SANDIA REPORT
SAND2020-10244
Printed Click to enter a date

How Low Can You Go? Using Synthetic 3D Imagery to Drastically Reduce Real-World Training Data for Object Detection

Zoe N. Gastelum
Timothy M. Shead

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

ABSTRACT

Deep convolutional neural networks (DCNNs) currently provide state-of-the-art performance on image classification and object detection tasks, and there are many global security mission areas where such models could be extremely useful. Crucially, the success of these models is driven in large part by the widespread availability of high-quality open source data sets such as Image Net, Common Objects in Context (COCO), and KITTI, which contain millions of images with thousands of unique labels. However, global security relevant objects-of-interest can be difficult to obtain: relevant events are low frequency and high consequence; the content of relevant images is sensitive; and adversaries and proliferators seek to obscure their activities. For these cases where exemplar data is hard to come-by, even fine-tuning an existing model with available data can be effectively impossible.

Recent work demonstrated that models can be trained using a combination of real-world and synthetic images generated from 3D representations; that such models can exceed the performance of models trained using real-world data alone; and that the generated images need not be perfectly realistic (Tremblay, et al., 2018). However, this approach still required hundreds to thousands of real-world images for training and fine tuning, which for sparse, global security-relevant datasets can be an unrealistic hurdle.

In this research, we validate the performance and behavior of DCNN models as we drive the number of real-world images used for training object detection tasks down to a minimal set. We perform multiple experiments to identify the best approach to train DCNNs from an extremely small set of real-world images. In doing so, we:

- Develop state-of-the-art, parameterized 3D models based on real-world images and sample from their parameters to increase the variance in synthetic image training data;
- Use machine learning explainability techniques to highlight – and correct through targeted training – the biases that result from training using completely synthetic images; and
- Validate our results by comparing the performance of the models trained on synthetic data to one another, and to a control model created by fine-tuning an existing ImageNet-trained model with a limited number (hundreds) of real-world images.

ACKNOWLEDGEMENTS

This project was funded by Sandia National Laboratories' Laboratory-Directed Research and Development program, through the Global Security Investment Area, project number: 218401. The authors with to thank the Lawrence Livermore National Laboratory Deep Learning Strategic Initiative for access to a curated open source image set that we incorporated into our training data, and Michael Higgins for support in collecting and labelling images.

CONTENTS

1. Introduction 9
2. Synthetic Image Development 10
3. Experimental Plan 12
   3.1. Parameter Variance Experiments 12
   3.2. Image Ratio Experiments 12
   3.3. Machine Learning Explainability Experiments 13
4. Object Detection 15
   4.1. Parameter Variance in Object Detection Results 15
   4.2. Image Ratio for Object Detection Results 15
   4.3. Machine Learning Explainability for Object Detection Results 15
5. Image Classification 16
   5.1. Parameter Varience in Image Classification Results 16
      5.1.1. Background Image Variance 16
      5.1.2. Modelled Object Variance 17
   5.2. Image Ratio for Image Classificaiton Results 19
   5.3. Model Thresholds 19
   5.4. Machine Learning Explainability for Image Classification Results 20
6. Discussion 21
   6.1. Data Proportions 21
   6.2. Larger Training Data Sets 21
   6.3. Other object detection models 21
7. Future Research 22
8. REFERENCES 23

LIST OF FIGURES

Figure 1 2D Rendering of Synthetic CRL-type Manipulator Arm 10

Figure 2 Synthetic remote manipulator arm against real-world industrial, black, ImageNet mean color, and noise backgrounds. Industrial background image credit: TerraPower. 12

Figure 3 CAM visualizations of two images classified as cooling towers. The image on the left is a correct classification, showing that the model has "learned" to look for the hyperboloid curve to classify a cooling tower. The image on the right s an incorrect classification, indicating that the curve of the guitar player's neck and strap were incorrectly used to classify the image as a cooling tower. (Image credit: (L) Ken Davis via Flickr, November 23, 2009. https://www.flickr.com/photos/ken_davis_archive/5351748461/in/photostream/; (R) johnny anguish via Flickr, April 1, 2012. https://www.flickr.com/photos/johnnyanguish/7037904109/in/photostream/) 14

Figure 4 Loss and Accuracy for Neutral Background Experiments 17

Figure 5 Loss and Accuracy for Single Tube Manipulators 18

Figure 6 Loss and Accuracy for CRL Design Models 18

Figure 7 Performance Metrics across Decision Thresholds 19

LIST OF TABLES

Table 1 Synthetic Datasets Rendered for this Research 11

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
COCO	Common Objects in Context
CRL	Central Research Laboratory
HDR	high-dynamic-range
IAEA	International Atomic Energy Agency

1. INTRODUCTION

For an introduction to computer vision – including the image classification and object detection tasks we will describe here – see our paper from the Institute of Nuclear Materials Management Annual Meeting (Gastelum, Shead, & Higgins, 2020).

The International Atomic Energy Agency (IAEA) Department of Safeguards is tasked with independently verifying that "nuclear facilities are not misuse and nuclear material not diverted from peaceful uses" using a set of technical measures applied according to each state's safeguards agreement (International Atomic Energy Agency, n.d.). The verification of IAEA safeguards involves the collection, integration, and analysis of multiple data streams, including data declared by state regulatory authorities, data collected in the field during on-site inspections or other technical visits, and other sources of information including that from open sources and satellite imagery. Many safeguards activities that require visual information inspection are facing an increase in data availability that out-paces human ability to review. Consequently, the IAEA is currently considering the application of DCNNs to support initial processing of several of its image-based data streams including open source information analysis, satellite imagery analysis, and analysis of surveillance camera data.

While the IAEA has a rich historical dataset from decades of safeguards implementation activities, they face training data challenges similar to many global security domains:

1) Relevant events are low frequency and high consequence. There have been historically few examples of nuclear proliferation, and we cannot presume that future cases of nuclear proliferation will follow similar patterns or use identical equipment to past examples.
2) Content of relevant images is sensitive. Some fuel cycle technologies or related equipment can be commercially proprietary, so there are limited photographic examples available in open sources.
3) Adversaries and proliferators seek to obscure their activities. Most nuclear proliferators try to obscure their activities until they reach a point where they determine that it is within their interest to expose their capabilities. In some historical cases, this was achieved only at the time of a nuclear test, or via an announcement once a program was dismantled.
4) Labelling data is expensive. Even in cases when sufficient imagery data is available to train DCNNs, the required human labelling of the data is time consuming, expensive, and prone to human error or disagreement among experts.

For these reasons, we are exploring the use of synthetic images to train computer vision models. Our synthetic images are computer-generated two-dimensional images that are rendered from three-dimensional CAD models of an object of interest. The images can be rendered with many different settings such as material type and lighting conditions. We are investigating how synthetic images can be used to reduce the number of real-world images required to train computer vision models. There has been recent research in this field, focusing on the level of realism needed in the synthetic images (Tremblay, et al., 2018), the integration of large real-world data set with synthetic data (Ekbatani, Pujol, & Segui, 2017) or the use of synthetic images for classes which approximate common objects already in pre-trained models such as tomatoes (Rahnemoonfar & Sheppard, 2017).

2. SYNTHETIC IMAGE DEVELOPMENT

In this work, we developed several synthetic image datasets to support our experimentation and collected a dataset of real-world imagery for testing. Our datasets were designed around two 3D computer models of remote manipulator arms: one using a visually simple design consisting of a single tube which contained most instrumentation, with a single pivot point attachment to the hot cell; and one using a mode advanced design of manipulator arm based on that developed by Central Research Laboratories (CRL).[1] See Gastelum, Shead & Higgins (2020) for examples of real-world an synthetic images of the single-tube manipulator design, and Figure 1 for an example synthetic image using the CRL design.

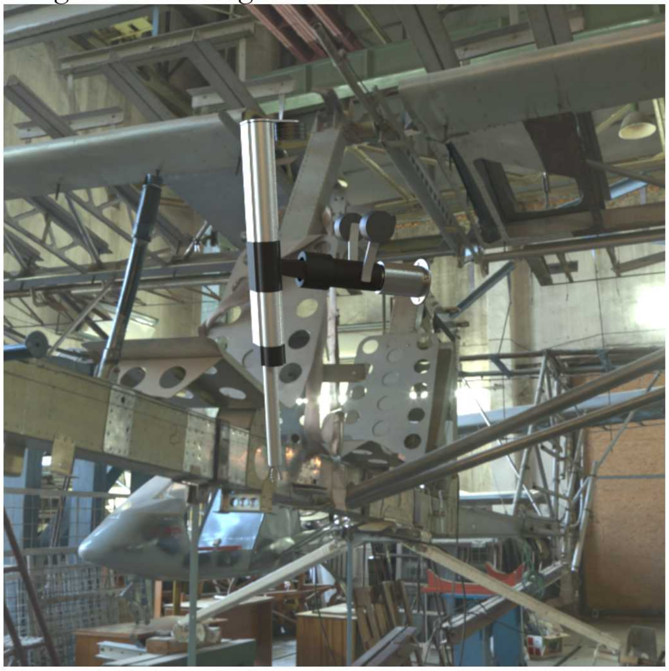


Figure 1 2D Rendering of Synthetic CRL-type Manipulator Arm

Our INMM paper (Gastelum, Shead, & Higgins, 2020) provides a detailed description of:

- Our development environment and capabilities of the SideFX Houdini software[2]
- The development of our synthetic imagery dataset of the single-tube manipulator arm design, including:
    - The single-tube design with backgrounds from 33 high-dynamic-range (HDR) panorama images
    - The single-tube design with backgrounds from our 191 2D distractor images
- Design decisions regarding background selection for our synthetic images, specifically as it pertains to avoiding the iconic yellow glow from hot cells.

We also collected 220 images of remote manipulator arms from open sources, and 191 distractor images of related environments such as spent fuel pools, lab space, and industrial equipment. A subset of the images was graciously provided by colleagues at Lawrence Livermore National Laboratory working on a nonproliferation and deep learning strategic initiative. [3]

[1] https://crlsolutions.com/products/telemanipulators/

[2] https://www.sidefx.com/products/houdini/

In Table 1, we provide an overview of our final experimental datasets used in this research, not including our 411 real-world images described above. Each dataset contains 1000 images with targets, and 1000 distractors.

Table 1 Synthetic Datasets Rendered for this Research

Dataset Name	Manipulator Design	Unique Qualities
Single Tube HDR Background	Single tube	Our original single tube dataset, using random selections from 33 3D HDR images as background and distractors
Single Tube 2D Background	Single tube	Uses the same 1000 target manipulators used in Single Tube HDR Background, but places them on a random selection of real-world distractor images
Single Tube Black Background	Single tube	Uses the same 1000 target manipulators used in Single Tube HDR Background, but placed on a fully black background
Single Tube Neutral Background	Single tube	Uses the same 1000 target manipulators used in Single Tube HDR Background, but placed on a background that represents the mean color from the ImageNet dataset
CRL Design HDR Background	CRL	Our newest manipulator design dataset based on the CRL manipulator design, rendered on random selections from 33 3D HDR images as background and distractors

[3] https://www.llnl.gov/news/researchers-developing-deep-learning-system-advance-nuclear-nonproliferation-analysis

3. EXPERIMENTAL PLAN

To determine how to reduce the number of real-world images needed to train deep learning computer vision models, we planned three experimental approaches: synthetic image parameter variance, real-world to synthetic image ratios, and machine learning explainability to identify biases. Each experimental focus is explained below. Results from these experiments are described in Sections 4 and 5. All experiments where the training and test data were drawn from a single dataset (such as experiments where we trained and tested on real-world data) used 5x2 cross validation. Since normal cross validation was meaningless where the training and test data were drawn from separate datasets (such as experiments where we trained on synthetic data and tested on real-world data), in those cases we repeated the training process ten times with training and validation data randomly drawn from the test set, and the results were averaged.

3.1. Parameter Variance Experiments

Our first set of experiments examined parameter variance. Parameter variance refers to the manipulations we created in the 2D images rendered from our 3D Houdini model. Our 3D model allowed us to manipulate background, size, position, lighting, and camera features in our 2D renderings. For our parameter variance experiments we compared results using a variety of image backgrounds, including panoramic HDR images, 2D industrial and laboratory environments from our test data, black backgrounds, backgrounds based on the mean color of the entire ImageNet dataset, and frequency-limited noise patterns (Figure 2).

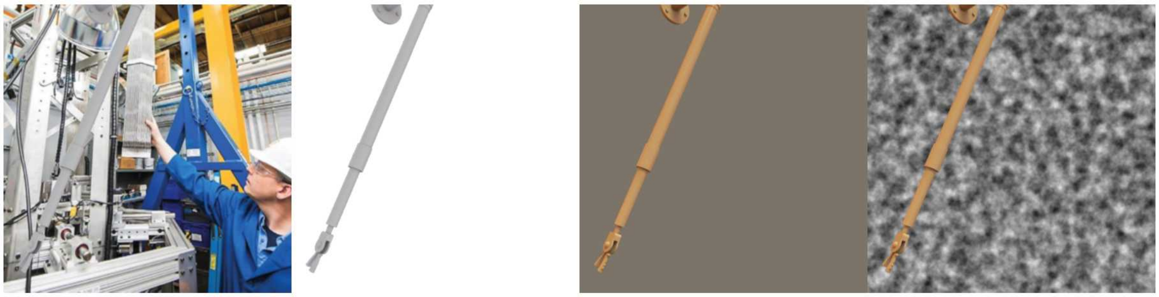


Figure 2 Synthetic remote manipulator arm against real-world industrial, black, ImageNet mean color, and noise backgrounds. Industrial background image credit: TerraPower.

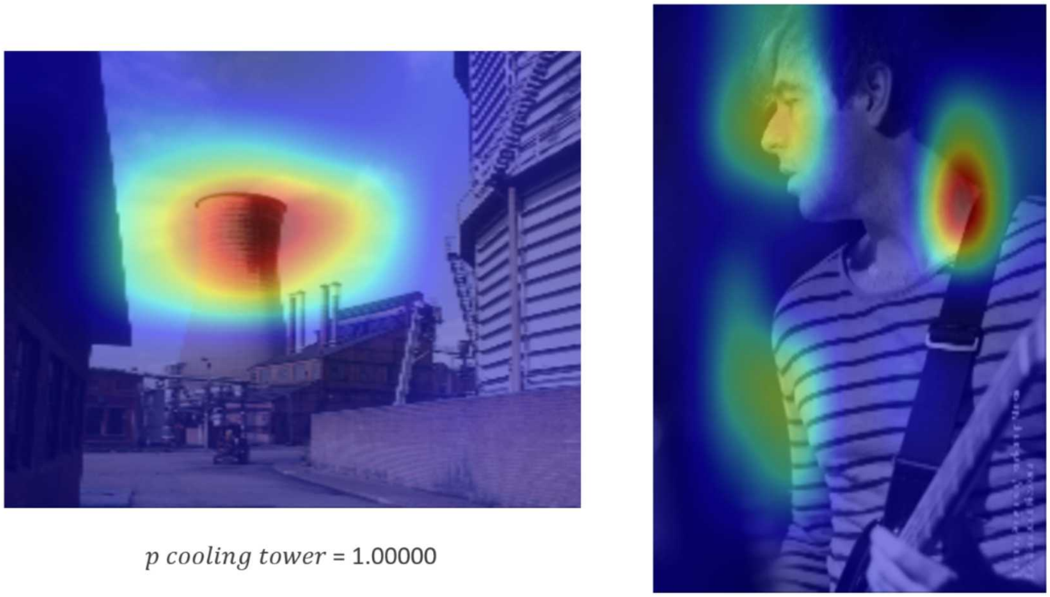
3.2. Image Ratio Experiments

Our second set of experiments focused on the ratio of real-world to synthetic images, with our research goal being to reduce the number of real-world images needed to the smallest possible number without sacrificing accuracy. We collected approximately 200 images from open sources, including the Lawrence Livermore dataset previously mentioned. Reserving 40 of those images for test, we trained a model with 160 real-world images using both transfer learning techniques and training from scratch. While 160 images represent a very small dataset for the deep learning community, it could be a significant image set for the international safeguards world. This model served as our baseline, to which we compared results of any model that incorporated synthetic training images. Our metric for success was that a model trained with some combination of real-world and synthetic data had to perform better than a model trained on real-world data alone. Our training sets for the image ratio experiments included:

- *All real-world images.* This model was trained on 160 real-world images and tested on 40 real-world images that were kept apart from the training set. This is the baseline to which all other model results would be compared.
- *Mixed data sets.* These datasets used all our synthetic data (because in data science, more data is always better!) and reducing numbers of real-world images. We halved the number of real-world data in a series of experiments, starting with 80 real-world images, then 40, then 20, then 10. Then we trained with just one real-world image along with the full synthetic image training set and tested against the 40 real-world images that were kept aside.
- *All synthetic images.* This model was trained on all 1000 of our synthetic images and tested on the 40 real-world test images. This model represents our most optimistic scenario, in which a model could be trained to recognize real-world manipulator arms without ever having "seen" one during training and validation.

3.3. Machine Learning Explainability Experiments

Our third set of experiments focused on using machine learning explainability techniques to visualize potential biases in our image data, then render new 2D images to counter those biases and re-train the models in an iterative approach to reducing incorrect classification and localizations in our models. Machine learning explainability techniques come in many forms, but our experiments used the Class Activation Mapping (CAM (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016)) algorithm that highlights pixels in an image that are highly correlated with the classification of an object. In a prior project that was developing a classification algorithm for nuclear cooling towers, we used CAM to identify what aspects of an image of a guitar player were causing incorrect classification of the image, shown in Figure 3. If we had sought to correct for this apparent bias in our model, we could have included more images of human necks approximating the curve of a hyperboloid cooling tower. For this work on manipulator arms, we intended to use a similar technique to identify these biases. Our intent was to use the findings from the explainability visualizations to identify biases in our training data, render new 2D images from our 3D model to counter those biases, and re-train our models with the updated data to improve performance.



$p\ cooling\ tower = 1.00000$

$p\ cooling\ tower = 0.96455$

Figure 3 CAM visualizations of two images classified as cooling towers. The image on the left is a correct classification, showing that the model has "learned" to look for the hyperboloid curve to classify a cooling tower. The image on the right s an incorrect classification, indicating that the curve of the guitar player’s neck and strap were incorrectly used to classify the image as a cooling tower. (Image credit: (L) Ken Davis via Flickr, November 23, 2009. https://www.flickr.com/photos/ken_davis_archive/5351748461/in/photostream/; (R) johnny anguish via Flickr, April 1, 2012. https://www.flickr.com/photos/johnnyanguish/7037904109/in/photostream/)

4. OBJECT DETECTION

We implemented two object detection models in the course of our research. The first object detection model was Detectron (Girschick, Radosavovic, Gkioxari, & Doll, 2018) which was developed by researchers at Facebook. Detectron is a multi-model architecture that includes object detection, background detection, and various types of image segmentation and masking. We conducted parameter variance and image ratio experiments using Detectron but saw little variance in results (see Results section below), and therefore switched to a simpler object detection model, You Only Look Once v3 (YOLO, (Redmon & Farhadi, 2018).

YOLO is an object detection model that is known for its quick processing times based on what the developers term a "unified architecture". It is pre-trained on the ImageNet dataset.
As mentioned above, we encountered significant difficulty in getting our object detection models to perform. The object detection models either failed to train, trained for long periods without ever reaching a turning point to over-training, or trained successfully but did not yield results that can be meaningfully differentiated.

Due to the unpredictable performance of the models, we are not highly confident in the results below. However, for the purpose of transparency and the documentation of challenges that we hope will lead to others starting from where we left off in this research, we tentatively and with significant caveats share our initial results below.

4.1. Parameter Variance in Object Detection Results

We first performed two parameter variance experiments: one comparing backgrounds to no-backgrounds, and one comparing existing backgrounds to an extended set of backgrounds that were intended to further train the models to recognize distractor objects. The experiments failed to yield distinguishable results.

4.2. Image Ratio for Object Detection Results

We observed an early trend that models trained solely on real-world images performed better than our models trained with a mix of real-world and synthetic, or just on synthetic, images. However, these results were not consistent among training trials and multiple failures to complete model training ultimately led to our abandoning object detection approaches.

4.3. Machine Learning Explainability for Object Detection Results

Anecdotally, in our early object detection results we noticed a significant number of misclassifications of people as remote manipulator arms. We posited that this could have been due to the angle at which their arms and legs were positioned in the images, but also realized without the use of machine learning explainability that we did not have any images of people in our synthetic data. However, due to the poor performance and erratic performance of the object detection models on prior experiments, we opted to forego the explainability experiments for object detection. Instead, we pursued our full suite of experiments using image classification models, which was not part of our original scope.

5. IMAGE CLASSIFICATION

In this research, we fine-tuned the ResNet-50 image classification model (He, Zhang, Ren, & Sun, 2016), pretrained on the ImageNet dataset, provided by the torchvision.models package included with PyTorch.[4]
we replaced the final 1000-output fully connected layer in each model, substituting a one-output fully connected layer with sigmoid activation function, which would yield a score between zero and one for a single class, manipulator. We used the value 0 for our ground truth for images without manipulators, and 1 for our ground truth of images with manipulators.

We conducted several baseline experiments to determine:

1) If we had sufficient real-world data and synthetic data to train deep-learning modes
2) If our models were architecturally sound (i.e., working)
3) Identify a baseline for later results

Our baseline experiments and results are provided in our INMM paper (Gastelum, Shead, & Higgins, 2020). While each model trained-and-tested on a single data type yielded reasonable performance, the model trained exclusively on synthetic data and tested on real data yielded results only slightly better than random performance, with a test accuracy around 55%. Loss while training indicated that the synthetic data may provide some very early benefit in training, but the model quickly becomes overfit.

5.1. Parameter Variance in Image Classification Results

5.1.1. Background Image Variance

Some of our parameter variance experiments focused on the background image variance. In one of our baseline experiments, we trained on a dataset which used random samples from 33 industrial looking HDR panoramic images as the background. We hypothesized that the low performance of the model when tested on real-world data could be attributed to low variance in background. We generated a new synthetic dataset of 1000 images in which the synthetic manipulators were placed atop 2D images that we had collected as our distractor" or "negative examples" – these were 191 images that were industrial, and sometimes nuclear, but did not include manipulator arms. Performance of this model was even lower than with the HDR backgrounds (approximately 50%, or random chance), with the loss curve never decreasing. Additional details are provided in our INMM paper (Gastelum, Shead, & Higgins, 2020).

We then opted to run the same experiment, but rather than using real-world backgrounds as in the prior experiment, or black backgrounds as we did in the object detection experiment, we used a shade of beige that represents the algebraic mean of all images from the ImageNet dataset. In this experiment, we again observed the trend in which the loss function shows an early decline, but quickly seems to be overfitting to the data. We found that the average accuracy using the neutral background resulted in even worse performance than our previous background variance experiments. Results are shown in Figure 4.

[4] https://pytorch.org/docs/stable/torchvision/models.html#classification

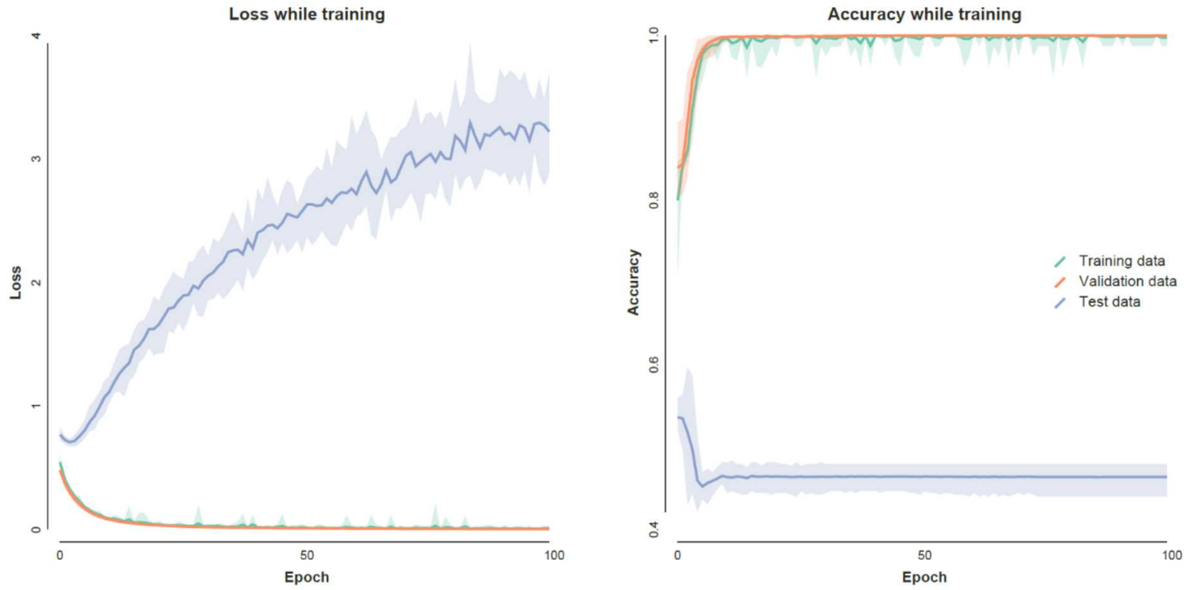


Figure 4 Loss and Accuracy for Neutral Background Experiments

5.1.2. Modelled Object Variance

The results of our background experiments indicated that we may need to have more realism in our images than we originally thought. We decided to test each of our manipulator model arms against only the real-world data that corresponded to that manipulator model, rather than the full dataset which represented approximately eight distinct models.

For the single tube design, we used the model trained exclusively on our single tube design synthetic dataset and tested it against only those real-world images that depicted single tube design manipulator arms. The accuracy went up from approximately 55% when tested on the full dataset, to 60% when using just the single tube images for test. As with previous experiments, this could indicate that realism plays a larger factor in image performance than we originally thought. See Figure 5.

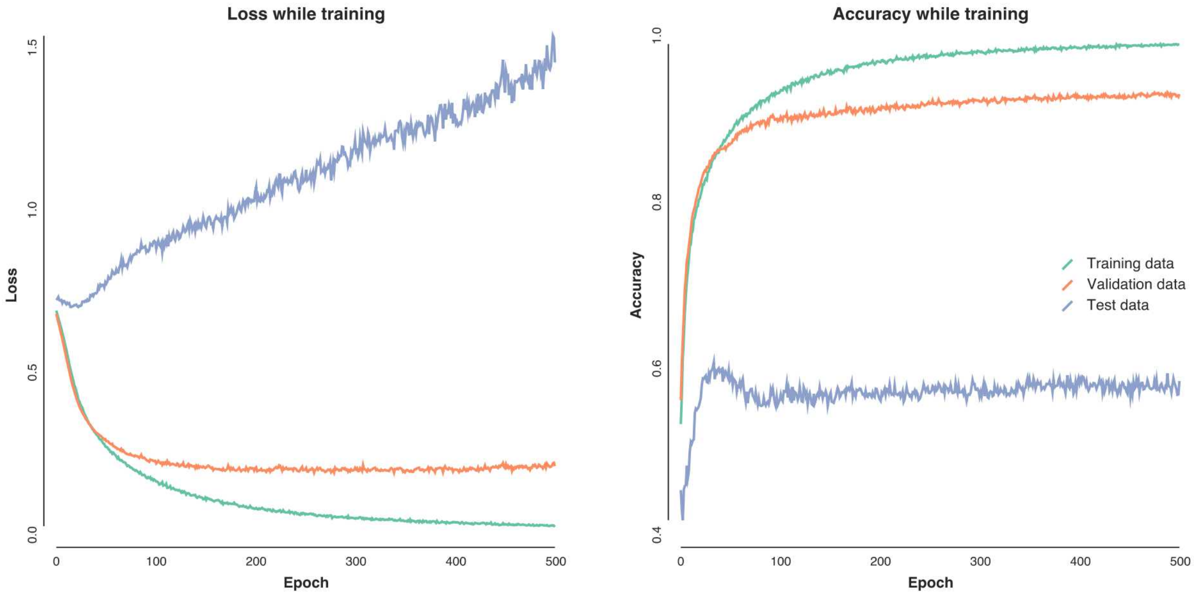


Figure 5 Loss and Accuracy for Single Tube Manipulators

For the CRL design, we used our model that was trained exclusively on our CRL design synthetic images and tested it only against those real-world images that depicted the CRL design manipulators. We observed that our test accuracy was as high as 70% for this experiment, perhaps due to the visually distinct features of the CRL design arms, and the limited number of real-world testing images that were available. Results are displayed in Figure 6.

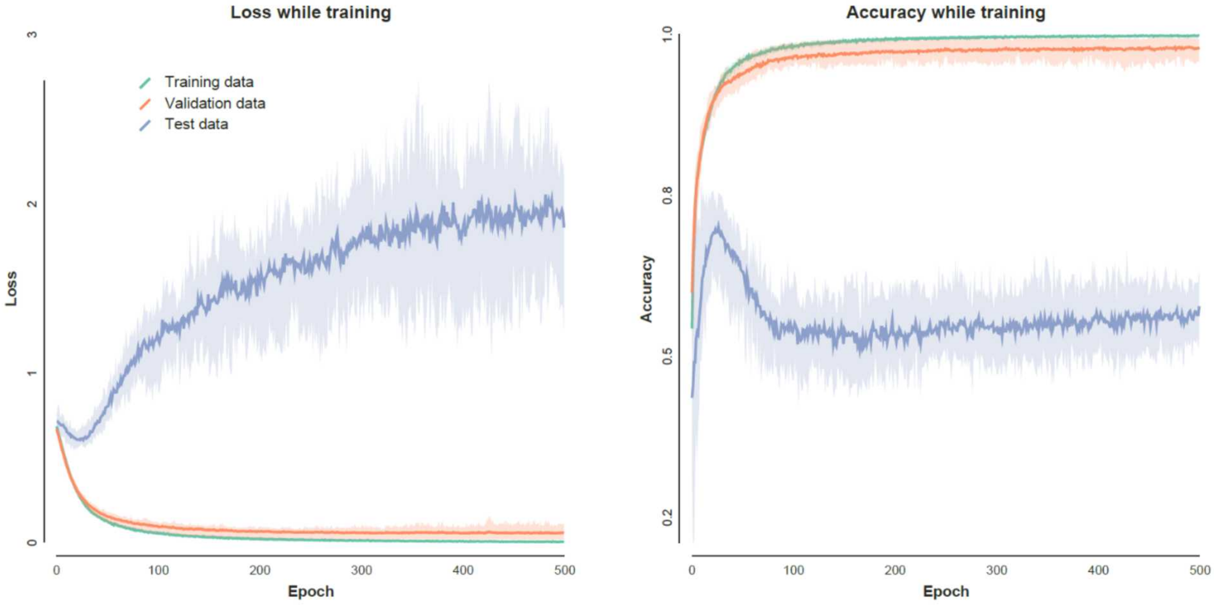


Figure 6 Loss and Accuracy for CRL Design Models

5.2. Image Ratio for Image Classification Results

We documented the results of our image ratio experiments in our INMM paper. In summary, we found that in using small numbers of real-world images (12, 51, and 205 real-world images, respectively), increasing the number of synthetic images used to train models did not impact average model performance but did result in a significant decrease of variance in model performance. That is, though the accuracy was approximately the same, the uncertainty in the results was smaller. Thus, the synthetic images do provide benefit, though not directly observable through accuracy.

5.3. Model Thresholds

All the image classification experiments above used 0.5 as a threshold sigmoid value at which to classify an image as having a remote manipulator arm. This value is a standard starting point in image classification algorithms but can be adjusted depending on the specific use case and requirements.

To understand our model performance as a function of threshold, we mapped model accuracy, precision, recall, and F1 scores across the full range of decision thresholds.[5] The performance map is shown in Figure 7. We can observe that as the decision threshold moves towards 1, accuracy of model increases. However, recall is a more significant factor for international safeguards, as missing an item of interest could have high consequences and analysts are accustomed to dismissing high numbers of false positives in the process of trying to identify relevant events.

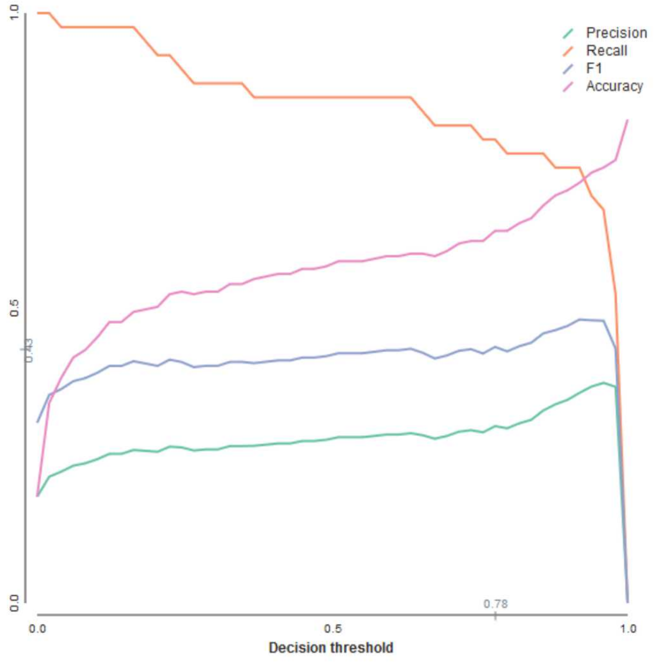


Figure 7 Performance Metrics across Decision Thresholds

[5] Accuracy is defined as the total number of correct responses, divided by the total number of potential responses. Precision refers to the ratio of selected items that are relevant and is calculated by dividing the number of true positives by the sum of true positives and false positives. Note that precision does not account for false negatives, or "misses" from the model. Recall refers to the ratio of relevant items that are selected and is calculated by diving the true positives by the sum of true positives and false negatives. Note that recall does not account for false positives, sometimes referred to as "false alarms." The F1 score is the harmonic mean of precision and recall scores.

Keeping recall as a priority, we can observe that from approximately 0.05 to 0.8, recall decreases considerably without a corresponding gain in precision. That is to say, the false positive rate of the model stays relatively constant while the false negative (i.e. misses) rate increases. As such, though overall model accuracy would be considered lower, a very low threshold between 0.05 and 0.1 could minimize the potential model misses (through maximizing on recall score) without significantly increasing the irrelevant materials that would be presented to an analyst for review.

5.4. Machine Learning Explainability for Image Classification Results

Our final experimental area was in machine learning explainability. We had intended to use explainability results to help identify biases in the synthetic data that could be negatively impacting our results. Then, we would correct for the biases in the 3D computer model of the manipulators, re-render images, and re-train the models in an iterative fashion.

Due to time limitations, we were unable to complete these experiments.

6. DISCUSSION

In this work, we found that the use of synthetic data does provide value in instances where there is low – or no – real world data. We found that for instances in which there are few real-world images, synthetic data can be useful in decreasing variance in model performance which could lead to higher confidence in model results by users. We found that when no real-world data is available, models performed better when tested against datasets that focused on the specific manipulator design on which they were trained. This could mean that our 3D computer models need more realism, and that additional 3D models need to be built to account for the wide variety present in the market of our selected technology. We also found that, for the CRL design model, very low thresholds led to higher degrees of detection of manipulator arms within the broad dataset without significantly increasing the number of false alarms that an analyst would have to review. How this finding will propagate across our other experiments remains an open question. Below, we will describe a few observations and recommendations for future work based on our findings and for the application of this work to international nuclear safeguards.

6.1. Data Proportions

Most safeguards-relevant information available in open sources is rare, especially in comparison to the volume of irrelevant information. Thus, image classification or object detection models deployed for open source information analysis would be looking for rare events.

In this work, our training approach followed a balanced approach, with approximately even numbers of positive and negative examples of manipulators used to train the model, and for testing. The exception was for the experiments in which we tested models only against the design of manipulator they represented and distractors, in which case there were 2-4 times the number of distractors. Even in the exception experiments, the test data was significantly more balanced than a potential stream of open source information for a model to classify. In future work, we would like to explore the idea ratio of positive-to-negative examples for training models that will be used to find rare events.

6.2. Larger Training Data Sets

For this research, we consistently used 1,000 images each of synthetic manipulators and distractors. In future research, we would like to explore the impact of increasing the size of our training data to determine its impacts on model performance, and hopefully decrease the impact of early overfitting.

6.3. Other object detection models

Our object detection model experiments were unsuccessful. Since this work began, new and refined object detection models have become available, including an update of the YOLO algorithm that we used in this research. We look forward to an opportunity to re-run some of our initial object detection models using these updated algorithms.

7. FUTURE RESEARCH

Significant research remains to be done on the use of synthetic images to train object detection and image classification algorithms in general, and for international safeguards specifically. We have received funding for a follow-on research project from the U.S. National Nuclear Security Administration's Defense Nuclear Nonproliferation Office to produce additional synthetic imagery and validate its ability to generalize across multiple classes of computer vision models, as well as between algorithms in a single class. As part of this follow-on research, we also intend to explore the use of synthetic data to counter the potential effects from adversarial machine learning examples.

8. REFERENCES

Ekbatani, H., Pujol, O., & Segui, S. (2017). Synthetic Data Generation for Deep Learning in Counting Pedestrians. *ICPRAM*, (pp. 318-323). doi:10.5220/0006119203180323

Gastelum, Z. N., Shead, T., & Higgins, M. (2020). Synthetic Training Images for Real-World Object Detection. *Proceedings of the Institute of Nuclear Materials Management.* Virtual.

Girschick, R., Radosavovic, I., Gkioxari, G., & Doll, P. (2018). *Detectron*. Retrieved from Githib: https://github.com/facebookresearch/Detectron

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer VIsion and Pattern Recognition*, (pp. 770-778). Retrieved from https://arxiv.org/abs/1512.03385

International Atomic Energy Agency. (n.d.). *Basics of IAEA Safeguards*. Retrieved September 2020, from IAEA: https://www.iaea.org/topics/basics-of-iaea-safeguards

Rahnemoonfar, M., & Sheppard, C. (2017). Deep Count: Fruit Counting Based on Deep Simulated Learning. *Sensors, 14*(4). doi:https://doi.org/10.3390/s17040905

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement.* Retrieved from https://arxiv.org/abs/1804.02767

Tremblay, J., Prakesh, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., . . . Birchfield, S. (2018). Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. *Proceedings of the CVPR 2018 Workshop on Autonomous Driving.* Retrieved from arXiv:1804.06516

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Conference on Computer Vision and Pattern Recognition.* Retrieved from http://cnnlocalization.csail.mit.edu/Zhou_Learning_Deep_Features_CVPR_2016_paper.pdf

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Dianna Blair	06800	dsblair@sandia.gov
Jonathan Salton	06833	jsalton@sandia.gov
Ladonna Wilk Martin	01971	lwmarti@sandia.gov
Technical Library	9536	libref@sandia.gov

This page left blank

This page left blank