Adversarial Learning Using Synthetic IR Imagery

James Uplinger[a], Derek Schesser[b], Christopher Meyer[b], Joseph Conroy[b], Celso de Melo[b]
[a]Huntington Ingalls Industries, 8350 Broad Street, Suite 1400 McLean, VA 22102; [b]Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, MD 20783

ABSTRACT

Recent years have seen impressive progress in Automatic Target Recognition (ATR) technology, both in the visible and non-visible spectra, which introduces an important challenge to the Army: understanding gaps in ATR algorithms' feature space for informed design methodology. To tackle this challenge, we look at a combination of synthetic data and adversarial learning techniques to explore the feature space of Machine Learning (ML) algorithms. Adversarial learning, however, requires large amounts of training data representing diversity in terms of target pose, lighting, and environmental conditions. Often the main bottleneck is collecting and labeling this real training data. The problem is exacerbated in infrared (IR) given unique challenges due to material and thermal variation. Here, we present a solution based on a simulator that supports generation of physically accurate custom synthetic IR training data; this data is then leveraged to systematically study weaknesses in a state-of-the-art ATR algorithm that is often used in practice, YOLOv5. We will present results showing that this approach can lead to critical insight on algorithm weaknesses with practical consequence for the design of defense mechanisms against ATR technology as well as improved training of ML algorithms to reduce feature space vulnerabilities.

Keywords: Infrared Imaging, Target Detection, Synthetic Images, Deep Learning, Adversarial Learning

1. INTRODUCTION

Research into deep learning has grown dramatically over the last decade and has resulted in a significant amount of insight into how to build and train models for applications such as Automatic Target Recognition (ATR). A persistent issue with these models is a demonstrated fragility that results in blind spots and non-intuitive detection characteristics. [1] Identification of these systematic vulnerabilities is an important challenge to the Army. Specifically, identification of fragility and non-intuitive detection characteristics of physically realizable variations provides valuable insight into how model training can be performed to create more robust training methods and data sets. Traditional methods of adversarial probes of deep learning models often introduce adversarial perturbations of minor pixel variations that are difficult to reproduce in real world situations, making data collection problematic. [2] This approach is difficult to implement consistently.

2. METHOD

We explored the use of a physics-based simulator to generate synthetic infrared (IR) images, combined with thermal perturbations of the target model with the use of the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model.[3,4] DIRSIG provided the ability to recreate scenes with differences only in the thermal perturbations, allowing us to confine the scene, atmospheric, camera, and lighting properties to reproduce complete datasets that are identical, save for the desired thermal perturbation. The synthetic IR data was to used for training of Automatic Target Recognition (ATR) models. We employed the widely used, state-of-the-art YOLOv5 deep learning object detection algorithm. [2] YOLOv5 is an evolution of the original YOLO architecture developed by Redmon et al., which frames object detection as a regression problem for object localization with associated class probabilities. [5] Differing from prior two-stage region proposal-

based object detection architectures, YOLO uses a single unified neural network to predict bounding boxes and class probabilities from the full image and is computationally efficient and suitable for at-the-edge computing with limited resources. YOLOv5 implements five differently sized variants, nano “n”, small “s”, medium “m”, large “l”, and extra-large “x”, ranging from 1.9M parameters in the “n” variant to 86.7.M parameters in the most accurate but largest “x” variant. This study employed the “s” and “x” variants to characterize and contrast the performance of different sized models trained with synthetic IR imagery. The “s” variant with 7.5M parameters is the most suitable for real-time ATR onboard a small UAV, for example where size weight and power (SWAP) are limited with the tradeoff being reduced detection accuracy.

1.1 Datasets

We used three separate datasets for training and testing in this work. The first dataset had 13,800 IR images of emplaced military vehicles collected by the Army Research Laboratory (ARL) in 2020 (ARL2020 dataset). The second dataset comprised 14,402 images generated from DIRSIG, comprising two target classes of a field-portable power generator and a pickup truck. Separate validation (5,918 real images and 3,084 DIRSIG images) were used in model training. Five labeled categories exist in the combined ARL2020 and DIRSIG datasets. The YOLOv5 analysis focused on one label category: generator.

1.2 DIRSIG Images and Perturbations

The DIRSIG model has been developed at the Rochester Institute of Technology (RIT) over decades of research and has been historically used for image generation and sensor development for a range of remote sensing modalities. Long wave infrared (LWIR) data was generated using the first principles-based model to predict radiance values received by a sensor. Object self-radiance values are calculated using surface temperatures and material properties assigned to the imported object. Surface temperatures were calculated using ANSYS Mechanical Finite Element Analysis (FEA) Software using the steady state thermal application to simulate equilibrium surface temperatures due to solid conduction, convective heat transfer, and radiative heat transfer. Emissive properties were measured using Surface Optics Corp. 410-Vis-IR portable emissometer and validation field temperature measurements were made with a FLIR camera. Material properties were assigned to the imported object from the DIRSIG material database representing materials similar to the imported targets actual properties. To import the material and thermal values to a geometric model file that DIRSIG uses, the original CAD was exported to the 3D object format OBJ which was then converted to a Geometric Database file (GDB) using DIRSIG’s object_tool module. The 3D GDB format supports three-point surface facet special values, temperature value, and material lookup definition for each facet. The temperature value for a facet was determined by finding the FEA mesh point closest to the center of each facet and assigning the temperature value of the mesh point to the facet in the GDB file. Similarly, the material properties were assigned using integer assigned temperatures in the results file to identify each material region and then using the same algorithm the material field was updated for each facet.

The sensor modeled in DIRSIG had a spectral response of 8-12 microns. The location of the sensor was mounted on an aerial platform that circled the imported object completing one circuit per hour for eight hours. The position of the sensor varied from orbits of 20 meters to 3000 meters in radius

and 100 meters to 1000 meters in altitude. Sample images generated by DIRSIG of a generator and pickup truck are shown in Figure 1.

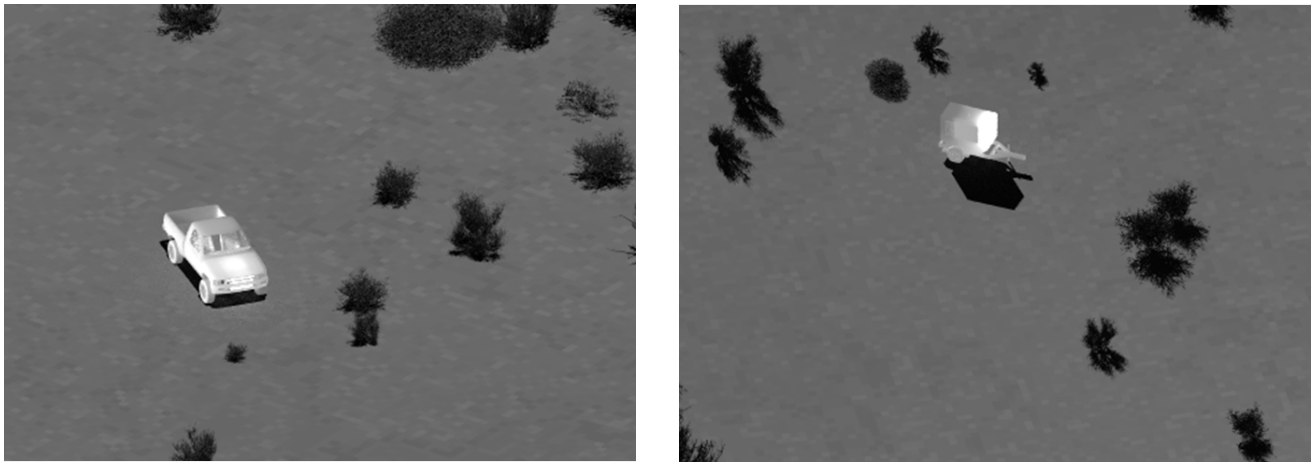

Figure 1. Left, a DIRSIG LWIR image of a pickup truck. Right, a DIRSIG LWIR image of a field portable power generator.

Thermal perturbations of generator features were achieved by assigning GDB facet temperatures assigned to specific model components. Specifically, thermal perturbations were made on four generator features: generator body, tires, wheels, and trim and fenders (TaF). Thermal perturbations were introduced on these features ranging from +/- 9ºC of ambient air temperature in increments of 3ºC. Once a thermal perturbation had been introduced to the GDB model file, a complete set of 1,443 images were rendered in DIRSIG that mirrored the rendered images of the nominal DIRSIG generator dataset in terms of scene composition, time of day, atmospheric conditions, sensor configuration, altitude, distance, and orientation in respect to the generator. Sample images from the thermal perturbation datasets are shown in Figure 2.

Figure 2. Sample images of thermal variants of the DIRSIG generator. Features were thermally modified in 3ºC step increments from ambient thermal solution.

1.3 Automatic Target Recognition Training and Evaluation

The YOLOv5 algorithm was re-trained using a training dataset containing DIRSIG pickup truck and DIRSIG generator images as well as 13,800 real IR images of military vehicles consisting of three classes starting with pre-trained weights. The validation dataset contained DIRSIG pickup truck and DIRSIG generator images as well as 5,918 real IR images of military vehicles consisting of three classes.

Table 1. Training and validation data information.

Dataset	Image Dimensions	Number of Classes	Number of Training Images	Number of Validation Images
ARL IR Images	640 × 512	3	13,800	5,918
DIRSIG Generator	640 × 512	1	7,201	1,542
DIRSIG Pickup Truck	640 × 512	1	7,201	1,542

Evaluation of the retrained YOLOv5 algorithm was performed with sets of 1,443 images consisting of the DIRSIG generator with nominal thermal profile as well as modified thermal profiles of the body, tires, wheels, and trim. Each modified thermal variant was modified in 3ºC increments from -9ºC to +9ºC of ambient temperature. In total, twenty-five test trials were performed.

3. RESULTS & DISCUSSION

We tested the widely used YOLOv5 object detector after retraining it with our training dataset described above. We started with pre-trained weights and assessed the detection performance of each of the twenty-five test sets described above to identify changes in detection performance. The “s” YOLOv5 model was trained to 30 epochs on the training set defined above.

Figure 3 shows the test results for the twenty-five sets of thermal variants of the DIRSIG generator datasets. The thermal variations targeted physical features of the generator such as the body, wheels, tires, and Trim and Fender (TaF). Each feature was modified by changing the temperature of the feature above and below the standard temperature of the nominal thermal solution. In Figure 3, the Standard Temperature is the test dataset representing the DIRSIG Generator at the ambient thermal solution.

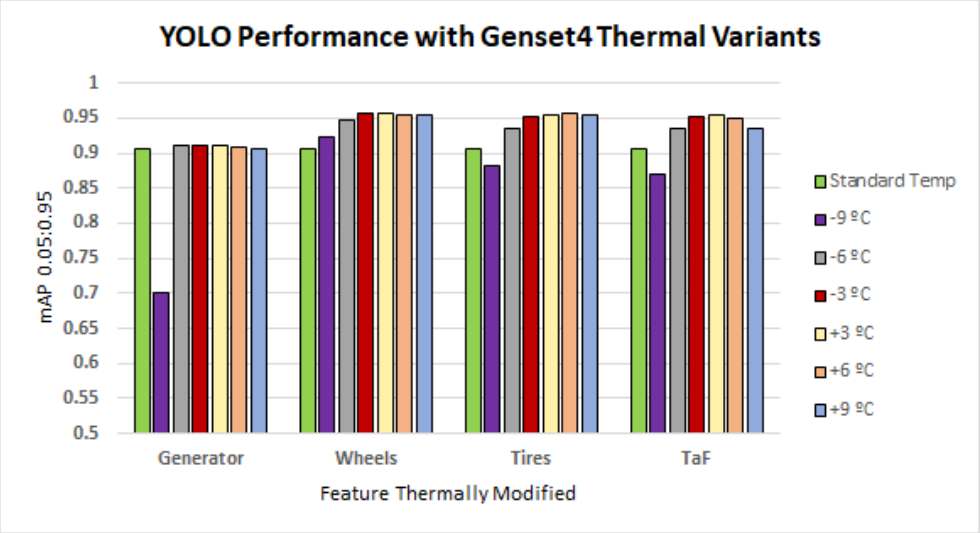

Figure 3. Experimental test results for thermal modifications of the Generator Body, Wheels, Tires, and Trim and Fender (TaF).

Results in Figure 2 show that by varying the temperature of identifiable features on the generator the detection performance of the YOLOv5 detector can be modified. Most significantly, reducing the temperature of the body significantly decreases the mAP from 0.906 to 0.701. While some thermal variants decreased the detection performance of the YOLOv5 detector, other variances increased the detection performance. Three thermal variants improved the mAP from 0.906 to 0.956, +3ºC modification to the wheels, +6ºC modification to the tires, and -3ºC to the wheels. Other modifications improved performance by smaller amounts.

Two images were selected from each of the twenty-five test datasets. Each image is the same across the twenty-five datasets with the exception of the thermal modification. These images are represented by two examples taken from the -9ºC generator body dataset, Scene 1 and Scene 2, shown in Figure 4. These two scenes were analyzed independently with the YOLOv5 detector to measure whether the variability in individual scenes were due to thermal variations.

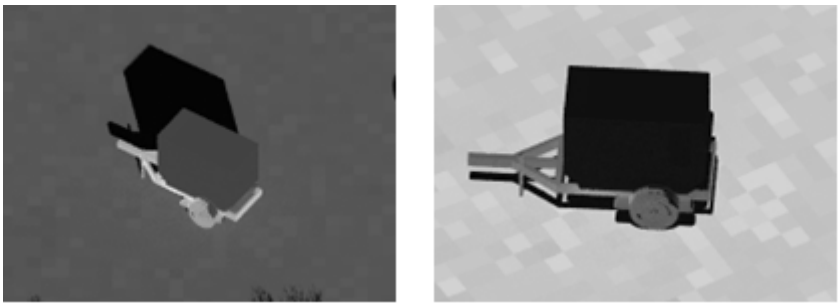

Figure 4. Two images from the -9ºC generator body dataset. Scene 1 (left) and Scene 2 (right).

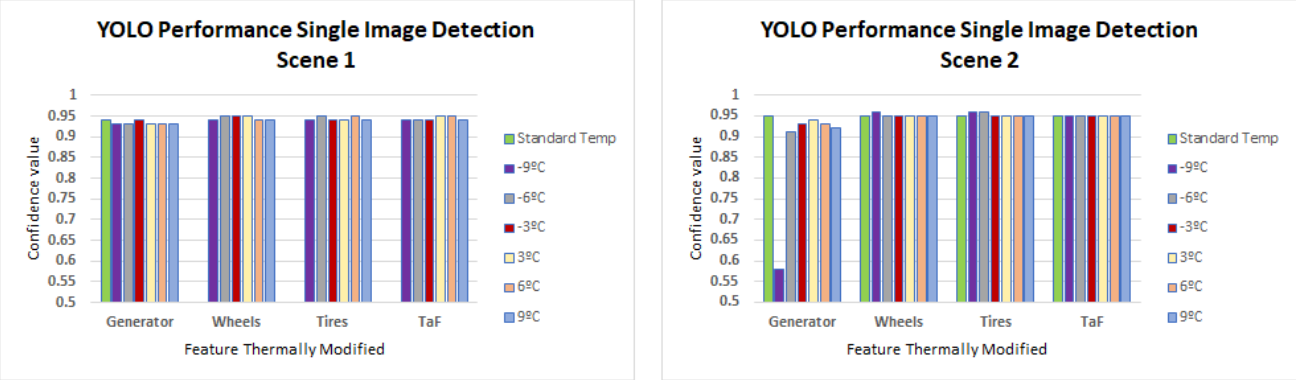

Figure 5. Confidence values for a single image across thermal variants for Scene 1 (left) and Scene 2 (right).

Figure 4 shows the results for YOLOv5 detection confidence in Scenes 1 and 2 for each of the twenty-five thermal variants. Each confidence value is the detection confidence for one image. Scene 1 shows virtually no change in detection confidence level based on thermal variations, while Scene 2 confidence values are significantly influenced by the thermal variations in the generator body.

These results illustrate that a bulk image dataset with a thermal variation does not necessarily result in a uniform drop in detection performance on each image of the dataset. The learned feature space of the YOLOv5 detector is not able to identify the specific shapes and shades of the generator in Scene 2 with all thermal variations with as high of confidence, when compared to all other thermal variants and Standard Temperature image.

4. CONCLUSION

AI/ML models for detection of targets are critical for automatic target recognition applications. However, these AI/ML models are known to be fragile and susceptible to real perturbations of imagery, due to lighting, angle, weather, etc. These real-world perturbations are difficult to collect in sufficient quantity to identify AI/ML model weaknesses. Synthetic data provides an important tool to explore AI/ML feature space and help to identify gaps in detector performance. Here, we show how the use of synthetic data can be used extract information about gaps in detector performance. These results show that synthetically generated datasets can be used to identify specific image characteristics that lead to poor detection performance. The results utilized the combined results of bulk dataset testing, as well as individual scene confidence evaluation to identify thermal modifications of interest to scene specific features that impact AI/ML model performance.

In practice, there is often limited amounts of labeled real data available for AI/ML training. In order to retain real labeled data for testing it becomes necessary to remove real data from the training datasets, which can impact model performance. Data collection events are often planned for a number of parameters, but often when collections take place specifics of variations and gaps in existing datasets are not thoroughly understood or accounted for. The work presented has shown that synthetic data can be used to inform data collection events for potential gaps in training data. In line with prior work showing that synthetic data can be used to supplement real datasets, this work can be used to generate small, targeted datasets to be included with training datasets to improve AI/ML performance.

5. FUTURE WORK

Whereas we are encouraged by the results reported here, there are several opportunities for future work. First, the AI/ML model used is a common detection algorithm used for automatic target recognition. However, there are multiple AI/ML model architectures available that may alternatively be used. It is important to understand the feature space similarities between these alternative model architectures and to identify any lessons that can be generalized across architectures. [7] Secondly, the research explored here was conducted in a virtual environment. The expansion of the work conducted to include field tests of the concepts to verify that the identified feature space weaknesses translate to real imagery must be conducted to real imagery. Finally, while this work focused on the thermal domain, it is also worth exploring the visible domain with simulated feature variations.

REFERENCES

6. **VERMA, D. C., VERMA, A., & MANGLA, U. (2021, DECEMBER). ADDRESSING THE LIMITATIONS OF AI/ML IN CREATING COGNITIVE SOLUTIONS. IN 2021 IEEE THIRD INTERNATIONAL CONFERENCE ON COGNITIVE MACHINE INTELLIGENCE (COGMI) (PP. 189-196). IEEE.**

7. Prakash, A., Moran, N., Garber, S., DiLillo, A., & Storer, J. (2018). Deflecting adversarial attacks with pixel deflection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8571-8580).

8. Sanders, J. S., & Brown, S. D. (2000, July). Utilization of DIRSIG in support of real-time infrared scene generation. In Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process (Vol. 4029, pp. 278-285). SPIE.

9. Goodenough, A. A., & Brown, S. D. (2012, May). DIRSIG 5: core design and implementation. In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII (Vol. 8390, pp. 124-132). SPIE.

10. Jocher, Glenn, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, yxNONG, Adam Hogan, lorenzomammana, AlexWang1900, Ayush Chaurasia, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug; Durgesh, Francisco Ingham, Frederik, Guilhen, Adrien Colmagro, Hu Ye, Jacobsolawetz, Jake Poznanski, Jiacong Fang, Junghoon Kim, Khiem Doan, and Lijun Yu. "Ultralytics/yolov5: v4.0—nn.SiLU() Activations, Weights & Biases Logging, PyTorch Hub Integration." Accessed 28 March 2021. https://doi.org/10.5281/zenodo.4418161.

11. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 26 June – 1 July 2016.

12. **YE, A. (2022, MARCH 20). VISUALIZING HOW DIFFERENT ML MODELS OPERATE IN THE FEATURE SPACE. RETRIEVED APRIL 3, 2023, FROM**

HTTPS://TOWARDSDATASCIENCE.COM/VISUALIZING-HOW-DIFFERENT-ML-MODELS-OPERATE-IN-THE-FEATURE-SPACE-C6CAA8A96375.