SAND2022-3231C

Synthetic threat injection using digital twin informed augmentation

Daniel Krofcheck*[a], Esther John[a], Hugh Galloway[a], Asael Sorensen[a], Carter Jameson[a], Connor Aubry[a], Arvind Prasadan[a], Robert Forrest[a]
[a]Sandia National Laboratories, 1515 Eubank Blvd SE, Albuquerque, NM 87123, USA

ABSTRACT

The growing x-ray detection burden for vehicles at Ports of Entry in the US requires the development of efficient and reliable algorithms to assist human operator in detecting contraband. Developing algorithms for large-scale non-intrusive inspection (NII) that both meet operational performance requirements and are extensible for use in an evolving environment requires large volumes and varieties of training data, yet collecting and labeling data for these enivornments is prohibitively costly and time consuming. Given these, generating synthetic data to augment algorithm training has been a focus of recent research. Here we discuss the use of synthetic imagery in an object detection framework, and describe a simulation based approach to determining domain-informed threat image projection (TIP) augmentation.

Keywords: Threat Image Projection; Non-Intrusive Detection; Synthetic Data; X-ray Detection; Digital Twin; Unity

1. INTRODUCTION

Roughly 5% of all vehicular traffic entering the United States through Ports of Entry is currently electronically scanned, and Customs and Border Protection aims to move that fraction to 100%[1,2]. The use of Non-Intrusive Inspection (NII) in large-scale commodity applications, most commonly using x-ray imaging, requires a human operator to visually inspect each scan for targets of interest. As the scan rate across ports of entry increases, so does the burden on both human adjudicators and the existing infrastructure [3,4,5].

The deployment of machine learning algorithms to augment human operators in the adjudication of large-scale NII tasks has been the focus of recent research, yet several challenges have so far precluded wide adoption[6]. Primarily, the issues stem from the large data volume requirements to train contemporary models (e.g., convolutional neural nets, transformers), contrasted with the apparent data scarcity in the large-scale commercial domain[1,5]. Specifically, algorithms designed to identify small metallic threats (SMTs) of interest require labeled examples of those targets in their common context, yet in-situ examples of SMTs are incredibly rare[7]. The high-consequence, low-probability nature of this domain presents a significant hurdle for algorithm development and training. Further, labeling data with sufficient variety and volume is often prohibitively time consuming and costly. In the case of large-scale NII using x-ray imagery, the challenges associated with collecting data become significantly more complex. Land ports of entry, for instance, are largely adopting drive-through x-ray imaging systems, and collecting training data in these environments is complex and time consuming.

The prevailing solution to these challenges is the generation of synthetic observations through the Threat Image Projection (TIP) approach, which has been used for this purpose in a variety of domains[8,9,10,11,12,13], including large-scale NII[6]. Additionally, the TIP method generally involves data augmentation strategies designed to increase the diversity of the TIPs within the context images, such as translations, rotations, and scale changes[7]. This represents an area of active research, as some specific augmentation regimes (e.g., scale change) represent the physical relationship between the TIP injection location and the specific sensor geometry being simulated. While several studies have described the potential for object detection models trained solely on TIP data to perform well in stream of commerce (SOC) contexts[1,7], few studies aim specifically to identify SMTs in large-scale NII, and to our knowledge, none have endeavored to characterize the potential impact of TIP on SMTs in a field driven experimental setting. Further, we are unaware of any experimental efforts aiming to constrain TIP augmentation as a function of TIP placement and sensor geometry.

Here we present a notional synthetic data generation workflow with two primary components: 1) threat injection, where an isolated image is combined with a context image to produce a new training image; and 2) an image augmentation approach informed by a 3D model digital twin of the x-ray instrument. Our geometric modeling approach describes a

specific set of transformations to be applied to the injected image to reproduce the geometric distortions of the threat based on its intended location inside the context image. While these are notional results, we believe that further research involving the use of TIP with domain informed augmentation in NII systems will increase the impact of data collection efforts, by increasing our ability to augment injected threats in a radiometrically and geometrically consistent manner.

2. DATA

We leveraged an empirical dataset collected during a multi-agency collaboration for this study. The complete dataset contains manually labeled large-scale NII scans produced at a Multi-Energy Portal (MEP) and contains a combination vehicle types, experimentally controlled to include a combination of commerce materials and SMTs. While the full dataset is described elsewhere[14], we constrained our analysis to a subset of the data: tractor trailers. The subset used in this study contains 387 tractor trailer images, as a combination of flat beds and containers. We also subset the SMT data to highlight two distinct targets, a geometrically simple SMT (Target A) and geometrically complex target (Target B). We split the dataset into a single train/test split (90/10) for our proof-of-concept analysis. We then built two experimental datasets for our analysis: No-TIP and With-TIP. The No-Tip set contained non-synthetic scans of Target A (n = 272) and Target B (n = 145) distributed throughout the train and test imagery. The With-TIP set included images to double the training observations of Target A (total n = 490) and Target B (total n = 270). Development of the TIP images is described below.

3. METHODOLOGY

3.1 Threat image projection (TIP)

We used hand-cropped and pseudo-isolated scans of each target (targets imaged with minimal structural support and against an open-air background) to serve as the source TIP. We then followed Rogers et al., 2016[7] to combine the target with a context image from the training fold described above. Because the context images were constrained to tractor trailers, we regionalized the location of the TIP to ensure a feasible placement in the context image (e.g., no floating in the air, or outside of the container).

3.2 Object detection

We leveraged a collection of existing deep learning object detection models using the Detectron2 API and Model Zoo[15]. We selected candidate architectures from the Model Zoo, 3 Faster RCNN and 3 Retinanet model architectures, listed in Table 1. Our dataset is very small for object detection purposes, and we did not fine tune the training regime or hyperparameters of the models, as optimizing performance is not a goal of this work. We trained each model to 3,500 epochs, with a learning rate of $1.2e^{-4}$. Both the No-TIP and With-TIP trained models were tested on the same non-synthetic test set using the COCO ground truth API (COCO citation)[16].

Table 1. Selected candidate architectures from Model Zoo used for object detection

Model Name	Model Family	Model Outputs
R50-FPN-1x	Faster RCNN	Class, Score, Box
R50-FPN-3x	Faster RCNN	Class, Score, Box
R101-FPN-3x	Faster RCNN	Class, Score, Box
RN-R50-1x	RetinaNet	Class, Score, Box
RN-R50-3x	RetinaNet	Class, Score, Box
R_101_C4_3x	Mask RCNN	Class, Score, Box, Mask
R_101_DC5_3x	Mask RCNN	Class, Score, Box, Mask
R_101_FPN_3x	Mask RCNN	Class, Score, Box, Mask

3.3 MEP digital twin

We used open-source descriptions of the geometry of a MEP to create a to-scale model of the system in the Unity game engine. For visualization purposes, we created an x-ray gantry using the Blender open-source modeling tool kit (Figure 1). The Unity simulation consists of a configurable fan beam emitter and an L-shaped detector, the geometry of which can be tuned to the specific system being simulated. We modeled the x-ray fan-beam source as a collection of raycasts originating from a range of heights and at a range of angles dictated by the system being simulated. Each ray is attenuated proportional to the path length of objects it intersects over its lifetime from emission to detection. The final attenuation of each ray is then converted to pixel values and written to a raster. In this way, first order approximations of object projection distortion can be explored by advancing a commodity or conveyance through the simulated gantry and analyzing the resulting synthetic x-ray image. We used notional angles, distances, and specific geometries in this example, to demonstrate the capability rather than make the results applicable to a specific platform.

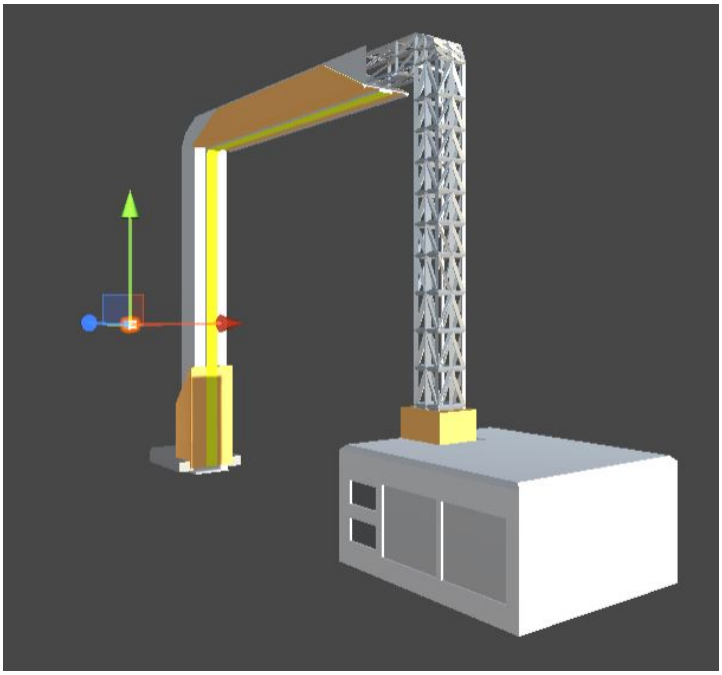

Figure 1. MEP x-ray gantry created in the Unity game engine. The x-axis (red) defined the direction of vehicular travel, the y-axis (green) dictates the height of objects as they pass through the detector, and the z-axis (blue) governs the distance from the xray source.

We used the digital twin to simulate scanning commodities at a variety of locations along the y-axis (green arrow in Figure 1) which represents the commodity height, the z-axis (blue arrow in Figure 1) which represents distance from the x-ray source, and rotation around the x-axis (red arrow in Figure 1). We ran two experiments, the rotation experiment which held z-axis displacement constant but progressively rotated objects around the x-axis, and the depth experiment which held rotation constant but varied the location of objects in the z-axis. In both experiments, we arrayed the target object along the y-axis to determine the interactions between target height on both x-axis rotation and z-axis displacement.

4. RESULTS

4.1 Synthetic data impact on model performance

Model results for the unmodified and synthetic datasets for each target type are shown in Table 2. When compared across models, there was no statistical difference between the unmodified and synthetic data (alpha = 0.05, target A $p = 0.17$, target B $p = 0.09$). Impact on target type was asymmetric, with the symmetric metallic target (Target A) seeing a small decrease in detection performance across the board, 0.03 on average, while also increasing model standard deviation from

0.03 without TIP to 0.04. The combination organic metallic target (Target B) saw one model slightly improve with the others slightly degrading. The difference for Target B was 0.07 on average, with the standard deviation increasing from 0.04 to 0.08. Given the paucity of data we didn't hold out a non-synthetic threat that greatly differed from its compatriots to see if the synthetic augmentation improves "surprise" location detections. We expect the slight decrease is due to improved generalization and leave testing this to future work.

Table 2. mAP comparison of a two-target detection experiment, trained with all in-situ observations (No Synthetic) and then augmented with an additional 100% synthetic TIP observations per target (Synthetic).

	No-TIP		**With-TIP**	
Model	**Target A**	**Target B**	**Target A**	**Target B**
R50-FPN-1x	0.77	0.87	0.74	0.68
R50-FPN-3x	0.75	0.90	0.71	0.87
R101-FPN-3x	0.79	0.94	0.77	0.87
RN-R50-1x	0.72	0.90	0.67	0.79
RN-R50-3x	0.76	0.84	0.74	0.86
X101-FPN-3x	0.80	0.95	0.78	0.90

4.2 Domain informed augmentation

Our digital twin simulation experiment produced an array of shadow projections that correspond to differences in TIP dimensions informed by intended injection location. Figure 2 shows the results of the x-axis rotation experiment, where the repeated columns vary by rotation in the axis orthogonal to the page. Figure 3 shows the results of the z-axis (blue arrow in figure 1) location experiment, where each column is arranged a set distance from the detection source.

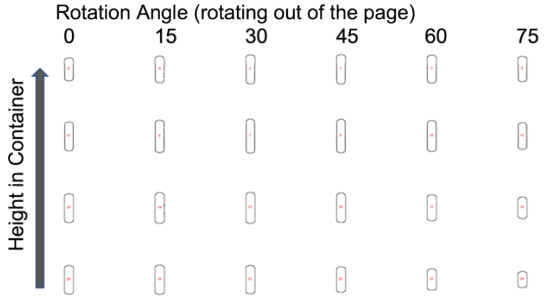

Figure 2. Shadow projections of objects with increasing rotation along the x-axis. All objects are identical, and each row distributed at equal heights in the synthetic container.

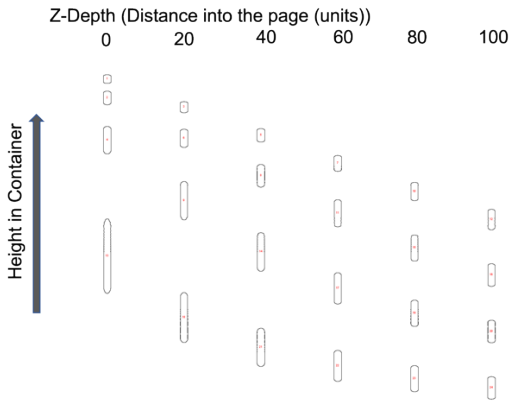

Figure 3. Shadow projections of objects with increasing distance from the detection source along the z-axis. All objects are identical, and each row distributed at equal heights in the synthetic container.

Using the results from our rotation (Figure 2) and depth (Figure 3) experiments, we described the projected area of the target objects as a function of distance from the x-ray source, and rotation in the x-axis (Figure 4).

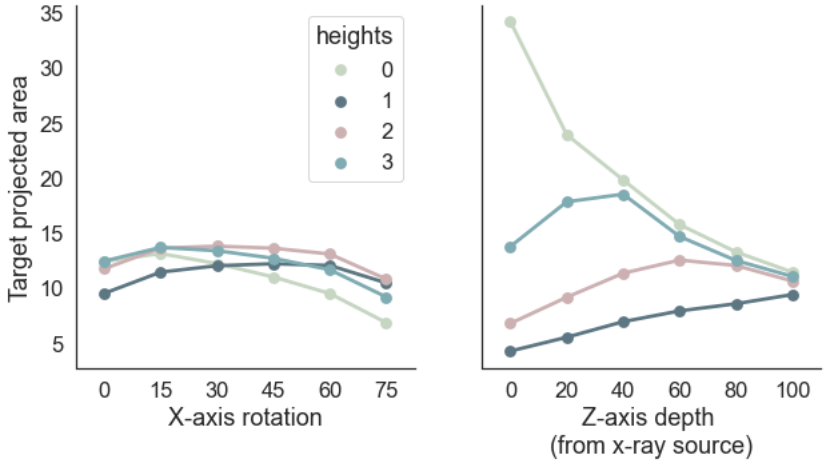

Figure 4. Projected area of the target objects as a function of rotation in the x-axis (left) and distance from the x-ray source (right).

The relative change in projected area due to rotation does have an interaction with placement height: as an object rotates toward the x-ray source, the angle that maximizes the surface area exposed to the beam will vary as a function of height. This effect is mild however compared to the magnification of the projected area due to depth relative to the x-ray source.

5. CONCLUSION

The growing need for effective synthetic data or TIP solutions is especially salient in high-consequence, low-probability domains such as large-scale NII. Given the logistical and security challenges associated with data collection and labeling, the community has been seeking to develop robust capabilities to help bridge the gap. However, while the TIP process is fairly well understood at this point, generating datasets for large-scale NII is still a complex task, in large part due to the requirement for isolated scans of targets of interest. In our stream of commerce experimentation, the inclusion of TIP data was largely non-effective, if not slightly detrimental to object detection performance. However, by exposing the training process to a larger corpus of target locations and contexts, we believe the detection capability would begin to generalize

more robustly, as seen in previous TIP and SMT studies. Notably, these previous works have access to datasets 2-3 orders of magnitude larger than our own.

Increasing observations using TIP alone, however, does not ensure that models are being trained on targets in such a way that the relationship between target geometry and context placement are constrained by the physics of the system of interest. Our work suggests this relationship can be exploited to learn a function that takes as input a TIP bounding box, height, and depth, and returns an augmentation can be prescribed prior to target injection. Subsequent TIP images would better adhere to the physical constraints imposed by the geometry of the system they are designed to mimic.

ACKNOWLEDGEMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy?s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government

REFERENCES

[1] Jaccard, N., Rogers, T. W., Morton, E. J. and Griffin, L. D., “Automated detection of smuggled high-risk security threats using Deep Learning,” 7th Int. Conf. for Imaging on Crime Detection and Preventation, 1-6 (2016).

[2] Archick, K., “U.S.-EU Cooperation Against Terrorism,” Congressional Research Service Report RS22030, Version 36 (2016).

[3] Orphan, V. J., Muenchau, E., Gormley, J. and Richardson, R, “Advanced γ ray technology for scanning cargo containers,” Applied Radiation and Isotopes 63(5–6), 723-732 (2005).

[4] Jaccard, N., Rogers, T. W., Morton, E. J. and Griffin, L. D., “Tackling the X-ray cargo inspection challenge using machine learning”, Proc. SPIE 9847, 98470N (2016).

[5] Rogers, T. W., Jaccard, N., Morton, E. J. and Griffin, L. D., “Automated X-ray image analysis for cargo security: critical review and future promise,” Journal of X-Ray Science and Technology 25(1), 33-56 (2017).

[6] Akcay, S. and Breckon, T., “Towards automatic threat detection: A survey of advances of deep learning with X-ray security imaging,” Pattern Recognition 122, 108245, (2022).

[7] Rogers, T. W., Jaccard, N., Protonotarios, E. D., Ollier, J., Morton, E. J. and Griffin, L. D., "Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines," 2016 IEEE International Carnahan Conference on Security Technology (ICCST), 1-7 (2016).

[8] Hofer, F. and Schwaninger, A., “Using threat image projection data for assessing individual screener performance,” WIT Transactions on the Built Environment 82, 417-426 (2005).

[9] Schwaninger, A., Hofer, F. and Wetter, O. E., “Adaptive computer-based training increases on the job performance of x-ray screeners,” 41st Annual IEEE International Carnahan Conference on Security Technology, 117–124 (2007).

[10] Schwaninger, A., Bolfing, A., Halbherr, T., Helman, S., Belyavin, A. and Hay, L., “The impact of image based factors and training on threat detection performance in x-ray screening,” Proc. 3rd International Conference on Research in Air Transportation, 317–324 (2008).

[11] Cutler, V. and Paddock, S., "Use of threat image projection (TIP) to enhance security performance," 43rd Annual 2009 International Carnahan Conference on Security Technology, 46-51 (2009).

[12] Mery, D. and Katsaggelos, A. K., “A logarithmic X-ray imaging model for baggage inspection: simulation and object detection”, Conference on Computer Vision and Pattern Recognition Workshops, 251–259 (2017).

[13] Bhowmik, N., Wang, Q., Gaus, Y. F. A., Szarek, M. and Breckon, T. P., “The good, the bad and the ugly: evaluating convolutional neural networks for prohibited item detection using real and synthetically composited X-ray imagery”, British Machine Vision Conference Workshops (2019).

[14] Krofcheck, D., John, E., Galloway, H., Sorensen, A., Jameson, C., Aubry, C., Prasadan, A., Galasso, J., Goodman, E., Blanchard, E. and Forrest, R., "Resilient adjudication in non-intrusive inspection with hierarchical object and anomaly detection," Proc. SPIE Defense + Commercial Sensing (2022).
[15] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. and Girshick, R., "Detectron2," 2019, <https://github.com/facebookresearch/detectron2> (9 March 2022).
[16] Lin, T.-Y., Patterson, G., Ronchi, M. R., Cui, Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, L. and Dollár, P., "COCO: common objects in cotext," < https://cocodataset.org/> (9 March 2022).