IET Radar, Sonar & Navigation

Research Article

IET Journals
The Institution of Engineering and Technology

Pose-informed deep learning method for SAR ATR

ISSN 1751-8784
Received on 16th December 2019
Revised 2nd March 2020
Accepted on 27th March 2020
E-First on 7th August 2020
doi: 10.1049/iet-rsn.2019.0615
www.ietdl.org

Carole Belloni[1,2] ✉, Nabil Aouf[3], Alessio Balleri[1], Jean-Marc Le Caillec[2], Thomas Merlet[4]

[1]*Centre for Electronic Warfare, Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA, UK*

[2]*Lab-STICC UMR CNRS 628, IMT Atlantique, Brest, France*

[3]*City University of London, London, UK*

[4]*Thales Optronique, Elancourt, France*

✉ *E-mail: c.d.belloni@cranfield.ac.uk*

Abstract: Synthetic aperture radar (SAR) images for automatic target classification (automatic target recognition (ATR)) have attracted significant interest as they can be acquired day and night under a wide range of weather conditions. However, SAR images can be time consuming to analyse, even for experts. ATR can alleviate this burden and deep learning is an attractive solution. A new deep learning Pose-informed architecture solution, that takes into account the impact of target orientation on the SAR image as the scatterers configuration changes, is proposed. The classification is achieved in two stages. First, the orientation of the target is determined using a Hough transform and a convolutional neural network (CNN). Then, classification is achieved with a CNN specifically trained on targets with similar orientations to the target under test. The networks are trained with translation and SAR-specific data augmentation. The proposed Pose-informed deep network architecture was successfully tested on the Military Ground Target Dataset (MGTD) and the Moving and Stationary Target Acquisition and Recognition (MSTAR) datasets. Results show the proposed solution outperformed standard AlexNets on the MGTD, MSTAR extended operating condition (EOC)1, EOC2 and standard operating condition (SOC)10 datasets with a score of 99.13% on the MSTAR SOC10.

1 Introduction

Synthetic aperture radar (SAR) images provide strategic information for military and civilian applications and they can be acquired day and night under a wide range of weather conditions. Since the interpretation of SAR images is a common challenge, automatic target recognition (ATR) algorithms can help assist with decision-making when the operator is in the loop or when the platforms are fully autonomous. Recent research on SAR ATR methods has focused on convolutional neural networks (CNNs) resulting from past research activities on deep learning for images. CNNs are already commonly used in the visual domain and in the existing literature, limited CNNs have been specifically proposed and developed for SAR ATR. Some of the CNNs proposed for SAR ATR are shallower than those used in the visual domain, where the image resolution is much higher than that of SAR and detailed features are commonly available [1, 2]. CNNs applied to SAR have shown high recognition rates on the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset with scores reaching 99.1, 96.12, 98.93 and 98.60%, respectively, on the MSTAR standard operating condition (SOC)10, MSTAR extended operating condition (EOC)1, MSTAR EOC2 and MSTAR EOC3 or even higher on the MSTAR SOC10 [2].

CNNs can be used directly for classification, but they have also been used in the past literature to extract complex features from SAR images and supply them to other classification solutions such as a support vector machines [3, 4]. For example, features have been extracted from intermediary layers of trained CNNs to provide improved results with respect to former feature based solutions that used a feature dictionary on the sparse representation of the target (up to 99.5% compared to 80% on the MSTAR SOC10 [3, 4]).

CNNs, whether used directly or as feature extractors, have to be trained to learn key features for classification. They can be trained entirely on SAR data, or through transfer learning process consisting of a pre-training in the visual domain to increase the total amount of training data before SAR training, or benefit from data augmentation techniques with added transformed images in the training set [3–5]. Other CNNs have been developed specifically to tackle the low number of training images typically characterising SAR ATR instead of working on the input data directly [6].

In addition to deep learning architecture and input data changes, improvements in classification performance have also been achieved by providing extra information to a classifier. Some methods to achieve this relied on information fusion, such as combinations of deep learning methods with Gaussian mixture models representating the SAR image, with texture information, with different features or with decisions of other classifiers [7–9].

Alternative methods introduced tackled time variations in SAR ATR by using several consecutive images of the target, rather than performing classification on a single image. Classification of a group of images (2–4 images) was carried out using a multiview deep learning network or a long short-term memory (LSTM) architecture [10, 11]. The LSTM is a recurrent neural network with several inputs. The processing is such that the information given by the first set of images is retained and used to process later images. This architecture achieved a score of 99.90% on the MSTAR SOC10 database [10]. However, these results were obtained with groups of images with significantly different target orientations that were not compatible with a realistic scenario for SAR images acquired during a straight and short flight. Thus, these scores cannot be directly compared to operational classification methods based on a single image of the target.

Additional information, such as knowledge on the environment during the SAR image acquisition, could help the classification process. Indeed, sensitivity of SAR images to environmental changes is a key challenge for SAR ATR algorithms. Changes of the acquisition scenario can modify the SAR image substantially due partly to the diffractive nature of the signal and lead to a drop of the classification rate [2, 12, 13]. For these reasons, previous

research has investigated ATR solutions that took into account a variety of viewing conditions.

In particular, it was shown that the orientation of the target had a strong influence on the target appearance [14, 15]. Improvement in feature based classification performance was shown when an estimate of the target orientation was provided to the classification algorithm [16, 17]. Deep learning work also benefited from the inclusion of target orientation knowledge to the network training. For example, in addition to the main objective to optimise a loss function on the target class, a secondary objective consisting in determining the correct target orientation was included [18]. Although neural network have been previously used in the visual domain to determine the orientation for face detection in the visual domain for example, 180° uncertainty was less for human faces in the visual domain than for rectangular targets in the SAR domain [19]. In this case, the potential face was rotated and fed to a unique detector not trained on the specific face orientation ranges. On the contrary, 360° orientation determination has been rarely tackled before for targets in the SAR domain [20, 21]. Most methods gave only an approximation of the orientation modulo 180° or ranges of possible orientation [22, 23]. Some methods were also based on prior target class information to obtain a precise orientation [24, 25].

Another proposed solution to improve classification rates was to generate and to add artificial images to the training set similar to those often misclassified by the CNN in order to improve robustness [26]. This approach was sufficient for few variables such as the depression angle and the target orientation, but a high number of variables cannot be addressed effectively in this manner.

In this paper, we propose a novel technique that, given a certain target orientation, assigns its classification to the best suited CNN, specialised in the recognition of targets in a similar range of orientations, out of a group of CNNs with various range orientation specialisations. Our approach relies on the determination of the viewing conditions before choosing the most adequate trained CNN for target classification. The technique presented here focuses on the orientation of the target but could be extended to a wider variety of conditions as long as enough examples with similar conditions are present and referenced in the training set. Several other parameters, e.g. the depression angle, background or configuration could be selected in place of target orientation for the neural network to focus on [12]. However, as target orientation has been shown to have a great impact on classification rates and is known with values ranging from 0° to 360° in the various datasets, it is chosen to be the focus point of the Pose-informed architecture [14–16, 18].

The Pose-informed method, first determines the target orientation. The determination of the target orientation required for the Pose-informed method is achieved over 360°. An alteration to the traditional Hough transform method is proposed that reduces the number of 90° errors (since the studied targets are mainly rectangular) by studying the direction of the electromagnetic energy backscattered by the target, giving a 180° angle information. Training is then required in order to determine the final orientation of the target with a CNN that distinguishes the front from the back of the target. The proposed method improves the precision of orientation determination compared to former methods and does not require prior information on the target class. Second, the image is supplied to a specialised CNN that has been previously trained on SAR images with targets with similar orientations. This CNN benefits from transfer learning from the visual domain to the SAR domain before being trained on last time on SAR images with targets in specific orientations.

In order to evaluate the Pose-informed method, a faire result representation is proposed for classification rates on small datasets. Indeed, in a small training dataset such as for the SAR ATR case and with an even smaller validation set, it is possible that several models reach the same best validation score with nonetheless different testing scores. With this in mind, the worst and best results on the testing set are given for all trained model achieving the same best validation score, to give a range of possible classification scores obtained on the testing set.

The orientation determination for the full 360° is determined with a mean average error (MAE) of 11.94° and 14.34° on the MSTAR SOC10 and Military Ground Target Dataset (MGTD), respectively. The Pose-informed architectures outperforms the simple CNN architecture on 4 out of the 5 datasets it was tested on with respective deltas of 3.09, 2.14, 8.26, 1.77 and − 1.70% on the MGTD, MSTAR SOC10, MSTAR EOC1, MSTAR EOC2 and MSTAR EOC3. There is no drop in the classification rates for targets with an orientation close to the border between the orientation ranges of the specialised Pose-informed CNNs.

Table 1 Description of the MGTD

Class	Training (21.8°–23.4°)		Testing (17.5°–20.3°)	
	Serial nb.	Image nb.	Serial nb.	Image nb.
T64	63	72	9	36
	64	72	10	36
	65	72	15	36
	66	72	16	36
T72	53	72	21	36
	54	72	22	36
	55	72	23	36
	56	72	24	36
BMP1	49	72	27	36
	50	72	28	36
	51	72	29	36
	52	72	30	36

2 Dataset

The proposed architecture is tested on five different datasets. One dataset is taken from the MGTD while the others are from the MSTAR database.

The MGTD has been generated at Cranfield University [27, 28]. The emitted signal spanned the frequency range from 13 to 18 to achieve a bandwidth of 5 GHz sampled with 4001 frequency points. The resolution obtained is of three in range and 3.3 in cross-range on model targets of around 1.5. Single polarised images (horizontal–horizontal (HH)) are generated using the backprojection algorithm [29]. Laboratory background is removed by substracting all zero-Doppler contributions from the range-profiles. A total of 1728 images are produced using a 20° integration angle. The images in the dataset are organised in 24 sequences separated into a training and testing set to facilitate a standard evaluation of SAR ATR algorithms. The training and testing sets are made with different target configurations, depression angles and laboratory backgrounds as presented in Table 1. This dataset contains three target models: a T64, a T72 and a BMP1.

The MSTAR database was developed by the US Defense Advanced Research Projects Agency (DARPA) and the US Air Force Research Laboratory (AFRL) [30]. All datasets were collected in HH polarisation in X-Band with a $30\,\text{cm} \times 30\,\text{cm}$ resolution. The MSTAR dataset acquired under SOC consists of ten targets. The training and testing sets were formed by selecting images with differences in depression angle of 2° as shown in Table 2. The three other MSTAR datasets have been acquired under EOC and include four different targets. There is a greater difference between the training and testing set for the EOC1, EOC2, EOC3 with a 13° depression angle offset, respectively, targets with target manufacturing differences and gear differences between targets such as the addition or removal of a fuel barrel, skirt or reactive armour.

In Table 3, the similarities and differences found in all datasets are summed up in order to better understand the proposed algorithm generalisation capabilities.

3 Deep learning pose-informed method

First, the CNN architecture retained in the Pose-informed recognition solution is presented in Section 3.1. Second, we present in Sections 3.2 and 3.3 the training of the CNNs, with transfer learning from the visual domain as well as classical and SAR-specific data augmentation to tackle the low number of images in SAR ATR datasets.

Table 2 Description of the MSTAR SOC10, MSTAR EOC1, MSTAR EOC2 and MSTAR EOC3

MSTAR SOC10				
Class	Training (17°)		Testing (15°)	
	Serial nb.	Image nb.	Serial nb.	Image nb.
BMP2	sn_9563	233	sn_9563	196
BTR70	sn_c71	233	sn_c71	196
T72	sn_132	232	sn_132	196
BTR60	sn_k10yt7532	256	sn_k10yt7532	195
2S1	sn_b01	299	sn_b01	274
BRDM	sn_E71	298	sn_E-71	274
D7	sn_92v13015	299	sn_92v13015	274
T62	sn_A51	299	sn_A51	273
ZIL	sn_E12	299	sn_E12	274
ZSU	sn_d08	299	sn_d08	274

MSTAR EOC1 – depression variant				
Class	Training (17°)		Testing (30°)	
	Serial nb.	Image nb.	Serial nb.	Image nb.
2S1	sn_b01	299	sn_b01	288
BRDM	sn_E-71	298	sn_E71	289
T72	sn_132	232	sn_A64	288
ZSU	sn_d08	299	sn_d08	288

MSTAR EOC2 – version variant				
Class	Training (17°)		Testing (15° and 17°)	
	Serial nb.	Image nb.	Serial nb.	Image nb.
BMP2	sn_9563	233	sn_9566	196 + 232 = 428
			sn_c21	196 + 233 = 429
BRDM	sn_E-71	298	—	—
BTR70	sn_c71	233	—	—
			sn_812	195 + 231 = 426
			sn_A04	275 + 299 = 573
T72	sn_132	232	sn_A05	274 + 299 = 573
			sn_A07	274 + 299 = 573
			sn_A10	271 + 296 = 567

MSTAR EOC3 – configuration variant				
Class	Training (17°)		Testing (15° and 17°)	
	SERIAL NB.	IMAGE NB.	SERIAL NB.	IMAGE NB.
BMP2	sn_9563	233	–	–
BRDM	sn_E-71	298	–	–
BTR70	sn_c71	233	–	–
			sn_s7	191 + 288 = 419
			sn_A32	274 + 298 = 572
T72	sn_132	232	sn_A62	274 + 299 = 573
			sn_A63	274 + 299 = 573
			sn_A64	274 + 299 = 573

3.1 Baseline CNN

The CNN used as a baseline for comparison is an AlexNet with five convolutional layers and three fully connected layers [32]. This architecture was selected as it has been successfully adapted to a variety of other applications in the past, such as human pose estimation, video classification and semantic segmentation [33–35]. The architecture of AlexNet is straightforward compared to other recent networks, such as ResNet or GoogleNet [36, 37]. The number of weights is lower than for deeper models such as VGGNet [38]. A simple architecture is selected for ease of implementation and the availability of pre-trained models. The effectiveness of the architecture on its own is not the prime focus of this study as our main goal is to investigate whether performances of a chosen network can be improved using the Pose-informed solution proposed and evaluated in the result sections of this paper. Thus the AlexNet is deemed a good candidate with its relative simplicity and limited number of weight.

As the original implementation of AlexNet is based on 1000 targets in ImageNet, the last fully connected layer is replaced to bring down the 1000 output classes to the number of classes in our dataset [32, 39]. The initialisation of the weights in the untrained last fully connected layer of the network is narrow-normal [40].

Further work once the Pose-informed model is validated could see the AlexNet replaced with shallower CNNs, trained to be robust to SAR characteristics such as speckle [41, 42].

3.2 Transfer learning training

In order to reduce training time and compensate for the low number of images available compared to usual deep learning training strategies, transfer learning is applied from a pre-trained AlexNet on the ImageNet [39] to the appropriate SAR (or ISAR) database presented in Section 2. The training method is the stochastic gradient descent with momentum method [43]. As per standard transfer learning, the network is first trained on the ImageNet visual database and is trained again on the SAR data. The second training focuses on the newly created last layers that were modified to deal with the number of final targets. Out of all training data described in Tables 1 and 2, 90% is allocated for training of the neural network while the remaining 10% form the validation set. The CNNs performing the best on the validation set are selected to classify the testing set.

The parameters seen in Table 4 were first initialised with empirical values and then refined with a random grid search. A decaying learning rate is chosen so that the learning rate diminishes as the loss becomes minimal. The learning rate is one of the most sensitive training parameters for a CNN: if the learning rate is too low, then the CNN is not able to learn the correct weights but if the learning rate is too high, then the weights cannot settle and the loss can increase. The learning rate is different for the various layers of the network. Indeed, for transfer learning the training is mainly focused on the deepest layers [44]. As the network is from a different modality, the lower layers still need light training. Layers higher than layer 9 have a higher learning rate with the highest learning rate for the last layer. The learning rate λ_0 is researched extensively using a random grid search and is given in Table 4. The learning rate is chosen so that it maximises the classification score on the validation set after five training epochs. An epoch consists of a training period while all training images have been through the network once. The values of the training parameters are summarised in Table 4.

All parameters are the same for all trainings in the different datasets, apart for the learning rates given in Table 4 in order to evaluate the generalisation capabilities of the method across varied datasets with fixed parameters. This also has the advantage of minimising the risk of overfitting by adjusting the CNNs parameters too finely.

3.3 Data augmentation

Deep learning performance is significantly dependent on the number and variety of the training images. This problem can be partially tackled with data augmentation, which consists of adding

Table 3 Databases comparison

Environment	MSTAR Outside	MGTD Laboratory
depression angle difference between training and testing	2°–13° (EOC1 only)	1.5°–5.9°
differences between training and testing	depression angle, target configuration and variant	depression angle, target configuration and laboratory background
polarisarion	HH	HH
resolution	30 cm	3.0 cm × 3.3 cm
approximate resolution to target length ratio	3×10^{-2}	2×10^{-2}
average SNR	35 dB [31]	57 dB [5]

Table 4 AlexNet training parameters

Parameter	Value
learning rate (up to layer 9)	$1 \cdot \lambda_0$
learning rate (after layer 9)	$6 \cdot \lambda_0$
learning rate (last layer)	$12 \cdot \lambda_0$
initial learning rate λ_0	MSTAR: 8.0×10^{-5},
—	MGTD: 1.2×10^{-5}
epochs number	75
learning rate dropping rate	0.75
number of periods before dropping the learning rate	7
batch size	15

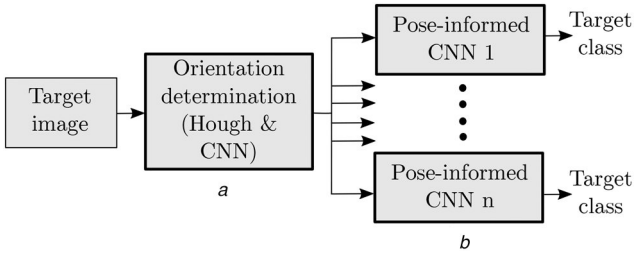

Fig. 1 *Overview of the Pose-informed architecture*
(a) Orientation determination (Hough and CNN), ***(b)*** Pose-informed CNN

new images to the training set, which are created by deforming original images from the training set and thus provide more training examples. Networks on all dataset benefit from translation data augmentation. The input images are first resized to 227×227 pixels from an original size of 128×128. Each image is then translated of a vector $[x, y] \in \mathbb{R}^2, [x, y] \in [[-100; 100], [-100; 100]]$. In practice, this means that the target is always entirely in the image even if the target was not exactly centred to begin with. The areas of the images not assigned to a value are zero-padded to retain the original size of the image.

Networks trained on the MGTD benefit from additional SAR specific data augmentation. This technique relies on the artificial addition of Weibull distributed noise to the radar returns before processing the SAR images to simulate a noisier acquisition [5]. Four different noise distributions with various signal to noise ratio (SNR) are investigated, increasing five times the original amount of training images of the MGTD.

4 Pose-informed method

The Pose-informed deep learning architecture in Fig. 1 is proposed to handle target classification. The first step consists in the determination of the target orientation using a Hough transform and a CNN, as described in Section 4.1. Once the orientation CNN is found, the appropriate Pose-informed CNN is used to determine the target class, as explained in Section 4.2. This Pose-informed CNN is specifically trained on an orientation range that includes the predicted orientation. There are n Pose-informed CNNs, each trained on $360/n$ degree wide orientation range, with n between 2 and 8.

4.1 Target orientation determination

The determination of the target orientation is the first-step of the Pose-informed method. To determine the orientation modulo 180° of the target, several methods can be used. Statistical methods are precise, however they require training which introduces additional randomness to the process, for example the initialisation of an expectation-maximisation algorithm [21, 24, 25]. Moreover, the values of the estimated distributions are target specific. This approach requires the target classification task to be carried out before the orientation determination. For these reasons, a direct pose estimator was selected, at the expense of a small precision loss. Methods not relying on the evaluation of a statistical distribution have already been investigated, for example by estimating a bounding box or a Hough transform considering several poses [22]. Here, a CNN is used to differentiate the front from the back of the target once the orientation modulo 180° is determined.

An alternative solution could be a CNN handling completely the orientation determination. A CNN on the AlexNet model was created with a regression layer to that end, but without success. As the discontinuity at 0° and 360° complicated the loss, another CNN was tested with two outputs representing the cosine and sine of the orientation angle so that the loss could be continuous. This solution did not give good results either. It could be because each output could independently be associated with two target orientations. Though, some deep learning methods have had some success retrieving the orientation of text in images in the visible domain [45]. This method relies however on the determination of an encapsulating box on the zone of interest that works well for text but has been shown to be a rather poor SAR target orientation determination (standard deviation of 14.02° against 8.12° for the encapsulating box and the Hough transform in [22], respectively).

The orientation is found in two steps. First, the orientation is determined modulo 180° with a Hough transform by relying on the rectangular shape of the target. Once the image is rotated with the determined angle, the image contains a horizontal target. This rotated image is fed to the CNN which determines the direction of the target by recognising the front from the back of the target. With these two steps, the full 360° orientation of the target is determined. The image of the target is then analysed by the appropriate Pose-informed CNN in Fig. 1*b*.

In this section, the pose estimation is investigated using a simple segmentation method and improving the orientation determined with a Hough transform by using prior knowledge that targets have a rectangular shape.

4.1.1 Segmentation and 180° target orientation determination: The objective of the segmentation for pose estimation consists in extracting a precise contour of the target so that the target orientation can be accurately estimated. In SAR images, one or two edges of the target are usually well defined depending on the electromagnetic wave illumination direction. Once detected with the Hough transform, the longest straight edge sets the target orientation. The segmentation process starts by applying a Gaussian filter to smooth the picture and obtain a simpler target shape to segment. After Gaussian filtering, the image is binarised using a threshold. This threshold is chosen to keep only 65% of the brightest pixels by computing the intensity cumulative distribution in images from the MSTAR database. It is increased to 88% in the MGTD as the target occupies more space and the average intensity is higher. The resulting binary image is shown in Fig. 2*b*. To smooth the edges of the target, morphological filtering is applied with two steps of dilation and one of erosion. Lastly, the smaller blobs are suppressed to keep only the largest before extracting its contour as shown in Fig. 2*c*.

Fig. 2 *Contour acquisition and orientation determination of the target*
(a) Original image, ***(b)*** Image after the Gaussian filtering, thresholding, morphological filtering and hole filling, ***(c)*** Small blobs removal, ***(d)*** Orientation of the target contour

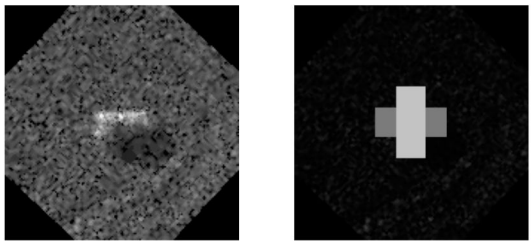

Fig. 3 *Location of the areas whose summed intensities are used to compute the vertical ratio and tackle* 90° *errors*
(a) Image of a rotated and centred target in the MSTAR dataset at 0°, ***(b)*** Location in the MSTAR of the rectangles used to compute the vertical ratio

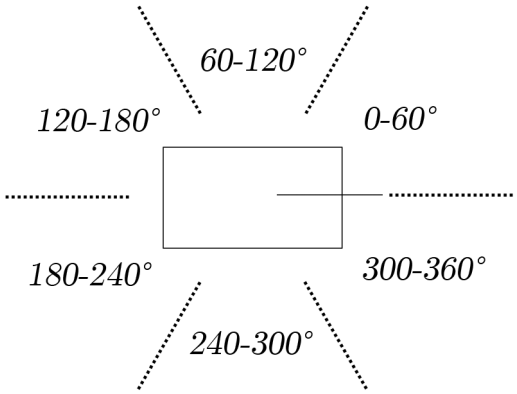

Fig. 4 *Orientation ranges for an architecture with six Pose-informed CNNs*

Once the image is segmented, a Hough transform is applied to the target contour. Only one peak, the brightest of the resulting matrix is kept. It corresponds to the longest line that can be superimposed on the target contour in Fig. 2*d*. The orientation of the resulting line is retained as the orientation modulo 180° of the target.

Owing to the difficulty to segment a target with poorly defined edges and a varying illumination, the long edges were not always detected as lines. These lines can be broken in several parts or not being detected at all when the illumination is focused on one of the small edges. If the small edge is the only one determined as a line by the Hough transform, then the estimated orientation is off by roughly 90°. We limit these recurring errors, by using prior knowledge on the rectangular shape of the target and that the intensity of the target is higher on average than that of the noise and clutter in both the MSTAR and the MGTD. Once the orientation of the target is computed by the Hough transform, the image is rotated to compensate the target orientation. If the orientation is off by 90°, then the target is vertical instead of horizontal. A vertical ratio of the sum of intensities of the pixels contained in two rectangles of fixed size is computed with either a 0° or 180° direction, respectively, as seen in Fig. 3. This ratio is expressed in (1).

$$H(I_{\hat{\theta}}) = \frac{\sum_{i=c_1-m}^{c_1+m}\sum_{j=c_2-s}^{c_2+s} I_{\hat{\theta}(i,j)}}{\sum_{i=c_1-s}^{c_1+s}\sum_{j=c_2-m}^{c_2+m} I_{\hat{\theta}(i,j)}} \quad (1)$$

where $\hat{\theta}$ is the estimation of the target orientation, $I_{\hat{\theta}}$ is the image resulting from the rotation of the original image of an angle $-\hat{\theta}$, $H(I_{\hat{\theta}})$ is the vertical ratio of the target computed for the image $I_{\hat{\theta}}$, c_1, c_2 are the abscissa and ordinate defining the image centred (once the target centred), s, m are half the length of the short and long side of the rectangle, respectively.

After some testing on the training sets, the cut-off value of the vertical ratio is determined to be 1.20 in the MGTD and 1.09 in the MSTAR database. The difference could be due to different target material as the MGTD targets are mainly in plastic. If the vertical threshold is higher than that threshold, it is assumed that the estimated orientation is off by 90° and this value is added to the original orientation estimation.

4.1.2 360° target orientation determination: The Hough transform provides an estimate of the orientation of the target modulo 180°. However, features are different for the front and the back of the target. Very few methods address the determination of the exact pose of the target over an entire sector of 360° [20, 21]. The Pose-informed classification method requires prior knowledge on the 360° orientation as the image will be distributed to a CNN trained on targets with similar orientations. In order to determine the direction of the target, a CNN similar to that proposed in Section 3.1 is used. This CNN, given a rotated input image with a horizontal target, determines if the final target orientation is α or $\alpha + 180$ with α the target orientation given by the Hough transform.

The CNN used for this analysis is the same AlexNet as presented in Section 3.1. The only difference is that the last fully connected layer provides only two classes, i.e. front or back of the target facing the right side of the image. The two classes are labelled 0° or 180°. The training parameters are the same as in Section 3.1. For training, rotated images with a horizontal target are supplied to the network from the appropriate training dataset. In order to maximise the training data, two types of images are supplied:

- *Images rotated with the ground truth orientations*: each image is rotated using the ground truth angle to produce two new images with two different target orientations consisting of a 0° or 180° direction.
- *Images rotated with the orientation found with the Hough transform*: the image is rotated according to the orientation determined with the Hough transform. The 0° or 180° labels are assigned according to the closest label orientation to the Hough transform orientation, as can be seen below

$$\text{direction label} \in \begin{cases} \{0°\} & \text{if } \left|\theta - \hat{\theta}\right| < 90 \text{ or } \left|\left|\theta - \hat{\theta}\right| - 360\right| < 90 \\ \{180°\} & \text{else.} \end{cases} \quad (2)$$

with $\hat{\theta}$ the estimated target orientation after a potential 90° correction as in (1).

Rotation data augmentation is also included with a random rotation between $-15°$ and 15° of the training data in order to make the CNN robust against potential orientation estimation errors made by the Hough transform.

4.2 Pose-informed architecture

Once the orientation of the target is estimated, the image is analysed by the appropriate CNN from the Pose-informed architecture. An example of the separation of the Pose-informed CNNs, each focusing on a specific orientation range, is shown in Fig. 4.

Instead of training each Pose-informed CNN directly, a parent CNN is trained on the full SAR training set with all possible target orientations as shown in Fig. 5*a*. The evolution of the validation loss reflecting the modality transfer learning is shown in Fig. 6. A second transfer learning step is the orientation-speciality transfer

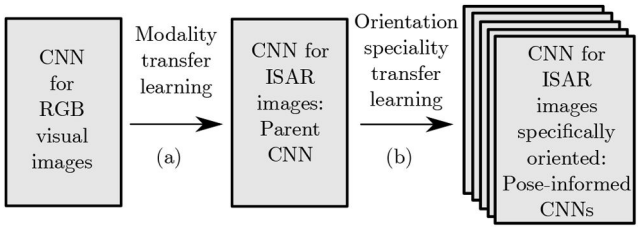

Fig. 5 *Training by stage of the Pose-informed CNN. A modality transfer learning followed by an orientation transfer learning*
(a) Modality transfer learning, ***(b)*** Orientation speciality transfer learning

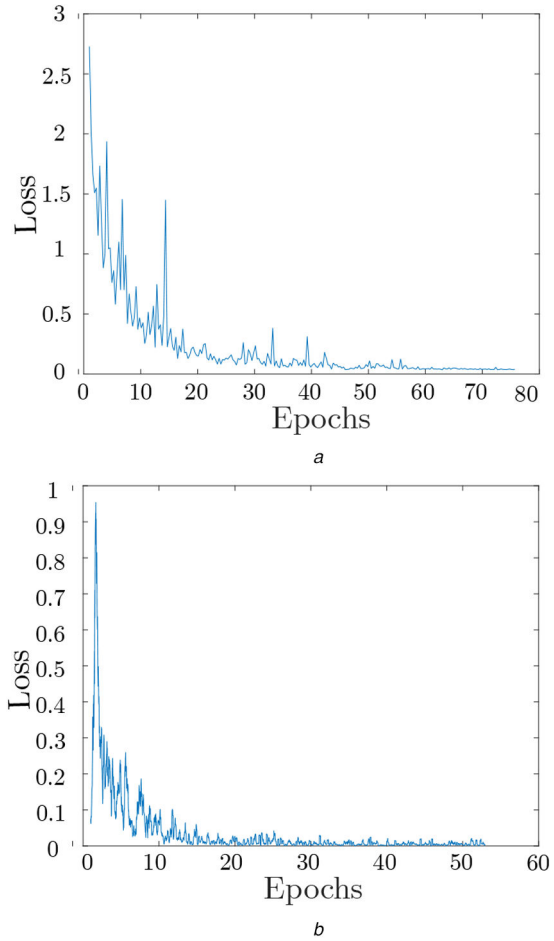

Fig. 6 *Evolution of the validation loss during transfer learning*
(a) Evolution of the validation loss during the modality transfer learning of the network, ***(b)*** Evolution of the validation loss during the orientation-speciality transfer learning of the network

learning shown in Fig. 5*b*. The parent CNN is re-trained on a specific orientation range by supplying SAR images with a targets in a specific orientation range to become a specialised Pose-informed CNN. The operation is repeated until all n Pose-informed CNNs composing the full architecture are obtained. An example of the validation loss during the orientation-speciality transfer learning is shown in Fig. 6.

This transfer learning by stage is a training strategy consisting of modality transfer learning followed by an environmental specific transfer learning, (in this case orientation-specific) to make the Pose-informed CNNs fully aware of the feature changes in a specific orientation. Transfer learning by stage optimises the number of images that the Pose-informed CNNs have at their disposal for training because the more common SAR features are learned by the parent CNN, while the Pose-informed CNNs focus on finer and more specific features during the orientation-speciality transfer learning.

The same AlexNet presented in Section 3.1 was used and trained with the same learning rate. With usual transfer learning, the Pose-informed CNN would be trained directly on the ISAR images in a specific orientation range. Instead, with transfer learning by stage, the Pose-informed CNN is trained on the full ISAR training set before training on the ISAR images in a specific orientation range. The evolution of the loss during transfer learning by stage with the first traditional transfer learning (from the visual domain to the SAR domain) and the orientation-speciality transfer learning is shown in Fig. 6. The loss corresponding to the orientation-speciality training has a lower starting value than that for the modality transfer learning, as the starting network already underwent an initial training in the SAR domain. At the end of the second training, the validation loss is lower than that at the end of the modality transfer learning (<0.01 compared to 0.05).

The first stage of transfer learning makes the Pose-informed CNNs learn standard SAR features. The second stage facilitates their specialisation in a particular orientation range. This method optimises the use of all training samples and promotes the learning of a specialised CNN. Results suggest it becomes possible to learn features that are present in the overall training set but are sparse in the specialisation areas.

4.2.1 Computation of the result range: It is not possible to have a proper separation between the training and the validation set in the MSTAR database because the MSTAR SOC10 and EOCs provide only one sequence of images for each target in the training set. Thus even if part of the training set is dedicated for validation only, a high validation score does not prevent overfitting as images are extremely similar. Indeed, images are formed with the same target, in the same configuration, with the same depression angle and were collected at the same time period. The MGTD instead provides four series of images, for each target with different configurations and taken at different times, of which one could be dedicated to validation purposes. However, the same procedure was applied to all datasets and 10% of the training set was randomly allocated as validation set to guarantee a uniform testing method across all datasets.

As a result, CNNs with the same validation score can have different testing results. This is especially true in datasets with a small number of images in the validation set, and under the EOCs for which the difference between the training and testing set is greater. In an attempt to report the results fairly, a range of the scores achieved on the testing set is given rather than a single percentage. The assumed best performing CNNs selected are those with the highest classification score on the validation set. The lowest and highest scores achieved on the testing set are then reported. The range result of the pose informed architecture has to take into account the different networks involved. For each orientation range, the best and worst performing CNN with the highest validation score are retained. The combination of the worst CNNs in each orientation range into one Pose-informed model gives the minimum of the classification rate achievable. The same process is adopted for the best CNNs in each orientation range.

In terms of computational complexity, it can be assumed that the segmentation, relying on thresholding mainly, has a low influence compared to the use of the two CNNs that enable the full orientation determination and classification. As a result, we can approximate the algorithm complexity to be close to $2.c$, c being the CNN complexity. In this paper, it is an AlexNet that is used that requires 725M floating point operations per seconds for a total of 61M parameters [46]. Further work could be dedicated to see how the Pose-informed architecture scales with shallower and less complex networks.

5 Results of the orientation determination

As described in Section 4, the orientation of the target has to be determined before it can be properly classified by the Pose-informed method. In this section, the target orientation estimation obtained with the improved Hough transform and the direction CNN is evaluated.

5.1 Target 180° orientation determination

The Hough transform achieves better results than those already reported (7.52° in Table 5 against 11.7° [47]) on the MSTAR SOC10 data, which shows the importance of a precise segmentation for this method to work well. The Hough transform scores could appear better if the best pose out of several potential poses was considered as in [22]. It achieves similar results to various geometrical pose estimation methods (MAE of 6.76° in Table 5 against 5.91° [47]).

The standard deviation is however higher for methods based on entropy or wavelets for a 180° estimation [24, 48]. The Hough transform method does not require training beforehand, whereas the entropy method needs training beforehand to optimise a maximum likelihood prediction of the entropy over the training images, and the wavelet method outperformed our method only when errors due to slant plane projection were compensated for with additional training. If such methods were chosen, then two trainings would be necessary in total as we would still need to lift the 180° ambiguity.

Overall, the best scores are achieved for the direct Hough transform with a vertical ratio. Indeed, the higher the error, the most likely the ratio will be over the threshold set in Section 4.1.1. Thus, the vertical ratio tackles strong errors around 90°.

The errors in the $0° - 40°$ range are, as in the MSTAR, mostly due to differences in illumination inside the target. Some areas are not included in the segmentation due to a too low intensity. This fake edge is picked up by the Hough transform as the line is longer than the real edge, being more or less on the target diagonal as seen in Fig. 7. This worsens the first estimation of the target orientation.

The best performing algorithm in Table 6 is the direct Hough transform with the vertical ratio.

5.2 Target 360° orientation determination

The orientation CNN distinguishing the target front from the back is evaluated on the testing data consisting of images rotated using the Hough transform estimation and thus with potential orientation errors. The target is not always horizontal in the images fed to the CNN.

When the difference between the final orientation and the ground truth angle is <90°, it is assumed that the CNN determined the correct target direction (front or back). On the contrary, if the error is >90°, then the error could be minimised by adding 180° to the target orientation. Metrics to evaluate the 360° orientation determination are given in Table 7 for the MSTAR and the MGTD.

In the two databases, the root mean square error (RMSE) increased compared to the previous section as the highest errors can now attain 180° instead of only 90°. The direction of the target is harder to determine in the MGTD images than in the MSTAR SOC10 AND EOC1 datasets. The full 360° orientation determination has rarely been investigated in the MSTAR datasets and, when it has been, it was often not on the standard datasets defined in Section 2. A statistical method has been tested on the MSTAR EOC2 and EOC3 [21]. Results were given with the Hilbert-Schmidt distance as in (3) with an equivalent error in degrees

$$d^2_{\mathrm{HS}} = 4 - 4\cos(\theta - \hat{\theta}) \tag{3}$$

where θ the same as in (1) and $\hat{\theta}$ is the 360° estimated target orientation.

However, as the cosine is not a linear function, the MAE cannot be obtained by a direct inversion, thus the squared Hilbert-Schmidt distance obtained with our method is calculated in order to be able to compare the results with this statistical method. A value of 0 corresponds to a perfect estimation of the orientation, 8 corresponds to the highest possible error equal to 180°. An average distance of 0.8 was achieved on the MSTAR EOC2 and 1.0 on the MSTAR EOC3 dataset, while the statistical method reported a distance of 1.7 on the MSTAR EOC2 and 2.0 on the MSTAR EOC3 dataset. The statistical method also assumed knowledge of the target type to achieve those results. The proposed method is thus more precise and requires less prior information.

6 Results of the Pose-informed architecture

6.1 Study of the potential border effects

In order to study the classification border effects between the different orientation ranges of the Pose-informed method, we plot the score of the correct class according to the difference between the target orientation and the closest orientation range border in Fig. 8 for each target in the MSTAR SOC10 testing set. Each point represents the classification result. When the distance between the target orientation and closest orientation range border is 0, the target is aligned with the border. For this analysis, a total of five Pose-informed CNNs are used as this number n of orientation ranges gives the best results overall in Section 6.2. The total span of each orientation range is thus 72°. It appears that there is no error increase, nor loss of confidence of the Pose-informed CNN while classifying targets with an orientation close to the border. We did not find a classification border effect for the Pose-informed method.

6.2 Results of the pose-informed architecture compared to the baseline CNN

Each table relates the scores achieved for the standard CNN and the proposed Pose-informed architecture. Results are reported for the MSTAR SOC10, MSTAR EOC1, MSTAR EOC2, MSTAR EOC3 and the MGTD in Table 8. The baseline CNN trained on the MSTAR SOC10 has no range for the standard CNN since the larger number of images in the validation set enabled a finer distinction between scores and only one CNN achieved the highest validation score.

The best rates for both methods on all datasets are: 97.56% for the standard CNN against 99.13% for the Pose-informed on the MSTAR SOC10, 85.06% against 88.97% on the MSTAR EOC1, 92.32% against 94.09% on the MSTAR EOC2, 95.24% against 93.54% on the MSTAR EOC3, 91.20% against 94.29% on the MGTD. Overall, the Pose-informed architecture outperforms the standard method, even though the amount of training data for the orientation-speciality transfer learning was very limited. Concerning the MSTAR EOC3, the Pose-informed architecture performs less than the standard CNN with a drop of 6% in the worst case scenario with two Pose-informed CNNs, and 3% for five Pose-informed CNNs.

The scores of the five CNNs Pose-informed method can also be compared with a two CNNs Pose-informed method to evaluate the importance of a higher number of CNNs and thus number of orientation ranges. In the MSTAR SOC10, the Pose-informed method with five CNNs performs 0.17% better than the Pose-informed method with only 2 CNNs. Similarly, it performs 4.86% better in the MSTAR EOC1, 0.79% better in the MSTAR EOC2, 3.14% better in the MSTAR EOC3, 3.09% better in the MGTD. The five CNNs Pose-informed architecture achieves better results than the two CNNs Pose-informed architecture. The two CNNs Pose-informed architecture has less possibility to adapt to a specific aspect angle as the images provided for training are less orientation specific. It would seem that the proposed method has indeed been able to learn extra information about specific orientations even without additional data by applying transfer learning in two stages. The Pose-informed method would probably significantly benefit from additional data because of the low number of images in the second training set resulting from the aspect angle partition of the training data.

It seems that the Pose-informed architecture with five CNNs performed the best overall. The highest score of the five orientation ranges Pose-informed architecture is always higher than that of the standard CNN with the exception of the MSTAR EOC3 dataset. The minimum score of the Pose-informed architecture is higher to that of the standard CNN in the MGTD, MSTAR SOC10, MSTAR EOC1. The lower scores of the Pose-informed architecture are less than that of the standard CNN in the MSTAR EOC2 and MSTAR EOC3. Thus, even if the worst performing CNNs from the Pose-

Table 5 Error statistics in the target 180° orientation determination in the MSTAR SOC10 database

Method to estimate the target orientation	Mean	σ	MAE	RMSE
direct Hough transform on the threshold segmented target	4.78	16.62	7.80	17.30
direct Hough transform on the threshold segmented target with vertical ratio	3.88	13.81	6.76	14.34

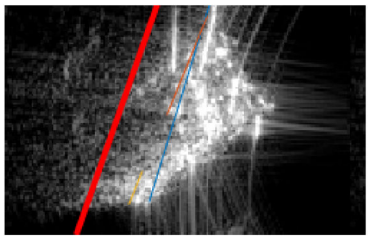

Fig. 7 *Potential drawback of the averaged Hough transform using the wrong edge of the target*

Table 6 Error statistics of the target 180° orientation determination in the MGTD

Method to estimate the target orientation	Mean	σ	MAE	RMSE
direct Hough transform on the threshold segmented target	2.98	16.59	6.79	16.85
direct Hough transform on the threshold segmented target with vertical ratio	2.31	7.86	4.37	8.14

Table 7 Error statistics of the target 360° orientation determination in the MSTAR database and MGTD

Database	Mean	σ	MAE	RMSE
MSTAR SOC10	−0.51	33.56	11.84	33.55
MSTAR EOC1	−0.08	24.71	11.23	24.70
MSTAR EOC2	−4.13	52.44	19.58	52.60
MSTAR EOC3	−5.84	60.97	24.51	61.25
MGTD	−1.46	45.29	14.34	45.28

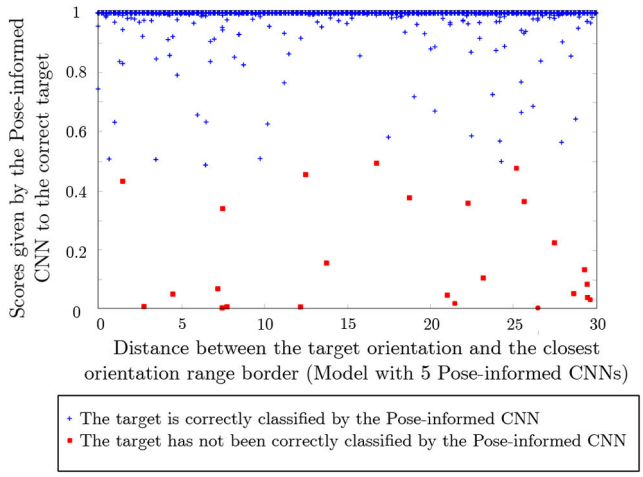

Fig. 8 *Analysis of the correct target score and errors according to the distance to the orientation range border on the MSTAR SOC10 testing set*

informed CNNs set are selected out of the CNNs with the best validation score, the method still achieves higher scores than the standard CNN method in three datasets out of 5. The worse results are achieved on the MSTAR EOC3 and this could be caused by the orientation determination results which achieved the worse results on this database in Section 5.2. If the target in one orientation is analysed by the CNN from another orientation range, the results could be worse than those of a standard CNN trained on the whole SAR training set.

6.3 Influence of the target orientation determination

In order to evaluate the influence of the precision of the orientation determination, the classification loss is evaluated when the classifying Pose-informed CNN is chosen randomly instead of being chosen according to the target orientation. Results are given on the MGTD in Table 9. The loss in the classification score increases with the number of Pose-informed CNNs. Indeed, as the number of Pose-informed CNNs increases, they become more and more specialised. If the full architecture contains only 2 Pose-informed CNNs, an error in the orientation range assignment results in the neighbouring range CNN classifying the image. Although with more Pose-informed CNN, the CNNs become less aware of the specificities of the further target orientations. Thus, the prospective studies could focus on the precision of the orientation determination which becomes crucial when it is an input for the classification.

7 Conclusion

In this paper, we propose a novel Pose-informed deep learning architecture which takes into account the target orientation in the classification process by SAR images. The target orientation is determined first, followed by the target classification using a CNN specialised in a certain target orientation range. This architecture is tested on five SAR and ISAR datasets that differs in terms of resolution, SNR, depression angle and acquisition environment. The CNNs, composing the architecture, are trained with nearly identical parameters and are generalised well over the two datasets.

The orientation determination is handled over 360° with a proposed association between a Hough transform, a study of the image intensity to limit 90° errors on rectangular targets and a CNN recognising the target direction. This orientation determination performs better over 360° and does not require prior knowledge on the target type, compared to existing statistically based methods.

The proposed Pose-informed architecture performs better than the standard CNN, except on the MSTAR EOC3 which has the poorest precision for orientation determination. It achieves respectively 99.01, 85.58, 94.09, 93.54 and 94.29% on the MSTAR SOC10, EOC1, EOC2, EOC3 and MGTD, with a delta compared to the standard CNN of + 2.14%, + 8.26%, + 1.77%, − 1.70% and + 3.09%.

It can be noted that in the five Pose-informed CNNs, classification errors do not depend on the closeness of the target orientation with the range border of the trained Pose-informed CNN. However, if the classification is systematically attributed to a CNN specialised in a non-relevant orientation range, the performances drop.

An area of improvement could be to take into account not only the orientation, but also other characteristics of the radar images. Further work could focus on the evaluation of the robustness of the model against occlusion by masking partly the target, against noise or the adaptation of the model to SAR characteristics such as speckle. In order to compare the robustness of several methods, the proposition of a standard dataset, either a new or an altered previous dataset, with different parameters such as resolution, speckle amount, occlusion would prove extremely useful.

8 Acknowledgments

The authors thank the MCM-ITP programme and Thales Optronique for funding this project. A special thanks to Odysseas Kechagias-Stamatis who provided technical help to this work.

Table 8 Range results (%) of the Pose-informed classification method compared to a standard CNN on the MSTAR SOC10, EOC1, EOC2 and EOC3 and MGTD

Dataset	Classification rate	Standard CNN	Number of Pose-informed CNN						
			2	3	4	5	6	7	8
MSTAR	maximum	96.87	98.80	99.13	98.85	98.97	98.97	98.60	99.01
SOC10	minimum		98.06	97.53	97.16	97.11	96.25	96.70	96.45
MSTAR	maximum	77.32	78.54	81.84	81.92	83.40	82.88	85.58	83.58
EOC1	minimum	66.81	69.85	71.85	70.72	71.24	71.07	70.37	69.68
MSTAR	maximum	92.32	93.30	93.50	93.64	94.09	94.00	93.89	94.03
EOC2	minimum	87.53	88.68	87.58	87.28	85.34	86.97	85.68	83.61
MSTAR	maximum	95.24	89.37	89.78	91.07	92.51	93.54	93.10	92.91
EOC3	minimum	89.48	79.37	78.67	77.60	72.99	72.21	74.95	71.62
MGTD	maximum	91.20	91.43	91.89	92.12	94.29	92.35	89.95	91.43
	minimum	82.19	89.38	88.35	84.02	86.07	84.36	82.42	85.84

Table 9 Evolution of the maximum classification scores on the MGTD depending on the choice of the classifying Pose-informed CNN

Assignment of the classifying Pose-informed CNN	Number of Pose-informed CNN						
	2	3	4	5	6	7	8
orientation determination	91.43	91.89	92.12	94.29	92.35	89.95	91.43
random	89.50	79.80	78.20	74.32	67.35	68.15	70.09

9 References

[1] Profeta, A., Rodriguez, A., Clouse, H.S.: 'Convolutional neural networks for synthetic aperture radar classification', in '*Algorithms for synthetic aperture radar imagery XXIII*', vol. **9843** (International Society for Optics and Photonics, USA., 2016), p. 98430M

[2] Chen, S., Wang, H., Xu, F., *et al.*: 'Target classification using the deep convolutional networks for SAR images', *IEEE Trans. Geosci. Remote Sens.*, 2016, **54**, (8), pp. 4806–4817

[3] Kechagias-Stamatis, O., Aouf, N., Belloni, C.: 'SAR automatic target recognition based on convolutional neural networks'. Int. Conf. on Radar Systems (Radar 2017), Belfast, 2017

[4] Al-Mufti, M., Al-Hadhramsi, E., Taha, B., *et al.*: 'SAR automatic target recognition using transfer learning approach'. 2018 Int. Conf. on Intelligent Autonomous Systems (ICoIAS), Singapore, Singapore, 2018, pp. 1–4

[5] Belloni, C., Aouf, N., Le-Caillec, J.M., *et al.*: 'SAR specific noise based data augmentation for deep learning'. 2019 IEEE Int. Radar Conf., Toulon, 2019

[6] Tian, S., Wang, C., Zhang, H., *et al.*: 'SAR object classification using the DAE with a modified triplet restriction', *IET Radar Sonar Navig.*, 2019, **13**, (7), pp. 1081–1091

[7] Kechagias-Stamatis, O., Aouf, N.: 'Fusing deep learning and sparse coding for SAR ATR', *IEEE Trans. Aerosp. Electron. Syst.*, 2018, **55**, (2), pp. 785–797

[8] Kang, M., Ji, K., Leng, X., *et al.*: 'Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder', *Sensors*, 2017, **17**, (1), p. 192

[9] Huan, R., Pan, Y.: 'Decision fusion strategies for SAR image target recognition', *IET Radar Sonar Navig.*, 2011, **5**, (7), pp. 747–755

[10] Zhang, F., Hu, C., Yin, Q., *et al.*: 'SAR target recognition using the multiaspect-aware bidirectional LSTM recurrent neural networks', arXiv preprint arXiv:170709875, 2017

[11] Pei, J., Huang, Y., Huo, W., *et al.*: 'SAR automatic target recognition based on multiview deep learning framework', *IEEE Trans. Geosci. Remote Sens.*, 2018, **56**, (4), pp. 2196–2210

[12] Keydel, E.R., Lee, S.W., Moore, J.T.: 'MSTAR extended operating conditions: A tutorial', in '*Aerospace/defense sensing and controls*' (International Society for Optics and Photonics, USA., 1996), pp. 228–242

[13] Novak, L.: 'State-of-the-art of SAR automatic target recognition'. The Record of the IEEE 2000 Int. Radar Conf., Alexandria, VA, USA., 2000, pp. 836–843

[14] Mossing, J.C., Ross, T.D.: '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, USA., 1998), pp. 554–566

[15] Yang, Y., Qiu, Y., Lu, C.: 'Automatic target classification—experiments on the MSTAR SAR images'. Sixth Int. Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS Int. Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005, Towson, MD, USA., 2005, pp. 2–7

[16] Belloni, C., Aouf, N., Le-Caillec, J.M., *et al.*: 'Comparison of descriptors for SAR ATR'. 2019 IEEE Radar Conf. (RadarConf19), Boston, 2019

[17] Doo, S.H., Smith, G.E., Baker, C.J.: 'Aspect invariant features for radar target recognition', *IET Radar Sonar Navig.*, 2016, **11**, (4), pp. 597–604

[18] Zhong, Y., Ettinger, G.: 'Enlightening deep neural networks with knowledge of confounding factors'. Proc. IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017, pp. 1077–1086

[19] Rowley, H.A., Baluja, S., Kanade, T.: 'Rotation invariant neural network-based face detection'. Proc. 1998 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), Santa Barbara, CA, USA., 1998, pp. 38–44

[20] Learn, A.W.: 'Target pose estimation from radar data using adaptive networks', in '*Air force Int of tech wright-pattersonafb OH school of engineering*', (Air force institute of Technology, Wright-Patterson AFB, OH, USA., 1999), pp. 1–116

[21] DeVore, M.D., Lanterman, A.D., O'Sullivan, J.A.: 'ATR performance of a rician model for sar images', in '*Automatic target recognition X*', vol. **4050** (International Society for Optics and Photonics, USA., 2000), pp. 34–46

[22] Voicu, L.I., Patton, R., Myler, H.R.: 'Multicriterion vehicle pose estimation for SAR ATR', in '*Aerosense'99*' (International Society for Optics and Photonics, USA., 1999), pp. 497–506

[23] Peng, L., Liu, X., Liu, M., *et al.*: 'SAR target recognition and posture estimation using spatial pyramid pooling within CNN'. 2017 Int. Conf. on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology, Beijing, China, vol. 10620, 2018, p. 106200W

[24] Principe, J.C., Xu, D., Fisher, J.W.: 'Pose estimation in SAR using an information theoretic criterion', in '*Algorithms for synthetic aperture radar imagery V*', vol. **3370** (International Society for Optics and Photonics, USA., 1998), pp. 218–230

[25] Zhao, Q., Xu, D., Principe, J.: 'Pose estimation of SAR automatic target recognition'. Proc. Image Understanding Workshop, Monterey, CA, USA., vol. 11, 1998

[26] Wagner, S.A.: 'SAR ATR by a combination of convolutional neural network and support vector machines', *IEEE Trans. Aerosp. Electron. Syst.*, 2016, **52**, (6), pp. 2861–2872

[27] 'Cranfield online research data, MGTD database', 2019. Available at https://doi.org/10.17862/cranfield.rd.7240742, accessed 1 September 2019

[28] Belloni, C., Balleri, A., Aouf, N., *et al.*: 'SAR image dataset of military ground targets with multiple poses for ATR', in '*Target and background signatures III*', vol. **10432** (International Society for Optics and Photonics, Warsaw, 2017), p. 104320N

[29] Gorham, L.A., Moore, L.J.: 'SAR image formation toolbox for MATLAB', in '*SPIE defense, security, and sensing*' (International Society for Optics and Photonics, USA., 2010), pp. 769906–769906

[30] Novak, L.M., Owirka, G.J., Weaver, A.L.: 'Automatic target recognition using enhanced resolution SAR data', *IEEE Trans. Aerosp. Electron. Syst.*, 1999, **35**, (1), pp. 157–175

[31] Nguyen, D.H., Kay, J.H., Orchard, B.J., *et al.*: 'Improving HRR ATR performance at low-SNR by multilook adaptive weighting', in '*Automatic target recognition XI*', vol. **4379** (International Society for Optics and Photonics, USA., 2001), pp. 216–228

[32] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks', in '*Advances in neural information processing systems*' (Curran Associates, Red Hook, NY, USA., 2012), pp. 1097–1105

[33] Toshev, A., Szegedy, C.: 'Deeppose: human pose estimation via deep neural networks'. Proc. IEEE Conf. on computer vision and pattern recognition, Columbus, OH, USA., 2014, pp. 1653–1660
[34] KarpathyA, T., Toderici, G., Shetty, S.*, et al.*: 'Large-scale video classification with convolutional neural networks'. IEEE Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA., 2014, p. 1725
[35] Girshick, R., Donahue, J., Darrell, T.*, et al.*: 'Rich feature hierarchies for accurate object detection and semantic segmentation'. Proc. IEEE Conf. on computer vision and pattern recognition, Columbus, OH, USA., 2014, pp. 580–587
[36] He, K., Zhang, X., Ren, S.*, et al.*: 'Deep residual learning for image recognition'. Proc. IEEE Conf. on computer vision and pattern recognition, Las Vegas, NV, USA., 2016, pp. 770–778
[37] Szegedy, C., Liu, W., Jia, Y.*, et al.*: 'Going deeper with convolutions'. Proc. IEEE Conf. on computer vision and pattern recognition, Boston, MA, USA., 2015, pp. 1–9
[38] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:14091556, 2014
[39] Deng, J., Dong, W., Socher, R.*, et al.*: 'Imagenet: a large-scale hierarchical image database'. CVPR09, Miami, FL, USA., 2009
[40] Kwak, Y., Song, W.J., Kim, S.E.: 'Speckle-noise-invariant convolutional neural network for SAR target recognition', *IEEE Geosci. Remote Sens. Lett.*, 2018, **16**, (4), pp. 549–553
[41] Ding, J., Chen, B., Liu, H.*, et al.*: 'Convolutional neural network with data augmentation for SAR target recognition', *IEEE Geosci. Remote Sens. Lett.*, 2016, **13**, (3), pp. 364–368
[42] Polyak, B.T.: 'Some methods of speeding up the convergence of iteration methods', *USSR Comput. Math. Math. Phys.*, 1964, **4**, (5), pp. 1–17
[43] Shin, H.C., Roth, H.R., Gao, M.*, et al.*: 'Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning', *IEEE Trans. Med. Imaging*, 2016, **35**, (5), pp. 1285–1298
[44] Jiang, Y., Zhu, X., Wang, X.*, et al.*: 'R2CNN: rotational region CNN for orientation robust scene text detection', arXiv preprint arXiv:170609579, 2017
[45] Dong, X., Huang, J., Yang, Y.*, et al.*: 'More is less: a more complicated network with less inference complexity'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA., 2017, pp. 5840–5848
[46] Jiang, Y., Zhao, X., Zhang, Y.*, et al.*: 'Pose estimation based on exploration of geometrical information in SAR images'. 2016 IEEE Radar Conf. (RadarConf), Philadelphia, PA, USA., 2016, pp. 1–4
[47] Kaplan, L.M., Murenzi, R.: 'Pose estimation of SAR imagery using the two dimensional continuous wavelet transform', *Pattern Recognit. Lett.*, 2003, **24**, (14), pp. 2269–2280