A comparison of target detection algorithms using DSIAC ATR algorithm development data set

Abhijit Mahalanobis, Bruce McIntosh
University of Central Florida, 4000 Central Florida Blvd, Orlando, FL, USA 32816

ABSTRACT

In this paper, we present preliminary results of infra-red target detection using the well-known Faster R-CNN network using a publicly available MWIR data set released by NVESD. We characterize the difficulty level of the images in terms of pixels on target (POT) and the local contrast. We then evaluate the performance of the network under challenging conditions and when the number of training images are varied.

Keywords: ATR, deep learning, classification, convolutional neural network, localization

1. INTRODUCTION

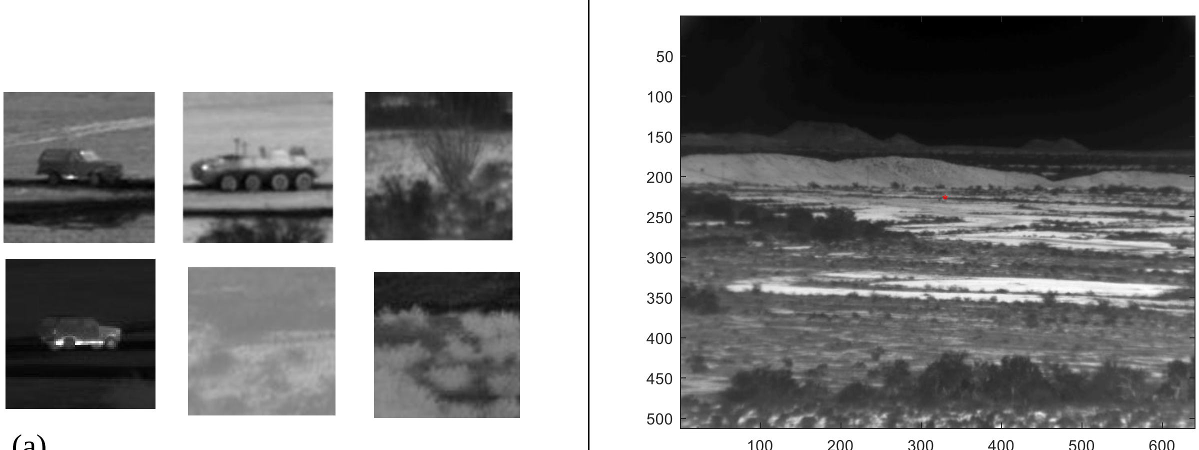
The problem of reliably detecting targets in infra-red imagery at acceptably low false alarm rates continues to be of significant interest to the research community. The challenge is that unlike common applications that use RGB cameras, the amount of labeled training data is very limited. Further, infra-red phenomenology differs significantly from that in the visible band as a result of which algorithms trained on RGB data cannot be readily incorporated into infra-red applications. Of course, there has been a tremendous surge in deep learning/convolutional neural nets (CNNs) in the field of computer vision at large. Hence, there is significant interest in determining if similar performance gains can be achieved by applying these techniques in the infra-red domain.

In this paper, our focus is to baseline the performance of a popular object detection algorithm called Faster RCNN [1], and report its performance on the medium wave infra-red (MWIR) data set [2] made available to the public by NVESD. Other researchers [3] have previously reported the use of the fast RCNN algorithm to find targets in this same data set. They exploited the dual-band (visible band and MWIR) and motion information of the data, and obtained promising overall performance. However, the data set contains different scenarios that vary in their level of difficulty. In fact, there is substantial variation in how well the targets are resolved at different ranges, and the competing clutter level of the background. Therefore, our goal is to characterize the performance of the Faster RCNN algorithm in terms of the level of difficulty of the images, in particular at longer ranges (when targets appear smaller) and at low contrast (i.e. weaker target signature compared to the surrounding background). Deep CNNs are highly data driven and require large amounts of data and long time to train. Unfortunately for infra-red applications, one of the biggest challenges is the lack of labeled infra- training data. For this reason, we are also interested in quantifying the degradation in performance when the number of training images is reduced.

The rest of the paper is organized as follows. We provide a brief description of the data set in Section 2. The results of characterizing the performance of target detection is provided in Section 3, followed by a summary and discussion of the future research on related topics.

2. DATA SET

The data based used is a collection of visible and MWIR imagery collected by the US Army Night Vision and Electronic Sensors Directorate (NVESD) that is intended to support the ATR algorithm development community. It contains 207 GB of MWIR imagery and 106 GB of visible imagery along with an image viewer, ground truth data, meteorological data, photographs of the targets, and other documentation to assist the user in correctly interpreting the imagery. The targets included in the database include people, foreign military vehicles, and civilian vehicles. All imagery was taken using commercial cameras operating in the MWIR and visible bands. The data also contains information about the sensors, target ground truth, scenario numbering concept, and other information that is necessary to utilize the database such as Target Delta-T, Weather, Target, and Image Metrics. Some examples of the targets in the images are shown in Figure 1. The data was collected during the day and at night, in range increments of 500 meters from 500 to 5000 meters during both day and night. The targets moved at constant velocity along a circle with a diameter of about 100 meters at each range that allowed the targets to be consistently imaged at the same locations and aspect angles. The path was frequently sprayed with water to avoid dust and heating. The 10 different vehicular targets in this data set are BTR70, BMP, BRDM, T62, T72, ZSU23, 2S3, MTLB (armored vehicle with artillery), Pickup truck, and SUV. Examples of some close-up views of targets and clutter are shown in Figure 1(a) while a full image frame of a target at 4000 m is shown in Figure 1(b).



(a) (b)

Figure 1: Examples of (a) target and clutter chips used for training the algorithms, and (b) a typical image frame with the target 4000m (indicated by the red *) used for testing.

A balanced training set was created from the data set with the intention of equally representing every target class at every nominal distance for both day and night. For each of these combinations, 100 frames were selected uniformly so as to capture a broad range of aspect angles as the target moves around the track. Two test sets, each containing 19,800 samples, were developed to help evaluate the network's performance. The first set contains samples randomly chosen (making sure not to include any training examples) across the same set of conditions as the training set and is considered the “easier test”.

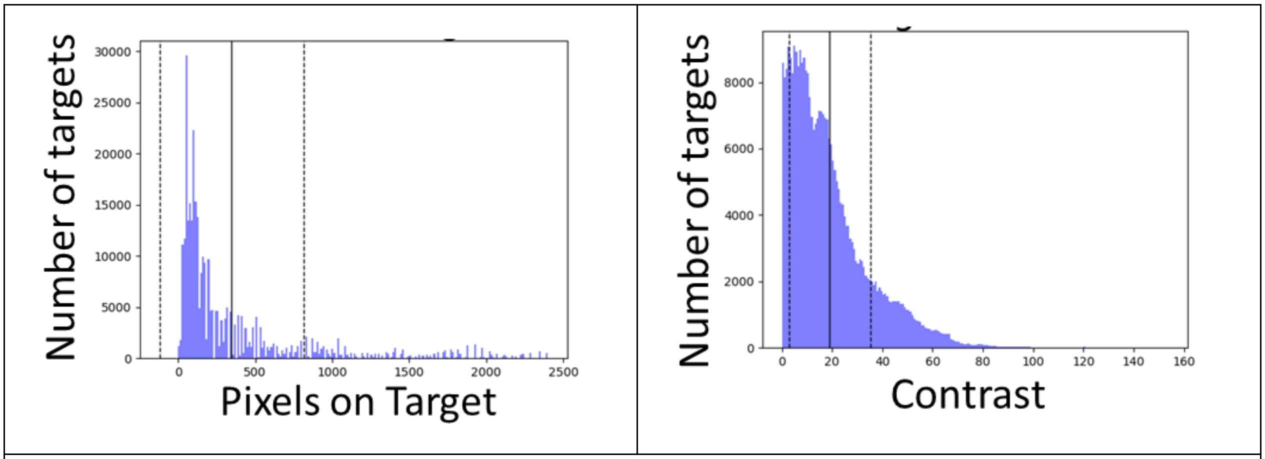


Figure 2: Distribution of POT and contrast. Images with lower values of these metrics are considered “difficult”

The second set, referred to as the ‘difficult’ set later, is composed of samples selected to be more difficult to detect and identify. Two important factors that impact the difficulty of detection are the contrast and number of Pixels on Target (POT). Contrast was calculated as the difference between the mean pixel value within the target ground truth bounding box and the mean pixel value of the surrounding pixels (the images are gray scale with values from 0 to 255). POT is the number of pixels within the ground truth bounding box. A histogram for each of these attributes was calculated for the entire set (see Figure 2), and samples were selected so that they were in the lowest quartile of both categories.

3. PERFORMANCE COMPARISONS

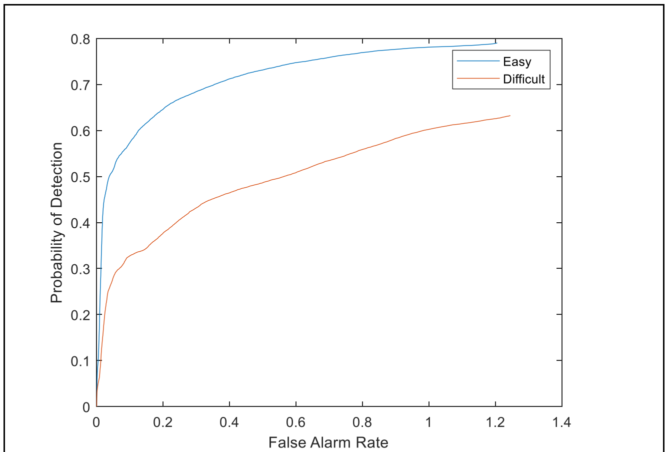


Figure 3: Performance of the QCF target detection algorithm on the “easy” and “difficult” test sets

We first evaluated a quadratic target detection algorithm known as the Quadratic Correlation Filter (QCF) [4] on the data set described in Section 2 which has been successfully used for detecting targets in background clutter in infra-red images. Essentially, the QCF uses a set of basis functions that are estimated using the second order statistics of the targets and clutter. The algorithm was trained on image chips (of both targets and clutter) such as those shown in Figure 1(a), and then tested on the full scene (such as shown in Figure 1(b)). The training images are all resized such that apparent distance to target is 2500m (this is done using the range information provided in the ground truth). The test images are also scaled so that the targets appear to be at 2500 m in the scene. The receiver operating characteristic (ROC) curves Figure 3 shows the performance of the QCF algorithm on the “easy” set and the more challenging “difficult” set. Here, the vertical axis is the Probability of Detection (or percentage of targets detected), while the horizontal axis represents the false alarm rate (FAR) in terms of “false alarms per square degree”. We see that there is considerable difference in performance between the two sets, with substantially poorer performance on the difficult set. For instance, on the easy set the QCF detects about 70% of the targets at a FAR of approximately 0.4 false alarms/square degree, whereas on the difficult set the algorithm achieves approximately 45% detection at the same value of FAR.

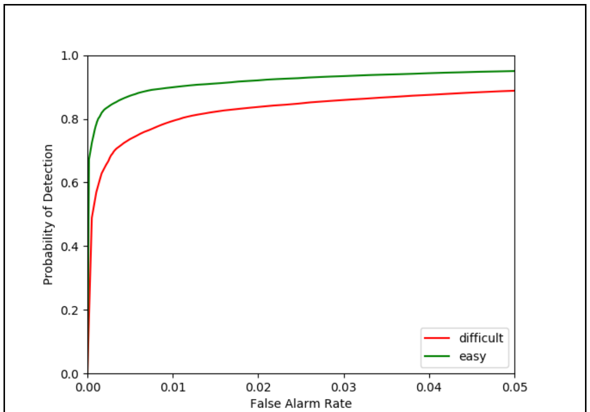


Figure 4: Comparison of the performance of the RCNN target detection algorithm on the “easy” and “hard” set which shows the relative difficulty of the “difficult” test set.

We then evaluated the Faster RCNN algorithm [1] which is composed of a region proposal network and an object detection network which are both deep, fully convolutional neural networks. They are trained end to end using stochastic gradient descent. The network is already pre-trained on the imagenet dataset, but we “fine tune” the weights on the training set described in Section 1. The last layer of the network was completely reset (so that algorithm learns this layer from our training data), but all layers of the network were updated during training. The goal is to let the lower layers “fine-tune” to the characteristics of the data while the output layer is completely learned on the balanced training set. The network was trained for 90,000 steps which took about 1 day on a Titan XP GPU.

After convergence, the network was tested on both the “easy” and “difficult” test sets, the results of which are shown in Figure 4. We see that Faster RCNN performs significantly better than QCF in both cases. Specifically, for the easy case (green line) the Faster RCNN achieved Pd=80% at FAR ≈ 0.002 FA/sq degree, while for the difficult case (red line) it achieved the same Pd=80% at FAR ≈ 0.01 FA/sq degree. While these curves also confirm the more challenging nature of the second test set, Faster RCNN yields an order of magnitude reduction in the false alarm rate.

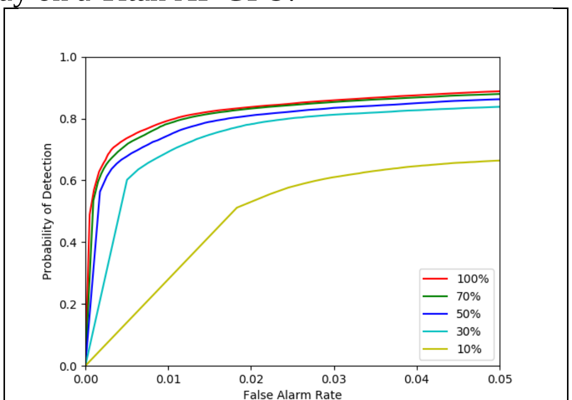


Figure 5: Impact of reducing the number of training images on the target detection performance

We then studied the effect of varying the training set size by using progressively smaller training sets. In order to preserve similarity to the original training set, samples were uniformly removed to create sets with 70%, 50%, 30% and 10% of the original. The network was then tested on the previously described 'difficult' set, and the results are shown in figure 5. The 70% set performs only marginally worse than the full set, but at 50% and lower, performance significantly degrades.

4. SUMMARY AND CONCLUSION

In this paper, we have presented our intial results of evaluating the performance of a Deep Learning techniques (i.e. the Faster RCNN algorithm) on the NVESD data set. We have systematically defined the training and test sets using metrics for characterizing the difficulty of the test conditions. Our initial conclusion is that even under difficult conditions, the Faster RCNN algorithm achieves noteworthy performance, detecting as much as 80% of the targets at a low false alarm rate of 0.01 false alarms/square degree.

One limitation of the data set is that it has limited diversity in the background. Although care was taken to separate the training and test images for the targets, the clutter/background may be similar between the training and test sets. We will therefore focus on removing this as a potential bias that can lead to optimistic results in future analysis. We also see that the performance is heavily dependent on the amount of training data available. Since large amounts of labeled training data is not readily available for such infra-red applications, our future work will also focus on finding ways to reduce the gap between the curves in Figure 5. For this purpose, we will consider deep learning techniques for generating different views of targets from a handful of images, and unsupervised (or weakly supervised) learning techniques for cases where unlabeled (or surrogate) training data is available.

Acknowledgement: The authors gratefully recognize the support of Leonardo DRS for funding parts of this research.

REFERENCES

1. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Advances in Neural Information Processing Systems 28 (NIPS 2015), arXiv:1506.01497
2. https://www.dsiac.org/resources/research-materials/cds-dvds-databases/atr-algorithm-development-image-database
3. Liu, S., & Liu, Z. (2017). Multi-Channel CNN-based Object Detection for Enhanced Situation Awareness. CoRR, abs/1712.00075. Published at the Sensors & Electronics Technology (SET) panel Symposium SET-241 on 9th NATO Military Sensing Symposium. arXiv:1712.00075
4. Abhijit Mahalanobis, Robert R. Muise, and S. Robert Stanfill, "Quadratic correlation filter design methodology for target detection and surveillance applications," Appl. Opt. 43, 5198-5205 (2004)