by humans, especially when features are generated by the deepest layers of the network [17]. Knowing and explaining the origin of CNN decisions is a very important problem that, if solved, would help to choose the most rational network among several solutions. As a result, deep networks would be more trustworthy and could take advantage of the impressive human visual understanding once the networks decision process matched the human reasoning and be less biased by the training dataset composition [18] . The understanding of the internal operations of deep learning solutions is a new research area and is essential to further improve deep learning methods, to validate them over former techniques and to increase the level of trust required for real-life deployment. Several approaches to explainability have been proposed in the visual domain with, for example, deconvolutional networks enabling the visualisation of high level features selected by the deep network [19], [20], the analysis of the role of features for each class respectively [21], the influence of training data selection over specific misclassifications [22]. However, recognising specific patterns is already challenging in the visual domain even if humans are used to it and is even more difficult for SAR images. Indeed, an untrained person is not able to distinguish different targets

in SAR images. In any case, as SAR images are based on vastly different physical processes than visual images, SAR and visual features are likely to be different and contribute differently to the network decisions. Deep learning network understanding in the SAR domain currently remains limited to the visualisation of low level deep learning features [23]-[25].

This paper aims at providing more detailed explanations about the decision process of a CNN regarding SAR ATR. Thus, three innovative streams of analysis of a trained deep learning network are proposed:

- Firstly, the individual contribution to classification performance of different image parts, respectively the target, the shadow and the background is assessed. SAR target classification relying only on the target shadow has already been investigated and results showed that the addition of features from the shadow can improve classification performance [26], [27]. Current SAR ATR algorithms are fed with full or segmented target images. The extent of the information loss incurred by target segmentation and thus the removal of the shadow and multipath should be investigated. This is achieved with a novel classification analysis technique using masked images and by studying the presence of critical features in each image zone as presented in Section III-B.

- Secondly, the location of the areas containing features critical to achieve correct classification is investigated and compared to the location of unessential features in Section III-C. Similar investigations, in the visual domain, studying the areas of images contributing to classification showed that, in the case of a biased dataset in which the tested CNN could base the differentiation between wolves and huskies by the presence of snow and grass, respectively, instead of physical characteristics [28]. For the study of critical feature location, occlusion maps are created to highlight the location of SAR features essential for correct classification by the tested deep learning method [19]. Guided backpropagation is another solution to visualize the patterns learned by the CNN to characterize an object [29], [30]. Occlusion maps are, in this paper, extended to novel classification maps to analyse a group of images rather than a single image. Applying this method on a group of images with common environmental factors clarifies the factors role on the choice of features learned by the CNN. After describing the generating process of classification maps, these maps are used to determine the location of features critical for classification of targets of a specific class or of targets within a specific orientation range. Results achieved are compared for a well-trained network and a less performing network, which did not benefit from data augmentation during training. Results show, amongst other things, the role of the target turret on classification performance. Such information could help isolate target areas robust to classification and inspire greater confidence for algorithms assisting operator decisions.

- Lastly, Section III-D investigates the specificity of features, that is how much distinct are the patterns that activate the corresponding trained filters. The network tends to develop features to be sensitive to a specific target class. A feature specific to a target class would not be activated by patterns issued from other target classes. On the contrary, a non-specific feature could be activated with any image analysed. In Section III-C, the specificity of features to the target class and to the target orientation is examined. The features investigated for specificity are the features activated the most when the network is presented with images of a target of a certain category or with an orientation in a particular range. The objective is to evaluate the power of discrimination of the network against target classes and target orientations along the network depth as the computed features grow in complexity. The specificity of the features is shown for both a high-performing network and a less performing one that did not benefit from a data augmented training. The specificity is shown using a proposed histogram feature representation on a specifically chosen group of images. The evaluation of the quality of such features is important as the same features could be used again by other SAR detection or classification algorithms.

The next section presents the datasets. Section III introduces the CNNs investigated and shows the analytical tools proposed. The associated results are presented in section IV.

II. DATASET

The analyses are carried out on two publicly available SAR ATR dataset. The first dataset is the Moving and Stationary Target Acquisition and Recognition (MSTAR) in Standard Operating Condition (SOC) which consists of 10 targets and that was developed by the U.S. Defense Advanced Research Projects Agency (DARPA) and the U.S. Air Force Research Laboratory (AFRL) [9], [31]. This database was collected under Horizontal Horizontal (HH) polarization in X-Band with a $30\,\mathrm{cm}\times30\,\mathrm{cm}$ resolution. There are ten targets in this dataset: a bulldozer (D7), a truck (ZIL), a rocket launcher (2S1), an air defence unit (ZSU), armoured personnel carriers (BRDM2, BTR60, BTR70, BMP2) and tanks (T62, T72). The images in the training and testing set are taken with a depression angle of respectively $17°$ and $15°$.

The second dataset is the Military Ground Target Dataset (MGTD), generated by our research group at Cranfield University [32], [33]. The emitted signal spans the frequency range from $13\,\mathrm{GHz}$ to $18\,\mathrm{GHz}$ to achieve a resolution of $3\,\mathrm{cm}$ in range and $3.3\,\mathrm{cm}$ in cross-range on model targets of 1.5-$1.7\,\mathrm{m}$. Single polarised images (Horizontal-Horizontal) are generated using a backprojection algorithm. This dataset contains three targets: a T64, a T72 and a BMP1. The training and testing sets are formed using different target configurations with depression angles ranging between $21.8°$-$23.4°$ for training and between $17.5°$-$20.3°$ for testing respectively, and using different laboratory backgrounds. For each target, 4 sequences of 72 images each are allocated for training and 4 other sequences acquired under different conditions are allocated for testing [32]. The target configurations consist of different turret orientations ($-90°$, $-45°$, $45°$, $90°$ for training and $-30°$, $0°$, $30°$ for testing). The gun is pointed up and down independently from training and testing. The data was acquired at three different times, with a different laboratory background. Series acquired at the same time period are either used in the training or testing set but not mixed.

In both cases, the proposed analyses are carried out on the resulting magnitude images.

III. METHODS

A. Deep learning networks evaluated

The evaluated Convolutional Networks (CNNs) are based on the AlexNet network with a new last fully connected layer fitting the number of output classes in the corresponding database [34]. This architecture is successful for numerous applications such as pose estimation, video classification or semantic segmentation [35]-[37]. It benefits from a straightforward architecture compared to ResNets with a limited number of weights compared to VGGNet [38], [39]. It is selected for ease of implementation and to benefit from the availability of pre-trained models.

These networks are trained using transfer learning from the visual domain to the appropriate SAR database. Training is achieved with a small learning rate for the first layers and a larger one starting from the 4th convolutional layer. The learning rate is specific to each dataset and determined through a random grid search. The main location of automatically chosen features, without which good CR cannot be achieved, are compared for two CNNs in order to evaluate what makes good target features. One CNN is trained with data augmentation (translation for the MSTAR SOC10, translation and additional synthetic Weibull noise based for the MGTD [40]) and one CNN is trained without any data augmentation. As the translation data augmentation can be applied on the fly, all of the targets are randomly translated at each epoch. However, the Weibull noise data augmentation requires to noise the range profiles before getting the augmented images through backprojection. Four different noise distribution parameters are chosen amongst typical SAR target detection SNRs to complement the original images, making 80% of the total training data.

In Sections III-B and III-D, the images are not pre-processed. The CNN trained with data augmentation achieves 98.17% on the MSTAR SOC10 and 92.47% on the MGTD while, without data augmentation, it only achieves 95.51% and 78.53%, respectively. In Section III-C, the training and testing images in the MSTAR dataset are rotated and translated so that all targets are in the same position, centered and facing the same direction. The position information is provided by the reference target segmentation [41]. The images in the MGTD are only rotated as no segmentation information is available. The well-trained network performs a CR of 98.47% on the MSTAR SOC10 and 91.32% on the MGTD. The second CNN is trained without any data augmentation and only achieves 95.12% and 67.00%, respectively.

B. Contribution of the target, shadow and clutter to the classification

The shadow of a target in SAR images contains information. For example, the target shape and height can be determined as the geometrical configuration of acquisition is known. The additional information provided could improve target detection and classification [26], [42]. The shadow and the target parts could also be sharpened to improve the quality of information extracted from SAR images [43]. Previous works in the literature investigating the target shadow do not use deep learning methods. The aim of this section is to propose an evaluation of the amount of information present in the pixels containing the target, its shadow and the clutter respectively used by the CNN. To evaluate their contribution, the change in classification scores following the occlusion of the target, shadow and clutter areas and their combinations is investigated. The results obtained from this method are shown in Section IV-A.

Classification scores achieved on masked images, and thus with incomplete information, are compared according to which zone is masked as seen in Table I. The impact of the information loss from a specific zone of the image gives insight about the importance of the features in that zone. The classification method used in this analysis is the CNN presented in Section III-A trained on the MSTAR SOC 10 dataset in Section II. Only the MSTAR SOC 10 dataset was used for this study since it is the only SAR dataset with a segmentation reference [41].

The method is here detailed to investigate the role of the target in the classification and is similar for the shadow and clutter areas. The earlier SARbake segmentation acts as groundtruth to locate the target, shadow and clutter area [41]. All images in the testing set of the MSTAR SOC 10 dataset have the target area set to black, so that all information directly from the target is removed. The CNN is then run on all masked images in the testing set. The result is a classification score for which the CNN could not rely on features from the target area. This process is repeated to obtain the classification scores corresponding to all possible segmentation combinations of the three areas as shown in Fig. 1.

| Target	X		X	X	X		
Shadow	X	X		X		X	
Clutter	X	X	X				X

Fig. 1: Images with segmented area(s) hidden.

C. Influence of the target class and target orientation in the location of the critical features

Occlusion maps are already used in the visual field [18], [19]. Their objective is to study the location of features contributing the most to correct classification in one image. The proposed classification map is an extension of the occlusion map applied to a group of images containing a target in a fixed position as seen in Fig. 2. Having a group of images rather than a single image highlights the role of factors shared within this group of images. Many variables can be presumed to have an influence on the activation of specific deep learning features such as, for example, the target class, target orientation, depression angle. The influence of one variable on the location of the most activated deep learning features is studied with classification maps obtained with images sharing the same

value for this variable. These groups of images in this section are images with the same target class or with similar target orientation.
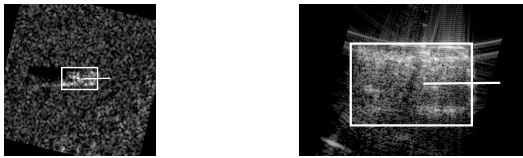
(a) Targets position in the MSTAR (Target length around $6-10$ m, range and cross-range resolution of 30 cm $\times$ 30 cm)

(b) Targets position in the MGTD (Target length around 1.5 m, range and cross-range resolution of 3.0 cm $\times$ 3.3 cm)

Fig. 2: Approximate target position in the classification maps after translation and rotation.

1) Computation of the classification map: The targets are first centered using the target center of mass calculated from the segmented images in SARBake and then rotated so that all targets are in a $0°$ position, facing the right side of the image. All targets are thus in the same location in each test image. Images with the same target or similar target orientation are then grouped together to evaluate respectively the influence of the target class or orientation on the location of the critical features learnt by the CNN. A 11×11 black square mask is applied to the top left part of all the images belonging to the studied group. The percentage of correctly classified images is used as the new intensity of the 5×5 pixels located in the center of the black square in the classification map. The black square is shifted on all the images by 5 pixels vertically and horizontally until the classification map is fully completed as shown in Fig. 3. A CR of 1, seen as white in the map, indicates that the features contained in the black square have few effects on the classification. The classification maps can be overlaid on the SAR images, with the same range and cross-range direction and resolution, respectively 30 cm $\times$ 30 cm for the MSTAR and 3.0 cm $\times$ 3.3 cm for the MGTD.
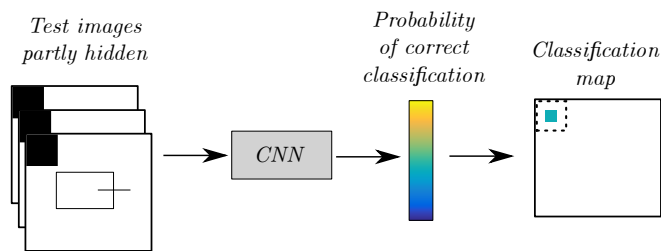

Fig. 3: Creation of the classification map.

The classification maps seen in Section IV-B are obtained both with a well-trained CNN and with a CNN trained without data augmentation and a result comparison is presented. This analysis highlights the location of the features leading to correct classification.

2) Influence of the target class in the location of the critical features: In order to achieve good classification, CNNs have to learn differences between targets. The characteristics of the targets in terms of location of critical features are studied for each target. The objective is to determine which zone, for each target, is the most important for classification. The most relevant features are expected to vary, especially between very different target types.

The images are grouped to produce classification maps according to the target they represent. Thus, respectively 10 and 3 classification maps (corresponding to each imaged target) are produced for the MSTAR and the MGTD. In all classification maps, all targets after being rotated have their front facing the right of the image.

To help with the interpretation of these maps, a black cross in the middle of the target and a contour around the lowest intensities, are added. The map intensity is indeed the classification score when the corresponding area is occluded. Thus the lowest intensity areas represent the most important areas for classification, The threshold is either an Otsu threshold or corresponds to the 10% lowest intensities if more than 50% of the image would have been retained otherwise [44].

3) Influence of the target orientation in the location of the critical features: The orientation or aspect angle of the target has an important impact on the appearance of the target in the image [45], [46]. This characteristic is thus isolated to investigate its influence and a study of the location of the zones related to the target orientation the CNN considers as important is proposed.

The images are grouped to produce classification maps according to the target orientation. Five bins are chosen to represent the target azimuth groundtruth provided with each image starting from $0°$ and equally distributed up to $360°$. The target looking to the right defines the new $0°$ arbitrarily in the rotated dataset. In this new frame of reference, the five groups of target orientations are as seen in Fig. 4. Each group of images represent all the images with a target orientation belonging to a single range bin. One classification map is computed for all these images, resulting in five classification maps representing the five different orientation bins.
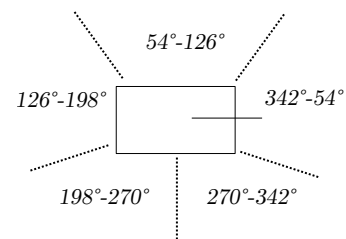

Fig. 4: Definition of the orientation ranges used to compute the orientation classification maps. Target at $0°$ is facing the right side of the image

In addition to the black contour and cross marking the lowest intensities of the classification map as in the previous subsection, arrows are added to represent the direction of the main illumination from the radar to the target.

D. Feature specialization along the CNN depth

The previous sections investigate the location of the critical target features. In this section, the specificity of the CNN features to a class and to a target orientation are examined.

Specificity of the feature is defined as the potential of a feature to be activated only in specific cases, for example only with a certain target type. This is conducted at different depth levels of the CNN as the complexity of features increases. An analysis based on the comparison of histograms summarising the most used features for specific targets or orientations is proposed. The histogram comparison shows the growing specificity of these features along the depth of the network as they become more complex.

1) Evaluation of the feature specialization: The images used are the original images from the testing set of the MSTAR SOC 10 and MGTD datasets. The complete images are fed to the CNN. The convolution between the input and one filter of the CNN results in an activation map, highlighting features associated with this filter. The input can be the original input image or a previous activation map. Activations maps obtained throughout the CNN after each convolutional layer are all extracted. Filters are then ranked, for each convolutional layer, according to the maximum intensity reached in their associated activation map. For each convolutional layer, each image results in a vector K (as seen in Fig. 5) containing the filter list ordered from the most to the least activated associated activation maps. The first filter in the vector leads to the strongest activation and is thus associated to features important for the classification while the last filter could lead to a totally black activation map. The filter number is attributed randomly between 1 and the number of filters in the studied convolutional layer.

a) Histogram of the most used features: After the computation of the ordered vector K of the filters leading to the strongest activations, the vector K is truncated to keep the $n = 20$ filters leading to the strongest activations as in Eq. (1).

$$\text{where } K(1:n) = \begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix} \quad (1)$$

Once the filter lists have been produced for a group of images, a histogram of the frequency at which filters are strongly activated by the network for a specific group of images is built as shown in Fig. 5. This histogram identifies the filters which are frequently used in a group of images. These images can be grouped by target or orientation.
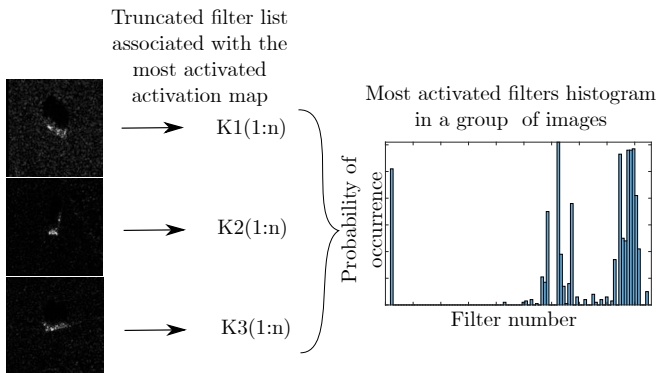

Fig. 5: Diagram representing the computation of the histogram of the most influential filters for a group of images.

b) Comparison of the features mostly used by the CNN for a specific class (target class or target orientation): The last step consists in an evaluation of the similarity or difference between the histograms produced with different groups of images. To that end, a normalised Chi-Square distance is introduced in Eq. (2).

$$D(H_{A,20}, H_{B,20}) = \frac{100}{2 \cdot m} \sum_{j=1}^{m} \frac{(H_{A,20}(j) - H_{B,20}(j))^2}{H_{A,20}(j) + H_{B,20}(j)} \quad (2)$$

where $H_{i,20}$ is the normalised histogram (maximum of 100) of the lists of top filters $K(1:20)$ for the images in the group i, m is the number of filters present in this convolutional layer and thus the bin number of $H_{A,20}$ and $H_{B,20}$.

The Chi-Square distance is a measure to evaluate histogram resemblance, and it is here normalised over the number of bins, so that this distance could be compared for histograms of different length as the number of filters increases with the network depth. The average distance expresses the difference of feature representation by the network for a specific class or for a specific orientation. For example, 3 histograms using the method represented in Fig. 5 are computed using all the test images in the MGTD respectively for the T64, the T72 and the BMP1. These histograms are produced by investigating the filters with the most stimulated activation maps. The average of all normalised Chi-Square distances between the histograms (T64 to T72, T64 to BMP1, BMP1 to T72) gives an insight of the specificity of the filters in the first layer to the target class. The bigger the distance, the more specific the features are for the concerned target.

This distance will be computed after each convolutional network between all histograms generated by images with a specific target class or orientation. For the orientation categories, five orientation ranges are used as in Section III-C3. The distance evolution along the network depth is used to evaluate the feature specialization to a specific category.

Distances are not only computed for specifically chosen groups of images but also with random groups of images of the same size to provide a control distance and ensure that the evolution of the distance is not only due to the feature complexification but really dependant on the common factor in the image group. The control distance helps to quantify how different the features really are. Full results are shown in Section IV-C.

IV. RESULTS

A. Global contribution of the target, shadow and clutter

The implementation of the method to evaluate the shadow, target and clutter contribution, described in Section III-B, results in the classification scores obtained with partially masked images with a well-trained CNN (see Table I). The results are target dependant and the relative contributions of the various areas change greatly from one target to another.

Results suggest that the shadow is rarely used by the CNN despite contributing significantly to the classification of the BRDM, D7 and ZIL. The BRDM has a score of 88% by combining the clutter and shadow areas while the D7

	Visible areas in the images fed to the CNN						
Target area	X		X	X	X		
Shadow area	X	X			X	X	
Clutter area	X	X	X				X
Target	Classification scores achieved						
2S1	97%	5%	97%	4%	3%	1%	8%
BMP	96%	29%	60%	11%	7%	3%	9%
BRDM	99%	88%	13%	9%	3%	3%	37%
BTR60	97%	5%	44%	23%	30%	2%	2%
BTR70	100%	29%	98%	5%	6%	2%	70%
D7	100%	0%	15%	87%	88%	80%	0%
T62	99%	1%	60%	18%	26%	8%	0%
T72	97%	0%	94%	17%	16%	0%	0%
ZIL	97%	97%	74%	34%	36%	48%	97%
ZSU	99%	4%	72%	99%	99%	4%	4%
All targets	98%	27%	65%	33%	34%	16%	23%

TABLE I: Classification scores attained with partly hidden images.

and ZIL reach respectively 80% and 48% using the shadow area only. Most of the time, the target and the clutter areas contain most of the information required for classification. The fact that the clutter area contains a lot of information for the CNN may mean that multipath information is used but the CNN could also be influenced by the database itself. There is indeed background correlation between clutter in the training and testing of the MSTAR SOC 10 dataset for some targets that can explain the high BTR70 classification performance [47]. Multipath present in the image background can also contain information on the electro-magnetic (EM) wave scatterers height, thus providing information on the target 3D structure [48]. The clutter influence could be investigated further with images of a segmented SAR dataset without background correlation. However, the SARBake segmentation is not supplied for the MSTAR EOCs.

The majority of targets (2S1, BMP, BTR60, BTR70, T62 and T72) main influence is a combination of the target and clutter areas with respective scores of 97%, 60%, 44%, 98%, 60% and 94%. The BRDM, D7 and ZIL are able to process more information from the shadow. The classification of the BRDM reaches 88% by combining the clutter and shadow areas while the D7 and ZIL reach respectively 80% and 48% using the shadow area only. The ZSU bases its decision on the target area alone, reaching 99%.

The higher scores obtained with only the shadow hidden could also be linked to its higher resemblance to the full original image, the shadow having lower intensities than the target and clutter areas. Further work on the shadow classification role could thus consist in replacing the occlusion zones with intensities following the clutter distribution rather than zero-padding.

B. Classification maps to understand the influence of the target class and orientation in the location of the critical features

1) Influence of the target class in the location of the critical features:

a) MSTAR dataset: Figs. 6(a), 6(b), 6(h) and 6(j) representing respectively the 2S1, BMP2, T72, ZSU show that the back-center area of the target (center left of the image)

(a) 2S1   (b) BMP [49]   (c) BRDM   (d) BTR60

(e) BTR70 [49]   (f) D7   (g) T62   (h) T72
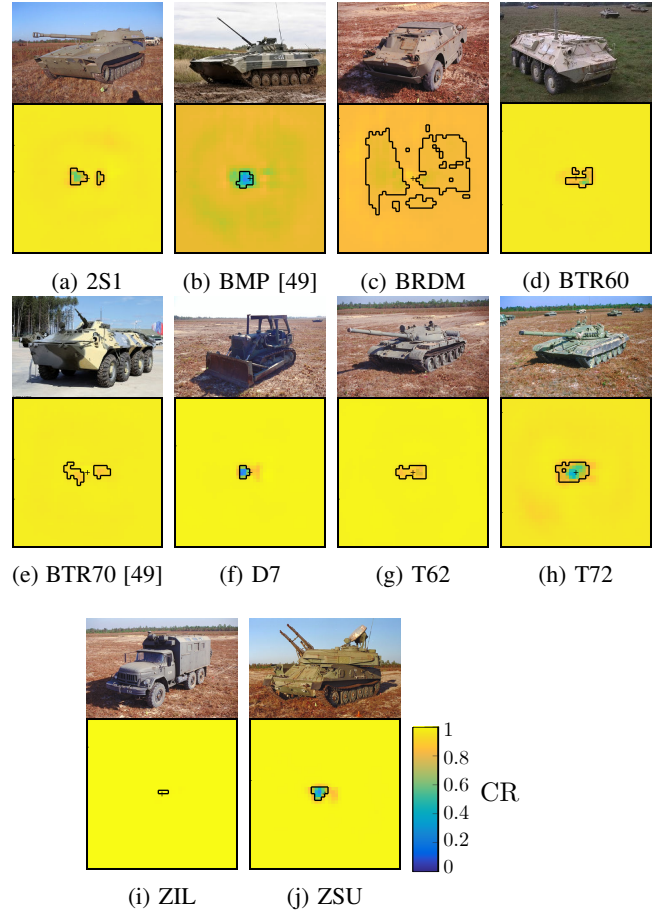
(i) ZIL   (j) ZSU

Fig. 6: Target classification maps with the original target images. BMP and BTR60 images found in [49].

is the darkest area for tanks and armoured personnel carriers with the exception of Fig. 6(g) representing a T62. It is the most critical area of the classification map and represents the highest and usually the most distinctive part of the target which corresponds to the turrets for the tanks. It is also true for the cabin of the D7 bulldozer in Fig. 6(f). However, it is not noticeable on some of the other target types that do not have such prominent features. The central darker spot is absent in Figs. 6(c) to 6(e) representing respectively the BRDM, BTR60 and BTR70.

Some targets are also recognised with the very front of the target, and this is expected for the bulldozer blade of the D7 as seen in Fig. 6(f). The same occurs for the 2S1, BTR60, T62 and ZSU as seen respectively in Figs. 6(a), 6(d), 6(g) and 6(j).

The darker background in Figs. 6(b) and 6(c) for the BMP2 and BRDM shows that the CNN is less confident in the classification of these targets in general.

The fact that the target appears brighter than the rest of the image in Figs. 6(c) and 6(e), representing the BRDM and BTR70, shows the background correlation in the MSTAR SOC 10 database [47]. In this case, the target is recognised using the background (clutter and shadow) rather than the target itself.

It seems that the ZIL in Fig. 6(i) has no critical features. As the ZIL falls in the longest targets of the database, the absence

of critical features could be linked to one of the shortcomings of the classification map computation: the impossibility to take into account combinations of several features. Indeed, only a part of the target is hidden and, if features in different locations enable the classification, hiding only one of these critical features could leave the score of the correct target unchanged. Another possibility is that the CNN chooses the ZIL in case of an uncertain prediction.
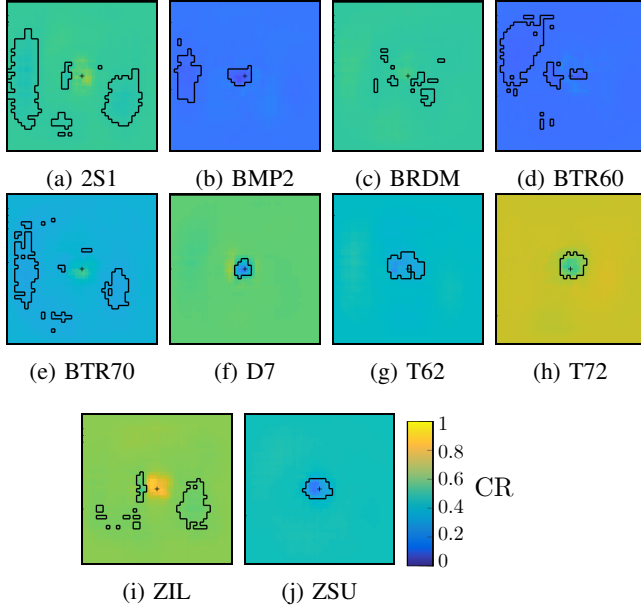

Fig. 7: Target classification maps using a CNN trained without data augmentation.

In order to better understand which areas are essential for classification, the location of the features critical for classification for a well-trained network are compared to the location deemed critical by a network trained without data augmentation and achieving lower classification scores. The classification maps of the less performing CNN can be seen in Fig. 7. The images are overall darker as the probability of correct classification is lower. The CNN trained without data augmentation seems to rely, in some cases, more on the background than the target itself as the lighter shade shows in the 2S1, BRDM, BTR70 and ZIL. The CNN did not narrow down the areas of importance like the CNN trained with data augmentation did. The darker areas on the targets are larger and blurry. They are not centered on specific areas of the target as could be seen for the well-trained CNN. Less explanations can be given for the classification choices of the CNN without data augmented training. The augmented training not only improves the classification score of the network but also improves its explainability. This is key, as the explainability of classification decisions is at least equally important to performances for implementing classification solutions under real conditions.

b) MGTD: The previous experiments are also conducted on the MGTD database and give different results as shown in Fig. 8. The CNN focused on different areas for each target unlike in the MSTAR database where the higher turret central area seemed to be a general focus points. For the BMP1, it is the top and bottom parts of the Fig. 8(a), corresponding to the target sides, which are darker. Fig. 8(b) which represents the T64 target, shows that the CNN is focused on the central part. The CNN highlights the front and back of the T72 target represented in Fig. 8(c) as the right and left part of the image.
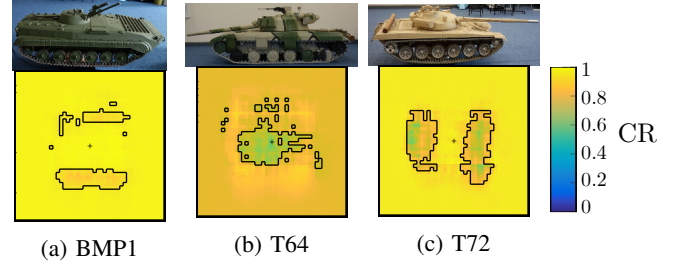

Fig. 8: Target classification maps with the original target image from the MGTD.

The darker classification maps for Figs. 8(b) and 8(c) representing the T64 and the T72, indicate that the confidence of the network in classifying these targets is lower and that they are harder to classify. The CNN indeed is less likely to mistake the BMP1 for another target than the T64 or the T72. Indeed, the T64 and T72 are very similar and differ largely from the BMP1. This explains the higher confidence of the network in the BMP1 classification.

To better understand the reasons of the location of critical features, the locations of critical areas found by a well-trained network are compared to those of a less performing one. The results can be seen in Fig. 9. The darker images overall are due to the lower CR on the testing set achieved by the CNN trained without data augmentation. The two CNNs, one trained with and the other without data augmentation, concentrate on different areas for each target. However, the reasons behind the difference of location of the critical area seem more uncertain than for the MSTAR database. The first difference with the results achieved on the MSTAR is that the critical areas for each target are different. The CNN without data augmentation focuses only on the front of the T72, the rear of the BMP1 and the sides and center of the T64 whereas the well-trained CNN focuses respectively on the sides of the BMP1, the center of the T64 and on both the front and the rear of the T72. This strategy could be because only 3 targets are present in the MGTD but this cannot be reproduced in the MSTAR which contains 10 targets.

2) Influence of the target orientation in the location of the critical features:

a) MSTAR dataset: The classification maps obtained are summarised in Fig. 10. Fig. 10(a) shows that the bottom right of the map is the most critical for a radar placed between 270° and 342°. This corresponds to the area with the best signal reflection. Because of the shape of the tank, the parts of the target facing the radar are likely to produce a specular reflection and therefore generate stronger reflections than the sides perpendicular to the radar. The target rear is not directly
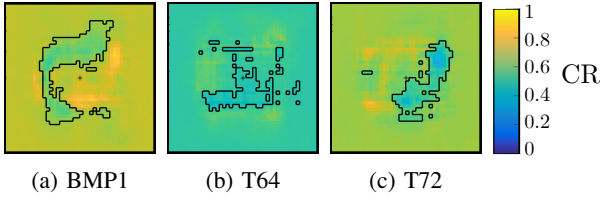
Fig. 9: Target classification maps from the MGTD with a CNN trained without data augmentation.
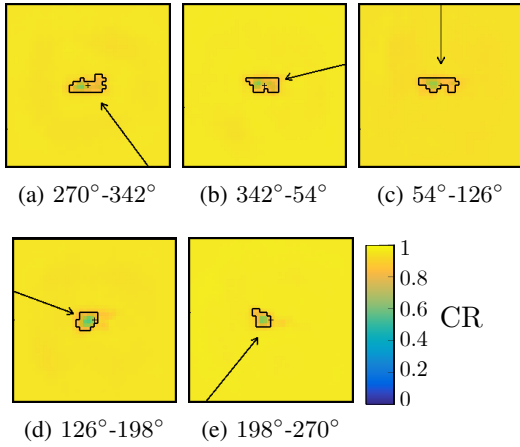

Fig. 10: Illumination direction and contour of the most critical areas in each orientation range classification map with a well-trained CNN for the MSTAR SOC 10.

illuminated but can be slightly visible through diffraction effects. Thus, the area surrounding the surface facing the radar, and the closest to the radar, is brighter in the SAR images. This area is also the critical area in most of classification maps with respectively Fig. 10(c) highlighting the top of the map, Fig. 10(d) focusing on the left of the map and Fig. 10(e) highlighting the bottom left of the map. It is however less clear for Fig. 10(b) that the most critical area is the front of the target on the right of the map, even though this part is still critical. The zones reflecting the strongest signal, usually in the area the closest to the radar as the front side faces the receiver, appear to be more critical.

It can also be noticed that the target rear, in the left part of the classification maps is always highlighted. It is indeed always inside the blue contour, which shows the darkest parts of the classification maps. This is the higher area that corresponds to the tanks turrets or the bulldozer cabin. It was also highlighted as a critical area in the target classification maps in Section IV-B1.

In order to better understand which areas are essential for classification, the location of the features critical for classification for a well-trained network are compared to the location deemed critical by a less performing network. The resulting classification maps can be seen in Fig. 11. A first observation is that the classification maps are overall darker, meaning that this CNN does not achieve the same quality of classification as

the CNN with data augmented training. Moreover, the intensity on the target is not a lot darker compared to the intensity seen in the background area. The network seems to less optimize the information present in the target even though it is still the most important area. It can be also seen that the darkest areas are not always on the area that is facing the radar in Fig. 11(c) as previously seen. The critical zone is smaller and the rear of the target is not used in all orientations as it was the case for the CNN trained with data augmentation.
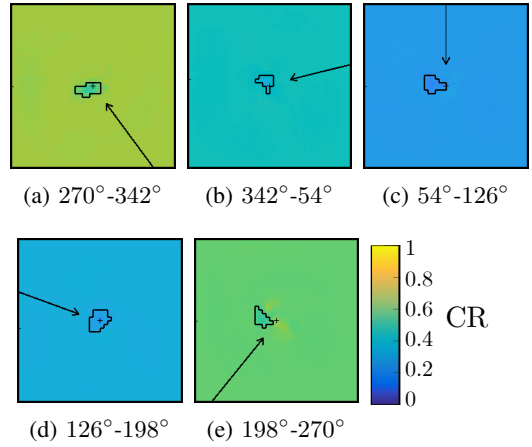

Fig. 11: Illumination direction and contour of the most critical areas in each orientation range classification map with a CNN trained without data augmentation for the MSTAR SOC 10 targets obtained.

b) MGTD: As for the MSTAR, classification maps relative to orientation ranges are produced using the data from the MGTD. Similarly to what was shown in the MSTAR in Section IV-B2, the critical areas are mostly located in the areas facing the radar as can be seen in Fig. 12. Indeed, Figs. 12(a) to 12(e) highlight respectively the bottom right, the right, the top, the left and the bottom-left of the map.
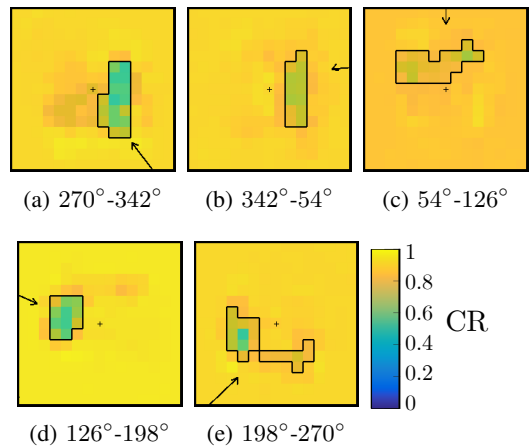

Fig. 12: Illumination direction and contour of the most critical areas in each orientation range classification map with a well-trained CNN for the MGTD.

The results in Fig. 12(a) show that, in this case, the classification relies on both the bottom-right but also the right and top-right of the map, that corresponds to the front of the target. The whole front of the target is used, even the further points that could be less illuminated. These further points are located around the areas containing corners. Corners can be modelled as trihedrals. Their EM response is known to be stable and strong over a wide range of azimuth and depression angles, easing further target recognition. Corners are present in the T72 and the T64 in the targets front as shown in Fig. 8.

However, unlike for the MSTAR, the highest part of the target, on the left of the map is not always a critical area. Indeed, it is not always inside the blue contour showing the most critical areas. This could be due to the different depression angle used to acquire both databases. Another cause for these differences could be linked to the turret material which is plastic in the MGTD and metal in the real targets in the MSTAR. Only the tracks of the model tank are in metal in the MGTD. Also, all targets in the MGTD have round turrets which minimize returns of the signal in the receiver direction compared to turrets with planar areas.

As for the MSTAR, classification maps related to orientation ranges in the MGTD are both created with a well-trained CNN and, here, with a CNN trained without data augmentation. The resulting classification maps can be seen in Fig. 13(e). Results show that the classification maps are overall darker, as it was for the MSTAR, suggesting that the classification quality dropped over the whole testing set.
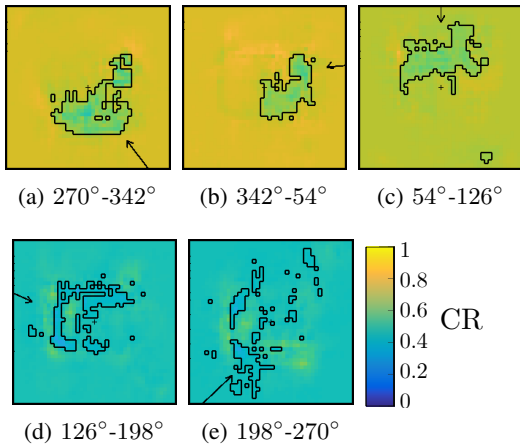

Fig. 13: Illumination direction and contour of the most critical areas in each orientation range classification map with a CNN trained without data augmentation for the MGTD.

The location of the most important parts of the classification maps are relatively comparable for the first three ranges in Figs. 13(a) to 13(c). However, they are quite different for the last two ranges in Figs. 13(d) and 13(e). Indeed, the CNN, in this case, does not seem to use the most illuminated areas which should contain most of the information on the target. In Fig. 13(e), parts of the background are also used. It is not known whether the background is used because of correlation or a multipath effect. The well-trained CNN focuses on the target unlike the CNN trained without data augmentation.

As stated for the MSTAR, the features of the CNN trained with data augmentation seem better learnt. Besides achieving higher classification scores, the CNN trained with data augmentation can also be better understood. Indeed, its critical areas are focused on the target and especially on the target area surrounding the surface facing the radar, thus reflecting well the signal because of the target geometry. The front surface facing the radar is more likely to reflect the signal towards the radar than the perpendicular surfaces or surfaces at the back. Having a better explainable network is essential when deep learning is implemented to operate in real scenarios.

### C. Feature specialization along the CNN depth

1) Feature specialization to the target class along the network depth:

a) Results obtained with the well-trained network: The average distance between histograms representing the most active filters specific to each target class grows constantly as shown in Fig. 14 and goes from 5 to 15 times the control distance in the MSTAR dataset and from 2 to 6 times the control distance in the MGTD. The CNN manages to increase the distance between targets class with features whose complexity reflects specificities of each target. The distance is more than 5 times higher in the MSTAR with 10 targets compared to the MGTD with only 3 targets. Various hypothesis can be made on the reasons of a greater distance between targets on the MSTAR trained CNN. The data from the MGTD could be harder to classify as the model targets are mainly made of hard plastic and not of metal, reflecting less clearly the emitted radar signal. If this causes the features to be less distinctive and thus less target specific, the resulting distance would indeed be lowered. The training includes more targets in the MSTAR dataset, enhancing the discrimination capability of the CNN after training. As the distance continues to increase even at the deepest layer, increasing the depth of the network could lead to better scores. The more complex the features become along the CNN depth, the more specific they are.

b) Results obtained with the network trained without data augmentation: This process is repeated with CNNs trained without data augmentation and the feature specialization can be seen in Fig. 15. Similarly to the more robust CNN, the distance between features dedicated to a specific target class grows larger in the network. However, it can be noticed that those features are less specific than those of the well-trained network as the ratios between the average distance between targets and the control distance are much lower for both the MSTAR and MGTD than in Fig.**??**. Data augmentation enhanced the specificity of features to the target class.

2) Feature specialization to the target orientation along the network depth:

a) Results obtained with the well-trained network: It can be seen in Fig. 16 that both for the MSTAR dataset and the MGTD, the average distance between specific orientation groups is a lot higher than the control distance. The maximum ratio is of 8 and 6 for the MSTAR and MGTD respectively. Images within each group represent targets of different types in similar orientations. This high distance cannot be only due
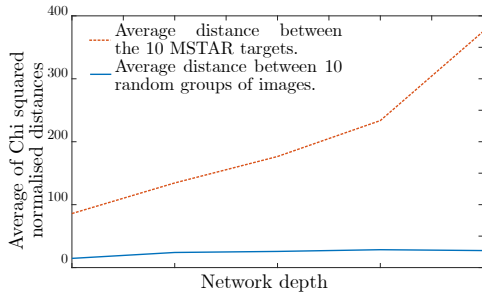
(a) Average distance between the histograms representing the most active filters used in each MSTAR target group.
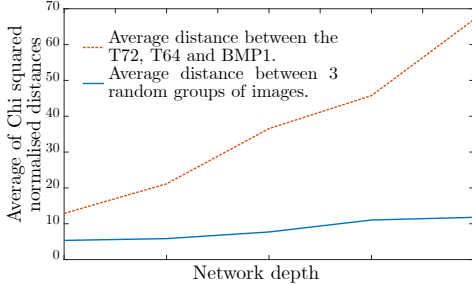
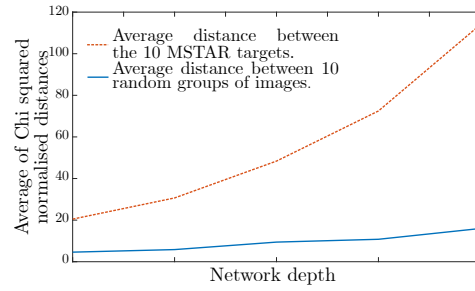(b) Average distance between the histograms representing the most active filters used in each MGTD target group.

Fig. 14: Average distance along the depth of the CNN between groups of images of different targets.

(a) Average distance between the histograms representing the most active filters used in each MSTAR target group.

(b) Average distance between the histograms representing the most active filters used in each MGTD target group.

Fig. 15: Average distance along the depth of the CNN between groups of images of different targets for the network trained without data augmentation.

to the difference of target classes, which could be the case when some target classes are more represented in a specific orientation range. The CNN specifically learned the orientation features. The network is thus able to learn environmental variable even when they are not included directly in the loss computation during training. The fact that the network is able to independently learn related environmental variables linked to the classification task is probably part of the success of neural networks on SAR images which are affected by many variables. It puts into perspective the inclusion of external variables in the loss to force the network learning regarding the target environment, as the network already carry this task to a certain extent on its own [50].
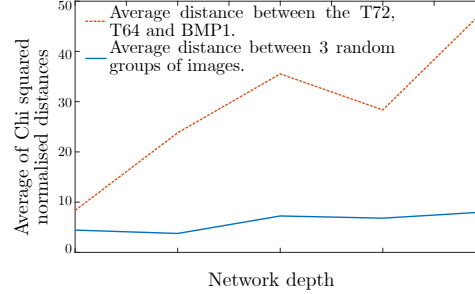
Features getting specific to environmental variables without dedicated training indicates that transfer learning could potentially be pushed further. Instead of retraining a CNN to only fit another database or different targets, the purpose of a network already trained on SAR data could be changed while still benefiting from its knowledge of the environment variables learned during training even if they were not included in the loss function. Indeed, the network has already learned potentially interesting features in addition to the features directly related to the initial task. In the presented case, a network determining target orientations could be learned from a network classifying target types.

b) Results obtained with the less performing network: All of the above is conducted again with a CNN trained without data augmentation as can be seen in Fig. 17. The same conclusion can be drawn looking at the distance between

orientation specific features learned by this CNN. The distance grows as the complexity increases with the depth of the network. It can be also noticed that the distance between features is lower at all depths for the CNN without data augmentation than for the well-trained CNN. The maximum ratios between the average distance between targets with different orientations compared to the control group are of only 7 and 5 for the MSTAR and MGTD respectively. The data augmentation creates a more challenging training set which means more specific features are needed to be able to tell the target classes apart. The features learnt by the CNN trained with data augmentation, because they are more specific to each target, enable better classification as the targets are more precisely described. Some of the learnt features relate also to the target orientation, hence the higher distances that can be seen for both the distance of features specific to targets and features specific to orientation ranges as seen in Fig. 16.

V. CONCLUSION

In this paper, some insights are given on the decision process of a trained CNN on both the MSTAR dataset and the MGTD. The analyses are carried out with two AlexNets [34]: trained with or without data augmentation in order to give an idea of the results variability for different neural networks.

The first investigation quantifies the influence of the target, shadow and background zone respectively in the deep learning classification process. It appears that the shadow area alone is mainly ignored by the CNN with an essential role in only two
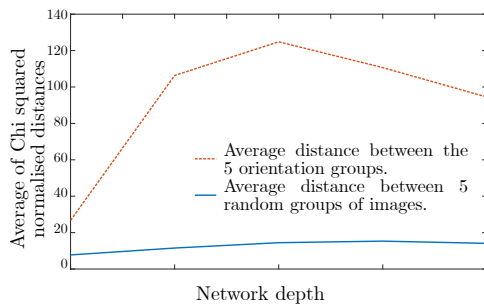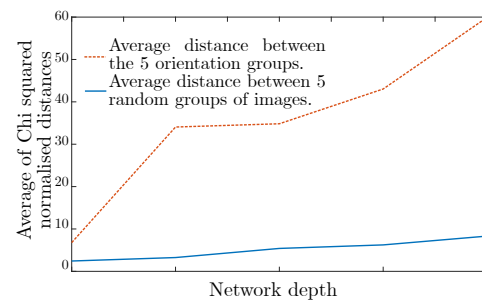
(a) MSTAR database.
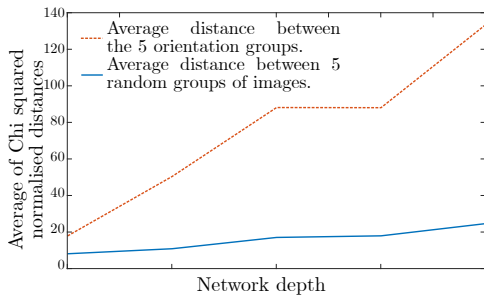
(b) MGTD.

Fig. 16: Average distance along the network's depth between histograms of filters activated the most for 5 different orientation bins.

(a) MSTAR database.

(b) MGTD.

Fig. 17: Average distance along the network's depth between histograms of filters activated the most for 5 different orientation bins with networks trained without data augmentation.

cases out of ten. However, the removal of the shadow incurs a classification loss of minimum 23% on 70% of the targets tested. The information deemed important for classification is mainly taken from the target and clutter zones. However, results on each zone contribution cannot be totally generalised as the impact of the shadow, target and background area can be very different from one target to another.
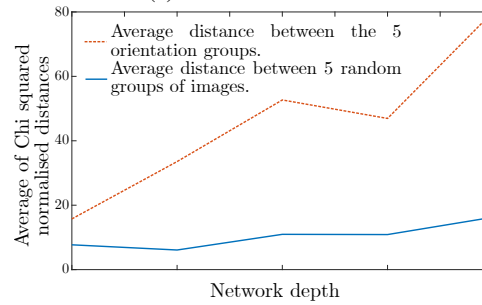
The second analysis with classification maps shows that the most important areas for the CNN are often located on specific parts of the target. The location of these areas are also influenced by the target orientation during the measurements with the areas facing the radar contributing more. The important features are located on zones specific to each target but the higher parts of the target such as the cabin or turret are often a focus point. A network that benefits from data augmentation during training not only performed better but also has a classification process more easily explained. This makes the usage of such networks for real solutions better from a reliability and acceptability point of view.

It is shown that the features become specific to a precise target as they increase in complexity. Classes become more easily distinguishable with the network depth which is coherent with the current trend of deep networks. The last analysis shows that without adapting the loss, which is focused on target classes only, the CNN still learns to build features specific to other environmental variables such as the target orientation.

The various approaches proposed give a better understanding of the significant target areas for deep learning classification such as the top target parts and the areas directly facing the radar. Information on the shadow seems not to be exploited as much as the other areas but supplementary work is required to further validate this hypothesis. Lastly, it can be seen that when correctly trained, CNNs are able to learn features specific to their environment without direct motivation from the loss function.

REFERENCES

[1] P. Tait, *Introduction to radar target recognition*. IET, 2005, vol. 18.
[2] C. Clemente, L. Pallotta, I. Proudler, A. De Maio, J. J. Soraghan, and A. Farina, "Pseudo-zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar," *IET Radar, Sonar & Navigation*, vol. 9, no. 4, pp. 457–466, 2015.
[3] C. Clemente, L. Pallotta, D. Gaglione, A. De Maio, and J. J. Soraghan, "Automatic target recognition of military vehicles with krawtchouk moments." *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 493–500, 2017.
[4] A. Agrawal, P. Mangalraj, and M. A. Bisherwal, "Target detection in SAR images using SIFT," in *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 90–94.
[5] Y. Sun, L. Du, Y. Wang, Y. Wang, and J. Hu, "Sar automatic target recognition based on dictionary learning and joint dynamic sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1777–1781, 2016.
[6] C. Yuan and D. Casasent, "A new SVM for distorted SAR object classification," in *Optical Pattern Recognition XVI*, vol. 5816. International Society for Optics and Photonics, 2005, pp. 10–23.
[7] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.
[8] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 1, pp. 125–138, 2015.

[9] Sensor data management system website, MSTAR database. [Online]. Available: https://www.sdms.afrl.af.mil/index.php?collection=mstar

[10] D. A. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXII, Baltimore, MD, USA*, vol. 23, p. 94750F, 2015.

[11] A. Profeta, A. Rodriguez, and H. S. Clouse, "Convolutional neural networks for synthetic aperture radar classification," in *Algorithms for Synthetic Aperture Radar Imagery XXIII*, vol. 9843. International Society for Optics and Photonics, 2016, p. 98430M.

[12] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[13] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sensing*, vol. 10, no. 6, p. 846, 2018.

[14] B. Lewis, J. Liu, and A. Wong, "Generative adversarial networks for SAR image realism," in *Algorithms for Synthetic Aperture Radar Imagery XXV*, vol. 10647. International Society for Optics and Photonics, 2018, p. 1064709.

[15] F. Liu, L. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE transactions on neural networks and learning systems*, 2019.

[16] S. K. Rogers, J. M. Colombi, C. E. Martin, J. C. Gainey, K. H. Fielding, T. J. Burns, D. W. Ruck, M. Kabrisky, and M. Oxley, "Neural networks for automatic target recognition," *Neural networks*, vol. 8, no. 7-8, pp. 1153–1184, 1995.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization." in *ICCV*, 2017, pp. 618–626.

[19] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[20] I. Rafegas and M. Vanrell, "Understanding learned CNN features through filter decoding with substitution," *CoRR*, vol. abs/1511.05084, 2015. [Online]. Available: http://arxiv.org/abs/1511.05084

[21] D. Garcia-Gasulla, F. Parés, A. Vilalta, J. Moreno, E. Ayguadé, J. Labarta, U. Cortés, and T. Suzumura, "On the behavior of convolutional nets for feature extraction," *CoRR*, vol. abs/1703.01127, 2017. [Online]. Available: http://arxiv.org/abs/1703.01127

[22] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," *arXiv preprint arXiv:1703.04730*, 2017.

[23] X. Li, C. Li, P. Wang, Z. Men, and H. Xu, "Sar atr based on dividing CNN into CAE and SNN," in *Synthetic Aperture Radar (APSAR), 2015 IEEE 5th Asia-Pacific Conference on*. IEEE, 2015, pp. 676–679.

[24] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and remote sensing letters*, vol. 13, no. 3, pp. 364–368, 2016.

[25] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *Data Science and Advanced Analytics (DSAA), 2014 International Conference on*. IEEE, 2014, pp. 541–547.

[26] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 969–980, 2012.

[27] J. Cui, J. Gudnason, and M. Brookes, "Automatic recognition of MSTAR targets using radar shadow and superresolution features," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. v–589.

[28] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 1135–1144.

[29] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[30] A. Mahendran and A. Vedaldi, "Salient deconvolutional networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 120–135.

[31] L. M. Novak, G. J. Owirka, and A. L. Weaver, "Automatic target recognition using enhanced resolution SAR data," *IEEE Transactions on Aerospace and Electronic systems*, vol. 35, no. 1, pp. 157–175, 1999.

[32] C. Belloni, A. Balleri, N. Aouf, T. Merlet, and J.-M. Le Caillec. Cranfield online research data, MGTD database. [Online]. Available: https://doi.org/10.17862/cranfield.rd.7240742

[33] ——, "SAR image dataset of military ground targets with multiple poses for ATR," in *Target and Background Signatures III*, vol. 10432. Warsaw: International Society for Optics and Photonics, Sep 2017, p. 104320N.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[35] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1653–1660.

[36] T. KarpathyA *et al.*, "Large-scale video classification with convolutional neural networks," *Computer Vision and Pattern Recognition (CVPR), IEEE*, vol. 2014, p. 1725.

[37] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

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

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[40] C. Belloni, N. Aouf, J.-M. Le Caillec, and T. Merlet, "SAR specific noise based data augmentation for deep learning," in *2019 IEEE International Radar Conference*. Toulon: IEEE, Sep 2019.

[41] D. Malmgren-Hansen, M. Nobel-J *et al.*, "Convolutional neural networks for SAR image segmentation," in *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 231–236.

[42] P. Lombardo, M. Sciotti, and L. M. Kaplan, "SAR prescreening using both target and shadow information," in *Radar Conference, 2001. Proceedings of the 2001 IEEE*. IEEE, 2001, pp. 147–152.

[43] T. Sparr, R. Hansen, H. Callow, and J. Groen, "Enhancing target shadows in SAR images," *Electronics letters*, vol. 43, no. 5, pp. 69–70, 2007.

[44] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[45] J. C. Mossing and T. D. Ross, "Evaluation of SAR ATR algorithm performance sensitivity to MSTAR extended operating conditions," in *Algorithms for Synthetic Aperture Radar Imagery V*, vol. 3370. International Society for Optics and Photonics, 1998, pp. 554–566.

[46] Y. Yang, Y. Qiu, and C. Lu, "Automatic target classification—experiments on the MSTAR SAR images," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on*. IEEE, 2005, pp. 2–7.

[47] R. Schumacher and J. Schiller, "Non-cooperative target identification of battlefield targets-classification results based on SAR images," in *Radar Conference, 2005 IEEE International*. IEEE, 2005, pp. 167–172.

[48] J.-M. Le Caillec, J. Habonneau, and A. Khenchaf, "Ship profile imaging using multipath backscattering," *Remote Sensing*, vol. 11, no. 7, p. 748, 2019.

[49] Photos of the BMP2 and BTR70 taken by vitaly kuzmin. (cc by-nc-nd 4.0). [Online]. Available: https://www.vitalykuzmin.net

[50] Y. Zhong and G. J. Ettinger, "Enlightening deep neural networks with knowledge of confounding factors," *CoRR*, vol. abs/1607.02397, 2016. [Online]. Available: http://arxiv.org/abs/1607.02397

Carole Belloni received in 2015 a general engineering degree from the Ecole des Mines de Saint-Etienne, France, and a M.Sc. degree in Digital Signal and Image Processing from Cranfield University, U.K. She pursued a Ph.D in Artificial Intelligence applied to SAR images in Cranfield University, U.K., and IMT Atlantique (Institut Mines Télécom), France, and graduated in 2019. She is currently working for the DGA (Direction Générale de l'Armement) in France as expert in Artificial Intelligence (AI) for defense systems. Her main research interests are Deep learning, Computer vision, Image and Signal Processing.

Prof. Nabil Aouf received his PhD from Mcgill University in 2002 at the Electrical and Computer Engineering Department. Currently, he is Professor of Autonomous Systems and Machine Intelligence at City University of London. He is the Director of the Systems, Autonomy and Control (SAC) Centre and the co-Director of the London Space Institute (LSI) at City University of London. He also leads the Robotics, Autonomy and Machine Intelligence (RAMI) group and works very closely with industries that have a strong heritage in autonomous systems and space research. He has authored over 180 high caliber publications in his domains of interest. His research interests are aerospace and defense systems, information fusion and vision systems, guidance and navigation, control, and autonomy of systems. He is an Associate Editor of 4 journals including IEEE Transactions of Intelligent Vehicles.

Thomas Merlet has graduated in 1993 from Ecole Supérieure de Physique et Chimie Industrielle, Paris, France and received his PhD degree on Opto-RF components from University of Paris XI, France in 1997. He is currently the head of the joint Lab LATERAL (Lab-STICC Thales Research Alliance) and managing Future Products & Innovation activities in the Seeker Division at Thales, Elancourt France.

Alessio Balleri received the Laurea degree (summa cum laude) (five legal years) from the University of Pisa, Pisa, Italy, in 2004, and the Ph.D. degree in electronic and electrical engineering from the University College London (UCL), London, U.K., in 2010. From February 2010 to March 2012, he was a Research Associate in radar systems with the Department of Electronic and Electrical Engineering, UCL. From June 2004 to December 2004, he was a Visiting Research Scholar with the Department of Electrical and Computer Engineering, University of Illinois at Chicago. He is currently a Reader in radar systems with Cranfield University, Shrivenham, U.K. His research interests include radar and sonar system design, biologically inspired radar and sonar systems, adaptive radar, radar and sonar target classification, target feature extraction, and modelling of radar clutter. He guest co-edited a special issue on "Biologically Inspired Radar and Sonar Systems" for the IET Radar, Sonar and Navigation in 2012 and a special issue on "Emerging Radar Techniques" for the EURASIP Journal on Advances in Signal Processing, in 2013. He was the Technical Program Committee Co-Chair for the IET International Radar Conference 2017 (Belfast, U.K.) and the Technical Co-Chair of the 2020 IEEE International Radar Conference (Washington, DC, USA).

Jean-Marc Le Caillec (S' 93, M' 99, SM 07) is a professor with the Ecole Nationale Supérieure des Télécommunications de Bretagne (Telecom Bretagne) in the 2IP (Information and Image Processing) department. He obtained the degree of Engineer in Telecommunications in 1992 from Telecom Bretagne. He received the Ph.D. degree in Mathematics and Signal Processing of the University of Rennes I in 1992. From 1997 to 1999, he worked for Thomson AirSys (now Thales AirSys). He joined Telecom Bretagne as associate professor in 1999 and became professor in 2007. He is in charge of the division CID (Knowledge, Information, Decision) of the Lab-STICC (Laboratory of sciences and technique for information, communication and knowledge, UMR 6285). His main research interests are statistics, nonlinear system modeling, mathematics and signal processing for applications in remote sensing and finance.