DEEP LEARNING BASED TARGET TRACKING AND CLASSIFICATION DIRECTLY IN COMPRESSIVE MEASUREMENT FOR LOW QUALITY VIDEOS

Chiman Kwan[1] , Bryan Chou[1], Jonathan Yang[2], and Trac Tran[3]

[1] Applied Research LLC, Rockville, Maryland, USA
[2]Google, Inc., Mountain View, California, USA
[3]Johns Hopkins University, Baltimore, Maryland, USA

ABSTRACT

Past research has found that compressive measurements save data storage and bandwidth usage. However, it is also observed that compressive measurements are difficult to be used directly for target tracking and classification without pixel reconstruction. This is because the Gaussian random matrix destroys the target location information in the original video frames. This paper summarizes our research effort on target tracking and classification directly in the compressive measurement domain. We focus on one type of compressive measurement using pixel subsampling. That is, the compressive measurements are obtained by randomly subsample the original pixels in video frames. Even in such special setting, conventional trackers still do not work well. We propose a deep learning approach that integrates YOLO (You Only Look Once) and ResNet (residual network) for target tracking and classification in low quality videos. YOLO is for multiple target detection and ResNet is for target classification. Extensive experiments using optical and mid-wave infrared (MWIR) videos in the SENSIAC database demonstrated the efficacy of the proposed approach.

KEYWORDS

Compressive measurements, target tracking, target classification, deep learning, YOLO, ResNet, optical videos, infrared videos, SENSIAC database

1. INTRODUCTION

Optical and infrared videos have been widely used for traffic monitoring, surveillance, and security monitoring applications [1]-[5]. Compared to radar based trackers [6][7], object features can be clearly seen in optical or infrared videos.

Compressive measurements [8]-[12] are normally collected by multiplying the original vectorized image with a Gaussian random matrix. Each measurement is a scalar and the measurement is repeated M times where M is much fewer than N (the number of pixels). Using compressive measurements for target tracking and classification, the image frames need to be reconstructed and then conventional trackers are then applied. There are two drawbacks in this conventional approach. First, the reconstruction process using L_0 [13] or L_1 [14]-[16] based methods is time consuming. Second, there may be information loss in the reconstruction process.

In the SENSIAC data [9], there are optical and mid-wave infrared (MWIR) videos containing multiple vehicles collected at ranges between 1000 m and 5000 m. The data are challenging because there are strong illumination variations and other environmental factors. Moreover, the target sizes are small and the video quality (resolution) is low. One active research area in target tracking is the use of compressive measurement directly without reconstruction because reconstruction requires a lot of time and thereby limits real-time applications. Some initial results have been presented in [10]-[12].

In the literature, there are some trackers such as [17] that used the term compressive tracking. However, those trackers are not using compressive measurements directly. Conventional approaches usually need to reconstruct the video frames and no one, except our team, has developed any high performance algorithms to deal with target tracking and classification directly in the compressive measurement domain, which has two key advantages. First, using compressive measurement directly enables faster processing and real-time tracking is then feasible. Second, our study showed that there will be no loss of information due to reconstruction if one uses compressive measurement directly [18]-[21].

Recently, we developed a residual network (ResNet) based tracking and classification framework using compressive measurements [12]. Pixel subsampling was used to obtain the compressive measurements. ResNet was used in both target detection and classification. The tracking is done by detection. Although the performance in [12] is much better than conventional trackers, there is still room for further improvement. The key area is to improve the tracking part, which has a significant impact on the classification performance. That is, if the target area is not correctly located, the classification performance will degrade.

In this paper, we propose an alternative approach, which aims to improve the both tracking performance and classification using compressive measurements. The idea is to deploy a high performance detector known as YOLO [22] for target tracking. YOLO is fast, accurate, and has comparable performance as other trackers such as Faster R-CNN [23]. The improved tracking results will further improve the classification performance. The classification is still using ResNet [24] because ResNet has better classification than the built-in classifier in YOLO [10][11][25]. Experiments using optical and MWIR videos in the SENSIAC database clearly demonstrated the performance of the proposed approach.

We would like to briefly review some state-of-the-art algorithms that performs action inference or object classification directly using compressive measurements. We will highlight the key differences between our approach and others.

The paper in [28] presents a reconstruction-free approach to action inference. Smashed filters are built using training samples that are affine transformed to a canonical viewpoint. It works very well even for 100 to 1 compression. However, the approach is for action inference, not for target detection, tracking, and classification in compressed measurement domain. Moreover, the smashed filter assumes that the camera is stationary and the angle is fixed. Extending the approach to target tracking and classification with moving cameras may be non-trivial.

In [29], a convolutional neural network (CNN) approach was presented to perform image classification directly in compressed measurement domain. The input image is assumed to be cropped and centered, and there is only one target in each image.

Papers [30,31] are similar in spirit to [29]. Both papers discussed direct object classification using compressed measurement. However, both papers assumed that the targets/objects are already centered. The approach in [32] is not reconstruction free. The integral image is one type of reconstructed image. In contrast, our paper does not require any image reconstruction.

Reference [33] used a random mask to conceal the actual contents of the original video. In addition, the key idea in [33] is about action recognition (similar to [28]), not object tracking and classification. Extending the idea in [33] to object tracking and classification may not be an easy task.

Reference [34] presents an object detection approach using correlation filters and sparse representation. There was no object classification. No reconstruction of compressive measurements is needed. The results are quite good. Different from [34], our paper focuses on object detection, tracking, and classification.

In [35], the authors present an approach to extracting features out of the compressed measurements and then uses the features to create a proxy image. This approach may not be considered as a reconstruction free approach. Similar to Refs. [29-31], the approach is suitable for stationary camera cases and also the objects are already centered in the images.

The paper in [36] presents an online reconstruction free approach to object classification using compressed measurements. Similar to [29]-[31] and [35], the approach assumes the object is already at the center of the image. The methods in [29]-[31], [35] and [36] also did not address the above mentioned issue.

Instead of using Gaussian random measurements to obtain the compressive measurements, we emphasize that we have proposed two alternative compressive measurements. One is called subsampling and the other is called coded aperture. The coded aperture case has been summarized and reported in our recent papers [37-38].

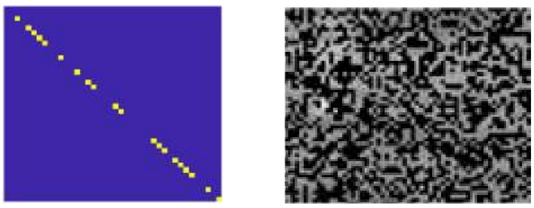
An earlier version of this paper was presented in an SPIE conference [25]. One major change is that significant amount of new experiments were performed. In our earlier paper, we had only the tracking and classification results for the optical videos. Here, we have included tracking and classification results for MWIR daytime and night-time videos. The earlier version had 11 pages and the current version has more than 20 pages.

This paper is organized as follows. Section 2 describes the compressive sensing via subsampling, YOLO tracker, ResNet, SENSIAC videos, and performance metrics. Section 3 presents the tracking and classification results directly in the compressive measurement domain using SENSIAC optical videos. Section 4 includes the tracking and classification results for the MWIR daytime videos. Section 5 then presents the tracking and classification results for the MWIR night-time videos. Finally, some concluding remarks and future research directions are included in Section 6.

2. BACKGROUND

2.1. Compressive Sensing Via Subsampling

Using Gaussian random measurement makes the target tracking very difficult. This is because the targets can be anywhere in a frame and the target location information is lost in the compressive measurements. Recently, we proposed an approach [12] using a random subsampling operator to perform compressive sensing. Figure 1 displays two examples of a random subsampling sensing matrices.



(a) (b)

Figure 1. (a) Visualization of the sensing matrix for a random subsampling operator with a compression factor of 2. The subsampling operator is applied to a vectorized image. This is equivalent to applying a random mask shown in (b) to an image.

2.2. YOLO

YOLO tracker [22] has similar performance as Faster R-CNN [23]. YOLO also comes with a classification module. However, based on our evaluations, the classification accuracy using YOLO was not good [10][11][25]. This is perhaps due to a lack of training data.

The training of YOLO is quite simple. Images with ground truth target locations are needed. The input image is resized to 448x448.
Figure 2 shows the architecture of YOLO version 1. There are 24 convolutional layers and 2 fully connected layers. The output is 7x7x30. We have used YOLOv2 because it is more accurate than YOLO version 1. The bounding box for each vehicle was manually determined using tools in MATLAB. For YOLO, the last layer of the deep learning model was re-trained. We did not change any of the activation functions. YOLO took approximately 2000 epochs to train.

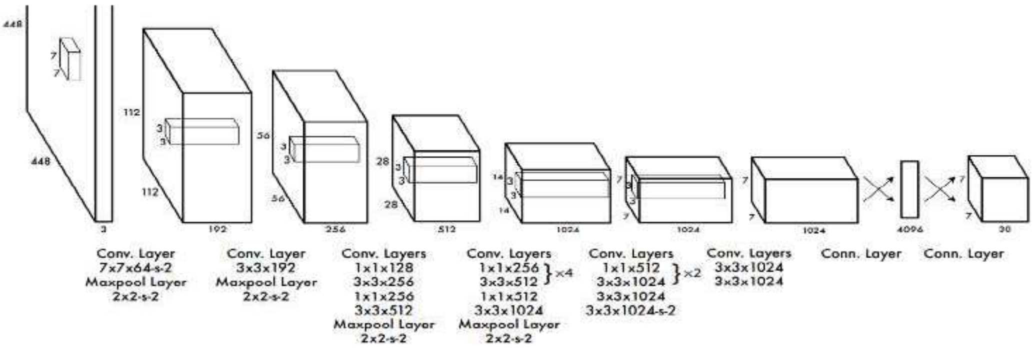


Figure 2. 24 convolutional layers followed by 2 fully connected layers for YOLO version 1 [22].

2.3. ResNet

The ResNet-18 model is an 18-layer convolutional neural network (CNN) that avoids performance saturation and/or degradation when training deeper layers.
Figure 3 shows the architecture of an 18-layer ResNet.

Training of ResNet requires target patches. The targets are cropped from training videos. Mirror images are then created. Data augmentation using scaling (larger and smaller), rotation (every 45 degrees), and illumination (brighter and dimmer) is usually deployed to create more training data. For each cropped target, we are able to create a data set with 64 more images.

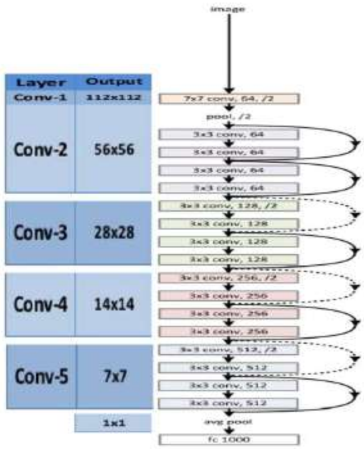


Figure 3. Architecture of ResNet-18. Figure from [24].

2.4. SENSIAC Optical and MWIR Videos

Our research objective is to perform tracking and classification of seven vehicles using the SENSIAC videos in compressive measurement domain. Both optical and mid-wave infrared (MWIR) videos were collected at distances from 1000 m to 5000 m with 500 m increments. Optical videos has only daytime videos. In contrast, there are both daytime and night-time MWIR videos. There are seven types of vehicles, which are shown in
Figure 4. These videos are challenging for several reasons. First, the target sizes are small due to long distances. Second, the target orientations also change drastically. Third, the illuminations in different videos are also different. Here, the compressive measurements are collected via directly sub-sampling. That is, 50% or 75% of the pixels are thrown away during the data collection process.

Figure 4. 7 targets in SENSIAC used: (a) Truck (b) SUV, (c) BTR70, (d) BRDM2, (e) BMP2, (f) T72, and (g) ZSU23-4.

2.5. Performance Metrics for Tracking and Classification

In our earlier paper [25], we have included some tracking results where conventional trackers such as GMM [26] and STAPLE [27] were used. The tracking performance was poor in the presence of missing data.
We experimented with a YOLO tracker, which has been determined to perform better tracking than our earlier ResNet based tracker described in [12]. We used the following metrics [9]-[12] for evaluating the tracker performance:

- Center Location Error (CLE): It is the error between the center of the bounding box and the ground-truth bounding box.
- Distance Precision (DP): It is the percentage of frames where the centroids of detected bounding boxes are within 20 pixels of the centroid of ground-truth bounding boxes.
- EinGT: It is the percentage of the frames where the centroids of the detected bounding boxes are inside the ground-truth bounding boxes.
- Number of frames with detection: This is the total number of frames that have detection.

For classification, we used confusion matrix and classification accuracy as performance metrics.

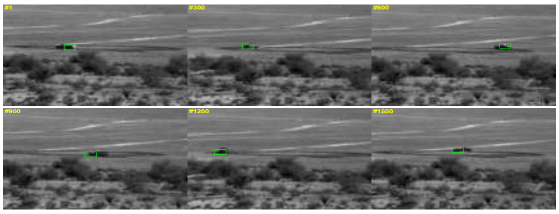
3. Tracking And Classification Results Using SENSIAC Optical Videos

Here, we focus on the optical videos.

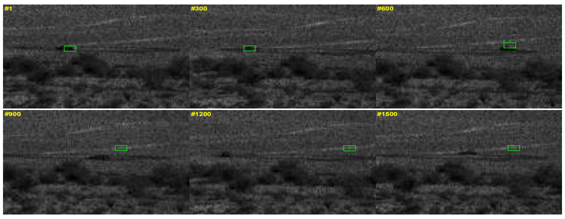
3.1. Conventional Trackers Performance

Figure 5 shows the tracking performance of a conventional tracker knows as GMM [26] for three compressive measurement cases: 0%, 50%, and 75%. It can be seen that the tracking is only effective for the 0% missing case. For the other two cases, the green bounding boxes are mostly off the target. Figure 6 shows the tracking results using another conventional tracker known as

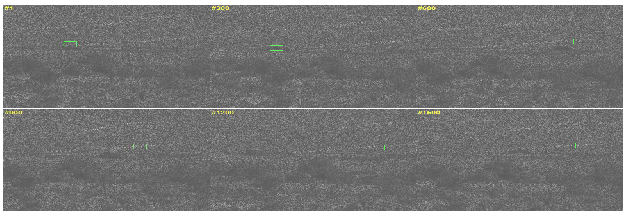
STAPLE [27]. STAPLE appears to yield reasonable tracking results up to 50% missing. However, for 75% missing case, the tracker loses all the targets. This is because STAPLE was not designed to handle videos with missing data.



(a) 0% missing



(b) 50% missing

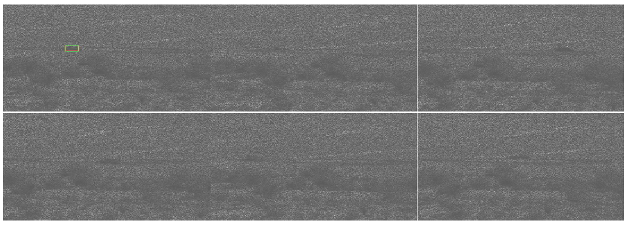


(c) 75% missing

Figure 5. Tracking using GMM for 3 missing cases of an optical video containing a truck from 1000 m range.

(a) 0% missing

(b) 50% missing



(c) 75% missing

Figure 6. Tracking using STAPLE for 3 missing cases of an optical video containing a truck from 1000 m range.

3.2. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

Here, we applied YOLO and used videos from ranges of 1500 m and 3000 m for training and videos from ranges of 1000m, 2000m, 2500m, and 3500m for testing. Table 1 - Table 3 summarize the performance metrics for 0%, 50%, and 75% missing cases, respectively. Our first observation is that the number of frames with detection decreases when we have more missing pixels. This is reasonable. For the same missing rate case, the tracking performance drops with increasing range, which is also reasonable. For 75% missing case, the tracking performance is only effective up to 1500 m. Figure 7 -

Figure 9 show the tracking results in some selected frames for 0%, 50%, and 75% missing cases, respectively. It can be seen that there are more missed detections in those cases of high missing rates.

Table 1. Tracking metrics for 0% missing case.

1000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	37.91	0.00	98.99
BRDM2	1.00	19.23	0.69	98.72
BTR70	1.00	28.24	0.00	100.00
SUV	1.00	25.26	0.00	100.00
T72	1.00	61.73	0.00	100.00
Truck	1.00	23.83	0.06	99.68
ZSU23-4	1.00	33.21	0.00	99.89

1500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	26.61	0.00	100.00
BRDM2	1.00	14.28	1.00	100.00
BTR70	1.00	20.46	0.41	100.00
SUV	1.00	18.56	0.94	100.00
T72	1.00	45.68	0.00	100.00
Truck	1.00	18.59	0.91	100.00
ZSU23-4	1.00	24.35	0.00	100.00

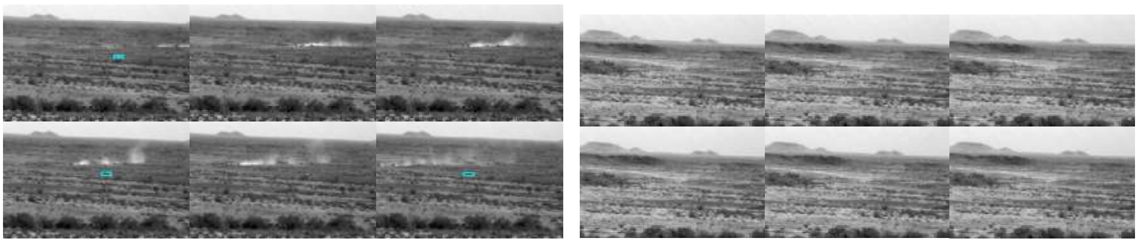
2000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	17.97	0.98	100.00
BRDM2	1.00	8.45	1.00	99.95
BTR70	1.00	14.25	1.00	100.00
SUV	1.00	11.73	1.00	99.47
T72	1.00	30.37	0.00	100.00
Truck	1.00	11.72	1.00	98.67
ZSU23-4	1.00	17.80	0.97	99.89

2500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	14.56	1.00	66.54
BRDM2	0.96	10.49	0.96	85.70
BTR70	0.99	11.26	0.99	54.11
SUV	1.00	8.74	1.00	84.53
T72	0.98	24.59	0.04	72.04
Truck	1.00	9.59	1.00	94.45
ZSU23-4	1.00	12.94	1.00	95.57

3000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	8.30	1.00	100.00
BRDM2	1.00	3.93	1.00	100.00
BTR70	1.00	6.13	1.00	99.79
SUV	1.00	5.35	1.00	99.68
T72	0.99	14.63	1.00	95.20
Truck	1.00	5.23	1.00	99.95
ZSU23-4	1.00	7.14	1.00	100.00

3500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.98	5.65	0.99	47.81
BRDM2	0.94	2.26	1.00	29.72
BTR70	1.00	4.24	1.00	29.24
SUV	0.50	2.81	0.98	8.38
T72	0.93	11.47	0.97	75.72
Truck	0.83	2.47	1.00	1.87
ZSU23-4	0.91	4.54	1.00	50.27

(a) 1000m (b) 2000m



(c) 2500m (d) 3500m

Figure 7. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 0% missing case. Optical – Subsampling Mode. Tracking at 0%; vehicle in the frames is SUV.

Table 2. Tracking metrics for 50% missing case.

1000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	37.96	0.00	96.10
BRDM2	1.00	17.92	0.80	88.05
BTR70	1.00	28.34	0.00	99.73
SUV	1.00	24.05	0.08	91.14
T72	1.00	60.35	0.00	98.29
Truck	1.00	22.64	0.20	88.74
ZSU23-4	1.00	31.19	0.00	97.44

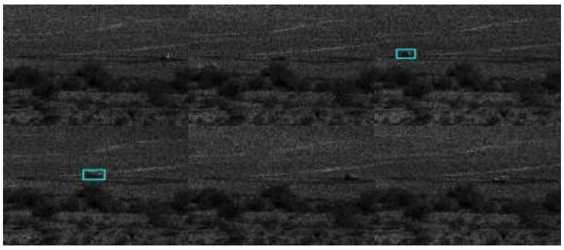
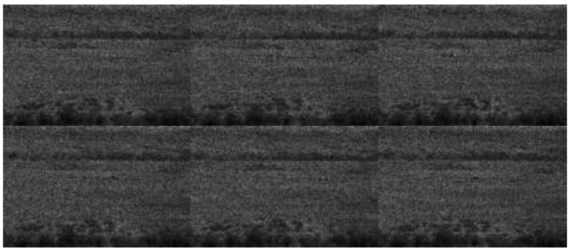
1500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	27.48	0.00	100.00
BRDM2	1.00	14.98	0.99	99.68
BTR70	1.00	21.91	0.16	100.00
SUV	1.00	18.89	0.84	100.00
T72	1.00	47.31	0.00	99.52
Truck	1.00	19.20	0.75	99.63
ZSU23-4	1.00	25.07	0.01	100.00

2000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	20.03	0.55	95.57
BRDM2	1.00	10.79	1.00	7.52
BTR70	1.00	16.37	1.00	98.35
SUV	1.00	13.60	1.00	50.37
T72	1.00	31.99	0.00	99.52
Truck	1.00	14.15	1.00	42.74
ZSU23-4	1.00	19.52	0.64	97.01

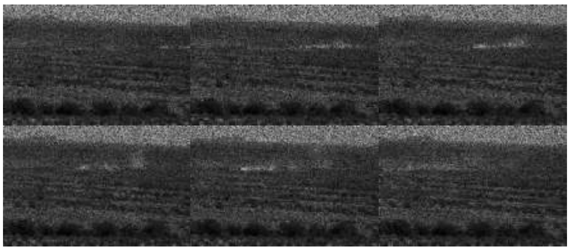
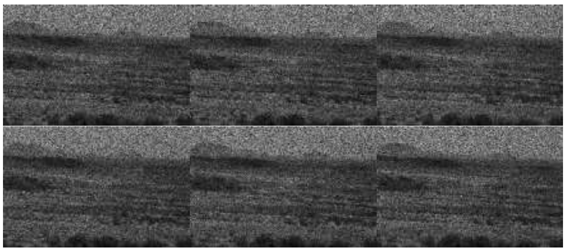
2500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.00	61.42	0.00	0.53
BRDM2	0.00	61.29	0.00	5.12
BTR70	0.00	84.98	0.00	1.76
SUV	0.00	79.45	0.00	3.84
T72	0.24	85.48	0.00	1.55
Truck	0.00	37.94	0.00	0.05
ZSU23-4	0.00	77.61	0.00	3.90

3000m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.94	12.11	0.95	92.37
BRDM2	0.99	5.32	1.00	96.05
BTR70	1.00	6.68	1.00	94.77
SUV	0.98	8.12	0.99	43.49
T72	0.99	15.63	0.98	93.06
Truck	0.97	8.44	0.98	84.26
ZSU23-4	0.98	8.78	0.98	92.90

3500m	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.00	58.30	0.05	9.50
BRDM2	0.00	61.01	0.08	2.08
BTR70	0.00	41.39	0.27	13.82
SUV	0.00	74.93	0.01	4.48
T72	0.00	59.56	0.25	16.81
Truck	0.00	73.13	0.01	7.63
ZSU23-4	0.00	64.78	0.05	15.80

(a) 1000 m (b) 2000 m

(c) 2500 m (d) 3500 m

Figure 8. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 50% missing case. Vehicle in the frames is SUV.

Table 3. Tracking metrics for 75% missing case.

1000m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	38.18	0.00	73.85
BRDM2	0.99	20.44	0.72	62.91
BTR70	1.00	28.34	0.00	95.14
SUV	0.97	28.91	0.07	54.16
T72	1.00	60.31	0.00	89.33
Truck	0.97	27.77	0.21	45.20
ZSU23-4	0.99	32.23	0.00	77.80

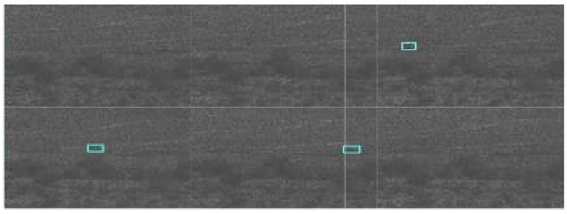
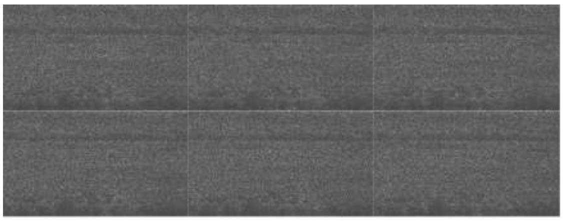
1500m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	27.68	0.00	6.87
BRDM2	0.99	16.96	0.92	6.38
BTR70	1.00	22.38	0.14	6.60
SUV	1.00	18.97	0.77	6.49
T72	1.00	47.56	0.00	7.90
Truck	1.00	19.50	0.67	6.49
ZSU23-4	1.00	25.12	0.02	6.73

2000m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	1.00	20.53	0.44	6.30
BRDM2	0.86	23.82	0.84	1.82
BTR70	1.00	17.28	0.94	6.33
SUV	0.97	17.15	0.95	3.76
T72	1.00	32.22	0.00	7.08
Truck	1.00	15.35	0.98	3.26
ZSU23-4	1.00	20.00	0.57	6.36

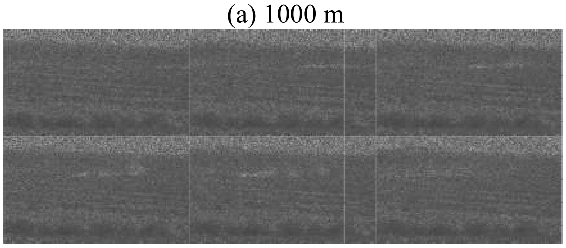
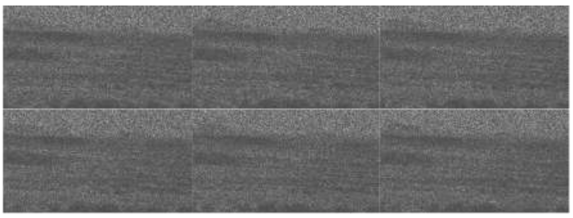
2500m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2				0.03
BRDM2	0.00	67.68	0.00	3.88
BTR70				0.09
SUV	0.00	93.08	0.00	5.17
T72				0.08
Truck				0.00
ZSU23-4	0.00	123.66	0.00	6.81

3000m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.45	38.27	0.49	7.07
BRDM2	0.81	17.71	0.83	6.21
BTR70	0.83	18.41	0.85	6.18
SUV	0.50	33.17	0.54	4.20
T72	0.70	35.23	0.68	6.97
Truck	0.67	28.95	0.71	6.17
ZSU23-4	0.65	28.16	0.67	6.58

3500m				
	EinGT	CLE	DP@20 pixels	% of frames with detections
BMP2	0.00	60.85	0.08	3.76
BRDM2	0.00	66.24	0.05	3.65
BTR70	0.00	51.50	0.07	3.50
SUV	0.00	58.57	0.03	3.37
T72	0.00	53.25	0.14	3.76
Truck	0.00	61.49	0.03	3.69
ZSU23-4	0.00	52.83	0.07	3.67

(a) 1000 m (b) 2000 m

(c) 2500m (d) 3500m

Figure 9. Tracking results for frames 1, 446, 892, 1338, 1784, and 2677. 75% missing case. Vehicle in the frames is SUV.

3.3. Classification Results

For vehicle classification, we deployed ResNet. For the ResNet classifier, we performed customized training where the training data are augmented with rotation, scaling, and illumination variations. We used videos from ranges 1500 m and 3000 m for training and videos from other ranges for testing.

Classification is only applied to frames with detection of targets from the tracker. Table 4-Table 6 summarize the ResNet classifiers for 0%, 50%, and 75% missing cases, respectively. The highlighted yellow cells indicate the biggest number of each row. The column labels in the confusion matrices are the ground truth and the row labels are the classifier output labels. For 0% missing, the classification performance is good up to 3000 m. For 50% missing case, the performance is still reasonable up to 2000m. However, for 75% missing case, the classification is only good up to 1500 m.

Table 4. Classification results for 0% missing case. Left columns are the confusion matrix; the last column is the classification results.

1000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	654	2	20	11	993	5	170	35%
BRDM2	0	1755	11	0	83	0	1	95%
BTR70	0	0	1441	0	420	1	12	77%
SUV	0	0	0	1862	12	0	0	99%
T72	0	0	0	0	1860	0	14	99%
Truck	12	0	9	304	338	1189	16	64%
ZSU23-4	0	2	142	1	1116	0	611	33%

1500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1826	0	1	0	47	0	0	97%
BRDM2	0	1865	0	0	9	0	0	100%
BTR70	2	0	1867	0	4	0	1	100%
SUV	0	0	0	1874	0	0	0	100%
T72	0	0	0	0	1871	0	3	100%
Truck	0	0	7	0	15	1852	0	99%
ZSU23-4	0	1	0	0	16	0	1857	99%

2000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1874	0	0	0	0	0	0	100%
BRDM2	0	1873	0	0	0	0	0	100%
BTR70	20	0	1854	0	0	0	0	99%
SUV	2	28	4	1687	6	137	0	91%
T72	1	0	0	0	1873	0	0	100%
Truck	3	41	0	0	6	1799	0	97%
ZSU23-4	0	6	0	0	2	0	1864	100%

2500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1090	4	47	8	59	28	11	87%
BRDM2	0	1513	9	22	36	26	0	94%
BTR70	109	2	800	0	37	10	56	79%
SUV	17	9	9	1076	115	349	9	68%
T72	15	3	47	37	1055	42	151	78%
Truck	72	20	12	119	78	1426	43	81%
ZSU23-4	1	0	0	4	14	5	1767	99%

3000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1550	17	3	0	262	6	36	83%
BRDM2	0	1865	1	0	0	8	0	100%
BTR70	45	0	1712	0	50	20	43	92%
SUV	63	41	14	1515	64	159	12	81%
T72	1	0	1	0	1780	0	2	100%
Truck	89	37	14	0	103	1552	78	83%
ZSU23-4	45	1	13	0	62	3	1750	93%

3500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	407	28	87	3	249	54	68	45%
BRDM2	11	524	2	2	1	14	3	94%
BTR70	96	5	350	0	69	10	18	64%
SUV	26	17	11	44	1	57	1	28%
T72	13	3	13	4	1338	43	5	94%
Truck	4	3	1	0	1	26	0	74%
ZSU23-4	215	16	48	0	139	24	500	53%

Table 5. Classification results for 50% missing case. Left columns are the confusion matrix; the last column is the classification results.

1000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1760	40	0	1	0	0	0	98%
BRDM2	2	1504	0	12	1	131	0	91%
BTR70	96	1	1740	4	0	27	1	93%
SUV	8	63	0	1615	5	17	0	95%
T72	0	3	0	9	1830	0	0	99%
Truck	118	249	2	61	35	1196	2	72%
ZSU23-4	16	82	0	8	178	0	1542	84%

1500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1838	32	0	4	0	0	0	98%
BRDM2	0	1861	3	3	1	0	0	100%
BTR70	39	1	1834	0	0	0	0	98%
SUV	17	13	0	1729	6	109	0	92%
T72	18	37	2	0	1806	2	0	97%
Truck	48	86	0	0	74	1659	0	89%
ZSU23-4	5	19	0	0	1	0	1849	99%

2000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1100	448	29	93	0	106	15	61%
BRDM2	0	67	26	23	3	21	1	48%
BTR70	1	15	1525	129	0	173	0	83%
SUV	0	250	9	459	7	219	0	49%
T72	120	111	30	860	329	370	45	18%
Truck	38	334	6	6	2	408	7	51%
ZSU23-4	45	866	146	127	43	29	562	31%

2500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	0	7	0	3	0	0	0	0%
BRDM2	6	10	4	9	63	4	0	10%
BTR70	0	1	1	2	29	0	0	3%
SUV	4	38	1	26	3	0	0	36%
T72	0	14	0	12	2	1	0	7%
Truck	0	1	0	0	0	0	0	0%
ZSU23-4	3	37	0	29	4	0	0	0%

3000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	11	145	8	897	626	41	3	1%
BRDM2	34	384	16	65	648	620	33	21%
BTR70	114	4	266	66	1253	29	44	15%
SUV	18	32	105	520	135	3	2	64%
T72	7	18	2	505	1074	138	0	62%
Truck	72	35	32	54	884	456	46	29%
ZSU23-4	2	17	2	136	1476	29	79	5%

3500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	0	13	0	165	0	0	0	0%
BRDM2	4	7	1	8	17	2	0	18%
BTR70	4	32	10	203	6	4	0	4%
SUV	0	18	7	34	21	0	4	40%
T72	3	26	2	157	127	0	0	40%
Truck	9	27	22	36	39	0	10	0%
ZSU23-4	0	13	1	282	0	0	0	0%

Table 6. Classification results for 75% missing case. Left columns are the confusion matrix; the last column is the classification results.

1000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1345	32	0	1	3	3	0	97%
BRDM2	3	1032	0	15	7	122	0	88%
BTR70	261	18	1481	1	6	16	0	83%
SUV	33	43	0	850	4	85	0	84%
T72	3	15	0	7	1644	4	1	98%
Truck	106	177	2	26	30	505	1	60%
ZSU23-4	27	133	0	5	427	12	854	59%

2500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	0	0	0	0	0	0	0	0%
BRDM2	0	0	0	1	5	0	0	0%
BTR70	0	0	0	0	0	0	0	0%
SUV	0	1	0	2	0	0	0	67%
T72	0	0	0	0	0	0	0	0%
Truck	0	0	0	0	0	0	0	0%
ZSU23-4	0	0	1	1	0	0	0	0%

1500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1783	61	10	2	8	0	0	96%
BRDM2	3	1508	6	6	43	4	3	96%
BTR70	163	6	1702	0	1	1	0	91%
SUV	85	27	0	1168	41	383	0	69%
T72	103	76	52	2	1586	5	8	87%
Truck	157	65	1	0	428	1013	3	61%
ZSU23-4	76	150	0	0	104	0	1507	82%

3000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	1	54	2	489	113	2	0	0%
BRDM2	51	87	19	98	836	431	14	6%
BTR70	65	61	154	169	1216	35	10	9%
SUV	1	9	6	168	44	0	1	73%
T72	46	83	26	371	839	34	4	60%
Truck	81	77	100	245	751	138	43	10%
ZSU23-4	14	103	32	372	833	17	4	0%

2000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	698	357	119	74	11	119	17	50%
BRDM2	1	51	17	33	28	17	1	34%
BTR70	11	32	1202	130	8	125	0	80%
SUV	1	56	2	53	7	54	0	31%
T72	95	97	66	667	361	291	5	23%
Truck	7	31	1	0	1	50	2	54%
ZSU23-4	85	716	184	129	215	21	119	8%

3500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	Accuracy
BMP2	0	1	0	61	0	0	0	0%
BRDM2	22	58	25	48	179	7	4	17%
BTR70	9	35	20	174	32	2	1	7%
SUV	21	150	60	415	119	3	9	53%
T72	0	35	7	185	31	0	0	12%
Truck	64	146	110	410	183	11	23	1%
ZSU23-4	0	23	1	413	4	0	0	0%

3.4. Discussions: Tracking and Classification at 75% missing rate: Optical Videos

Table 7 summarizes the tracking (YOLO) and classification (ResNet) results for 75% missing data case. By looking at
Table 7, we have the following observations:

- If we have trained models for some specific ranges, then the tracking accuracy is good. For example, we have trained models at 1500 m and 3000 m and the tracking results are good at 1500 m, but tracking results deteriorate at 3000 m. This means that, in theory, tracking can perform well if we have enough training data at various ranges.
- The classification is only effective up to 1500 m. This is because the target size is so small at farther ranges. The target information is further reduced due to missing pixels.

Table 7. Summary of Tracking and Classification Results at 75% missing rate.

	75% summary	
Range	Average % of frames with detections	Average Accuracy
1000	71.20	81%
1500	94.15	83%
2000	48.54	40%
2500	0.08	10%
3000	63.65	23%
3500	23.64	13%

4. Tracking And Classification Results Using SENSIAC MWIR Day Time Videos

There are two types of MWIR videos: daytime and night-time. Here, we focus on daytime MWIR videos.

4.1. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

For the YOLO models, we used 1500 m and 3000 m videos to train two separate models. The 1500m model will be used for 1000 m and 2000 m ranges and the 3000 m model will be for ranges of 2500 m and 3500 m. Table 8-Table 10 show the metrics. We can see that the numbers of frames with detection are high for low missing rates. For frames with detection, the CLE values generally increase whereas the DP and EinGT values are relatively stable.

Figure 10-Figure 12 show the tracking results visually. The tracking results are good up to 2000 m at 0% missing. For longer ranges such as 2500 m and beyond, the detection is not good. For example, at 50% missing rate, the tracking is only good up to 1000 m. For 75% missing, the tracking is poor even at 1000 m range. This means MWIR is not suitable for daytime surveillance.

Table 8. Tracking metrics for 0% missing case. MWIR Day time videos.

1000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	29.64	0.00	46.11
BRDM2	1.00	25.00	0.13	52.72
BTR70	1.00	27.19	0.85	44.83
SUV	1.00	17.51	0.85	85.39
T72	0.99	38.89	0.00	80.72
Truck	1.00	23.15	0.25	81.56
ZSU23-4	1.00	28.64	0.03	61.33

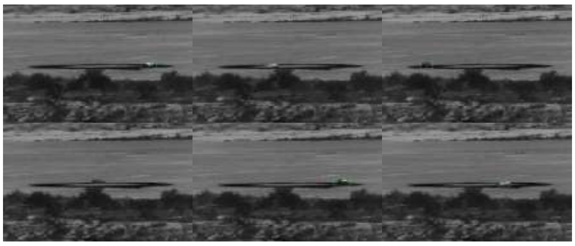
1500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	21.55	0.16	99.39
BRDM2	1.00	18.94	0.86	99.72
BTR70	1.00	17.11	0.98	99.94
SUV	1.00	12.99	1.00	99.11
T72	1.00	25.71	0.00	95.61
Truck	1.00	16.29	0.98	100.00
ZSU23-4	1.00	20.25	0.52	100.00

2000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	15.11	1.00	99.67
BRDM2	1.00	12.29	1.00	96.50
BTR70	1.00	10.90	1.00	98.28
SUV	1.00	9.78	1.00	70.11
T72	1.00	18.00	0.93	88.67
Truck	1.00	10.77	1.00	73.11
ZSU23-4	1.00	13.07	0.99	89.56

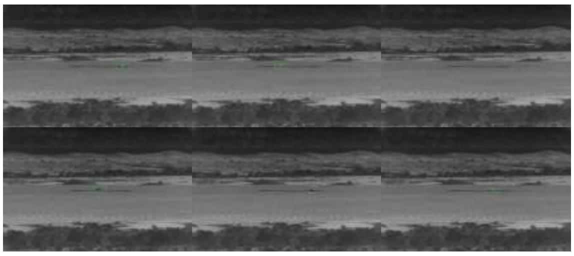
2500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	10.36	1.00	83.00
BRDM2	0.91	15.36	0.91	94.67
BTR70	0.41	49.83	0.44	70.78
SUV	0.39	57.76	0.40	51.89
T72	0.74	30.61	0.76	57.00
Truck	0.65	34.39	0.69	56.33
ZSU23-4	0.65	20.31	0.72	81.44

3000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	7.30	1.00	100.00
BRDM2	1.00	6.20	1.00	100.00
BTR70	1.00	5.88	1.00	100.00
SUV	1.00	4.79	1.00	100.00
T72	1.00	8.42	1.00	100.00
Truck	1.00	5.57	1.00	99.89
ZSU23-4	1.00	6.56	1.00	100.00

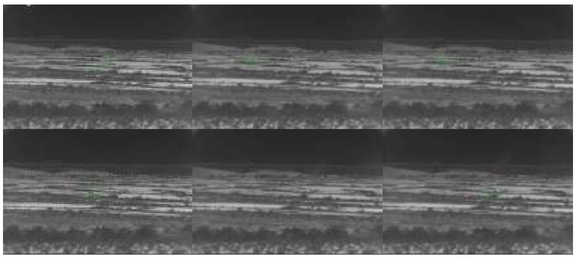
3500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	4.09	1.00	28.33
BRDM2	0.69	2.85	1.00	56.56
BTR70	0.92	4.33	0.99	41.17
SUV	0.68	52.73	0.69	22.61
T72	0.63	7.29	0.95	75.06
Truck	0.93	4.02	0.99	32.78
ZSU23-4	0.96	4.53	0.98	37.50



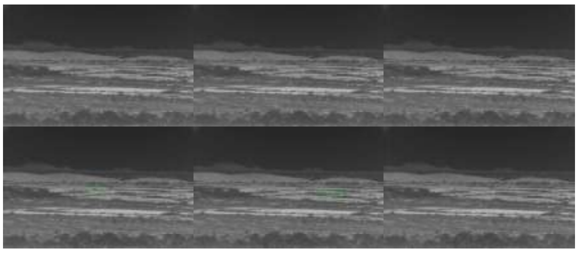
(a) 1000m



(b) 2000m



(c) 2500m



(d) 3500m

Figure 10. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 0% missing case. Vehicle: SUV.

Table 9. Tracking metrics for 50% missing case. MWIR Day time videos.

1000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.99	31.95	0.02	66.50
BRDM2	1.00	26.90	0.12	83.00
BTR70	1.00	24.50	0.20	60.83
SUV	1.00	19.51	0.62	76.06
T72	1.00	36.01	0.00	100.00
Truck	1.00	24.66	0.18	50.22
ZSU23-4	1.00	31.79	0.00	93.28

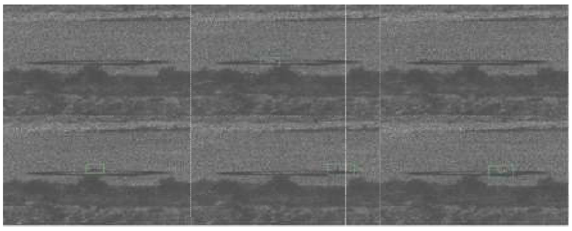
1500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	22.55	0.13	97.83
BRDM2	1.00	19.51	0.65	99.94
BTR70	1.00	18.65	0.77	99.61
SUV	1.00	14.65	0.99	91.17
T72	1.00	25.50	0.01	98.17
Truck	1.00	18.18	0.86	84.89
ZSU23-4	1.00	21.17	0.32	99.33

2000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	16.63	0.91	78.72
BRDM2	1.00	12.73	0.99	67.11
BTR70	1.00	13.35	1.00	37.06
SUV	0.98	11.02	0.98	3.17
T72	1.00	18.55	0.83	63.06
Truck	0.92	14.66	0.88	10.89
ZSU23-4	1.00	15.04	0.94	69.11

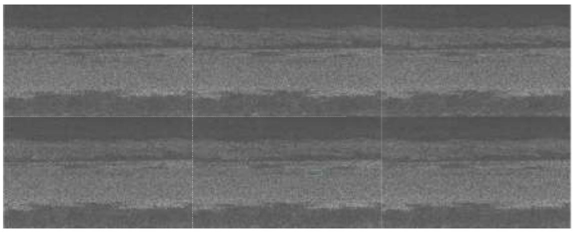
2500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.08	16.58	0.88	1.39
BRDM2	0.00	52.14	0.01	10.39
BTR70	0.00	55.84	0.00	7.78
SUV	0.00	51.86	0.00	2.94
T72	0.00	55.47	0.00	0.28
Truck	0.04	42.91	0.04	1.33
ZSU23-4	0.00	32.63	0.12	0.94

3000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.95	9.45	0.98	15.00
BRDM2	0.92	9.55	0.94	54.44
BTR70	0.95	7.60	0.97	64.89
SUV	0.80	14.52	0.85	18.39
T72	0.91	6.93	0.90	50.11
Truck	0.87	11.34	0.90	29.56
ZSU23-4	0.93	9.44	0.96	69.44

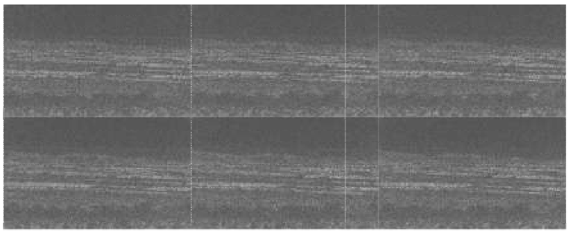
3500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.00	34.39	0.33	0.33
BRDM2	0.00	52.42	0.17	1.33
BTR70	0.00	59.81	0.06	2.83
SUV	0.00	52.73	0.10	6.72
T72	0.00	5.92	1.00	14.28
Truck	0.00	63.58	0.17	3.56
ZSU23-4	0.00	49.06	0.16	8.89



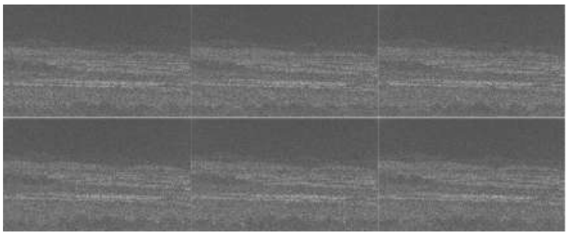
(a) 1000m



(b) 2000m



(c) 2500m



(d) 3500m

Figure 11. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 50% missing case. Vehicle: SUV.

Table 10. Tracking metrics for 75% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges. MWIR Day time videos.

1000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.89	41.58	0.04	4.11
BRDM2	0.97	31.05	0.21	8.83
BTR70	0.98	19.62	0.67	3.22
SUV	0.94	25.85	0.69	2.67
T72	1.00	37.53	0.00	43.94
Truck	0.90	28.63	0.39	2.28
ZSU23-4	0.95	35.71	0.08	23.89

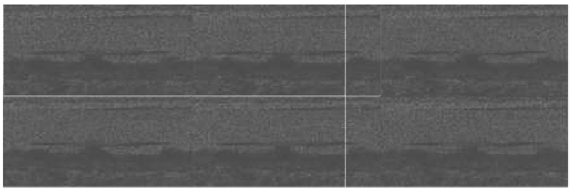
1500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.94	30.18	0.29	44.83
BRDM2	0.99	20.04	0.71	82.17
BTR70	0.98	19.94	0.75	68.78
SUV	0.87	19.38	0.89	15.28
T72	0.99	27.58	0.04	87.94
Truck	0.96	20.67	0.72	20.72
ZSU23-4	0.99	21.77	0.44	92.39

2000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.89	24.95	0.85	21.17
BRDM2	0.98	13.14	0.96	21.67
BTR70	0.94	20.21	0.91	1.83
SUV	0.50	62.33	0.50	0.22
T72	0.99	18.53	0.80	21.22
Truck	0.97	15.44	0.88	5.22
ZSU23-4	1.00	13.27	0.97	13.78

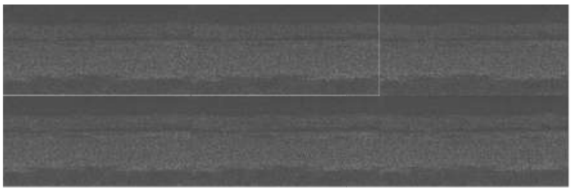
2500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00
BTR70	0.00	0.00	0.00	0.00
SUV	0.00	0.00	0.00	0.00
T72	0.00	0.00	0.00	0.00
Truck	0.00	0.00	0.00	0.00
ZSU23-4	0.00	0.00	0.00	0.00

3000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00
BTR70	0.00	0.00	0.00	0.00
SUV	0.00	0.00	0.00	0.00
T72	0.00	0.00	0.00	0.00
Truck	0.00	0.00	0.00	0.00
ZSU23-4	0.00	0.00	0.00	0.00

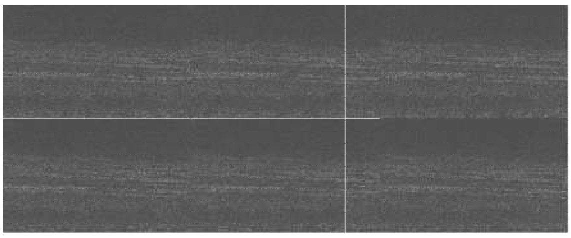
3500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.00	0.00	0.00	0.00
BRDM2	0.00	0.00	0.00	0.00
BTR70	0.00	0.00	0.00	0.00
SUV	0.00	0.00	0.00	0.00
T72	0.00	0.00	0.00	0.00
Truck	0.00	0.00	0.00	0.00
ZSU23-4	0.00	0.00	0.00	0.00



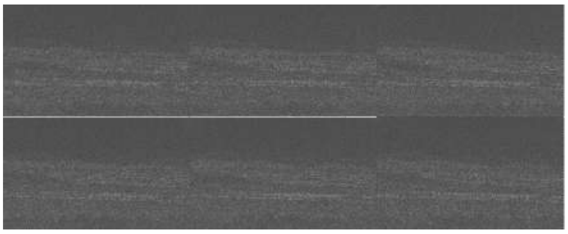
(a) 1000m



(b) 2000m



(c) 2500m



(d) 3500m

Figure 12. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 75% missing case. Vehicle: SUV.

4.2. Classification Results

Classification is only applied to frames with detection of targets from the tracker. Table 11-Table 13 summarize the ResNet classifiers for 0%, 50%, and 75% missing cases, respectively. The highlighted yellow cells indicate the biggest number of each row. The column labels in the confusion matrices are the ground truth and the row labels are the classifier output labels. As shown in the tables, the classification results using ResNet is in general not so good even at 1000 m with 75% missing. This means MWIR is not suitable for daytime surveillance.

Table 11. Classification results for 0% missing case. Left is the confusion matrix; the last column is the classification accuracy.

1000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	19	349	176	101	40	38	107	2%
BRDM2	1	859	51	9	12	5	12	91%
BTR70	1	33	738	16	8	7	4	91%
SUV	0	62	10	1425	1	39	0	93%
T72	71	236	22	474	479	78	93	33%
Truck	2	241	121	135	0	968	1	66%
ZSU23-4	3	72	9	53	53	114	800	72%

1500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	1300	377	0	0	0	112	0%
BRDM2	0	1687	0	6	0	102	0	94%
BTR70	2	18	1702	13	1	53	10	95%
SUV	0	387	4	1128	5	256	4	63%
T72	0	662	154	142	480	235	48	28%
Truck	0	503	7	71	2	1063	154	59%
ZSU23-4	0	41	1	0	0	0	1758	98%

2000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	1088	584	52	0	46	24	0%
BRDM2	205	221	0	0	199	994	118	13%
BTR70	6	1186	377	0	63	0	137	21%
SUV	8	107	5	60	884	103	95	5%
T72	0	415	18	0	824	269	70	52%
Truck	4	60	2	2	143	1053	52	80%
ZSU23-4	0	2	13	0	9	0	1588	99%

2500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	20	139	1	111	10	1056	157	1%
BRDM2	0	33	107	51	93	1417	3	2%
BTR70	0	179	48	121	106	789	31	4%
SUV	1	31	85	277	138	397	5	30%
T72	0	71	250	63	88	503	51	9%
Truck	0	34	79	196	65	639	1	63%
ZSU23-4	0	19	138	16	37	1221	35	2%

3000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	608	8	250	7	897	30	0%
BRDM2	10	83	412	611	26	657	1	5%
BTR70	0	153	1	859	114	653	20	0%
SUV	0	39	251	394	64	1052	0	22%
T72	5	353	1	112	179	903	247	10%
Truck	0	36	3	170	570	1017	2	57%
ZSU23-4	0	111	317	117	2	1215	38	2%

3500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	453	0	0	3	54	0	0%
BRDM2	36	42	64	659	0	217	0	4%
BTR70	0	40	1	682	4	14	0	0%
SUV	29	38	1	332	7	0	0	82%
T72	3	104	0	3	186	776	279	14%
Truck	6	27	3	516	9	29	0	5%
ZSU23-4	0	82	5	299	3	281	5	1%

Table 12. Classification results for 50% missing case. Left is the confusion matrix; right is the classification results.

1000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	153	273	105	117	0	6	543	13%
BRDM2	23	793	631	8	0	18	21	53%
BTR70	16	4	1057	15	0	1	2	97%
SUV	4	216	17	1014	0	118	0	74%
T72	23	187	183	1113	30	28	236	2%
Truck	45	131	49	193	0	483	3	53%
ZSU23-4	33	146	198	124	0	16	1162	69%

1500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	13	1144	78	279	0	84	163	1%
BRDM2	1	1449	23	4	0	304	18	81%
BTR70	3	112	1059	78	10	472	59	59%
SUV	0	496	3	897	6	228	11	55%
T72	0	535	191	263	144	591	43	8%
Truck	0	502	1	63	0	838	124	55%
ZSU23-4	0	50	59	5	0	1	1673	94%

2000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	75	399	6	570	11	316	40	5%
BRDM2	131	290	2	0	187	497	101	24%
BTR70	4	578	14	0	15	1	55	2%
SUV	0	26	0	0	26	3	2	0%
T72	2	689	20	0	334	69	21	29%
Truck	3	16	1	2	65	105	4	54%
ZSU23-4	7	53	19	0	7	0	1158	93%

2500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	0	0	17	0	6	2	0%
BRDM2	0	18	101	7	7	54	0	10%
BTR70	0	20	81	11	2	26	0	58%
SUV	0	2	35	2	1	13	0	4%
T72	0	0	5	0	0	0	0	0%
Truck	0	3	16	0	2	3	0	13%
ZSU23-4	0	0	5	0	0	12	0	0%

3000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	62	0	4	1	202	1	0%
BRDM2	13	105	154	386	13	309	0	11%
BTR70	0	146	6	435	117	462	2	1%
SUV	0	3	2	45	15	266	0	14%
T72	1	201	0	23	103	478	96	11%
Truck	0	22	1	86	75	348	0	65%
ZSU23-4	0	218	63	105	4	860	0	0%

3500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	2	0	4	0	0	0	0%
BRDM2	0	4	0	16	0	4	0	17%
BTR70	0	6	0	24	11	10	0	0%
SUV	0	1	22	39	57	2	0	32%
T72	0	74	0	0	81	27	75	32%
Truck	0	3	0	20	11	30	0	47%
ZSU23-4	0	14	1	97	10	38	0	0%

Table 13. Classification results for 75% missing case. Left is the confusion matrix; right is the classification results.

1000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	11	34	7	6	0	0	16	15%
BRDM2	1	107	48	1	0	0	2	67%
BTR70	0	0	55	2	0	0	1	95%
SUV	0	13	1	34	0	0	0	71%
T72	14	220	40	373	51	35	58	6%
Truck	0	11	2	10	0	18	0	44%
ZSU23-4	3	200	40	53	6	5	123	29%

1500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	648	17	48	1	74	19	0%
BRDM2	3	1250	39	6	1	170	10	85%
BTR70	4	287	626	26	4	266	25	51%
SUV	1	187	1	65	2	11	8	24%
T72	1	790	114	151	184	284	59	12%
Truck	0	97	0	23	2	221	30	59%
ZSU23-4	0	562	133	25	1	8	934	56%

2000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	3	192	0	85	2	83	16	1%
BRDM2	27	110	1	0	119	77	56	28%
BTR70	0	31	0	0	2	0	0	0%
SUV	0	2	0	0	2	0	0	0%
T72	1	201	3	0	98	65	14	26%
Truck	1	11	0	0	48	32	2	34%
ZSU23-4	1	59	2	0	7	6	173	70%

2500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	0	0	0	0	0	0	0%
BRDM2	0	0	0	0	0	0	0	0%
BTR70	0	0	0	0	0	0	0	0%
SUV	0	0	0	0	0	0	0	0%
T72	0	0	0	0	0	0	0	0%
Truck	0	0	0	0	0	0	0	0%
ZSU23-4	0	0	0	0	0	0	0	0%

3000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	0	0	0	0	0	0	0%
BRDM2	0	0	0	0	0	0	0	0%
BTR70	0	0	0	0	0	0	0	0%
SUV	0	0	0	0	0	0	0	0%
T72	0	0	0	0	0	0	0	0%
Truck	0	0	0	0	0	0	0	0%
ZSU23-4	0	0	0	0	0	0	0	0%

3500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	0	0	0	0	0	0	0	0%
BRDM2	0	0	0	0	0	0	0	0%
BTR70	0	0	0	0	0	0	0	0%
SUV	0	0	0	0	0	0	0	0%
T72	0	0	0	0	0	0	0	0%
Truck	0	0	0	0	0	0	0	0%
ZSU23-4	0	0	0	0	0	0	0	0%

4.3. Discussions: Tracking and Classification at 75% rate at MWIR Day-Time

Since our research objective is to perform tracking and classification using compressive measurements, we focus on 75% missing pixel scenario where there are savings in both storage and transmission. From Table 14, we can see that both tracking and classification performance using MWIR videos is not good. Hence, MWIR is not suitable for target tracking and classification for day time videos.

Table 14. MWIR Day Time Summary at 75% missing rate.

Range	Average % of frames with detections	Average Accuracy
1000	12.71	47%
1500	58.87	41%
2000	12.16	23%
2500	0.00	0%
3000	0.00	0%
3500	0.00	0%

5. Tracking And Classification Results Using SENSIAC MWIR Night-Time Videos

5.1. Train using 1500 m and 3000 m videos and Test using 1000 m, 2000 m, 2500 m, 3500 m videos

Now, we will focus on tracking and classification for MWIR night-time videos. Table 15-Table 17 show the metrics when we used videos from 1500 m and 3000 m for training and other videos from other ranges for testing. We can see that the numbers of frames with detection are high for low missing rates. For a given missing rate, the percentage of frames with detection decreases with longer ranges.

Figure 13-Figure 15 show the tracking results visually. It can be seen that we have good tracking up to 3500 m at 0% missing rate. At 50%, the tracking performance drops, as we can barely see some bounding boxes around vehicles up to range 2000 m. At 75% missing rate, the tracking is only good up to 1000 m.

Table 15. Tracking metrics for 0% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.

1000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	26.21	0.01	31.11
BRDM2	1.00	24.60	0.13	88.00
BTR70	1.00	15.36	0.88	80.89
SUV	1.00	12.91	1.00	99.67
T72	1.00	31.50	0.00	86.83
Truck	1.00	24.81	0.01	94.89
ZSU23-4	1.00	25.52	0.08	66.89

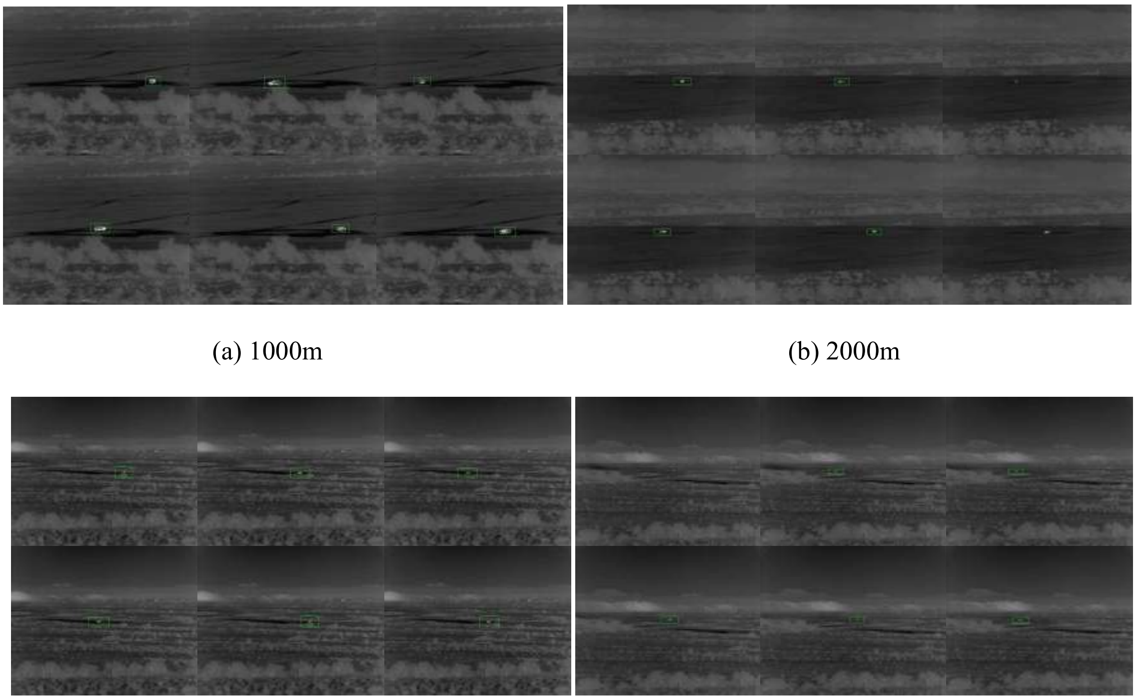
1500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	18.65	0.95	99.83
BRDM2	1.00	18.15	0.93	98.83
BTR70	1.00	10.54	1.00	100.00
SUV	1.00	9.29	1.00	99.83
T72	1.00	23.04	0.03	99.06
Truck	1.00	19.09	0.86	99.78
ZSU23-4	1.00	18.76	0.91	99.67

2000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	12.80	1.00	80.56
BRDM2	1.00	12.42	1.00	82.17
BTR70	1.00	6.62	1.00	91.44
SUV	1.00	5.66	1.00	73.50
T72	1.00	15.62	1.00	94.67
Truck	1.00	12.92	1.00	49.61
ZSU23-4	1.00	12.51	1.00	93.61

2500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.99	9.96	0.99	96.72
BRDM2	1.00	8.71	1.00	100.00
BTR70	0.99	5.19	1.00	99.89
SUV	0.95	4.55	1.00	100.00
T72	1.00	10.28	1.00	99.67
Truck	1.00	8.18	1.00	91.17
ZSU23-4	0.94	8.33	1.00	91.33

3000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	5.92	1.00	100.00
BRDM2	1.00	5.55	1.00	100.00
BTR70	1.00	3.01	1.00	100.00
SUV	1.00	2.87	1.00	100.00
T72	1.00	7.36	1.00	100.00
Truck	1.00	5.86	1.00	100.00
ZSU23-4	1.00	5.68	1.00	100.00

3500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.87	2.38	1.00	65.83
BRDM2	0.99	2.22	1.00	96.56
BTR70	0.78	1.71	1.00	97.06
SUV	0.66	1.29	1.00	96.39
T72	0.89	3.32	1.00	99.22
Truck	0.63	7.14	0.97	94.17
ZSU23-4	0.90	2.84	1.00	99.94



(a) 1000m (b) 2000m

(c) 2500m (d) 3500m

Figure 13. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 0% missing case. Vehicle is SUV.

Table 16. Tracking metrics for 50% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.

1000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	25.71	0.05	74.83
BRDM2	1.00	23.77	0.16	98.83
BTR70	1.00	14.79	0.92	94.17
SUV	1.00	12.81	1.00	100.00
T72	1.00	30.92	0.00	95.78
Truck	1.00	25.34	0.00	95.94
ZSU23-4	1.00	26.10	0.00	95.83

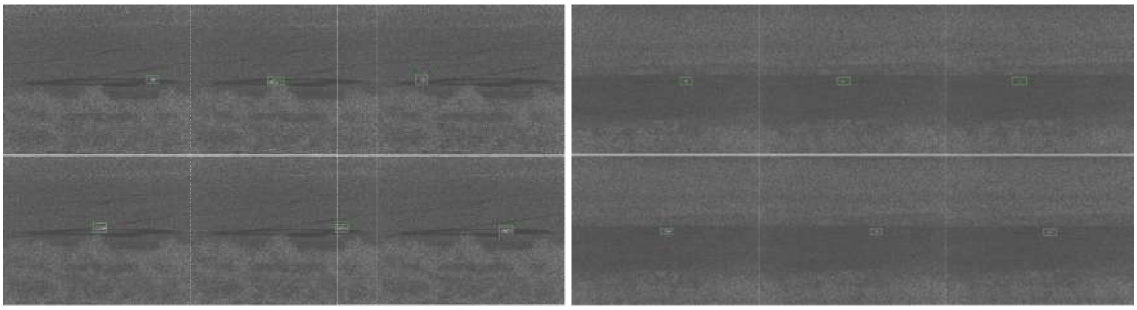
1500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	18.90	0.81	99.83
BRDM2	1.00	18.14	0.93	100.00
BTR70	1.00	10.63	1.00	100.00
SUV	1.00	9.57	1.00	100.00
T72	1.00	23.13	0.03	99.28
Truck	1.00	18.71	0.88	99.94
ZSU23-4	1.00	18.63	0.91	100.00

2000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	13.35	1.00	39.67
BRDM2	1.00	13.13	1.00	99.83
BTR70	1.00	7.21	1.00	98.17
SUV	1.00	6.37	1.00	95.89
T72	1.00	15.78	0.99	97.83
Truck	1.00	13.35	0.99	65.17
ZSU23-4	1.00	12.86	1.00	94.17

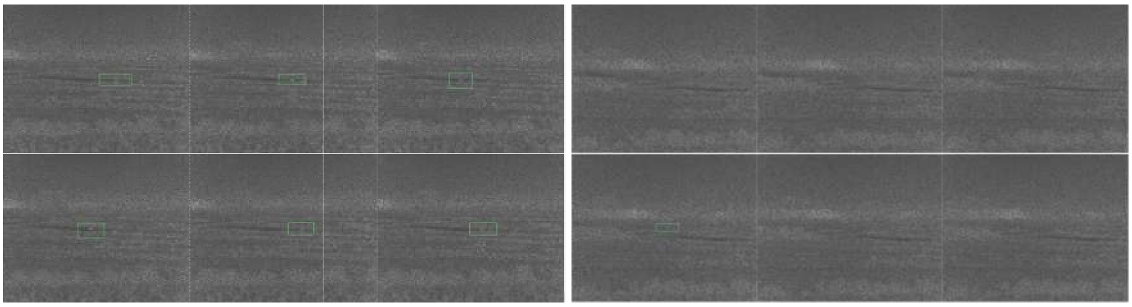
2500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.80	10.87	0.97	29.89
BRDM2	0.98	7.36	1.00	99.06
BTR70	0.97	5.71	1.00	99.94
SUV	0.93	3.75	1.00	75.33
T72	0.97	8.04	1.00	85.39
Truck	0.90	7.25	1.00	66.17
ZSU23-4	0.97	8.23	1.00	86.22

3000m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.90	5.07	1.00	90.67
BRDM2	0.93	4.54	1.00	56.28
BTR70	0.95	3.50	1.00	86.00
SUV	1.00	2.79	1.00	58.94
T72	0.97	7.36	1.00	98.83
Truck	1.00	6.25	1.00	78.56
ZSU23-4	0.91	5.74	1.00	93.56

3500m	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.00	14.11	0.94	13.83
BRDM2	0.22	7.08	1.00	12.39
BTR70	0.81	2.65	1.00	9.11
SUV	0.85	1.52	1.00	7.17
T72	0.02	7.29	1.00	69.56
Truck	0.46	3.18	0.99	11.28
ZSU23-4	0.65	5.27	0.99	19.61



(a) 1000m (b) 2000m



(c) 2500m (d) 3500m

Figure 14. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 50% missing case. Vehicle is SUV.

Table 17. Tracking metrics for 75% missing case. Train using videos from 1500 m and 3000 m ranges and test using videos from other ranges.

1000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	27.83	0.05	46.44
BRDM2	1.00	24.60	0.31	93.22
BTR70	1.00	15.87	0.74	80.50
SUV	1.00	12.43	1.00	99.89
T72	1.00	31.11	0.00	51.06
Truck	1.00	25.30	0.07	96.00
ZSU23-4	1.00	25.88	0.07	92.72

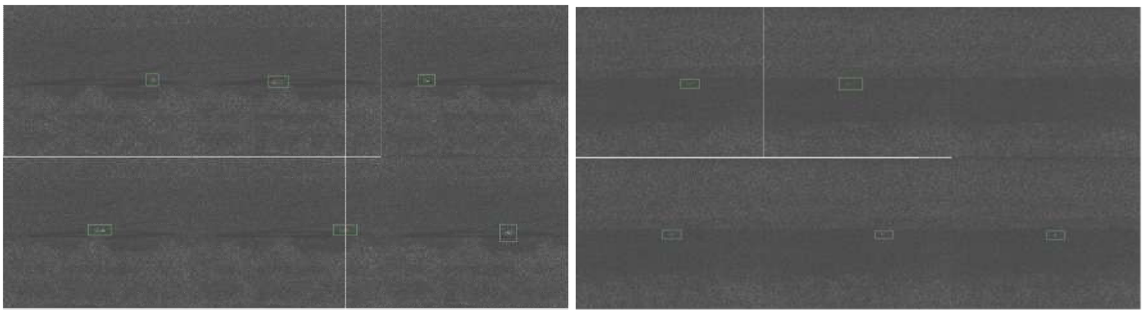
1500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	18.92	0.72	97.56
BRDM2	1.00	18.52	0.77	99.61
BTR70	1.00	10.83	1.00	99.83
SUV	1.00	9.80	1.00	99.56
T72	1.00	23.06	0.10	99.89
Truck	1.00	18.86	0.72	96.33
ZSU23-4	1.00	19.65	0.70	96.28

2000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	1.00	13.19	0.99	28.39
BRDM2	1.00	13.51	0.98	95.89
BTR70	1.00	7.46	1.00	82.89
SUV	1.00	6.47	1.00	84.17
T72	1.00	15.69	0.97	96.44
Truck	0.98	14.06	0.91	30.39
ZSU23-4	1.00	12.98	1.00	78.61

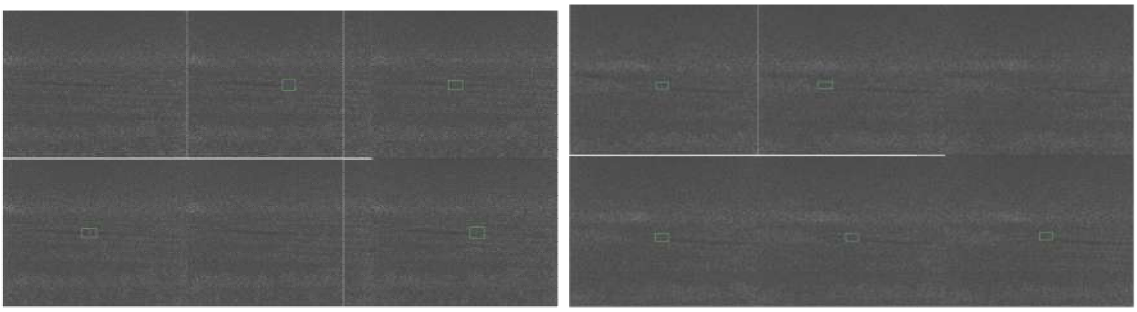
2500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.30	162.08	0.30	12.89
BRDM2	1.00	9.66	1.00	85.11
BTR70	0.89	18.33	0.89	52.89
SUV	0.96	6.81	0.97	71.50
T72	0.97	16.95	0.97	33.28
Truck	0.92	14.53	0.94	83.56
ZSU23-4	0.95	12.41	0.95	17.33

3000m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.98	6.87	0.99	98.17
BRDM2	1.00	5.66	1.00	99.94
BTR70	0.98	3.51	1.00	99.83
SUV	0.94	5.57	0.97	96.39
T72	1.00	6.79	1.00	99.94
Truck	0.96	8.66	0.98	97.72
ZSU23-4	0.99	5.87	1.00	99.61

3500m				
	EinGT	CLE	DP@20 pixels	% of detections
BMP2	0.01	61.63	0.16	48.83
BRDM2	0.38	17.65	0.57	21.50
BTR70	0.28	23.91	0.60	51.50
SUV	0.01	58.81	0.10	37.94
T72	0.40	6.48	0.93	59.61
Truck	0.01	57.26	0.13	59.17
ZSU23-4	0.12	49.29	0.36	14.67



(a) 1000m (b) 2000m



(c) 2500m (d) 3500m

Figure 15. Tracking results for frames 1, 300, 600, 900, 1200, and 1500. 75% missing case. Vehicle is SUV.

5.2. Classification Results

The classification is still done with ResNet. As shown in Table 18-Table 20, we have the following observations:

- At 0% missing case, classification performance was reasonable for videos up to 2000 m. Some vehicles at 2000 m could be correctly classified even they could be correctly tracked at high percentage. The classification performance was not good for videos beyond 2500 m. For example, the accuracy at 3000 m was poor even using the same training videos for testing.
- At 50% missing case, we have the same observations as those at 0% missing rate. However, the percentages of correct classification were lower.
- Similarly, at 75% missing case, the percentages of correct classification dropped to even lower values even for ranges below 2000m.

Table 18. Classification results for 0% missing case. Left is the confusion matrix; the last column is the classification accuracy.

1000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	474	19	0	0	59	5	3	85%
BRDM2	14	1492	5	0	4	43	26	94%
BTR70	10	2	1281	6	18	3	136	88%
SUV	0	0	8	1366	1	0	419	76%
T72	5	1	7	2	1458	4	86	93%
Truck	66	0	0	5	238	662	737	39%
ZSU23-4	0	0	24	0	34	0	1146	95%

1500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	1364	0	55	0	133	171	74	76%
BRDM2	0	1772	1	0	6	0	0	100%
BTR70	0	0	1764	0	17	0	19	98%
SUV	0	0	124	905	40	4	724	50%
T72	0	0	0	0	1783	0	0	100%
Truck	0	0	61	0	711	75	949	4%
ZSU23-4	0	0	15	0	20	0	1759	98%

2000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	23	3	6	0	877	536	5	2%
BRDM2	1	1146	3	2	293	13	21	77%
BTR70	17	10	1236	15	229	50	89	75%
SUV	0	0	345	305	401	7	265	23%
T72	248	0	4	0	1398	3	51	82%
Truck	0	0	196	0	327	4	366	0%
ZSU23-4	32	4	85	7	399	121	1037	62%

2500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	137	208	0	172	553	667	4	8%
BRDM2	44	297	0	17	1161	173	108	17%
BTR70	119	6	0	1	1012	455	205	0%
SUV	400	34	2	0	772	359	233	0%
T72	232	14	17	1	405	361	764	23%
Truck	430	47	1	39	541	283	300	17%
ZSU23-4	399	0	62	1	315	619	248	15%

3000m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	232	511	0	0	525	532	0	13%
BRDM2	257	0	6	4	832	332	369	0%
BTR70	39	40	0	13	1372	336	0	0%
SUV	5	233	0	418	925	187	32	23%
T72	10	1	0	236	825	242	486	46%
Truck	114	441	0	221	782	237	5	13%
ZSU23-4	164	6	223	28	435	212	732	41%

3500m								
	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	57	274	0	70	446	323	15	5%
BRDM2	210	0	1	211	754	556	6	0%
BTR70	251	212	0	355	823	27	79	0%
SUV	8	294	0	20	984	426	3	1%
T72	61	0	1	43	473	419	789	26%
Truck	82	372	26	430	669	106	10	6%
ZSU23-4	321	0	0	31	1038	123	286	16%

Table 19. Classification results for 50% missing case. Left is the confusion matrix; the last column is the classification accuracy.

1000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	1275	2	0	0	47	0	23	95%
BRDM2	0	1774	3	0	1	0	1	100%
BTR70	1	2	1658	0	10	5	19	98%
SUV	0	0	0	1746	0	0	54	97%
T72	1	6	0	0	1696	1	20	98%
Truck	116	6	8	12	245	939	401	54%
ZSU23-4	4	0	0	0	3	1	1717	100%

2500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	190	77	2	1	196	48	24	35%
BRDM2	238	234	0	0	1233	67	11	13%
BTR70	56	6	4	6	1059	666	2	0%
SUV	428	53	2	1	516	355	1	0%
T72	197	35	34	14	582	298	377	38%
Truck	354	129	1	4	426	270	7	23%
ZSU23-4	183	1	11	9	562	677	109	7%

1500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	1221	2	40	0	214	137	183	68%
BRDM2	0	1771	1	0	27	1	0	98%
BTR70	0	0	1799	0	0	0	1	100%
SUV	0	0	35	1491	1	26	247	83%
T72	2	0	2	0	1750	0	33	98%
Truck	9	5	106	8	506	546	619	30%
ZSU23-4	1	0	15	1	17	0	1766	98%

3000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	808	52	0	0	617	81	74	50%
BRDM2	125	0	0	44	153	246	445	0%
BTR70	72	41	3	2	1081	343	6	0%
SUV	71	116	0	94	712	67	1	9%
T72	27	49	0	112	788	221	582	44%
Truck	105	430	0	21	777	79	2	6%
ZSU23-4	70	2	80	11	558	267	696	41%

2000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	11	14	21	0	519	135	14	2%
BRDM2	5	1505	11	11	198	48	19	84%
BTR70	0	7	1439	1	105	68	147	81%
SUV	2	0	905	195	285	5	334	11%
T72	246	0	2	1	1313	2	197	75%
Truck	3	0	517	8	252	28	365	2%
ZSU23-4	6	3	167	4	294	69	1152	68%

3500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	126	4	0	1	93	8	17	51%
BRDM2	93	1	4	15	63	25	22	0%
BTR70	31	3	0	5	124	1	0	0%
SUV	1	12	0	3	68	45	0	2%
T72	142	10	1	35	400	86	578	32%
Truck	5	39	8	11	136	0	4	0%
ZSU23-4	208	0	1	4	80	52	8	2%

Table 20. Classification results for 75% missing case. Left is the confusion matrix; the last column is the classification accuracy.

1000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	752	6	0	0	47	0	31	90%
BRDM2	7	1646	20	0	0	2	3	98%
BTR70	22	10	1342	2	33	10	30	93%
SUV	12	2	47	993	24	12	708	55%
T72	5	4	1	0	846	4	59	92%
Truck	116	17	44	12	213	487	839	28%
ZSU23-4	4	0	23	0	12	2	1628	98%

2500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	38	16	1	9	107	49	12	16%
BRDM2	99	516	0	2	731	103	81	34%
BTR70	112	33	1	67	453	191	95	0%
SUV	175	66	6	14	625	374	27	1%
T72	34	45	25	1	119	29	346	20%
Truck	434	178	13	45	476	253	105	17%
ZSU23-4	76	1	0	2	118	99	16	5%

1500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	737	23	83	13	352	437	111	42%
BRDM2	16	1661	20	2	73	8	13	93%
BTR70	0	1	1732	0	28	1	35	96%
SUV	19	3	556	195	413	21	585	11%
T72	21	1	8	1	1610	6	151	90%
Truck	29	6	411	8	721	99	460	6%
ZSU23-4	0	2	113	1	88	0	1529	88%

3000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	237	289	0	0	879	352	10	13%
BRDM2	278	19	20	3	856	403	220	1%
BTR70	9	154	0	5	1280	346	3	0%
SUV	47	425	5	142	947	140	29	8%
T72	16	46	0	103	780	275	579	43%
Truck	148	641	0	38	817	97	18	6%
ZSU23-4	74	16	70	62	708	232	631	35%

2000m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	5	19	1	0	346	134	6	1%
BRDM2	33	788	18	1	697	151	38	46%
BTR70	0	4	472	1	657	316	42	32%
SUV	6	0	786	14	496	6	207	1%
T72	172	0	22	21	998	44	479	57%
Truck	6	1	259	0	176	3	102	1%
ZSU23-4	5	7	211	7	476	236	473	33%

3500m	BMP2	BRDM2	BTR70	SUV	T72	Truck	ZSU23-4	% of accuracy
BMP2	189	3	0	0	677	3	7	22%
BRDM2	77	2	4	4	99	76	125	1%
BTR70	240	139	0	39	411	90	8	0%
SUV	28	10	1	1	486	149	8	0%
T72	31	9	18	25	222	197	571	21%
Truck	15	78	2	7	839	121	3	11%
ZSU23-4	73	0	0	5	125	34	27	10%

5.3. Discussions: Tracking and Classification at 75% rate at MWIR Night-Time

Tracking and Classification

Our main interest is the 75% missing data case. From Table 21, one can observe that tracking is good up to 3000 m and classification is good up to 1500 m. MWIR videos should be applied in night time.

Table 21. MWIR Night Time Summary at 75% missing rate.

Range	Average % of frames with detections	Average Accuracy
1000	79.98	79%
1500	98.44	61%
2000	70.97	24%
2500	50.94	13%
3000	98.80	15%
3500	41.89	9%

Computational Efficiency

Here we briefly emphasize one key advantage of the proposed approach. For compressive measurements using the random subsampling approach, conventional trackers will require the missing pixels to be reconstructed. For randomly missing pixels in the videos, simple interpolation algorithms can be used to reconstruct the videos. However, even the simple interpolation methods still take time. We ran some experiments to compare the execution times in terms of number of frames per second (fps). From Table 22, it can be seen that reconstruction slows down the execution time considerably. Tracking directly can achieve 7 fps for 50% and 75% missing cases, but the reconstruction based algorithm can only achieve 2 and 1 fps for 50% and 75% missing cases, respectively.

Table 22. Comparison of fps with and without missing data reconstruction for one video from 1000 m range.

	Missing rates	fps
No pixel reconstruction	0	5.90
	50	7.61
	75	7.47
With reconstruction	50	1.98
	75	1.29

6. CONCLUSIONS

In this paper, a deep learning approach for multiple target tracking and classification directly in the compressive measurement domain is presented for low quality videos where the targets are small and the image resolution is poor. The compressive measurements are obtained via subsampling of the original image pixels. Experiments clearly demonstrated the performance of the proposed approach under different conditions even when the training data are limited.

YOLO is not suitable for classification of vehicles in low quality videos. However, it is good for tracking by detection. ResNet can be customized for classification and has achieved reasonable performance.

Optical camera is good for daytime up to 2000 m. Of course, if training data are available, tracking can be good up to 3000 m, but classification is not good. MWIR imager is not suitable for daytime tracking and classification. However, it is good for night-time up to 3000 m for tracking, but classification is not good even with training data at 3000m.

One future direction is to integrate the proposed approach with video cameras and perform real-time tracking and classification.

ACKNOWLEDGEMENTS

This research was supported by the US Air Force under contract FA8651-17-C-0017. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] Li, X., Kwan, C., Mei, G. and Li, B., (2006) "A Generic Approach to Object Matching and Tracking," *Proc. Third International Conference Image Analysis and Recognition, Lecture Notes in Computer Science*, pp 839-849,

[2] Zhou, J. and Kwan, C., (2018) "Tracking of Multiple Pixel Targets Using Multiple Cameras," *15th International Symposium on Neural Networks*.

[3] Zhou, J. and Kwan, C., (2018) "Anomaly Detection in Low Quality Traffic Monitoring Videos Using Optical Flow," *Proc. SPIE 10649, Pattern Recognition and Tracking XXIX, 106490F*.

[4] Kwan, C., Zhou, J., Wang, Z. and Li, B., (2018) "Efficient Anomaly Detection Algorithms for Summarizing Low Quality Videos," *Proc. SPIE 10649, Pattern Recognition and Tracking XXIX, 1064906*.

[5] Kwan, C., Yin, J. and Zhou, J., (2018) "The Development of a Video Browsing and Video Summary Review Tool," *Proc. SPIE 10649, Pattern Recognition and Tracking XXIX, 1064907*.
[6] Zhao, Z., Chen, H., Chen, G., Kwan, C. and Li, X. R., (2006) "IMM-LMMSE Filtering Algorithm for Ballistic Target Tracking with Unknown Ballistic Coefficient," *Proc. SPIE, Volume 6236, Signal and Data Processing of Small Targets*.
[7] Zhao, Z., Chen, H., Chen, G., Kwan, C. and Li, X. R., (2006) "Comparison of several ballistic target tracking filters," *Proc. American Control Conference*, pp 2197-2202.
[8] Candes, E. J. and Wakin, M. B., (2008) "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30.
[9] Kwan, C., Chou, B. and Kwan, L. M., (2018) "A Comparative Study of Conventional and Deep Learning Target Tracking Algorithms for Low Quality Videos," *15th International Symposium on Neural Networks*.
[10] Kwan, C., Chou, B., Yang, J. and Tran, T., (2019) "Compressive object tracking and classification using deep learning for infrared videos," *Pattern Recognition and Tracking XXX (Conference SI120)*.

[11] Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J. and Etienne-Cummings, R., (2019) "Target Tracking and Classification Directly Using Compressive Sensing Camera for SWIR videos," *Journal of Signal, Image, and Video Processing*.
[12] Kwan, C., Chou, B., Echavarren, A., Budavari, B., Li, J. and Tran, T., (2018) "Compressive vehicle tracking using deep learning," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference*.
[13] Tropp, J. A., (2004) "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp 2231–2242.
[14] Yang, J. and Zhang, Y., (2011) "Alternating direction algorithms for l1-problems in compressive sensing," *SIAM journal on scientific computing*, 33, pp 250–278.
[15] Dao, M., Kwan, C., Koperski, K. and Marchisio, G., (2017) "A Joint Sparsity Approach to Tunnel Activity Monitoring Using High Resolution Satellite Images," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference*, pp 322-328,
[16] Zhou, J., Ayhan, B., Kwan, C. and Tran, T., (2018) "ATR Performance Improvement Using Images with Corrupted or Missing Pixels," *Proc. SPIE 10649, Pattern Recognition and Tracking XXIX, 106490E*.
[17] Yang, M. H., Zhang, K. and Zhang, L., (2012) "Real-Time Compressive Tracking," *European Conference on Computer Vision*.
[18] Applied Research LLC, Phase 1 Final Report, 2017.
[19] Kwan, C., Gribben, D. and Tran, T. (2019) "Multiple Human Objects Tracking and Classification Directly in Compressive Measurement Domain for Long Range Infrared Videos," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference*, New York City.
[20] Kwan, C., Chou, B., Yang, J., and Tran, T. (2019) "Deep Learning based Target Tracking and Classification for Infrared Videos Using Compressive Measurements," *Journal Signal and Information Processing*.
[21] Kwan, C., Gribben, D. and Tran, T. (2019) "Tracking and Classification of Multiple Human Objects Directly in Compressive Measurement Domain for Low Quality Optical Videos," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference*, New York City.
[22] Redmon, J. and Farhadi, A., (2018) "YOLOv3: An Incremental Improvement," arxiv, April.
[23] Ren S., He, K., Girshick, R. and Sun, J., (2015) "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*.
[24] He, K., Zhang, X., Ren, S. Ren and Sun, J., (2016) "Deep Residual Learning for Image Recognition," *Conference on Computer Vision and Pattern Recognition*.
[25] Kwan, C., Chou, B., Yang, J., and Tran, T., (2019) "Target Tracking and Classification Directly in Compressive Measurement Domain for Low Quality Videos," *Pattern Recognition and Tracking XXX (Conference SI120)*.
[26] Stauffer, C. and Grimson, W. E. L., (1999) "Adaptive Background Mixture Models for Real-Time Tracking," *Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 2246-252.
[27] Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O. and Torr, P., (2016) "Staple: Complementary Learners for Real-Time Tracking," *Conference on Computer Vision and Pattern Recognition*.
[28] Kulkarni, K. and Turaga, P. K. (2016) "Reconstruction-Free Action Inference from Compressive Imagers," *IEEE Trans. Pattern Anal. Mach. Intell*. 38(4), pp 772-784.

[29] Lohit, S., Kulkarni, K. and Turaga, P. K. (2016) "Direct inference on compressive measurements using convolutional neural networks," *Int. Conference on Image Processing*, pp 1913-1917.

[30] Adler, A., Elad, M. and Zibulevsky, M. (2016) "Compressed Learning: A Deep Neural Network Approach," arXiv:1610.09615v1 [cs.CV].

[31] Xu, Y. and Kelly, K. F. (2019) "Compressed domain image classification using a multi-rate neural network," arXiv:1901.09983 [cs.CV].

[32] Kulkarni, K. and Turaga, P. K. (2016) "Fast Integral Image Estimation at 1% measurement rate," arXiv:1601.07258v1 [cs.CV].

[33] Wang, Z. W., Vineet, V., Pittaluga, F., Sinha, S. N., Cossairt, O. and Kang, S. B. (2019) "Privacy-Preserving Action Recognition Using Coded Aperture Videos," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[34] Vargas, H., Fonseca, Y. and Arguello, H. (2018) "Object Detection on Compressive Measurements using Correlation Filters and Sparse Representation," *26th European Signal Processing Conference (EUSIPCO)*, pp 1960-1964.

[35] Değerli, A., Aslan, S., Yamac, M., Sankur, B. and Gabbouj, M. (2018) "Compressively Sensed Image Recognition," *7th European Workshop on Visual Information Processing (EUVIP)*, Tampere, pp. 1-6.

[36] Latorre-Carmona, P., Traver, V. J., Sánchez, J. S. and Tajahuerce, E. (2019) "Online reconstruction-free single-pixel image classification," *Image and Vision Computing*, Vol. 86.

[37] Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J. and Etienne-Cummings, R., (2019) "Target Tracking and Classification Using Compressive Measurements of MWIR and LWIR Coded Aperture Cameras," *Journal Signal and Information Processing*, vol. 10, no. 3.

[38] Kwan, C., Chou, B., Yang, J., Rangamani, A., Tran, T., Zhang, J. and Etienne-Cummings, R., (2019) "Deep Learning based Target Tracking and Classification for Low Quality Videos Using Coded Aperture Camera," *Sensors*, vol. 19, no. 17, 3702.

AUTHORS

Chiman Kwan received his BS with honours in Electronics from the Chinese University of Hong Kong in 1988, and MS and Ph.D. in Electrical Engineering from the University of Texas at Arlington in 1989 and 1993, respectively. Currently, he is the President of Applied Research LLC, leading research and development efforts in image and video processing, and remote sensing. He has 15 patents, 60 invention disclosures, 350 papers in journals and conferences, and over 500 proprietary technical reports. Over the past two decades, he has secured more than 120 diverse research projects totalling more than 36 million dollars from various government agencies and private companies. He also received numerous awards and recognitions from IEEE, NASA, and other agencies.

Bryan Chou graduated in Electrical Engineering from the George Mason University. In the past three years, he has worked on image demosaicing, image super-resolution, pansharpening, image fusion, and multi-target tracking.

Jonathan Yang graduated from the Computer Science Department of the University of Maryland. He was with Applied Research LLC and is now with Google Inc.

Trac D. Tran received his BS from MIT and Ph.D. from U. Wisconsin Madison. He is a Professor in the Department of Electrical and Computer Engineering, The Johns Hopkins University. His pioneering research on integer-coefficient transforms and pre-/post-filtering operators has been adopted as critical components of Microsoft Windows Media Video 9 and JPEG XR. He was the recipient of the NSF Career Award and is a Fellow of IEEE.