PRACTICAL APPROACHES TO TARGET DETECTION IN LONG RANGE AND LOW QUALITY INFRARED VIDEOS

Chiman Kwan and David Gribben

Applied Research, LLC, Rockville, Maryland, USA

ABSTRACT

It is challenging to detect vehicles in long range and low quality infrared videos using deep learning techniques such as You Only Look Once (YOLO) mainly due to small target size. This is because small targets do not have detailed texture information. This paper focuses on practical approaches for target detection in infrared videos using deep learning techniques. We first investigated a newer version of You Only Look Once (YOLO v4). We then proposed a practical and effective approach by training the YOLO model using videos from longer ranges. Experimental results using real infrared videos ranging from 1000 m to 3500 m demonstrated huge performance improvements. In particular, the average detection percentage over the six ranges of 1000 m to 3500 m improved from 54% when we used the 1500 m videos for training to 95% if we used the 3000 m videos for training.

KEYWORDS

Deep learning; YOLO v4; infrared videos; target detection; training strategy

1. INTRODUCTION

For infrared videos, people have normally applied two groups of target detection algorithms. The first group of methods [1]-[6] requires target locations in the first frame of the videos to be known and then those target features in the first frame are used to detect the targets in subsequent frames. The second group uses deep learning algorithms such as You Only Look Once (YOLO) for target detection in optical and infrared videos [7]-[21]. Target locations in the first frame are no longer required. However, some training videos are needed in these algorithms. Some deep learning algorithms [6]-[16] use compressive measurements directly without time consuming reconstruction of compressive measurements for target detection and classification. As a result, fast target detection and classification can be achieved.

In long range infrared videos, the target size is small, the resolution is low, and the video quality such as contrast is also poor. It is therefore extremely important to develop practical methods that can improve the detection performance using deep learning methods. In [22], we proposed the incorporation of video super-resolution (VSR) techniques to enhance target detection performance in infrared videos. The resolution of the video frame is improved by two to four times. The target detection and classification performance using actual infrared videos was observed to be improved. In another paper [23], low contrast videos were enhanced using 16-bit videos and we also observed improved performance in target detection and classification. In [24], target motion information has been utilized using optical flow techniques. We observed that target detection performance has been improved significantly. In some videos, if the targets are moving towards or away from the imager, target motion may be difficult to extract. In such

DOI: 10.5121/sipij.2021.12301

scenarios, target detection using single frames can be applied. In [25][26], target detection techniques based on single frames were proposed and evaluated. There are also many recent papers discussing various methods for target detection (no classification) in infrared images/videos [27]-[39].

In our early target detection and classification papers [7]-[10], YOLO v3 [40] was used. Recently, there are new YOLO versions in the public domain. In this paper, we present practical approaches to enhancing the detection performance in long range infrared videos. First, we investigated target detection performance enhancement using YOLO v4 [42]. YOLO v4 has more layers than YOLO v3. However, after running some experiments, the improvement in target detection is observed to be slight, but noticeable. For instance, for 2000 m videos, the average precision improved from 52% to 56%. Second, we investigated target detection performance enhancement using a new training strategy. In our previous papers, we used videos at 1500 m to train the YOLO. Although target detection performance is good for videos at 1000 m and 2000 m ranges, the performance drops significantly for 2500 m to 3500 m videos. After some extensive experiments, we observed that using videos at 2500 m or 3000 m to train the YOLO model actually performed the best across all ranges. For example, the average detection percentage is 54% over the six ranges of 1000 m to 3500 m when we used the 1500 m video for training. However, if we used the 3000 m videos for training, the average detection percentage over six ranges became 95%. This is a dramatic improvement.

Our key contribution is the following. We propose a practical training strategy to improve target detection in long range infrared videos. Instead of using near range videos to train the YOLO model, the best strategy is to use far range videos for training. Although the idea is simple, we observed dramatic performance improvement.

The remainder of this paper is organized as follows. Section 2 briefly summarizes the target detection algorithm, performance metrics for target detection, and data. Section 3 shows that a new YOLO version 4 improves over an older version 3 using actual long range infrared videos. Section 4 summarizes a new training strategy that can improve detection performance in all ranges. We also compared with a two-model approach. Section 5 concludes the paper with a few remarks.

2. ALGORITHMS, METRICS, AND DATA

2.1. YOLOv3 versus YOLOv4

Here, we would like to briefly compare the YOLO v3 and v4 models. In general, they are largely the same. They share the same general framework around it to encourage an accurate model. The Input and Backbone as shown in Fig. 1, are exactly the same except for the Darknet model used. In v3, the Darknet model used is Darknet53, a 53 layer model depicted in Fig. 2, while the Darknet model used for v4 is a 137 layer model.

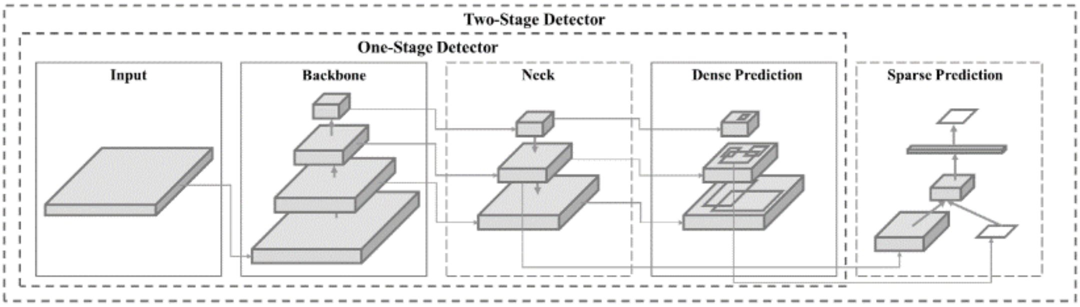
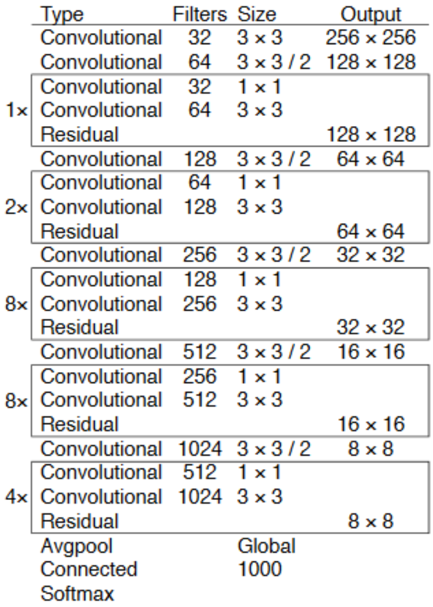


Figure 1. YOLO V4 Block diagram. Input is the images; Backbone is the base Darknet model; Dense Prediction is the YOLO model; Sparse Prediction is Faster R-CNN [41].



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 2. Diagram showing construction of the layers in the Darknet53 model.

According to paper [42] put out by the authors of YOLO v4, it is faster and more accurate than YOLO v3. The improvement denoted in Fig. 3 is about a 10 percent improvement in Average Precision at the same frame per second (FPS).

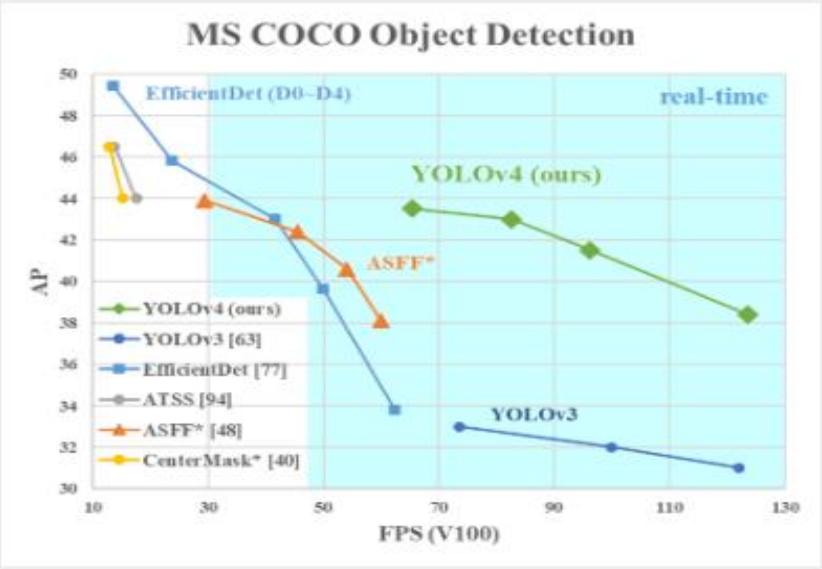


Figure 3. Comparison of YOLO v4, YOLO v3, and other strong object detectors [42].

2.2. Performance Metrics for Assessing Target Detection Performance

The five different performance metrics to assess the detection performance are:Center Location Error (CLE), Distance Precision at 10 pixels (DP@10), Estimates in Ground Truth (EinGT), Intersection over Union (IoU), and number of frames with detection. These metrics are summarized as follows:

- Center Location Error (CLE): This is the error between the center of the bounding box and the ground-truth bounding box. Smaller means better. CLE is calculated by measuring the distance between the ground truth center location $(C_{x,gt}, C_{y,gt})$ and the detected center location $(C_{x,est}, C_{y,est})$. Mathematically, CLE is given by

$$CLE = \sqrt{\left(C_{x,est} - C_{x,gt}\right)^2 + \left(C_{y,est} - C_{y,gt}\right)^2}. \tag{1}$$

- Distance Precision (DP): This is the percentage of frames where the centroids of detected bounding boxes are within 10 pixels of the centroid of ground-truth bounding boxes. Close to 1 or 100% indicates good results.

- Estimates in Ground Truth (EinGT): This is the percentage of the frames where the centroids of the detected bounding boxes are inside the ground-truth bounding boxes. It depends on the size of the bounding box and is simply a less strict version of the DP metric. Close to 1 or 100% indicates good results.

- Intersection over the Union (IoU): It is the ratio of the intersected area over the union of the estimated and ground truth bounding boxes.

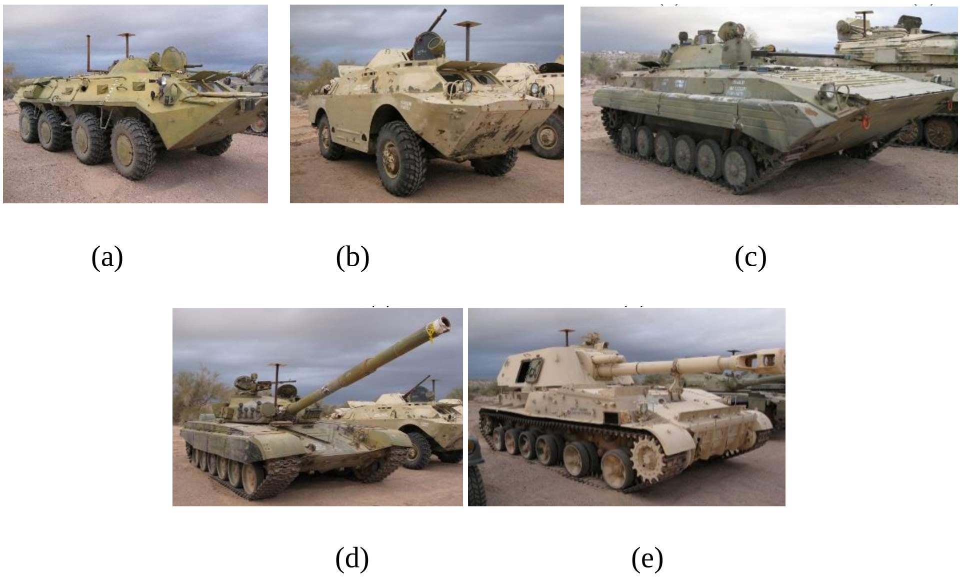
$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{2}$$

Number of frames with detection: This is the total number of frames that have detection.

2.3. Defense Systems Information Analysis Center (DSIAC) Data

DSIAC is more challenging than MOT [43] because the targets in MOT are large and the image resolution is also better in MOT. We selected five vehicles in the DSIAC videos for experiments. There are optical and mid-wave infrared (MWIR) videos collected at distances ranging from 1000 m to 5000 m with 500 m increments. The five types of vehicles are shown in Figure 4. These videos are challenging for several reasons. First, the target sizes are small due to long ranges. This is very different from some benchmark datasets such as MOT Challenge [43] where the range is short and the targets are big. Second, the target orientations also change drastically. Third, the illuminations in different videos are also different. Fourth, the cameras also move in some videos.

In this research, MWIR night-time videos were used because MWIR is more effective for surveillance during the nights than optical videos. The video frame rate is 7 frames/second and the image size is 640x512. The total number of frames is 1800 per video. All frames are contrast enhanced using some reference frames in the 1500 m range videos. We also used 16-bit videos in all of our experiments in this paper due to improved performance of using 16-bit videos in our earlier paper [23].



(a) (b) (c)

(d) (e)

Figure 4. Five vehicles in DSIAC: (a) BTR70; (b) BRDM2; (c) BMP2; (d) T72; and (e) ZSU23-4.

2.3.1. Further Subsections

Further sub-sectioning, if required, is indicated using 1.1.1. Qqq, etc. headings with 11 pt. bold Times New Roman font with a 6pt line spacing following.

3. Enhancing Target Detection Performance using a New Version of Yolo (Yolov4)

In our past papers [7]-[10], we used YOLO v3 in our experiments. In the course of our research, we found that there is a new version of YOLO. The objective of this section is to summarize our study on how YOLO v4 is better than YOLO v3.

Statistically, the comparison will be done by using several performance metrics. Center Location Error (CLE) is the average pixel distance of the guessed center of the vehicle bounding box from the ground truth center location. Distance Precision (DP) is the percentage of detections that fall within a certain number or pixels of the ground truth center location, for this project that distance is 20 pixels. Estimated in Ground Truth (EinGT) is the percentage of centroid values that fall within the ground truth bounding box. Intersection over Union (IoU) is the average area of intersection between the detected and ground truth bounding box divided by the area of union of that bounding box, that value is then reduced to a percentage. The final metric, Percent Detection (%det), is the percentage of frames that have any kind of detection within the frame. Table 1 and Table 2 contain the metrics used to judge how well the model is trained. The red numbers indicate the detection metrics using 1500 m videos, which were also used for training. Table 1 has the average results for each distance for YOLO v3 while Table 2 has the average results for each distance for YOLO v4 trained at the 1500 meter distance. Detection metrics for individual vehicles at various ranges can be found in Appendix 1.

We used 1500 m videos for training the YOLOs. There are five videos in this range and 1800 frames per video. Comparing the two versions of YOLO shows in general that there is not a huge difference between the two. For YOLO v4, the first three distances are largely improved in reference to YOLO v3. Therefore, distances adjacent to the trained distance have a higher degree of accuracy. However, the further distances, 2500 through 3500 m ranges, have a much steeper degradation of detection. The accuracy metrics still improve but only slightly and it is not universally better performing.

Table 1. Average performance metrics for each distance using a 1500 m trained YOLO v3 model. Detailed metrics for each individual vehicle at various ranges can be found in Table 6 of Appendix 1.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	3.645	100.00%	100.00%	71.79%	94.15%
1500	1.368	100.00%	100.00%	83.99%	100.00%
2000	2.009	99.99%	99.99%	52.24%	90.64%
2500	6.810	98.37%	98.20%	20.14%	44.03%
3000	3.924	100.00%	78.33%	11.54%	11.48%
3500	3.481	100.00%	49.83%	2.94%	6.19%
Avg.	3.540	99.73%	87.73%	40.44%	57.75%

Table 2. Average performance metrics for each distance using a 1500 m trained YOLO v4 model. Detailed metrics for each individual vehicle at various ranges can be found in Table 7 of Appendix 1.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	2.968	100.0%	100.0%	72.59%	100.0%
1500	1.100	100.0%	99.99%	89.40%	100.0%
2000	1.557	100.0%	100.0%	56.47%	100.0%
2500	8.612	97.43%	97.43%	27.21%	54.33%
3000	2.148	100.0%	100.0%	15.78%	6.32%
3500	1.022	40.00%	24.42%	4.43%	0.76%
Avg.	2.901	89.57%	86.97%	44.31%	50.20%

4. Vehicle Detection Performance Improvement using a New Training Strategy

In the previous section, we observed that the target detection performance has not improved much in the long ranges, except in the ranges of 1000 and 2000 m even we used a new version of YOLO. After analyzing the earlier results, we speculate that the YOLO model trained using a certain range will be effective only for videos collected close to the range of the training data. This observation led us to a new training strategy for long range videos.

Here, we summarize a new training strategy for training YOLO for target detection. The goal is to minimize false positives and missed detections.

4.1. A New Strategy for Target Detection

The reason for this investigation stems from a pattern we see in Table 1 and Table 2. That pattern is that the 1000 m distance is better performing than the other adjacent distance of 2000 m. The hypothesis is that the YOLO model is better at accurately detecting the vehicles when the trained distance is greater than the tested distance. We can see if this is true in the next few tables.

The results for the 2000 m trained model in Table 3 show that clearly the 2000 m distance greatly improved. This is expected because that is the same distance the model was trained on. However, there is little degradation in performance to the 1000 m distance as well as a vast improvement to the 3000 m results. The improvement for 3500 m is much more tepid but it is clear that there is improved performance across all distances when the YOLO model is trained on the further distance videos.

Table 3. Average performance metrics for 2000 m YOLO v4 trained model. Detailed metrics for each individual vehicle at various ranges can be found in Table 8 of Appendix 2. Red numbers indicate testing results using the same training data.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	3.501	99.9%	100.0%	64.30%	99.6%
1500	2.077	99.88%	99.89%	63.25%	99.6%
2000	1.023	100.0%	100.0%	89.19%	100.0%
2500	13.955	94.75%	94.75%	49.14%	80.44%
3000	1.951	99.5%	99.3%	31.20%	45.40%
3500	1.742	99.67%	94.66%	9.07%	10.90%
Avg.	4.042	98.95%	98.10%	51.03%	72.66%

One anomaly that should be noted is the poor value for CLE at the 2500 m distance. At first, it looks like an error or poor testing results. Further inspection of the BMP2 video revealed that for some reason there is a second vehicle driving in circles in the background of that video. Fig. 5 contains a frame of that video, showing both vehicles—the correct BMP2 vehicle in the center of the frame and another vehicle. YOLO detects the background vehicle and, because all detections are measured and considered when generating metrics, it greatly affects the CLE results. This will become very apparent in Table 4 when the training distance is 2500 m.

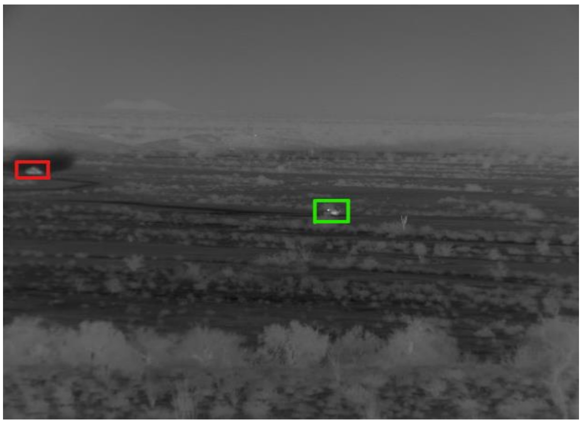


Figure 5. One frame of 2500 m video of BMP2 vehicle. Red box around background vehicle and green box around BMP2 vehicle.

Table 4. Average performance metrics for 2500 m YOLO v4 trained model. Detailed metrics for each individual vehicle at various ranges can be found in Table 9 of Appendix 2. Red numbers indicate testing results using the same training data.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	3.148	99.9%	100.0%	49.89%	97.3%
1500	2.024	100.0%	100.0%	49.26%	100.0%
2000	1.665	100.0%	100.0%	71.93%	99.5%
2500	1.071	100.0%	100.0%	71.77%	100.0%
3000	1.603	100.0%	100.0%	42.19%	91.42%
3500	1.881	99.79%	94.45%	9.28%	47.29%
Avg.	1.899	99.96%	99.08%	49.05%	89.25%

We continue to see little losses for the closer distances as we move further away in training distance. There is also a huge improvement in the CLE results in Table 4 for the 2500 m distance. This was expected because, now that the model is trained on that distance, only the foreground and relevant vehicle is detected as a possible match. In relation to average statistics, this trained distance is the best performing distance. While 3500 m video still does not have great detection results and IoU is still quite poor, each other value is close to the best results we see in other training distance.

Table 5. Average performance metrics for 3000 m YOLO v4 trained model. Detailed metrics for each individual vehicle at various ranges can be found in Table 10 of Appendix 2. Red numbers indicate testing results using the same training data.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	4.071	99.2%	100.0%	27.29%	83.8%
1500	1.954	100.0%	100.0%	30.44%	97.2%
2000	1.522	100.0%	100.0%	49.85%	95.4%
2500	24.089	90.5%	90.5%	65.25%	96.7%
3000	0.950	100.0%	100.0%	80.41%	100.0%
3500	1.330	99.95%	96.20%	23.38%	95.11%
Avg.	5.653	98.28%	97.78%	46.10%	94.7%

There is, however, one more training distance to observe, 3000 meters. In Table 5, we see the greatest number of detections for any trained model. There is a large decrease in the CLE metric for 2500 m, 24.089 is by far the worst result for any YOLO model. If the explanation for why that number is so poor is accepted, then the results for every other value in comparison to each other trained model is the most consistent. It is not perfect, as seen by the 23% values for IoU, but it is certainly better performing than most if not all others.

When looking at all distance models trained using YOLO v4, we certainly see a pattern of improvement as we move the training dataset further and further away. An argument can be made for which distance is best whether it be the 2500 m or 3000 m model. The biggest negative of the 2500 m model is that the 3500 m videos have an average detection of just 47.29% and the IoU is rather low. The biggest negative of the 3000 m model is that the CLE for 2500 m is extremely poor because of the vehicle interfering with the detections for the BMP2 vehicle.

Our study shows that using longer range videos (2500 or 3000 m) to train YOLO can achieve good detection results in all ranges. A simple explanation for this is that YOLO has built-in data augmentation capability, which can generate targets with different sizes. In our experiments, it is clear that YOLO is more capable of generating training images that are larger than the original images than the opposite case.

4.2. Comparison with a Two-Model Approach

In Section 3 and Section 4.1, we used a single model trained on videos from a single range. One may wonder what the performance will be if one utilizes a two-model approach, which essentially uses two separate models trained using videos from two separate ranges.

Here, we first show the detection performance of using two models. One model is obtained by using videos from the 1500 m range and another model is using videos from the 3000 m range. The 1500 m model is only used for 1000 to 2000 m ranges and the 3000 m model is only for 2500 m to 3500 m ranges. The combined model results are shown in Table 6. Comparing with the detection results in Table 5, the two-model performance is indeed better. The cost is that two models are needed that takes longer time to train.

Table 6. Average performance metrics using two trained models at 1500 m and 3000 m YOLO v4 trained model. Red numbers indicate testing results using the same training data.

Distance (m)	CLE	DP	EinGT	IoU	% det.
1000	2.968	100.0%	100.0%	72.59%	100.0%
1500	1.100	100.0%	99.99%	89.40%	100.0%
2000	1.557	100.0%	100.0%	56.47%	100.0%
2500	24.089	90.5%	90.5%	65.25%	96.7%
3000	0.950	100.0%	100.0%	80.41%	100.0%
3500	1.330	99.95%	96.20%	23.38%	95.11%
Avg.	5.333	98.41%	97.78%	64.5%	98.64%

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we have presented some practical and high performance methods to enhance target detection performance in long range infrared videos. We observed that using a new version of YOLO improves slightly the performance. Moreover, we demonstrated that using a new training strategy, which simply uses longer range videos for training, can significantly improve the detection performance in all ranges. Finally, we observed that if we used a two-model approach, the performance will be even better at the cost of requiring more training data and training time.

One future research direction is to investigate target detection based on changes between frames. Such a detection strategy is useful when there is motion in the targets. Another future direction is to integrate super-resolution videos with the idea proposed in this paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

This research was supported by the US Army under contract W909MY-20-P-0024. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

APPENDIX 1: SUPPORTING MATERIALS FOR SECTION 3.1

Table 7. Accuracy statistics for 1500 m trained YOLO v3.

1000 m	CLE	DP	EinGT	IoU	% det.
BTR70	3.677	100.00%	100.00%	67.89%	99.89%
BRDM2	3.829	100.00%	100.00%	72.02%	92.22%
BMP2	3.762	100.00%	100.00%	70.52%	99.22%
T72	3.638	100.00%	100.00%	72.90%	85.17%
ZSU23-4	3.316	100.00%	100.00%	75.61%	94.28%
Avg	3.645	100.00%	100.00%	71.79%	94.15%
1500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.401	100.00%	100.00%	83.00%	100.00%
BRDM2	1.266	100.00%	100.00%	83.16%	100.00%
BMP2	1.293	100.00%	100.00%	86.42%	100.00%
T72	1.491	100.00%	100.00%	85.90%	100.00%
ZSU23-4	1.387	100.00%	100.00%	81.45%	100.00%
Avg	1.368	100.00%	100.00%	83.99%	100.00%
2000 m	CLE	DP	EinGT	IoU	% det.
BTR70	2.039	100.00%	100.00%	46.21%	91.72%
BRDM2	2.328	100.00%	100.00%	52.79%	99.39%
BMP2	2.005	100.00%	100.00%	59.22%	68.61%
T72	1.467	100.00%	100.00%	51.99%	94.44%
ZSU23-4	2.208	99.95%	99.95%	50.97%	99.06%
Avg	2.009	99.99%	99.99%	52.24%	90.64%
2500 m	CLE	DP	EinGT	IoU	% det.
BTR70	2.804	99.89%	99.89%	16.85%	36.61%
BRDM2	3.193	100.00%	99.12%	18.85%	71.44%
BMP2	22.030	91.97%	91.97%	21.60%	28.78%
T72	2.978	100.00%	100.00%	24.18%	45.44%
ZSU23-4	3.046	100.00%	100.00%	19.23%	37.89%
Avg	6.810	98.37%	98.20%	20.14%	44.03%
3000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.843	100.00%	100.00%	8.44%	10.33%

BRDM2	4.367	100.00%	98.52%	11.16%	13.94%
BMP2	5.242	100.00%	0.00%	11.80%	0.11%
T72	5.033	100.00%	93.14%	14.12%	18.00%
ZSU23-4	3.137	100.00%	100.00%	12.15%	15.00%
Avg	3.924	100.00%	78.33%	11.54%	11.48%
3500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.860	100.00%	71.05%	2.50%	1.83%
BRDM2	3.795	100.00%	45.24%	2.79%	2.28%
BMP2	n/a	n/a	n/a	n/a	0.00%
T72	4.692	100.00%	25.43%	3.51%	16.06%
ZSU23-4	3.578	100.00%	57.61%	2.98%	10.78%
Avg	3.481	100.00%	49.83%	2.94%	6.19%

Table 8. Accuracy statistics for 1500 m trained YOLO v4.

1000 m	CLE	DP	EinGT	IoU	% det.
BTR70	2.970	100.0%	100.0%	79.40%	100.0%
BRDM2	2.423	100.0%	100.0%	65.20%	100.0%
BMP2	3.633	100.0%	100.0%	72.20%	100.0%
T72	3.149	100.0%	100.0%	61.07%	100.0%
ZSU23-4	2.667	100.0%	100.0%	85.07%	100.0%
Avg	2.968	100.0%	100.0%	72.59%	100.0%
1500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.233	100.0%	99.94%	89.14%	100.0%
BRDM2	0.999	100.0%	100.0%	90.66%	100.0%
BMP2	1.014	100.0%	100.0%	90.61%	100.0%
T72	1.182	100.0%	100.0%	89.85%	100.0%
ZSU23-4	1.074	100.0%	100.0%	86.75%	100.0%
Avg	1.100	100.0%	99.99%	89.40%	100.0%
2000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.166	100.0%	100.0%	54.98%	100.0%
BRDM2	2.154	100.0%	100.0%	54.47%	100.0%
BMP2	1.286	100.0%	100.0%	69.69%	100.0%
T72	1.366	100.0%	100.0%	51.63%	100.0%
ZSU23-4	1.811	100.0%	100.0%	51.60%	100.0%
Avg	1.557	100.0%	100.0%	56.47%	100.0%
2500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.331	100.0%	100.0%	25.93%	64.44%
BRDM2	1.804	100.0%	100.0%	25.74%	86.56%
BMP2	36.233	87.13%	87.13%	29.64%	35.89%
T72	2.243	100.0%	100.0%	28.97%	71.17%
ZSU23-4	1.451	100.0%	100.0%	25.75%	13.61%
Avg	8.612	97.43%	97.43%	27.21%	54.33%
3000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.211	100.0%	100.0%	22.03%	8.61%
BRDM2	2.268	100.0%	100.0%	10.76%	10.78%
BMP2	n/a	n/a	n/a	n/a	0.00%
T72	1.997	100.0%	100.0%	15.43%	5.78%
ZSU23-4	3.118	100.0%	100.0%	14.89%	0.11%
Avg	2.148	100.0%	100.0%	15.78%	6.32%
3500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.207	100.0%	100.0%	6.03%	0.11%
BRDM2	n/a	n/a	n/a	n/a	0.00%
BMP2	n/a	n/a	n/a	n/a	0.00%

T72	3.904	100.0%	22.08%	2.82%	3.67%
ZSU23-4	n/a	n/a	n/a	n/a	0.00%
Avg	1.022	40.00%	24.42%	4.43%	0.76%

APPENDIX 2: SUPPORTING MATERIALS FOR SECTION 4

Table 9. YOLO v4 accuracy statistics for 2000 m trained model.

1000 m	CLE	DP	EinGT	IoU	% det.
BTR70	3.862	100.0%	100.0%	71.19%	99.7%
BRDM2	3.939	100.0%	100.0%	50.13%	100.0%
BMP2	3.368	100.0%	100.0%	67.32%	98.7%
T72	3.294	99.70%	100.0%	62.72%	99.8%
ZSU23-4	3.039	100.0%	100.0%	70.15%	99.9%
Avg	3.501	99.9%	100.0%	64.30%	99.6%
1500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.503	100.0%	100.0%	69.37%	100.0%
BRDM2	2.362	100.0%	100.0%	55.15%	100.0%
BMP2	1.401	100.0%	100.0%	54.41%	98.1%
T72	2.527	100.0%	100.0%	64.85%	100.0%
ZSU23-4	2.590	99.41%	99.46%	72.49%	100.0%
Avg	2.077	99.88%	99.89%	63.25%	99.6%
2000 m	CLE	DP	EinGT	IoU	% det.
BTR70	0.910	100.0%	100.0%	90.44%	100.0%
BRDM2	0.965	100.0%	100.0%	87.92%	100.0%
BMP2	0.984	100.0%	100.0%	90.32%	100.0%
T72	1.082	100.0%	100.0%	87.45%	100.0%
ZSU23-4	1.176	100.0%	100.0%	89.84%	100.0%
Avg	1.023	100.0%	100.0%	89.19%	100.0%
2500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.414	100.0%	100.0%	42.32%	97.94%
BRDM2	1.413	100.0%	100.0%	44.56%	95.33%
BMP2	62.823	73.93%	73.93%	52.13%	41.00%
T72	2.440	99.8%	99.8%	60.33%	96.94%
ZSU23-4	1.683	100.0%	100.0%	46.34%	71.00%
Avg	13.955	94.75%	94.75%	49.14%	80.44%
3000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.312	100.0%	100.0%	30.63%	65.44%
BRDM2	1.820	100.0%	100.0%	26.79%	76.56%
BMP2	1.876	100.0%	100.0%	37.47%	0.33%
T72	1.546	100.0%	100.0%	33.29%	55.11%
ZSU23-4	3.201	97.6%	96.4%	27.79%	29.56%
Avg	1.951	99.5%	99.3%	31.20%	45.40%
3500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.266	100.0%	99.5%	10.98%	10.89%
BRDM2	2.011	98.68%	98.41%	8.47%	20.50%
BMP2	n/a	n/a	n/a	n/a	0.00%
T72	2.313	100.0%	81.33%	6.65%	3.50%
ZSU23-4	1.378	100.0%	99.39%	10.19%	8.72%
Avg	1.742	99.67%	94.66%	9.07%	10.90%

Table 10. YOLO v4 accuracy statistics for 2500 m trained model.

1000 m	CLE	DP	EinGT	IoU	% det.
BTR70	2.809	100.0%	100.0%	63.52%	96.7%
BRDM2	3.221	100.0%	100.0%	44.58%	98.1%
BMP2	3.526	99.6%	100.0%	40.01%	99.6%
T72	3.462	100.0%	100.0%	45.37%	98.0%
ZSU23-4	2.720	100.0%	100.0%	55.95%	94.2%
Avg	3.148	99.9%	100.0%	49.89%	97.3%
1500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.607	100.0%	100.0%	60.92%	100.0%
BRDM2	2.011	100.0%	100.0%	44.12%	100.0%
BMP2	1.710	100.0%	100.0%	40.03%	100.0%
T72	1.937	100.0%	100.0%	43.46%	100.0%
ZSU23-4	2.857	100.0%	100.0%	57.77%	99.9%
Avg	2.024	100.0%	100.0%	49.26%	100.0%
2000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.242	100.0%	100.0%	73.35%	99.8%
BRDM2	1.721	100.0%	100.0%	74.87%	99.3%
BMP2	1.582	100.0%	100.0%	64.41%	99.3%
T72	2.062	100.0%	100.0%	67.95%	100.0%
ZSU23-4	1.717	100.0%	100.0%	79.04%	99.3%
Avg	1.665	100.0%	100.0%	71.93%	99.5%
2500 m	CLE	DP	EinGT	IoU	% det.
BTR70	0.920	100.0%	100.0%	66.68%	100.0%
BRDM2	1.138	100.0%	100.0%	79.39%	100.0%
BMP2	0.795	100.0%	100.0%	77.29%	100.0%
T72	1.356	100.0%	100.0%	67.68%	100.0%
ZSU23-4	1.143	100.0%	100.0%	67.80%	100.0%
Avg	1.071	100.0%	100.0%	71.77%	100.0%
3000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.399	100.0%	100.0%	32.26%	98.17%
BRDM2	1.556	100.0%	100.0%	43.22%	81.89%
BMP2	1.363	100.0%	100.0%	52.86%	98.50%
T72	1.948	100.0%	100.0%	42.75%	84.56%
ZSU23-4	1.750	100.0%	99.9%	39.85%	94.00%
Avg	1.603	100.0%	100.0%	42.19%	91.42%
3500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.817	100.0%	84.0%	6.54%	66.44%
BRDM2	1.726	98.97%	98.97%	8.38%	26.67%
BMP2	1.897	100.0%	95.36%	13.41%	25.17%
T72	2.650	100.0%	94.70%	9.14%	48.56%
ZSU23-4	1.316	100.0%	99.20%	8.93%	69.61%
Avg	1.881	99.79%	94.45%	9.28%	47.29%

Table 11. YOLO v4 accuracy statistics for 3000 m trained model.

1000 m	CLE	DP	EinGT	IoU	% det.
BTR70	3.412	99.9%	100.0%	34.79%	83.3%
BRDM2	6.965	97.8%	100.0%	21.12%	93.9%
BMP2	2.874	100.0%	100.0%	23.44%	64.2%
T72	3.996	98.3%	100.0%	26.86%	86.8%
ZSU23-4	3.107	100.0%	100.0%	30.22%	91.0%
Avg	4.071	99.2%	100.0%	27.29%	83.8%

1500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.936	100.0%	100.0%	39.93%	93.6%
BRDM2	1.430	100.0%	100.0%	27.19%	99.6%
BMP2	1.510	100.0%	100.0%	24.40%	94.4%
T72	2.168	100.0%	100.0%	27.67%	100.0%
ZSU23-4	2.724	100.0%	100.0%	33.01%	98.4%
Avg	1.954	100.0%	100.0%	30.44%	97.2%
2000 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.221	100.0%	100.0%	54.21%	98.3%
BRDM2	1.748	100.0%	100.0%	44.17%	100.0%
BMP2	1.476	100.0%	100.0%	43.86%	91.1%
T72	1.482	100.0%	100.0%	45.70%	100.0%
ZSU23-4	1.683	100.0%	100.0%	61.34%	87.8%
Avg	1.522	100.0%	100.0%	49.85%	95.4%
2500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.155	100.0%	100.0%	63.71%	99.8%
BRDM2	1.920	100.0%	100.0%	73.95%	95.4%
BMP2	113.347	52.6%	52.6%	41.28%	100.0%
T72	2.305	100.0%	100.0%	75.16%	99.9%
ZSU23-4	1.720	100.0%	100.0%	72.16%	88.4%
Avg	24.089	90.5%	90.5%	65.25%	96.7%
3000 m	CLE	DP	EinGT	IoU	% det.
BTR70	0.922	100.0%	100.0%	83.74%	100.00%
BRDM2	1.027	100.0%	100.0%	72.76%	100.00%
BMP2	0.885	100.0%	100.0%	82.95%	100.00%
T72	1.100	100.0%	100.0%	79.87%	100.00%
ZSU23-4	0.817	100.0%	100.0%	82.75%	100.00%
Avg	0.950	100.0%	100.0%	80.41%	100.00%
3500 m	CLE	DP	EinGT	IoU	% det.
BTR70	1.047	100.0%	99.1%	21.36%	99.61%
BRDM2	1.535	99.73%	99.73%	23.91%	97.39%
BMP2	1.036	100.0%	100.00%	28.05%	96.17%
T72	1.568	100.0%	98.38%	20.80%	87.83%
ZSU23-4	1.463	100.0%	83.78%	22.77%	94.56%
Avg	1.330	99.95%	96.20%	23.38%	95.11%

REFERENCES

[1] C. Kwan, B. Chou, and L. M. Kwan, “A Comparative Study of Conventional and Deep Learning Target Tracking Algorithms for Low Quality Videos,” 15th International Symposium on Neural Networks, 2018; DOI: 10.1007/978-3-319-92537-0_60.

[2] L. Bertinetto, et al., “Staple: Complementary Learners for Real-Time Tracking,” In CVPR. 2016.

[3] C. Ma, X. Yang, C. Zhang, and M. H. Yang, “Long-term correlation tracking,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp. 5388-5396, 2015.

[4] Z. Kalal, K. Mikolajczyk, and J. Matas, “Tracking-Learning-Detection,” TPAMI, 34(7), 2012.

[5] C. Stauffer and W. E. L. Grimson, “Adaptive Background Mixture Models for Real-Time Tracking, Computer Vision and Pattern Recognition,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 2246-252, 1999.

[6] H. S. Demir and A. E. Cetin, “Co-difference based object tracking algorithm for infrared videos,” IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 434-438

[7] C. Kwan, D. Gribben, and T. Tran, “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. 2019.

[8] C. Kwan, B. Chou, J. Yang, and T. Tran, “Deep Learning based Target Tracking and Classification Directly in Compressive Measurement for Low Quality Videos,” Signal & Image Processing: An International Journal (SIPIJ), November 16, 2019.

[9] C. Kwan, D. Gribben, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, “Detection and Confirmation of Multiple Human Targets Using Pixel-Wise Code Aperture Measurements,” J. Imaging. 6(6), 40, 2020.

[10] C. Kwan, B. Chou, J. Yang, and T. Tran, “Deep Learning based Target Tracking and Classification for Infrared Videos Using Compressive Measurements,” Journal Signal and Information Processing, November 2019.

[11] S. Lohit, K. Kulkarni, and P. K. Turaga, “Direct inference on compressive measurements using convolutional neural networks,” Int. Conference on Image Processing. 2016. 1913-1917.

[12] A. Adler, M. Elad, and M. Zibulevsky, “Compressed Learning: A Deep Neural Network Approach,” arXiv:1610.09615v1 [cs.CV]. 2016.

[13] Y. Xu and K. F. Kelly, “Compressed domain image classification using a multi-rate neural network,” arXiv:1901.09983 [cs.CV]. 2019.

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

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

[16] A. Değerli, S. Aslan, M. Yamac, B. Sankur, and M. Gabbouj, “Compressively Sensed Image Recognition,” 7th European Workshop on Visual Information Processing (EUVIP), Tampere, 2018.

[17] P. Latorre-Carmona, V. J. Traver, J. S. Sánchez, and E. Tajahuerce, “Online reconstruction-free single-pixel image classification,” Image and Vision Computing, 86, 2018.

[18] C. Li and W. Wang, “Detection and Tracking of Moving Targets for Thermal Infrared Video Sequences,” Sensors, 18, 3944, 2018.

[19] Y. Tan, Y. Guo, and C. Gao, “Background subtraction based level sets for human segmentation in thermal infrared surveillance systems,” Infrared Phys. Technol., 61: 230–240, 2013.

[20] A. Berg, J. Ahlberg, and M. Felsberg, “Channel Coded Distribution Field Tracking for Thermal Infrared Imagery,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; Las Vegas, NV, USA. pp. 1248–1256, 2016.

[21] C. Kwan, D. Gribben, B. Chou, B. Budavari, J. Larkin, A. Rangamani, T. Tran, J. Zhang, R. Etienne-Cummings, “Real-Time and Deep Learning based Vehicle Detection and Classification using Pixel-Wise Code Exposure Measurements,” Electronics, June 18, 2020.

[22] C. Kwan, D. Gribben, and B. Budavari, “Target Detection and Classification Performance Enhancement Using Super-Resolution Infrared Videos,” Signal & Image Processing: An International Journal (SIPIJ), vol. 12, no. 2, April 30, 2021.

[23] C. Kwan and D. Gribben, “Target Detection and Classification Improvements Using Contrast Enhanced 16-bit Infrared Videos,” Signal & Image Processing: An International Journal (SIPIJ), vol. 12, no. 1, February 28, 2021. DOI: 10.5121/sipij.2021.12103.

[24] C. Kwan and B. Budavari, “Enhancing Small Moving Target Detection Performance in Low Quality and Long Range Infrared Videos Using Optical Flow Techniques,” Remote Sensing, 12(24), 4024, December 9, 2020.

[25] Y. Chen, G. Zhang, Y. Ma, J. U. Kang, and C. Kwan, “Small Infrared Target Detection based on Fast Adaptive Masking and Scaling with Iterative Segmentation,” IEEE Geoscience and Remote Sensing Letters, January 2021.

[26] C. Kwan and B. Budavari, “A High Performance Approach to Detecting Small Targets in Long Range Low Quality Infrared Videos,” arXiv:2012.02579, 2020.

[27] X. Kong, C. Yang, S. Cao, C. Li and Z. Peng, “Infrared Small Target Detection via Nonconvex Tensor Fibered Rank Approximation,” in IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2021.3068465.

[28] D. Ma, L. Dong and W. Xu, “A Method for Infrared Sea-Sky Condition Judgment and Search System: Robust Target Detection via PLS and CEDoG,” in IEEE Access, vol. 9, pp. 1439-1453, 2021, doi: 10.1109/ACCESS.2020.3047736.

[29] Y. Dai, Y. Wu, F. Zhou and K. Barnard, "Attentional Local Contrast Networks for Infrared Small Target Detection," in IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2020.3044958.

[30] P. Yang, L. Dong and W. Xu, "Infrared Small Maritime Target Detection Based on Integrated Target Saliency Measure," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2369-2386, 2021, doi: 10.1109/JSTARS.2021.3049847.

[31] D. Pang, T. Shan, P. Ma, W. Li, S. Liu and R. Tao, "A Novel Spatiotemporal Saliency Method for Low-Altitude Slow Small Infrared Target Detection," in IEEE Geoscience and Remote Sensing Letters, doi: 10.1109/LGRS.2020.3048199.

[32] Q. Hou, Z. Wang, F. Tan, Y. Zhao, H. Zheng and W. Zhang, "RISTDnet: Robust Infrared Small Target Detection Network," in IEEE Geoscience and Remote Sensing Letters, doi: 10.1109/LGRS.2021.3050828.

[33] S. Du, P. Zhang, B. Zhang and H. Xu, "Weak and Occluded Vehicle Detection in Complex Infrared Environment Based on Improved YOLOv4," in IEEE Access, vol. 9, pp. 25671-25680, 2021, doi: 10.1109/ACCESS.2021.3057723.

[34] Z. Song, J. Yang, D. Zhang, S. Wang and Z. Li, "Semi-Supervised Dim and Small Infrared Ship Detection Network Based on Haar Wavelet," in IEEE Access, vol. 9, pp. 29686-29695, 2021, doi: 10.1109/ACCESS.2021.3058526.

[35] M. Wan, X. Ye, X. Zhang, Y. Xu, G. Gu and Q. Chen, "Infrared Small Target Tracking via Gaussian Curvature-Based Compressive Convolution Feature Extraction," in IEEE Geoscience and Remote Sensing Letters, doi: 10.1109/LGRS.2021.3051183.

[36] M. Zhao, W. Li, L. Li, P. Ma, Z. Cai and R. Tao, "Three-Order Tensor Creation and Tucker Decomposition for Infrared Small-Target Detection," in IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2021.3057696.

[37] H. Sun et al., "Fusion of Infrared and Visible Images for Remote Detection of Low-Altitude Slow-Speed Small Targets," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2971-2983, 2021, doi: 10.1109/JSTARS.2021.3061496.

[38] A. Raza et al., "IR-MSDNet: Infrared and Visible Image Fusion Based On Infrared Features and Multiscale Dense Network," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3426-3437, 2021, doi: 10.1109/JSTARS.2021.3065121.

[39] W. Xue, J. Qi, G. Shao, Z. Xiao, Y. Zhang and P. Zhong, "Low-Rank Approximation and Multiple Sparse Constraints Modelling for Infrared Low-Flying Fixed-Wing UAV Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, doi: 10.1109/JSTARS.2021.3069032.

[40] J. Redmon and A. Farhadi, YOLOv3: An Incremental Improvement, arxiv, 2018.

[41] S. Ren, K., He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems, (2015)

[42] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[43] MOT Challenge, motchallenge.net/, accessed May 18, 2021.

AUTHORS

Chiman Kwan received his Ph.D. degree in electrical engineering from the University of Texas at Arlington in 1993. He has written one book, four book chapters, 15 patents, 75 invention disclosures, 380 technical papers in journals and conferences, and 550 technical reports. Over the past 25 years, he has been the PI/Program Manager of over 120 diverse projects with total funding exceeding 36 million dollars. He is also the founder and Chief Technology Officer of Signal Processing, Inc. and Applied Research LLC. He received numerous awards from IEEE, NASA, and some other agencies and has given keynote speeches in several international conferences. He is an Associate Editor of IEEE Trans. Geoscience and Remote Sensing and a Japan Society for Promotion of Science (JSPS) Fellow.

David Gribben received his B.S. in Computer Science and Physics from McDaniel College, Maryland, USA, in 2015. He is a software engineer at ARLLC. He has been involved in diverse projects, including mission planning for UAVs, target detection and classification, and remote sensing.