SAR Target Classification using Deep Learning

A thesis submitted in partial fulfillment of
the requirements for the degree of

Bachelor of Technology

by

Sai Bhaskar Devatha & Ratikant Patil
(Roll No. 150108012 & 150108027)

Under the guidance of
Dr. Hanumant Singh Shekhawat

DEPARTMENT OF ELECTRONICS & ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
November 2018

Abstract

Synthetic aperture radar (SAR) is an imaging radar that produces high-resolution images. Classifying targets that are detected within a SAR image find many applications in environment monitoring and mapping. In this project an approach is proposed to classify targets that are detected in a SAR image. In this approach, dataset is passed through CNN to automatically extract features. These feature vectors are then passed to softmax classifier for its training. Examinations on the general MSTAR informational collection demonstrates that an exactness of 99% can be accomplished when classification is done on two kinds of targets, and a precision of 98% on seven sorts of targets.

Contents

List of Figures

Nomenclature

SAR	Synthetic Aperture Radar
CNN	Convolutional Neural Network
MSTAR	Moving and Stationary Target Acquisition and Recognition
SAIP	Semi-automated Image Intelligence Processing System
CONV	Convolutional Layer
POOL	Pooling Layer

Chapter 1

Introduction

A synthetic-aperture radar (SAR) is an imaging radar stationed on a moving airborne vehicle (such as an aircraft or a spacecraft). Like a typical radar, SAR transmits and receives electromagnetic waves sequentially and the corresponding echoes are collected. Transmission and reception occur at successive intervals and hence are mapped to different positions (due to fast moving aircraft). This creates a virtual perception of an antenna that is much longer than the actual size of the antenna. The distance travelled by the aircraft in simulation of this antenna is known as synthetic aperture. The system stores the data for further processing. Signal processing of this recorded data creates higher-resolution images that would otherwise be impossible with a physical antenna. SAR images have wide applications in environment monitoring, remote sensing and mapping of the surfaces. One of the major monitoring application lies in military surveillance. It is of high value due to its persistent operation in any weather condition or any time of the day.

Interpretation of SAR images requires trained interpreters because of the noise and speckled SAR imagery. This manual method is too time consuming and identification of smaller sized targets is impractical. This poses the need for target classification algorithms for SAR images.

Different approaches have been made to implement target classification in SAR images [1]. One of the first methods was SAIP (semi-automated image intelligence processing system) program [2]. In this approach, mean squared error was used to find the best match between target data and testing database. The testing accuracy proved out to be satisfactory when tested under similar conditions as in training database environment, but varied drastically when the targets were in extended varied conditions. These conditions include variations in background environment, target camouflaging, image articulation, etc.

To solve this problem, a model-based Moving and Stationary Target Acquisition and Recognition (MSTAR) system was developed [3]. The main characteristics of the model-based methods are it's front-end and back-end. Given a SAR image chip, first, the front-end of the MSTAR system runs a typical pattern recognition check. If not found in the front-end, the target is sent to back-end of the system, where the features of the image chip are compared with the predicted features.

With the beginning of the age of machine learning, classifier models have been used in speech recognition and computer vision [4]. Machine learning algorithms have been applied in SAR target classification studies. The classification problem was now reduced to handcrafting set of features which can be used to represent targets and then training a classifier model using these feature vectors. In machine learning, most of the work is to choose the correct set of features. Since it is task specific, the designed features must well discriminate among different targets.

To tackle this problem, a more robust method was required. With great results of deep learning theory in fields of computer vision and speech recognition, efforts have been made to apply deep convolutional neural networks (CNNs) to SAR target classification study [5] [6]. CNN has achieved remarkable classification accuracies on different tasks, such as handwritten digits or Latin and Chinese character recognition, Traffic sign recognition, Face detection and recognition [7] [8] [9] [10]. Deep neural networks are shown to be powerful for image recognition tasks because they automatically extract high level features that serve as representation for the targets.

1.1 Proposed Approach

The methodology we pursued as appeared in fig.1. At first, the dataset we have is divided into two sections for training and testing the neural network. The training data is then expanded and sent to the CNN where the highlights are extricated and after that the highlights are sent to a softmax classifier. At that point testing data was then used to assess the model.

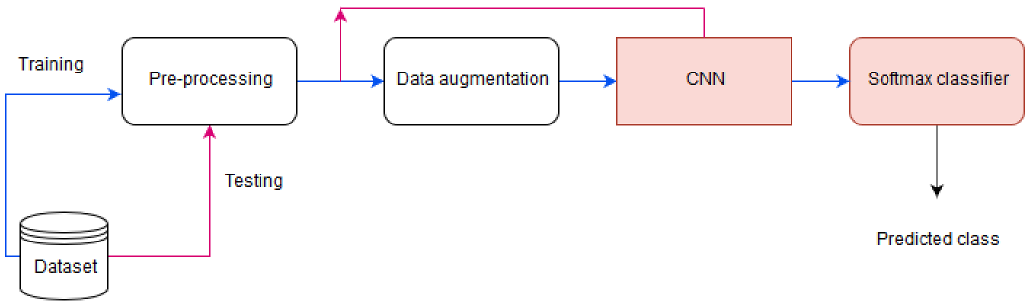


Figure 1.1: Proposed Approach

1.2 Outline of the Report

In the second chapter, we have described the background of our proposed approach. We have introduced the basic application of CNNs to SAR image target classification. We have also described the scalability of the used method to improve classification. In the third chapter, we have described the process by which the extracted SAR target data is classified using CNNs and results have been discussed based on experiments. In chapter four, we have concluded our report and described the future scope of the project.

Chapter 2

Background

In this chapter, we shall discuss the core concepts required for this project.

2.1 Synthetic Aperture Radar

Synthetic Aperture Radar (SAR), is an imaging radar system which uses the motion of airborne vehicle (over which it is mounted) over a target to simulate a long, virtual antenna. SAR generates high-resolution images and these images can be utilized for mapping and monitoring. Generally, SAR images are 2-dimensional. One dimension is called range or cross-track. Range is a measure of distance between the radar and target along the line of sight. Range measurement and resolution in SAR is same as in any typical radar. The other dimension is called azimuth or along-track. This is perpendicular to range. Azimuth resolution in SAR is finer than that in conventional radars. This is due to two reasons. First reason: Longer antennas give higher range and azimuth resolutions. Since SAR operates while flying, this creates a scenario as if the data is accumulated from a physically long antenna. The distance travelled by the aircraft in simulation of this antenna is known as synthetic aperture. Second reason: Doppler processing of the echoes. The target's Doppler frequency is the measure of it's azimuth location. In the line of motion of the SAR carrying aircraft, if the target lies in front of the aircraft, it produces positive Doppler offset, and if it lies behind the aircraft, it produces negative Doppler offset.

Many factors influence the properties of reflected electromagnetic wave, and hence also the properties of resulting SAR image. Physical factors include dielectric constant of the surface materials, surface type (man-made, vegetation, soil). Geometric factors are slope, orientation of the objects relative to the radar beam direction and roughness. Other factors are microwave frequency, polarization and incident angle.

Microwave Frequency

Microwave frequency is the measure of penetrating power of an electromagnetic wave. Lower the microwave frequency, higher is it's penetrating power.

Microwave Polarization

Microwave polarization refers to the electric field component of an electromagnetic wave. Electric field can be either horizontal (H polarized) or vertical (V polarized).

Incident Angle

The incident angle refers to the angle between incident radar beam and the direction perpendicular to the ground surface. This corresponds to the orientation between target and radar beam, which is an important factor in the resulting SAR image.

2.2 Augmentation

Plentiful high-quality data is the key to machine learning. Unfortunately, acquiring more data is always costly in terms of time or money. The scarcity of data can hinder the development of a good model. Data augmentation is a solution to this. The model becomes more robust and understandable after augmentation when done right, due to improved training set. Several simple approaches include:

Gaussian Noise

Over-fitting usually happens when the neural network tries to learn patterns that occur a lot that may not be useful. Gaussian noise, has data points in all frequencies and has zero mean, thus effectively distorting the high frequency features. The low frequency desired features are also distorted, but the distortion is insignificant.

Flipping

To avoid the biased assumption that certain features are being located in specific parts of an image, flipping has to be done. Images can be flipped vertically or horizontally.

Rotation

Images are rotated at fine angles to acquire data corresponding to different orientations. Image size is preserved after rotation if the original image is a square.

Translation

Translation involves moving the image along X direction or Y direction or both. Since, in an image objects can be located anywhere, we require the model to look everywhere.

Scaling

The real data collected for training and testing can be tiny or large. Hence scaling needs to be done. Scaling can be done outwards or inwards.

Imputation and dimensional reduction can be used to add samples in the dataset. Advanced techniques require simulation of the database based on dynamic systems. Generative Adversarial Networks (GAN's) can also be used for data augmentation.

2.3 Convolutional Neural Networks

A Neural Network is a registering framework comprised of various basic, exceptionally interconnected handling components, which process data by their dynamic state reaction to outside information sources. This is made up of layers, it starts with input layer and runs through different hidden layers. The last layer is generally the output layer. Each layer is comprised of neurons and every neuron is associated with neurons in the past layer, whereas neurons in the single layer are not associated.

The problem we deal with is images and each image has very high number of weights that have to be trained. This makes nearly impossible to build deep neural networks which helps in classification. Image data is mostly correlated as neighbour pixels does not contribute more

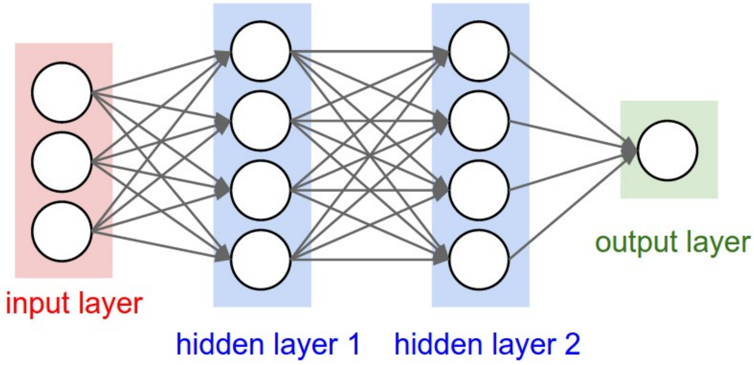


Figure 2.1: Neural Network

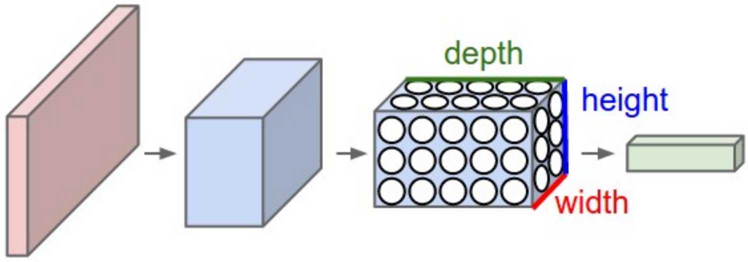


Figure 2.2: Convolutional Neural Network

towards training. In late 90's Yann LeCun proposed a Convolutional Neural Networks which is inspired from human visual perception of recognizing things.

Convolutional Neural Networks (CNN) are kind of Neural Networks which is used effectively in Image Classification and Tracking. CNN's are arranged in 3 Dimensions (Length, Width, Depth) and not all pixels in an image are connected to next layer. As described in Fig 2.2, Each layer of CNN transforms activation of layer to another using a differentiable function for next layer. The CNN Architectures are generally made up of four kinds of layers namely, Convolutional Layer, Pooling Layer, Fully-connected Layer and Non-linear Layer. Stacking these four layers in different configurations will act as Image classifier.

2.3.1 Convolution Layer

Convolutional Layer (CONV) has set of learnable filters as parameters. This layer in a CNN does the most computations. Every filter has length, width as hyper-parameters to be set and depth of filter is equal to the depth of input volume. In every layer, the filter slides through length and width of input volume and computes the dot product to generate a 2-D activation map which gives the responses of that filter at every position of input volume. This particular activation map

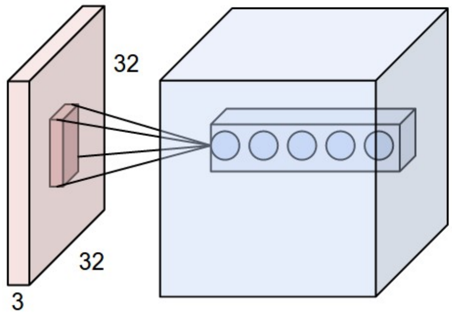


Figure 2.3: Convolutional Layer

is used to extract a feature from the input volume and to extract different features, we use set of filters and each of them generates a 2-D activation map. Output volume is produced by stacking these activation maps across depth.

Hyper-parameters

No. of filters(n) : This will effect the depth of the output volume.
Filter size(f) : this effects the length and width of the output volume.
Padding(p) : The boundaries of an image is not give equal importance as a pixel inside the boundaries. To give importance to boundaries we use a zero padding around the boundary which also helps in size of response.
Stride(s) : The number of pixels a filter can skip while convolution is taking place on the image. Generally a stride of 1 or 2 is used.

$$Output\ Volume = \frac{N - f + 2p}{s} + 1$$

where, N is input length or width.

2.3.2 Pooling Layer

Pooling Layer(POOL) is generally used in CNN to reduce the size of responses which will help to reduce amount of computations and parameters. There are no learning parameters in POOL as its used only to reduce the size. There are two kinds of pooling layers, AVG and MAX. Filter used, slides through the input volume, according to the filter set it either get the Maximum value in the numbers (MAX) or Average value of the numbers (AVG). The depth dimension remains unchanged.

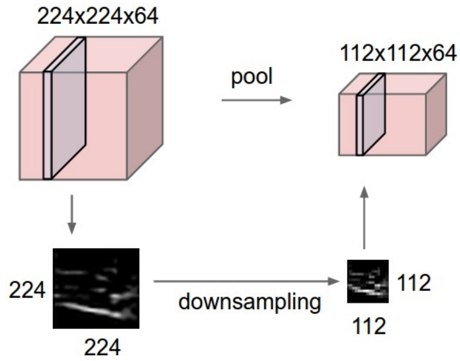


Figure 2.4: Pooling Layer

Hyper-parameters

Filter size(f) : this effects the length and width of the output volume.

Padding(p) : We generally don't use padding as our main cause is to reduce size.

Stride(s) : The number of pixels a filter can skip while convolution is taking place on the image. Generally a stride of 1 or 2 is used.

$$Output\ Volume = \frac{N - f}{s} + 1$$

where, N is input length or width.

2.3.3 Fully-connected Layer

Fully connected layer has associations from every one of the enactments in the past layer. This is a regular Neural Network Layer which can calculated by matrix multiplications along with bias addition. The trainable parameters will be more compared to CONV Layer. This layer is used to take high level features from previous CONV and POOL Layers and classify into various classes by passing it through Activation.

2.3.4 Non-linear Layer

Every Activation or Non-linear Layer performs a mathematical function on each and every number of the volume. This layer is generally used to introduce non-linearity into the neural network. After every CONV layer, a activation(ReLu or Tanh) is used to speed up the convergence of stochastic gradient descent. A sigmoid or softmax layer is used after the fully-connected layer to normalize all the values between 0 and 1.

$$SoftmaxFunction - f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

2.4 Regularization

Overfitting is one of the major problems in any machine learning model. Regularization is a form of regression that discourages learning a more complex or flexible model, so as to avoid overfitting. There are several ways of regularization:

L2 regularization

It is executed by adding squared values of every weight to the loss function. Using this every weight is linearly decayed towards zero but not exactly zero.

L1 regularization

It can be executed by adding the term λ w to each weight w. This regularization can lead the weight vectors become very close to zero during optimization. L2 regularization gives better results than L1 regularization.

Max Norm regularization

It can be executed by scaling the magnitude of the weight vector for every neuron not more than a particular magnitude.

Dropout regularization

It can be executed by ignoring some of the neurons during training. This means the neurons won't contribute during forward pass.

Chapter 3

Methodology and Results

In this chapter, we will initially discuss the dataset used and pre-processing done on it and later discuss more about the model and results.

3.1 Dataset

This dataset was gathered by Sandia National Laboratory, they utilized a X-Band sensor at a one-foot resolution in spotlight mode to gather information at 15 and 17-degree depression angles. This gathering was consolidated exertion Defense Advanced Research Projects Agency and Air Force Research Laboratory as a piece of MSTAR Program. The datasets available for public consists of 7 classes of different ground military targets (BRDM_2, 2S1, BTR_60, T62, D7, ZSU_23_4, ZIL131). The descriptions of the training and test dataset is shown in Table 1.

	Train		**Test**	
Class	**Depression Angle**	**No. of Images**	**Depression Angle**	**No. of Images**
BRDM_2	17°	298	15°	274
2S1	17°	299	15°	274
BTR_60	17°	232	15°	195
T62	17°	270	15°	274
D7	17°	225	15°	274
ZSU_23_4	17°	265	15°	274
ZIL131	17°	299	15°	274

Table 3.1: Train and Test Data Description

For purpose of comparison between military targets and SAR images, *Fig 3.1* is presented. To train the model, we used 17° SAR images and for testing we used 15° SAR images.These images are gray scale with different sizes and resolutions depending on the class of the military target. You can observe by SAR images, even a human finds it difficult to classify the classes as all look similar.

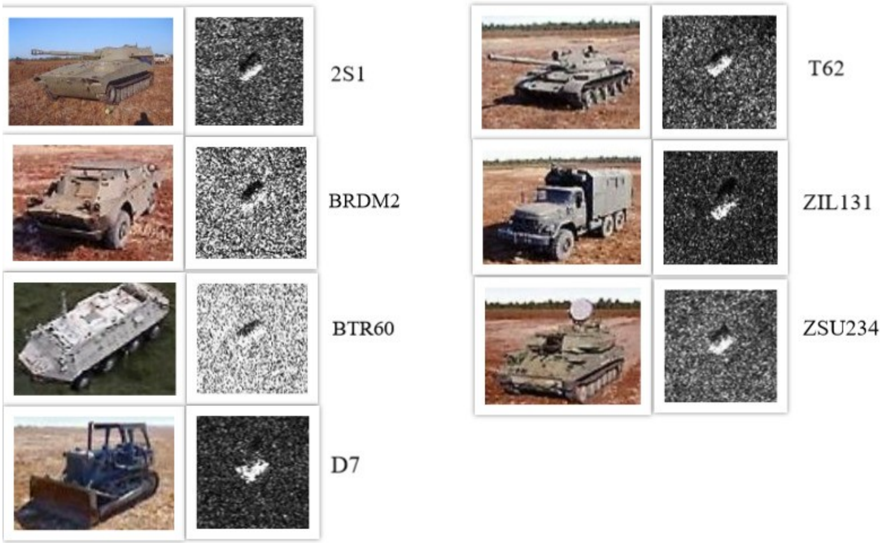


Figure 3.1: Different classes along with SAR Image

3.2 Data Preprocessing

The original MSTAR dataset images come in different sizes. Although images are centered by the target as seen in Fig 3.1. It is important that images passed to CNN should be of equal size for convolutions. So, all images are re-sized to *150 X 150* pixels before passing to CNN. Since the neighboring pixel values are highly correlated and to reduce the redundancy. ZCA Whitening is applied on all the SAR images to decorrelate the images and all input images to have same variance. Most important aspect for training any Neural Network is data, having small amounts of data can lead to over-fitting of the model and might work differently with unseen data. Augmentation is a technique to increase the data samples by flipping, scaling, shearing, rotating and adding noise. In our case, we augmented the data by shearing, scaling and flipping the original SAR images.

3.3 Model

The CNN Model we built was a basic stack of three convolutional layers with ReLu activation after each CONV layer, Max POOL Layer is followed. The Hyper-parameters of each layer is described in Table 3.1 and architectural overview is described in Fig 3.2.

Layer Name	**Hyper-parameters**
CONV 1	(32, 3x3)
POOL 1 (MAX)	(2, 2)
CONV 2	(32, 3x3)
POOL 2 (MAX)	(2, 2)
CONV 3	(64, 3x3)
POOL 3 (MAX)	(2, 2)

Table 3.2: Hyper-parameters for CNN

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_6 (Conv2D)            (None, 148, 148, 32)      320

activation_10 (Activation)   (None, 148, 148, 32)      0

max_pooling2d_6 (MaxPooling2 (None, 74, 74, 32)        0

conv2d_7 (Conv2D)            (None, 72, 72, 32)        9248

activation_11 (Activation)   (None, 72, 72, 32)        0

max_pooling2d_7 (MaxPooling2 (None, 36, 36, 32)        0

conv2d_8 (Conv2D)            (None, 34, 34, 64)        18496

activation_12 (Activation)   (None, 34, 34, 64)        0

max_pooling2d_8 (MaxPooling2 (None, 17, 17, 64)        0

flatten_2 (Flatten)          (None, 18496)             0

dense_4 (Dense)              (None, 64)                1183808

activation_13 (Activation)   (None, 64)                0

dropout_2 (Dropout)          (None, 64)                0

dense_5 (Dense)              (None, 7)                 455

activation_14 (Activation)   (None, 7)                 0
=================================================================
Total params: 1,212,327
Trainable params: 1,212,327
Non-trainable params: 0
```

Figure 3.2: Details of the Model

The Padding and Stride for every layer is 0 and 1 respectively. A Flatten layer is added after three sets of convolutional layers and two fully-connected layers are added on the top of this. A dropout is added just before the last fully-connected layer to regularize the model and Soft-max Classifier is added as an activation function to the last fully-connected layer with 7 categorical outputs. We get the predicted class after this classifier and there are a total of 1,212,327 parameters that has to be trained through back propagation.

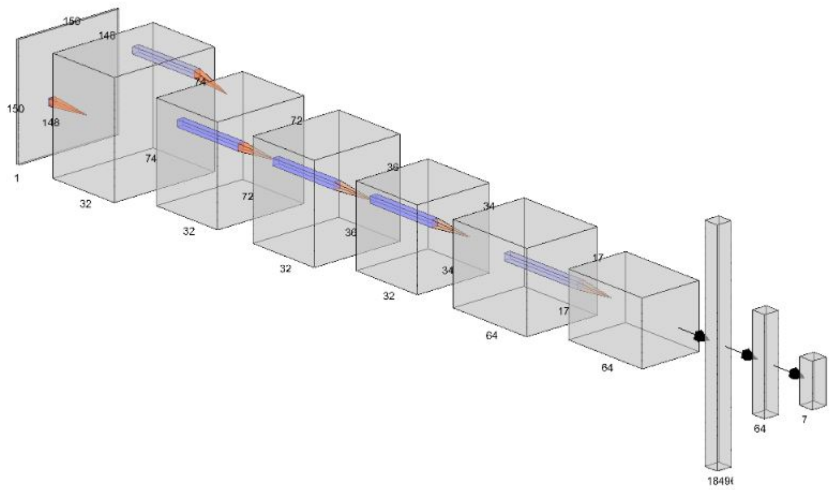
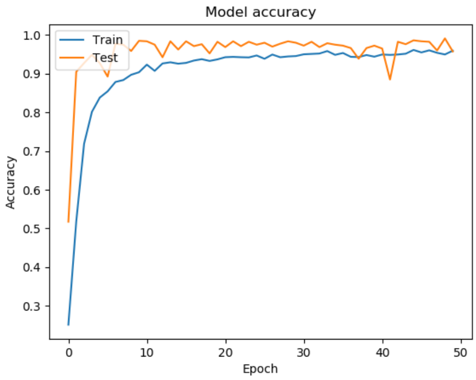


Figure 3.3: Neural Network Architecture

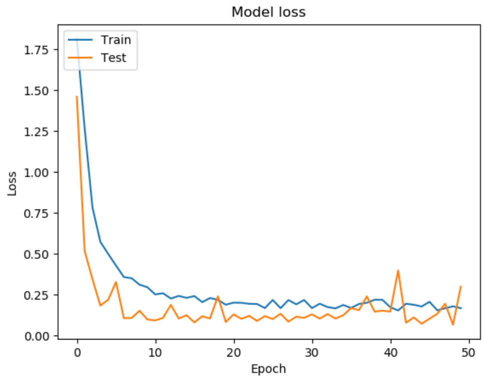
We have randomly initialized the weights using Gaussian Distribution and bias as zero vector to increase the learning rate. To train our model we took categorical crossentropy loss, as we are dealing with categorical classification and RMSprop is selected as the optimizer because it restricts oscillations in vertical direction and increase movement in horizontal direction which helps us in converging faster.

3.4 Results

To evaluate the CNN and the proposed approach, two different experiments are held. First experiment by considering only two different classes, in which an accuracy of 99% was achieved. We considered two entirely different classes for this experiment (D7 and T62). Second experiment is conducted on seven different military targets and results were plotted in Fig 3.3. We observe that a accuracy of approximate 98% is reached on test dataset and a loss of 10% on train dataset.



(a) Accuracy Vs Epochs



(b) Loss Vs Epochs

Figure 3.4: Plots of Accuracy and Loss

Chapter 4

Conclusion and Future Work

4.1 Conclusion

As a initial study, three layers of CNN have been applied to the SAR data set, as feature extractors and pooling has been done. A simple softmax classifier is used to achieve an exactness of 99% for two kinds of targets and 98% for seven distinct targets. Simple data augmentation technique has been implemented due to the scarcity of SAR image samples. Regularization technique has been used to avoid overfitting.

4.2 Future Work

It can be seen that simple CNN used in this approach can classify SAR image targets with an accuracy of 98% but to make this model more robust, tests have to be made at different depression angles and different kinds of SAR Imagery targets. Instead of augmentation, Transfer learning approach can be applied to obtain better classification with small amounts of data. In this approach, multiple pre-trained classifiers are used to extract useful features, along with a classifier.

Bibliography

[1] Sizhe Chen, Haipeng Wang, "SAR Target Recognition Based on Deep Learning," Dec. 1992.

[2] Novak L M, Owirka G J, Brower W S, et al. "The automatic target recognition system in SAIP," *Lincoln Laboratory Journal*, 1997, 10(2).

[3] Wissinger J, Ristroph R, Diemunsch J R, et al, "MSTAR's extensible search engine and model-based inferencing toolkit," AeroSense'99.International Society for Optics and Photonics, 1999: 554-570

[4] Krizhevsky A, Sutskever I, Hinton G E. "ImageNet Classification with Deep Convolutional Neural Networks," NIPS. 2012, 1(2): 4.

[5] Sizhe Chen, Haipeng Wang, Feng Xu and Ya-Qiu Jin, "Target Classification Using the Deep Convolutional Networks for SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 8, August 2016.

[6] Zhijun Sun, Lei Xue and Yangming Xu, "Recognition of SAR Target based on Multilayer Auto-Encoder and SNN," *ICIC International*, Mar. 2013.

[7] Ciresan D C, Meier U, Gambardella L M, et al, "Deep, big, simple neural nets for handwritten digit recognition," Neural computation, 2010, 22(12): 3207-3220.

[8] Ciresan D C, Meier U, Schmidhuber J. "Transfer learning for Latin and Chinese characters with deep neural networks," Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012: 1-6.

[9] Cirean D, Meier U, Masci J, et al. "Multi-column deep neural network for traffic sign classification," Neural Networks, 2012, 32: 333-338.

[10] Taigman Y, Yang M, Ranzato M, et al."Deep-Face: Closing the Gap to Human-Level Performance in Face Verification," IEEE CVPR. 2014.

[11] MSTAR Dataset : https://www.sdms.afrl.af.mil/index.php?collection=mstar