Comparative Analysis of Convolutional Neural Networks and Support Vector Machines for Automatic Target Recognition

Ievgen M. Gorovyi and Dmytro S. Sharapov
Department of Microwave Electronics, Institute of Radio Astronomy of NAS of Ukraine
4 Chervonopraporna Str., Kharkov 61002, Ukraine
ceo@it-jim.com, sharapov@it-jim.com

***Abstract*—nowadays automatic methods based on artificial intelligence are rapidly growing. In the paper, a problem of automatic target recognition in synthetic aperture radar images is described. It is demonstrated, that two different machine learning instruments can provide very high classification accuracy. In particular, support vector machines with proper optimization and developed local feature set gives competitive results. Secondly, a novel architecture of convolutional neural network is proposed. Important practical aspects of both methods are analyzed. Experimental results for MSTAR are given.**

Keywords—automatic target recognition; feature extraction; support vector machines; convolutional neural networks, radar image

I. INTRODUCTION

Various intelligence systems are very popular nowadays. A key of automatic decision making is related with application of proper artificial intelligence (AI) algorithms. Computer vision and machine learning are two most used groups of applied techniques.

In the paper, the problem of automatic target recognition (ATR) in synthetic aperture radar (SAR) images is analyzed. Unlike to optical imagery, radar imaging has some peculiarities. In particular, speckle noise significantly complicates the information extraction process [1].

Brief analysis of existing literature indicates that ATR is quite old [2]-[3], but still very popular topic [4]-[6]. Different feature types, classifiers, image processing methods have been developed [7]. In [8] principal component analysis (PCA), independent component analysis (ICA) and Hu moment invariants were tested together with several classifiers: linear discriminant classifier (LDC), quadratic discriminant classifier (QDC), k-nearest neighbors (k-NN) and support vector machines (SVM). Comparative analysis was performed. In [9] cepstrum coefficients features were utilized as a feature vectors for ATR. In [10] Bayesian compressive sensing (BCS) technique was applied with scattering centers features. High-resolution range profiles were used in [11]. An example of appearance-based model was proposed in [12]. It was shown that competitive accuracy can be achieved with low dimensional feature vectors.

A special group of methods for SAR ATR is related with convolutional neural networks (CNNs). There is a high research trend in this field nowadays [4]-[6]. In this case, proper network architecture creation is the main question. In [5] it as shown that CNN can be used for target detection as well.

There are three main steps for SAR ATR (Fig. 1).

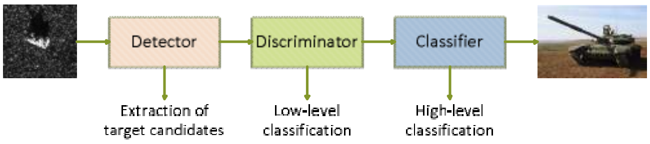


Figure 1. Key steps of ATR.

Firstly, detector is applied for extraction of target candidates. [2]-[3]. After that discriminator is applied. This is considered as a low-level binary classification. As a result, false positives such as buildings, trees and clutter are rejected. Finally, high-level classification is performed. As a result, target types are automatically determined. In this study, the classification step is investigated. In particular, we comprehensively analyze the potential of two different machine learning techniques: SVM and CNN. We demonstrate how to achieve high recognition accuracy using both techniques. It is shown, that proper tuning of features and SVM classifier optimization allows to increase the classification potential. In addition, we propose a novel CNN architecture giving outstanding ATR performance.

Section II contains the information about SVM and CNN. In particular, proposed feature sets and network layers description is given. Optimization steps and experimental results are discussed in section III.

II. MACHINE LEARNING TOOLS AND FEATURE EXTRACTION

This section contains an information about chosen machine learning tools: SVM and CNN.

978-1-5090-5390-2 /17/$31.00 ©2017 IEEE

A. Support Vector Machines and Custom Features

SVM is widely used technique for such tasks as pedestrian detection, handwritten text recognition [13], face recognition [14], etc. Initially this method was developed for a binary classification problem. The principle of SVM is to find the optimal separating hyperplane between two classes (Fig. 2).

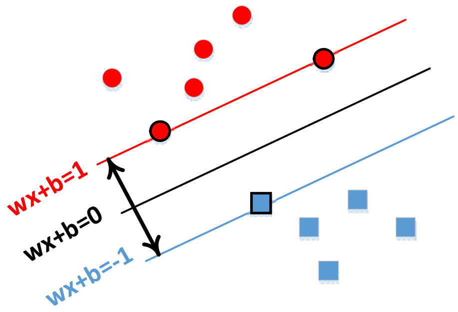


Figure 2. SVM principle.

The optimization problem is considered in a so-called soft-margin formulation [15].

$$\min(\|w\|^2) + C\sum_i \varepsilon_i, y_i(w^T x_i + b) \geq 1 - \varepsilon_i \qquad (1)$$

where x_i is a feature vector, ε_i are so-called slack variables for avoiding the overfitting, C is a penalty term. It is known, that SVM can be easily adopted for multiclass problem via construction of a set of one-against all binary classifiers [15]. Data samples close to the decision boundary play a crucial role in such consideration (support vectors).

A feature selection is an important step in target recognition algorithm. In the paper, we comprehensively analyze two types of local features: gray level co-occurrence matrices (GLCM) [16] and 1-D target profiles [17].

GLCM or Haralick features are commonly used for texture segmentation and classification problems. For a given image $I(x, y)$, the co-occurrence matrix P is defined as

$$P_{ij} = \sum_{x=1}^{N}\sum_{y=1}^{N}\begin{cases}1, & \text{if } I(x,y)=i \text{ and } I(x+\Delta_x, y+\Delta_y)=j,\\ 0, & \text{otherwise.}\end{cases} \qquad (2)$$

where the offset $(\Delta x, \Delta y)$ corresponds to the distance between the target pixel and corresponding neighbor. A default configuration of GLCM feature vector contains 3 offsets, 4 directions and averaged features among directions. For each co-occurence matrix, 13 statistical features are calculated. As a result, a default Haralick feature vector is 195-dimensional (5 directions * 3 offsets* 13 features). A proposed extesnsion of GLCM feature vector will be analyzed in the experimental section.

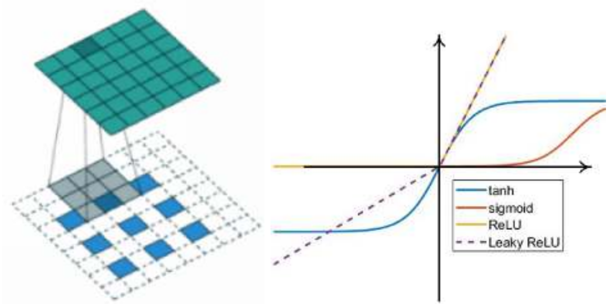
In addition, we analyze developed feature vector based on a fusion of 1-D target profiles (azimuth and range) as an input SVM classifier. There are several advantages of such local features. Firstly, a simple calculation as follows

$$I_A^x = \frac{1}{NY}\sum_{y=1}^{NY} I(x,y), I_R^y = \frac{1}{NX}\sum_{x=1}^{NX} I(x,y), \qquad (3)$$

where NX, NY are dimensions of image region of interest (ROI). Secondly, above feature are compact and suitable for real-time applications. Performance of the proposed fusion profiles and GLCM will be discussed in the next section.

B. Convolutional Neural Networks and Custom Architecture

Interesting peculiarity of CNN is automatic feature extraction [4]-[6]. Thus, a key question is a construction of a proper network architecture. Several building blocks (layers) are typically utilized. A convolutional layer (Fig. 3) contains a set of 2-D kernels of a particular sizes (typically the same). Parameter stride (distance between blue squares in Fig. 3a) controls the decimation of the outputs. In addition, zero padding (dotted lines in Fig. 3a) can be used to control the boundary effects and size of convolution outputs (green square). Filter size is 3*3 in Fig. 3.



(a) (b)

Figure 3. Typical CNN building blocks
(a – convolution, b – activation functions)

Commonly, convolution layers are connected with nonlinear activation functions. Fig. 3b contains several most widely used examples. It was shown [18] that rectified linear unit (RelU) provides significantly faster training convergence and demonstrates nonsaturating nonlinearity. LeakyRelU (Fig. 3b) is also often used modified version of this activation function.

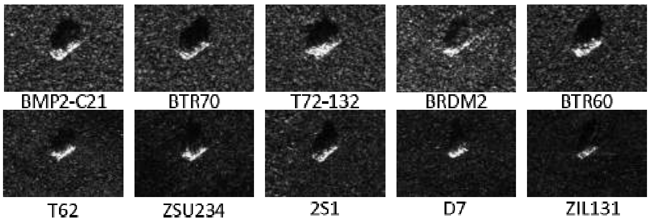
Two more layer types should be mentioned during CNN description. The max or average pull layer, which is applied for data reduction. It acts similarly to convolutional one, however instead of direct multiplication of image with kernel, the maximum of average value is returned as an output. Thus, the data volume is controlled. Another important layer type is dropout [19]. A key idea is to randomly drop out some units and their corresponding connections from network during training process. This helps to prevent CNN overfitting and works well as a regularization method. As for the training itself, stochastic gradient descent (SGD) is a good option.

Next section contains the results of ATR using both machine learning tools.

III. Experimental Results

A. Local Features Analysis

In this study, we used the moving and stationary target acquisition and recognition (MSTAR) database [20]. This is a public dataset of SAR images of 0.3m by 0.3m resolution. The images were taken over 360 degrees covering various target orientations. Dataset was obtained in two acquisitions with 3671 and 3203 samples respectively (10 classes). Fig. 4 illustrates corresponding sample examples from each class.



(b)

Figure 4. Images of 10 targets from MSTAR database.

Image samples have 128*128 size with centered targets.

Initially, we have examined GLCM features. Default configuration has provided quite low accuracy (around 79%). In order to increase the performance, we have analyzed an extended configuration with larger offsets $(\Delta x, \Delta y)$. As a result, fused feature vectors were formed. Fig. 5 illustrates the dependence of classification accuracy on the number of combined co-occurrence matrices.

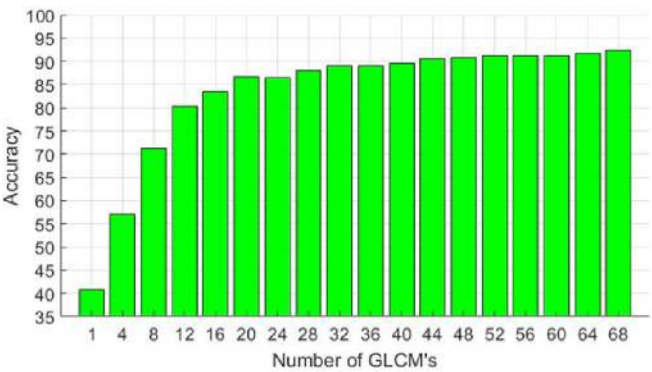


Figure 5. Recognition accuracy for extended GLCM features.

One can see that accuracy is improving for higher amount of co-occurrence matrices. The peak value is 92%, but the drawback is that feature vector dimensionality is very high, which complicates the real-time application of the above feature set.

At the next step, 1-D target profiles were tested. It was found that fusion of azimuth and range profiles gives better classification results. Moreover, optimal image preprocessing and ROI size have been determined. Fig. 6 contains the recognition results for different local feature sets. At first, one can see that effect of GLCM feature vector extension is noticeable. Also it outperforms a common local-binary patterns (LBP) [21]. It was found, that azimuth profile does not provide good discrimination power. Nevertheless, effect

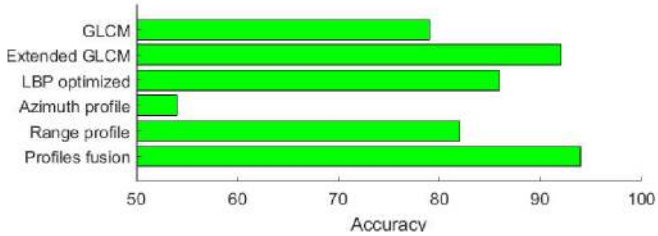


Figure 6. Recognition accuracy for different feature sets with SVM.

of its fusion with the range profile is positive, giving the performance improvement from 82% to 94%. One should notice that SVM parameters were optimized for each feature set. In particular, cross-validation of classifier was accomplished [15]. In addition, the best SVM kernel function was found among several alternatives. The radial basis function (RBF) [15] has provided the maximum accuracy for all feature sets.

Thus, two different feature sets provide competitive accuracy for MSTAR dataset. However, proposed profiles fusion scheme is better option due to high compactness and simple construction.

B. CNN Testing

After experimental analysis of CNN layers, kernel parameters and sizes, we have built a deep architecture for SAR ATR (Fig. 7).

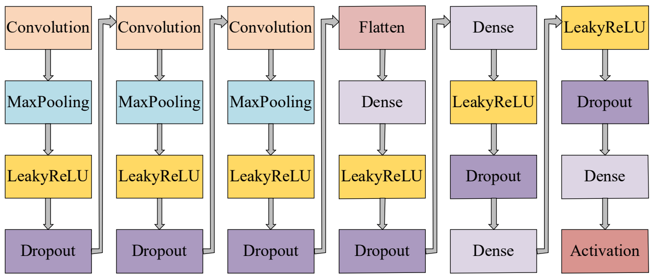


Figure 7. Proposed CNN architecture.

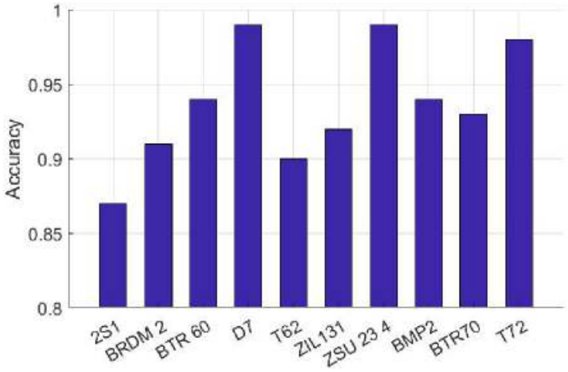
Proposed CNN contains three convolutions layers with kernel sizes 7*7, 5*5 and 3*3 respectively. Max pooling has shown higher efficiency than average pooling. Six dropout layers were used to carefully control overfitting.

In addition, data augmentation was performed. Additional image samples (128 samples) with random rotation were added into the training set. This led to minor accuracy improvement (around 0.3%). Also, it was found that ROI size has no effect of CNN accuracy. Target segmentation did not give any benefits.

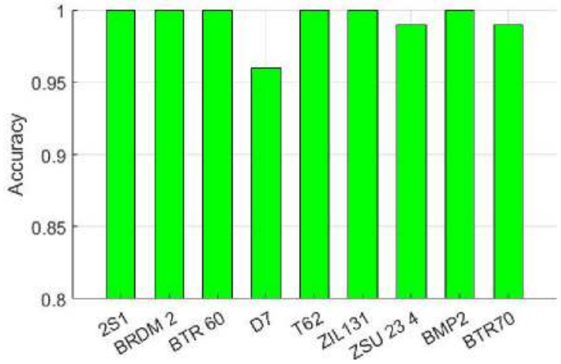
C. Results Discussion

After testing of SVM and CNN for target recognition, the final step is to compare the accuracies. Fig. 8 contains comparative results for 10 classes from MSTAR database.

The overall accuracy is 94.1% and 99.5% for SVM and CNN respectively. One can observe that CNN outperforms SVM



(a)



(b)

Figure 8. Recognition results
(a – SVM, b – CNN).

for 9 classes, while SVM provides 99% only for D7. This gives an idea to analyze the classifiers fusion effect. We are planning to study this problem in the near future.

IV. Conclusion

In the paper, a problem of automatic target recognition was examined. It was demonstrated, that two different machine learning methods provide promising classification results. It was shown, that optimized local features and properly tuned SVM classifier works well for quite challenging MSTAR dataset. A novel feature vector based on target profiles fusion was proposed. As a result, it outperformed an extended GLCM feature vector in terms of speed and accuracy. Moreover, a novel CNN architecture was constructed specifically for a given real database with radar targets. As a result, outstanding recognition accuracy of 99.5% was achieved.

References

[1] G. Franceschetti and R. Lanari, Synthetic Aperture Radar Processing. CRC Press, 1999.

[2] L. Novak, G. Owirka and C. Netishen, “Performance of a high-resolution polarimetric SAR automatic target recognition system”. Lincoln Laboratory Journal, Vol. 6, No. 1, pp. 11–24, 1997.

[3] D. Dugeon and R. Lacoss, “An overview of automatic target recognition”. Lincoln Laboratory Journal, Vol. 6, No. 1, pp. 3–9, 1993.

[4] E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification, IEEE Transactions on Geoscience and Remote Sensing, Vol. 55, No. 2, pp. 645-657, 2017.

[5] C. Sizhe; W. Haipeng; X. Feng and J. Ya-Qiu, “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.

[6] K. Du, Y. Deng, R. Wang, T. Zhao and N. Li, SAR ATR based on displacement- and rotation-insensitive CNN, Remote Sensing Letters, No. 7, Vol. 9, pp. 895-904, 2016.

[7] K. El-Darymli et al., Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review, IEEE Access, Vol. 4, pp. 6014-6058, 2016

[8] J. Park and K.-T. Kim, “Modified polar mapping classifier for SAR automatic target recognition”. IEEE Transactions on Aerospace and Electronic Systems, Vol. 50, No. 2, pp. 1092 – 1107, 2014.

[9] A. Eryildirim, Methods for automatic target classification in radar. Master thesis, Ankara, Bilkent University, 76 p, 2009.

[10] X. Z. Zhang, J. H. Qin, and G. J. Li, “SAR target classification using Bayesian compressive sensing with scattering centers features”. Progress in Electromagnetics Research, vol. 136, pp. 385–407, 2013.

[11] Wong, S. K., “High range resolution profiles as motion-invariant features for moving ground targets identification in SAR-based automatic target recognition”. IEEE Transactions on Aerospace and Electronic Systems, Vol. 45, No. 3, 1512–1524, 2009.

[12] J. Pei et al., “SAR imagery feature extraction using 2DPCA-Based two-dimensional neighborhood virtual points discriminant embedding”. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, Vol. 9, No. 6, pp. 2206-2214, 2016.

[13] S Ahmadkhani, and P Adibi, Face recognition using supervised probabilistic principal component analysis mixture model in dimensionality reduction without loss framework. IET Computer Vision, Vol. 10, No. 3, pp. 193-201, 2016.

[14] J. Ryu, H. Koo, N.Cho, Word Segmentation Method for Handwritten Documents based on Structured Learning, IEEE Signal Processing Letters, Vol. 22, No. 8, pp. 1161-1165, 2015.

[15] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.

[16] R .M. Haralick, “Statistical and structural approaches to texture”. Proc. IEEE, Vol. 67, No. 5, pp. 786-804, 1969.

[17] Ievgen M. Gorovyi and Dmytro S. Sharapov, Efficient Object Classification and Recognition in SAR Imagery, Proceedings of the 18th International Radar Symposium (IRS 2017), Prague, Czech Republic, June 28-30, pp. 1-7, 2017.

[18] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier networks,”in Proc. Int. Conf. Artif. Intell. Statist, pp. 315–323, 2011.

[19] N. Srivastava et. al, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, The Journal of Machine Learning Research, Vol. 15, No. 1, pp. 1929-1958, 2014.

[20] Y. Yang, Y. Qiu and C. Lu, “Automatic target classification experiments on the MSTAR SAR images”. Proceedings of the 6th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and 1st ACIS International Workshop on Self-Assembling Wireless Networks; pp. 2–7, 2005.

[21] . T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution gray-scale and rotation ¨ invariant texture classification with local binary pattern”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, pp. 971–987, 2002.