

Analysis On Radar Image Classification Using Deep Learning

B.Yamini Pushpa¹, S.Vishnupriya Chowdhary²

¹Dept. Of ECE Sri Padmavathi Mahila Viswavidyalayam, Tirupati, India

²Dept. Of ECE Sri Padmavathi Mahila Viswavidyalayam, Tirupati, India

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: The progress of the last 10 years of deep learning technology has inspired many fields of research, such as the processing of radar signal, speech and audio recognition, etc. Data representation acquired with Lidar or camera sensors are used for most prominent deep learning models, leaving automotive radars seldom used. Despite their vital potential in adverse weather conditions and their ability to seamlessly measure the range of an object and radial speed. Since radar signals have still not been used, the available benchmarking data is lacking. In the recent past, however, the application of radar data to various profound learning algorithms has been very interesting, since more datasets are being provided. This article aims to describe a new method of grading applied for the synthetic aperture radar (SAR), followed by fine tuning in such a grading scheme; Pre-trained architectures in the ImageNet database were used; the VGG 16 had actually been used as a feature extractor and the new classifier was trained based on the extracted features. The Dataset used is the data acquisition and recognition (MSTAR) of the Moving and Stationary Target; for ten (10) different classes we have achieved a final accuracy of 97.91 percent.

1. Introduction

The recent growth of data from different satellites has given rise to an enormous interest in advanced remote sensing techniques used for data mining to compute the extraction of data from their massive datasets from remote sensing [1]. A variety of specifications for remote sensing systems are spread between different satellite operators and manufacturers. Different products and their remote sensing applications exist. There are several applications to high-resolution satellite images. The applications include planning and mapping (tech, natural resources, urban, infrastructure), detection of exchanges, land use, tourism, crop management, military and environmental surveillance. They also enable us to resolve various problems, such as environmental monitoring and influencing anthropogenic factors, the detection of contaminated territories, unapproved buildings, forest planting status estimates, operational land resources surveillance, urban buildings, and other[2].

In nearly every computer vision problem, including in the remote sensing domain the classification of visual data is one of the most important steps. High resolution satellite image classification has many new topics in the field of remote sensing. The solution to the problem of the classification of VHR has played a decisive role in advanced methodologies[3] in recent years. There are two main questions in the classification process; firstly, how to identify the goal features; secondly, how to apply this ID to the new one. Since these features are mostly unusual manually, considerable efforts over the years have been made to develop automatic and discriminatory visual feature descriptors. The essence of machine learning techniques[4] is matching between old and new objects.

Over the last decade, autonomous driving and Advanced Driver Assistance Systems (ADAS) have been among the leading research domains explored in deep learning technology. Important progress has been realized, particularly in autonomous driving research, since its inception in 1980 and the DARPA urban competition in 2007 [1,2]. However, up to now, developing a reliable autonomous driving system remained a challenge [3]. Object detection and recognition are some of the challenging tasks involved in achieving accurate, robust, reliable, and real-time perceptions [4]. In this regard, perception systems are commonly equipped with multiple complementary sensors (e.g., camera, Lidar, and radar) for better precision and robustness in monitoring objects. In many instances, the complementary information from those sensors is fused to achieve the desired accuracy [5].

Synthetic opening radar (SAR) can operate under multiple conditions and produce large-scale images. A multiple noise known as speckle noise affects the images produced, It is very difficult and extremely complex to interpret and understand SAR images. A variety of approaches are being developed to make it less time consuming and more practical for SAR images to be understood and therefore overcome associated difficulties[1]. Over the last few years Deep learning algorithms have contributed in a number of successful computer vision tasks[3] such as classification, detection and localization[4] and especially Deep Convolutional Neural Networks (Convnet). Deep Convnets are able to automatically extract features from images using convolutions and pooling layers, as opposed to traditional classification tasks[1].

2. Related work

When processing very high-dimensional data, such as satellite images, deep learning algorithms are computationally cheap. This is due probably to the slow learning process associated with an increased number of structured layered learning data hierarchies. This structure comprises abstractions and depictions from a lower to a higher layer. Deeper learning techniques have become an active research theme in remote sensing communities to help them to classify satellite images. The recent accessibility of high spatial and spectral resolutions acquired by the new generation of satellites is particularly encouraging. These techniques are used for all satellite image classification applications. The following work overview addresses the issue of data extraction and representation with various deep learning techniques and revolutionary neural networks.

Multi-sensor fusion refers to the technique of combining different pieces of information from multiple sensors to acquire better accuracy and performance that cannot be attained using either one of the sensors alone. Readers can refer to for detailed discussions about multi-sensor fusion and related problems. Based on the conventional fusion algorithms using radar and vision data, a radar sensor is mostly used to make an initial prediction of objects in the surroundings with bounding boxes drawn around them for later use. Then, machine learning or deep learning algorithms are applied to the bounding boxes over the vision data to confirm and validate the presence of earlier radar detections. Moreover, other fusion methods integrate both radar and vision detections using probabilistic tracking algorithms such as the Kalman filter or particle filter, and then track the final fused results appropriately.

Over the recent years, some radar signal datasets are being reported for public usage. As a result, many researchers have begun to apply radar signals as inputs to various deep learning networks for object detection, object segmentation, object classification, and their combination with vision data for deep-learning-based multi-modal object detection. This paper specifically reviewed the recent articles on deep learning-based radar data processing for object detection and classification. In addition, we reviewed the deep learning-based multi-modal fusion of radar and camera data for autonomous driving applications, together with available datasets being used in that respect

In an end-to-end processing schema, Marmanis et al.[3] introduced pre-trained ImageNet networks to address the limited-data problem. Zhang et al. developed a hierarchical discriminatory learning algorithm for the deformation to the spatial-pyramid-matching model of hyperspectral image classification. Chan et al. used the PCA to learn multi-stage filtering banking and block histogram indexing and pooling of the filter. Kussul et al. introduced a multi-level architecture to target land cover and multi-source imagery classification.

Yao et al. proposed the stacked sparse autoencoder for the purpose of learning high-level features in an auxiliary satellite image set and then transferring the high-level features learned in semantime annotation. Mei et al. have used a five-layer CNN to learn classification features using advances in the deep learning sector, such as standardization, drop-out and the activation function Parametric Rectifiable Linear Unit (PReLU). Ferreira et al. introduces an Enhanced Regional Classification (RSI) Technology which allows encoding of features from different spectral and spatial domains.

3. Methodology

Convnets consists of several Convolution and Pooling layers and a final completely connected classification layer (Fig. 1). A large amount of data is required to train such a deep architecture. For the MSTAR information we used 2746 training pictures and 2425 test images, each with ten different classes (Tab. 1)

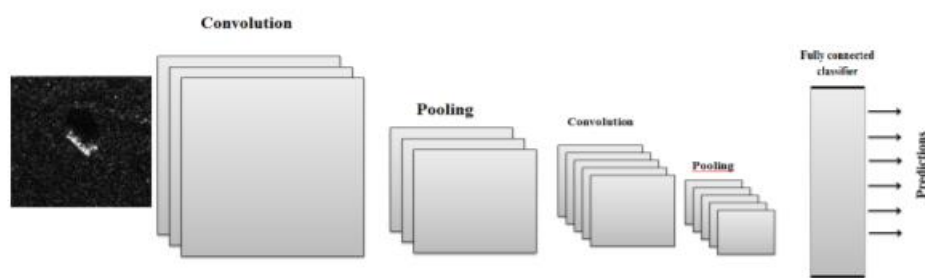


Fig. 1. Example of a Convolution Neural Network

Classes	training samples	testing samples
2S1	299	274
BMP2	233	195
BRDM_2	298	274
BTR60	256	195
BTR70	233	196
D7	299	274
T62	298	273
T72	232	196
ZIL131	299	274
ZSU_23_4	299	274

Table 1. MSTAR Data

Fig.2 presents visual samples of the MSTAR data. First, there are no pre-processing algorithms for the loaded images on the VGG16 convnet.

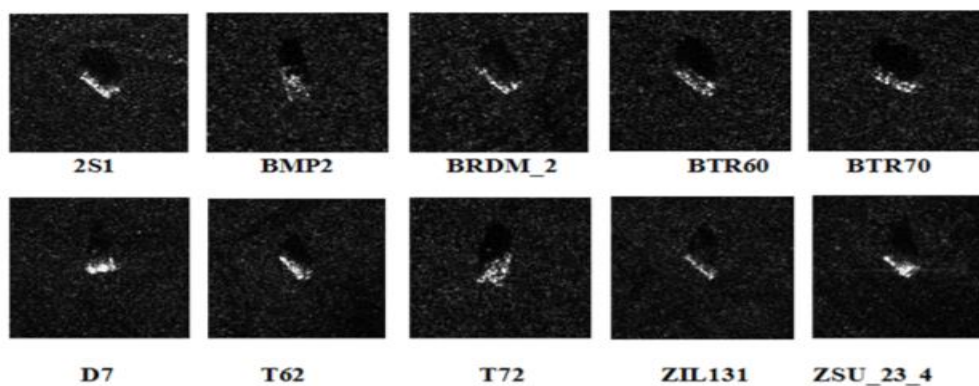


Fig. 2. Images from the MSTAR data

The steps to extract functions from the SAR images follow: The following:

- 1- Load the ImageNet training network of the VGG16 Network [7] (Fig.3)
- 2- Remove the complete classifier
- 3 - Use the remaining layers for training and test sets as a functional extractor (Fig.4)
- 4- Train a new fully integrated classifier based on the features extracted (Fig.4).

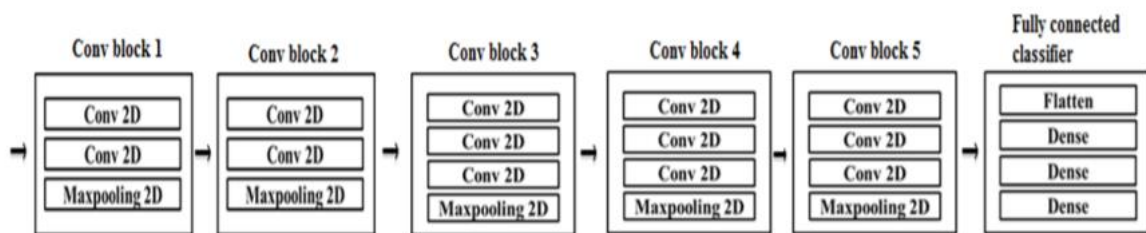


Fig. 3. VGG 16 Architecture

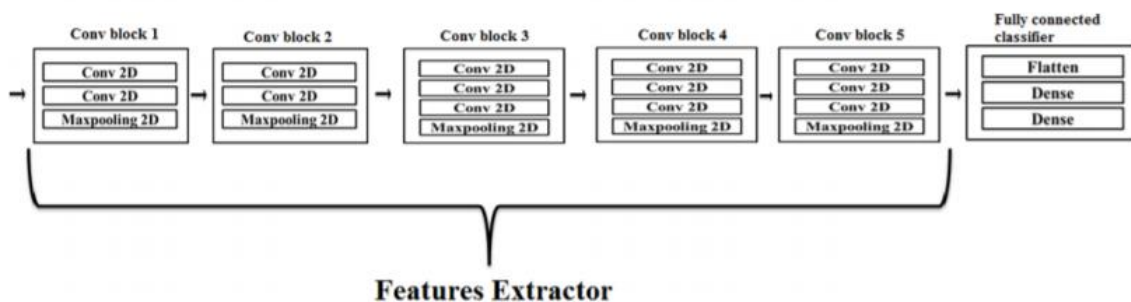


Fig. 4. Feature Extractor + New classifier

4. Overview of deep learning

This section provides an overview of the current neural network frameworks widely employed in computer vision and machine learning-related fields that could also be applied for processing radar signals. This spans across different models on object detection and classification.

Over the last decade, computer vision and machine learning have seen tremendous progress using deep learning algorithms. This is driven by the massive availability of publicly accessible datasets, as well as the graphical processing units (GPUs) that enable the parallelization of neural network training. Overwhelmed by its successes across different domains, deep learning is now being employed in many other fields, including signal processing, medical imaging, speech recognition, and much more challenging tasks in autonomous driving applications such as image classification and object detection.

However, before we dive into the deep learning discussion, it is important to talk about the traditional machine learning algorithm briefly, as it is the foundation of deep learning models. While deep learning and machine learning are specialized research fields in artificial intelligence, they have significant differences. Machine learning utilizes algorithms to analyze a given data, learn from it, and provide the possible decision based on what it has learned. One of the famous problems solved by machine learning algorithms is classification, where the algorithm provides a discrete prediction response. Usually, the machine algorithm uses feature extraction algorithms to extract notable features from the given input data and subsequently make a prediction using classifiers. Some examples of machine learning algorithms include symbolic methods such as support vector machines (SVM), Bayesian networks, decision trees, etc. and nonsymbolic methods such as genetic algorithms and neural networks.

On the other hand, a deep learning algorithm is structured based on the multiple layers of artificial neural networks, inspired according to the way neurons in the human brain function. Neural networks learn from the input data high-level feature representations, which are used to make intelligent decisions. Some common deep learning networks include deep convolutional neural networks (DCNNs), recurrent neural networks (RNNs), autoencoders, etc.

The most significant distinction between deep learning and machine learning is its performance, given the large amount of data available. However, when the training data is less, the deep learning performance is not that much. This is because they do need a large volume of datasets to learn perfectly. On the other hand, the classical machine learning methods perform significantly well with small data. Deep learning network functionality depends on powerful high-end machines. This is because deep learning models are composed of many parameters that require a longer time for training. Thus, they perform complex matrix multiplication operations that can be easily realized and optimized using GPUs, while, on the contrary, machine learning algorithms can work efficiently well even on low-end machines such as CPUs.

Training Deep Learning Models

Deep learning employs the Backpropagation algorithm to update the weights in each of the layers during the course of the learning process. The weights of the network are usually initialized randomly using small values. Given a training sample, the predictions are obtained based on the current weight's values, and the outputs are compared with the target variable. An objective function is utilized to make the comparisons and estimate the error. The error obtained is fed back into the network for updating the network weights accordingly. More information on Backpropagation can be found

Deep Neural Network Models

Here, we provide an overview of some of the popular deep neural networks utilized by the research communities, which include the deep convolutional neural networks (DCNNs), recurrent neural networks (RNNs), long short-term memory (LSTM), encoder-decoder, and the generative adversarial networks (GANs).

5. Detection and classification of radar signals using deep learning algorithms

This section provides an in-depth review of the recent deep learning algorithms that employ various radar signal representations for object detection and classification in both ADAS and autonomous driving systems. One of the most challenging tasks in using radar signals with deep learning models is representing the radar signals to fit in as inputs to the various deep learning algorithms.

In this respect, many radar data representations have been proposed over the years. These include radar occupancy grid maps, Range-Doppler-Azimuth tensor, radar point clouds, micro-Doppler signature, etc. Each one of these radar data representations has its pros and cons. With the recent availability of accessible radar data, many studies have begun to explore radar data to understand them extensively. Thus, we based our review article on this direction. Figure 9 illustrates an example of the various types of radar signal representations

5.1. Radar Occupancy Grid Maps

For a host vehicle equipped with radar sensors and drives along a given road, radar sensors can collect data about its motion in that environment. At every point in time, radars can resolve the object's radial distance, the azimuth angle, and the radial velocity that falls within its field of view. Distance and angle (both elevation and azimuth) entail more about the target's relative position (orientation) concerning the ego vehicle coordinate system. Simultaneously, the target's radial velocity obtained from the Doppler frequency shift will aid in detecting the moving targets.

Hence, based on the vehicle pose, radar return signals can be accumulated in the form of occupancy grid maps from which algorithms in machine learning and deep learning can be utilized to detect the objects surrounding the ego vehicle. In this way, both static and dynamic obstacles in front of the radar can be segmented, identified, and classified. The authors of discussed different radar occupancy grid map representations. The grid map algorithm's sole purpose is to determine the probability of whether each of the cells in the grids is empty or occupied.

$$P(M_t | Z_{1:t}, X_{1:t}) = \prod_i P(m_i | Z_{1:t}, X_{1:t})$$

A Bayes filter is typically used to calculate the occupancy value for each cell. Mainly, a posterior log formulation is used to integrate each of the new measurements for convenience. Even though CNNs function extraordinarily well on images, they can also be tried and applied to other sensors that can yield image-like data. The two-dimensional radar grid representations accumulated according to different occupancy grid map algorithms have already been exploited in deep learning domains for various autonomous system tasks, such as static object classification and dynamic object classification. In this case, the objects denote any road user within an autonomous system environment, like the pedestrian, vehicles, motorcyclists, etc.

5.2 Radar Range-Velocity-Azimuth Maps

Having talked about radar grid representations in the previous section, as well as their drawbacks, especially in detecting moving targets. It will be essential to explore other ways to represent the radar data so that more information can be added to achieve a better performance. A radar image created via multidimensional FFT can preserve more informative data in the radar signal, as well as conforms to the required 2D grid data representation applicable to the deep learning algorithms like CNNs

Many kinds of radar image tensors can be generated from the raw radar signals (ADC samples). This includes the range map, the Range-Doppler map, and the Range-DopplerAzimuth map. A range map is a two-dimensional map that reveals the range profile of the target signal over time. Therefore, it demonstrates how the target range changes over time and can be generated by performing one-dimensional FFT on the raw radar ADC samples. In contrast, the Range-Doppler map is generated by conducting 2D FFT on the radar frames. The first FFT (also called range FFT) is performed across samples in the time domain signal, while the second FFT (the velocity FFT) is performed across the chirps. In this way, a 2D image of radar targets is created that resolves targets in both range and velocity dimensions.

Conclusion

Object detection and classification using Lidar and camera data is an established research domain in the computer vision community, particularly with the deep learning progress over the years. Recently, radar signals are being exploited to achieve the tasks above with deep learning models for ADAS and autonomous vehicle applications. They are also applied with the corresponding images collected with camera sensors for deep learning-based multi-sensor fusion. This is primarily due to their strong advantages in adverse weather conditions and their ability to simultaneously measure the range, velocity, and angle of moving objects seamlessly, which cannot be achieved or realized easily with cameras. This review provided an extensive overview of the recent deep learning networks employing radar signals for object detection and recognition. In addition, we also provided a summary of the recent studies exploiting different radar signal representations and camera images for deep learning-based

multimodal fusion. Radar point clouds are also projected onto the image plane or birds-eye view using the coordinate relationships between radar and camera sensors creating pseudo-radar images. As a result, the generated images are used as the input for the deep learning networks. However, to the best of our knowledge, and at the time of writing this paper, we did not come across a study that uses this type of radar signal representation as the input for any deep learning model, whether for detection or classification. Many of the existing papers with this type of radar signal projections were about a multi-sensor fusion of radar and camera.

References

1. M. Das and S. K. Ghosh, "Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1984–1988, 2016.
2. D. M. M. Hordiiuk and V. V. V. Hnatushenko, "Neural network and local laplace filter methods applied to very high resolution remote sensing imagery in urban damage detection," in *2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF)*, 2017, pp. 363–3.
3. D. Marmanis, M. Datcu, T. Esch, U. Stilla, and S. Member, "Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 1–5, 2015.
4. Y. Yang and S. Newsam, "Bag-of-visual-words and Spatial Extensions for Land-use Classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.
5. A. Romero, C. Gatta, G. Camps-valls, and S. Member, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1–14, 2015.
6. G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, 2016.
7. A. C. Correlation, "Class-Specific Random Forest With Cross-Correlation Constraints for Spectral – Spatial Hyperspectral Image Classification," vol. 14, no. 2, pp. 257–261, 2017.
8. Z. Wu, W. Lin, Z. Zhang, A. Wen, and L. Lin, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis," *22017 IEEE Int. Conf. Comput. Sci. Eng. IEEE Int. Conf. Embed. Ubiquitous Comput.*, vol. 5, pp. 531–536, Jul. 2017.
9. T. L. M. Barreto et al., "Classification of Detected Changes From Multitemporal High-Res Xband SAR Images: Intensity and Texture Descriptors From SuperPixels," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 12, pp. 5436–5448, 2016.
10. B. Zheng, S. W. Myint, P. S. Thenkabail, and R. M. Aggarwal, "A support vector machine to identify irrigated crop types using time-series Landsat NDVI data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 34, no. 1, pp. 103–112, 2015.