

SAR Automatic Target Recognition based on Convolutional Neural Networks

Odysseas Kechagias-Stamatis, Nabil Aouf, Carole Belloni

**Centre for Electronic Warfare Information and Cyber, Cranfield University, Defence Academy of the UK, Shrivenham, SN6 8LA, UK, o.kechagiasstamatis@cranfield.ac.uk*

Keywords: Automatic Target recognition, Convolutional Neural Networks, Deep Learning, Support Vector Machine, Synthetic Aperture Radar

Abstract

We propose a multi-modal multi-discipline strategy appropriate for Automatic Target Recognition (ATR) on Synthetic Aperture Radar (SAR) imagery. Our architecture relies on a pre-trained, in the RGB domain, Convolutional Neural Network that is innovatively applied on SAR imagery, and is combined with multiclass Support Vector Machine classification. The multi-modal aspect of our architecture enforces the generalisation capabilities of our proposal, while the multi-discipline aspect bridges the modality gap. Even though our technique is trained in a single depression angle of 17° , average performance on the MSTAR database over a 10-class target classification problem in 15° , 30° and 45° depression is 97.8%. This multi-target and multi-depression ATR capability has not been reported yet in the MSTAR database literature.

1 Introduction

The Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset is the most cited Synthetic Aperture Radar (SAR) imagery database used for Automatic Target Recognition (ATR) purposes. ATR is attempted based on various concepts such local feature matching [1], Compressive sensing [2], Sparse classification [3], Convolutional Neural Networks (CNN) [4]–[8] or Neural Networks (NN) [9]. CNN's, have been proven an extremely valuable asset for target recognition achieving state of the art performance. A downside of current approaches is that they rely on handcrafted CNN that have to be trained based on a limited number of public available SAR images. To compensate that, most current CNN based solutions populate the training images either by creating variants e.g. rotated versions of the existing templates or by sampling patches.

In contrast to the limited SAR imagery available, RGB images are easily found and in fact, literature suggests a great number of sophisticated and multi-layered pre-trained CNNs. Urged from that, we propose a multi-discipline and multi-modal architecture that combines the concepts of CNN and multiclass Support Vector Machine (SVM) classification. Our intention is to transfer the already proven classification capability of the AlexNet from the RGB domain to the X-

band SAR. This operation is not straightforward as directly activating the AlexNet with data of a different modality i.e. SAR imagery, is a suboptimal solution. Therefore, we bridge the data-modality gap by introducing a multi-discipline architecture.

2 Proposed architecture

Krizhevsky [10] with his pioneering work, has brought a new era in machine vision, setting new standards in pattern recognition. His CNN, named AlexNet is a 23-layered network that encapsulates from an RGB image features that vary from low-level corners and blobs, in the initial stages, up to high-level RGB oriented features in the last layers. Although AlexNet is powerful, it has been trained on RGB images that are completely different to SAR based imagery. In fact, AlexNet is trained on RGB colour bands while SAR imagery contains radar reflections of the scene. Therefore directly applying AlexNet on SAR imagery is not an optimum solution.

Thus, we group the 23 layers of AlexNet into eight clusters of varying feature description capability, introducing the *clustered AlexNet* architecture that is presented in Figure 1.

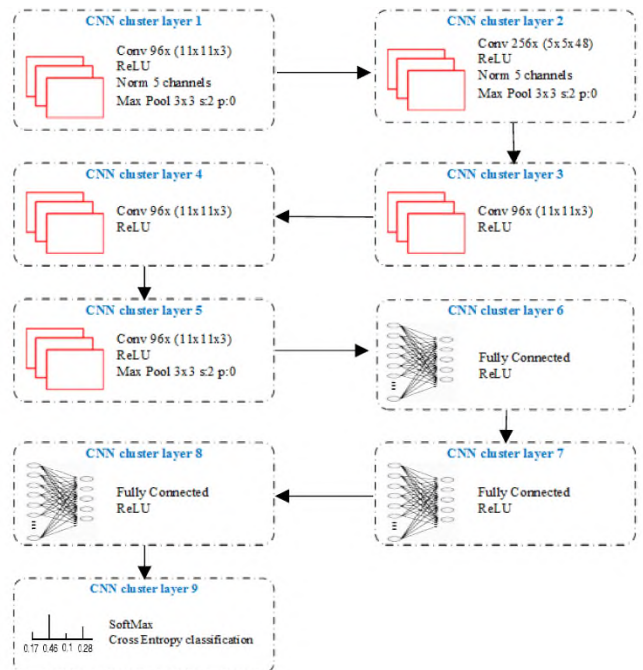


Figure 1: Clustered AlexNet architecture

Our *clustered AlexNet* CNN has a variable feature description capability that extends from low-level corners and blobs up to RGB specific features, depending on the cluster activated. Cluster nine includes the last two layers of the original AlexNet i.e. the CNN’s decision-making process which we will not exploit. Our CNN architecture uses the same parameters as in the original AlexNet [10]. Since AlexNet requires as input RGB imagery i.e. tri-layered data, we replicate each SAR image such as to imitate each colour-layer.

Then given a training set of the MSTAR dataset, we activate one of the clustered layers of the AlexNet, based on their pre-trained weights in the RGB domain. These activation responses are then used to train a multiclass SVM classification scheme. Final classification is based on a SoftMax function which is applied on the response of the SVM. Our ATR pipeline is depicted in Figure 2. An analysis of the optimum clustered layer selection is presented in section 4.1.

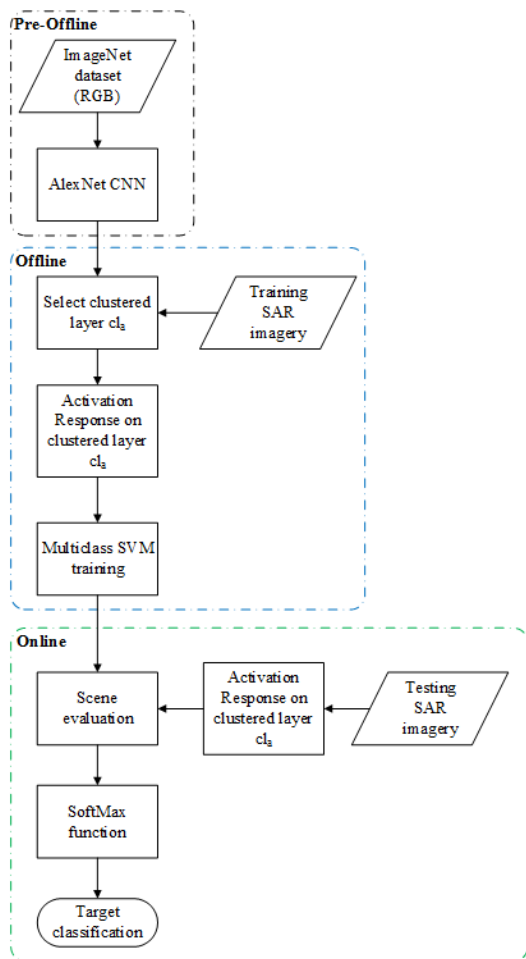


Figure 2: Proposed architecture

3 MSTAR database

We use the publicly available subset of the MSTAR database [11] that includes 10 different classes of ground targets as presented in Figure 3. Each class contains chips of 15° and 17° depression angles using an X-band SAR sensor, while some classes contain additional 30° and 45° viewings.

In any case, all chips cover a full 0°-360° azimuth orientation. For compatibility reasons with the current literature we follow the methodology of [11] and establish a training set based on the 17° chips and a testing set containing the 15°. Even though up to date literature [4]–[7], [12], [9] strictly follows that evaluation scheme, we perform an extra batch of trials in which we evaluate the performance on all available depression angles i.e. 15°, 30° and 45°. This is challenging because from Figure 3(b) it can be clearly seen that the target’s X-band reflection and thus its SAR fingerprint changes substantially.

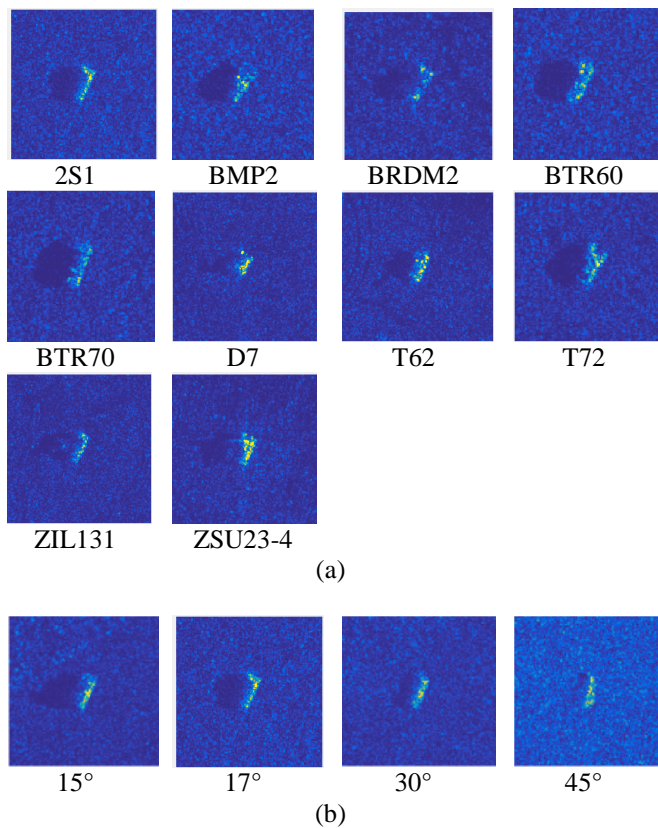


Figure 3: (a) 10 classes of the public MSTAR database at 17° depression angle (b) the 2S1 target at various depression angles while at same azimuth

4 Experiments

The 10-target classes have an unevenly distributed amount of images. Hence, we avoid settling the neural network into a local minimum by randomly duplicating the images of each class to meet the class having most instances. This

methodology populates our training set from 2747 SAR chips to 2890.

4.1 Optimising the Clustered AlexNet performance

Intuitively selecting a low-level cluster layer is not optimum and therefore we evaluate the performance of each cluster layer cl_a by activating it on the 17° training set and then evaluating it on the two testing sets i.e. 15° and $15^\circ, 30^\circ$ and 45° . Notation cl_a refers to the cluster layer activated with $a \in \{1, 2, 3, 4, 5, 6, 7, 8\}$. This means, for instance, cl_4 activates AlexNet’s clustered layer 4 while the remaining layers $\{5, 6, 7, 8, 9\}$ are discarded. Equally to [4]–[7], [12], performance is measured by the *mean Accuracy* metric (mA) over the 10-target classes.

The performance our our proposal is shown in Figure 4. As expected, exploiting all layers of the AlexNet does not pose an optimum solution as this CNN has been trained on encoding high level features belonging to 2D data of a different modality. As we are moving towards lower level feature extraction layers, performance improves because the basic nature of these encoders can handle the SAR modality. Peak performance of 98.1% mean Accuracy is obtained by activating cl_2 . Although activating cl_1 gains high performance, the features encoded are too generic and thus not distinctive among the 10-target classes. Similarly to the standard evaluation set (chips of 15°) our proposal can also handle well depressions that exceed quite a lot the training images (chips of $15^\circ, 30^\circ$ and 45°) which is unique in the MSTAR literature. This happens, because the low-level features of cluster layer cl_2 in combination with the relatively small training set and the multiclass SVM, enforce generalisation and avoid overfitting. In specific, cl_2 manages 97.8% mean Accuracy, while the overall performance has the same trend as for the 15° evaluation set enforcing the rigidity of your proposal. That multi-target multi-depression angle capability of our architecture is very significant because it substantially extends the ATR capabilities.

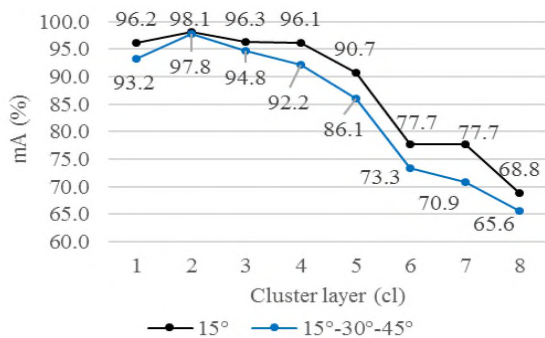


Figure 4: Recognition performance per cluster layer of the modified AlexNet

4.2 Comparison with state-of-the-art solutions

We compare and contrast the performance of the suggested clustered AlexNet at cl_2 with current state-of-the-art architectures. To the best of our knowledge, none of the current solutions tackles the simultaneous multi-target and multi-depression problem. In addition, in contrast to the vast majority of current SAR ATR oriented CNN’s or NN’s that use a handcrafted network¹, our proposal is based on the well established AlexNet. Although the ATR performance of each architecture is presented in Table 1, a direct comparison is not trivial. For example, even though AFRLNet [8] achieves 99% while our clustered AlexNet 98.1% (on the same testing set), the former includes only 210 testing images while our trials involve 3203. Another important point is the amount of SAR imagery data required during the training stage of each architecture. For example, A-ConvNet [7] manages 99.1% but it uses 27000 training images compared to 2890 of our suggested CNN architecture, which reveals that our suggestion is more effective as it requires fewer SAR images. Urged from that, we introduce the network *effectiveness* metric that exposes under a single value the ATR performance of a CNN in relation to the training and testing images used:

$$Effectiveness = mA \frac{\#testing\ images}{\#training\ images} \quad (1)$$

Algorithm	training images	testing images	Depression angle			mA (%)
			15°	30°	45°	
Chen [5]	2747	2425	✓			84.7
A-ConvNet [7]	27000	2425	✓			99.1
DCNN [4]	12146	5378	✓			99.5
Morgan [6]	3671	3203	✓			92.3
SGD [12]	3671	3203	✓			97.1
AdaGrad [12]	3671	3203	✓			88.0
AdaDelta [12]	3671	3203	✓			97.4
AFRLNet [8]	1410	210	✓			99.0
SAE [9]	2747	2426	✓			95.4
Ours	2890	3203	✓			98.1
Ours	2890	3794	✓	✓	✓	97.8

Table 1: mA performance along with training, testing and the depression angle of the test scenarios evaluated. Our architecture is the only that reports such a wide depression angle ATR capability

¹ AFRLNet [8] is also a mainstream CNN, but it has a low effectiveness as discussed in the current section

Figure 5 depicts the network effectiveness of each CNN presented in Table 1, where the higher the value the better. For fairness, we take into account the core SAR training set before manipulating it for better performance. That is, 2747 for the A-ConvNet², 6073 for the DCNN³ and 2747 for our proposal. From Figure 5 we clearly observe that our CNN architecture (green bar) is the most effective in the 15° trials as it achieves high performance, while is the only one that has a smaller SAR chip training set compared to the evaluation set. Considering our multi-depression evaluation (blue bar), this has an even greater effectiveness as it uses the same training dataset of 17° while the testing one is 18.5% larger including chips from three depression angles instead of only one. Least efficient is the AFRLNet because even though it performs well (in absolute numbers), its evaluation is based on only 210 images.

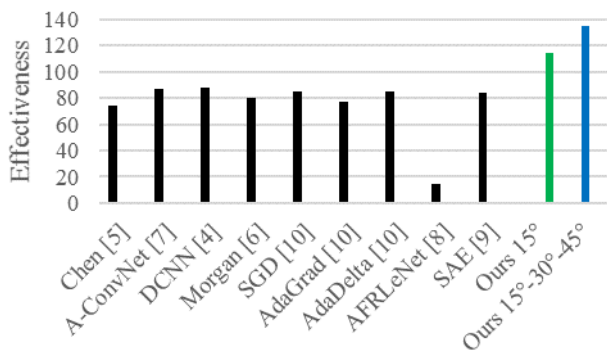


Figure 5: Network effectiveness per CNN architecture

5 Conclusion

We present a CNN architecture that is appealing for SAR ATR applications. Our strategy efficiently extends the pre-trained AlexNet from the RGB domain into the X-band SAR by clustering its neuron layers, obtaining the activation response from a training subset of the MSTAR dataset, and finally exploit these responses to train a multiclass SVM classification scheme.

Compared to current state-of-the-art CNN based architectures, ours is the only one that has multi-target multi-depression ATR capabilities. Depression angles extend from 15° up to 45°. This is feasible because our CNN strategy has the highest network effectiveness of all competitor CNN's which reveals its strong generalisation capabilities.

References

- [1] P. Mangalraj, M. A. Bisherwal, and A. Agrawal, "Target detection in SAR images using SIFT," in *2015 IEEE*

- International Symposium on Signal Processing and Information Technology, ISSPIT 2015*, 2015, pp. 90–94.
- [2] X. . Zhang, J. . Qin, and G. . Li, "SAR target classification using Bayesian compressive sensing with scattering centers features," *Progress in Electromagnetics Research*, **volume** 136, no. January, pp. 385–407, 2013.
- [3] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Transactions on Aerospace and Electronic Systems*, **volume** 48, no. 3, pp. 2481–2497, 2012.
- [4] Y. Zhong and G. Ettinger, "Enlightening Deep Neural Networks with Knowledge of Confounding Factors," *arXiv preprint arXiv:1607.02397*, pp. 1–10, 2016.
- [5] 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.
- [6] D. A. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," *SPIE Defense + Security*, **volume** 9475, p. 94750F, 2015.
- [7] H. Wang, S. Chen, F. Xu, and Y. Q. Jin, "Application of deep-learning algorithms to MSTAR data," *International Geoscience and Remote Sensing Symposium (IGARSS)*, **volume** 2015–Novem, pp. 3743–3745, 2015.
- [8] A. Profeta, A. Rodriguez, and H. S. Clouse, "Convolutional Neural Networks for Synthetic Aperture Radar Classification," in *SPIE Defense+ Security International Society for Optics and Photonics*, **volume** 9843, pp. 1–10, 2016.
- [9] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic Aperture Radar Target Recognition with Feature Fusion Based on a Stacked Autoencoder," *Sensors*, **volume** 17, no. 1, p. 192, Jan. 2017.
- [10] I. S. Alex Krizhevsky Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *the Neural Information Processing Systems Foundation 2012 conference*, pp. 1–9, 2012.
- [11] T. Ross, S. Worrell, V. Velten, J. Mossing, and M. Bryant, "Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set," *SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery V*, **volume** 3370, no. April 1998, pp. 566–573, 1998.
- [12] M. Wilmanski, C. Kreucher, and J. Lauer, "Modern Approaches in Deep Learning for SAR ATR," *Proc. SPIE*, **volume** 9843, p. 98430N–98430N–10, 2016.

² The 27000 training images are populated based on randomly sampling patches from the core training set of 2747 SAR chips

³ The 12146 training images contain 6073 SAR chips along with their left-right flipped version