Leveraging Synthetic Imagery
for Collision-at-sea Avoidance

Chris M. Ward
Josh Harguess
Alexander G. Corelli

SSC Pacific
San Diego, California 92152-5001

M. K. Yokoyama, CAPT, USN
Commanding Officer

W. R. Bonwit
Executive Director

ADMINISTRATIVE INFORMATION

Released by
Matthew C. Bowes, Head
Advanced Analysis Systems Branch

Under authority of
Craig A. Wilgenbusch, Head
ISR Division

RJP

EXECUTIVE SUMMARY

Maritime collisions involving multiple ships are considered rare, but in 2017 several United States Navy vessels were involved in fatal at-sea collisions that resulted in the death of seventeen American Service members. The experimentation introduced in this report is a direct response to these incidents. To prevent collisions in the future we propose a shipboard Collision-At-Sea avoidance system, based on video image processing, that will ensure the safe stationing and navigation of maritime vessels. Our system leverages a convolutional neural network trained on synthetic maritime imagery in order to detect nearby vessels within a scene, perform heading analysis of detected vessels, and provide an alert in the presence of an inbound vessel. Additionally, we present the Navigational Hazards - Synthetic (NAVHAZ-Synthetic) dataset. This dataset, is comprised of one million annotated images of ten vessel classes observed from virtual vessel-mounted cameras, as well as a human "Topside Lookout" perspective. NAVHAZ-Synthetic includes imagery displaying varying sea-states, lighting conditions, and optical degradations such as fog, sea-spray, and salt-accumulation. In this report we present our results on the use of synthetic imagery in a computer vision based collision-at-sea warning system with promising performance.

CONTENTS

Figures

Tables

1. INTRODUCTION

Availability of well-annotated data is a critical component in the successful application of modern machine learning techniques to a problem domain. In this report our team leverages openly available tools to generate synthetic data that can be used to train models in problem domains with sparse data. Additionally, we focus on the problem of maritime-traffic analysis as an experimentation vehicle.

This report is broken down into five sections. Section 2 provides a brief background on the work and utility of using synthetically generated imagery in machine learning applications. Section 3 introduces the synthetic dataset we created for this study, and Section 4 details the use of our synthetic data applied to the problem of martime collision-avoidance. Section 5, summarizes the results of the study and proposes future work solutions.

2. PREVIOUS WORK

Synthetic imagery has been used before across a wide variety of problem spaces as a way to get past the limitations of having small real-world annotated datasets. As an example, the Virtual KITTI [1] and Synthia [2] datasets have demonstrated that for vehicular-focused imagery systems, incorporation of synthetic data outperforms using only real-world data [1–7]. It has also been shown that using only synthetic data yields results on par with using a limited set of real-world training data [3, 4, 6]. Additionally, in a study conducted by G Rose [2] it is demonstrated that the combination of synthetic data with real-world data outperforms using either set exclusively. A study conducted by de Souza [7] further emphasizes this claim, stating that, at least for their datasets, increasing the amount of the real-world data to match that of the synthetic outperforms other models. The best results came from training on a mix of real-world and synthetic data, and then fine-tuning with examples from the target dataset [2, 6]. Fine-tuning using the target data helps the model account for the realities of the given domain, such as image noise, that might not be fully accounted for in the synthetic data [8].

3. NAVIGATIONAL HAZARDS - SYNTHETIC (NAVHAZ-SYNTHETIC) DATASET

In this section we introduce the Navigational Hazards - Synthetic (NAVHAZ-Synthetic) dataset, and describe our methodology for creating it. This dataset is comprised of one-million labeled images of ten generic vessel classes, observed from ten shipboard points. In order to capture the high level of scene variance in maritime imagery, NAVHAZ-Synthetic contains images with diverse sea-states, weather conditions, lighting conditions, sensor noise, and optical degradations including sea-spray, and salt-accumulation. Each image contains a single target vessel and is labeled with the class and heading.

$$\left|\text{Classes}\right| * \left|\text{Vessel Headings}\right| * \left|\text{Sun Positions}\right| * \left|\text{Sky Conditions}\right| * \left|\text{Sea State}\right| * \left|\text{Observers}\right| = 1 x 10^6 \text{ images}$$

3.1 UNITY GAMING ENGINE

Based on work in using video games a basis for training machine learning models [6], our team selected the Unity game engine as our modeling environment. Unity provides several advantages over proprietary modeling solutions: it is free, it is has a strong online support community, and it supports a wide variety of model formats and platforms. Running on an Intel i7 notebook, we used the Unity engine to generate the NAVHAZ-Synthetic dataset at a rate of 100,000 images/hour.

3.2 CLASSES

NAVHAZ-Synthetic is comprised of ten commercial vessel classes: Liquid natural gas (LNG) Tanker, Oil Tanker, Container Ship 1 (Ballast Condition), Container Ship 1 (Full Load), Container Ship 2 (Ballast Condition), Container Ship 2 (Full Load), Cargo Ship 1, Cargo Ship 2, Cargo Ship 3, and Barge. While we did not perform any class recognition in our experimentation, each image in NAVHAZ-Synthetic is labeled with its respective vessel class.

3.3 LIGHT SOURCE

Lighting conditions were varied over four hemispherical paths, and five positions along each path, as depicted in Figure 1. We selected light-source positions that represent four daytime intervals: dawn, morning, noon, afternoon, and dusk. Additionally, we programmatically vary sky coloration to simulate light scattering in dawn and dusk conditions.
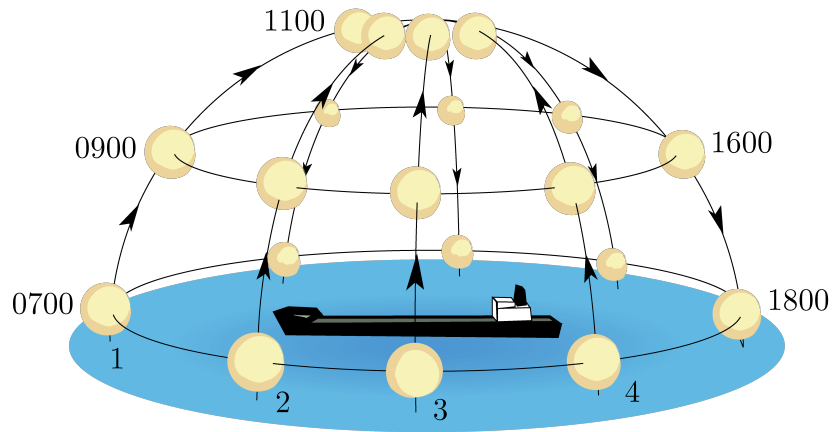

Figure 1. Scene lighting positions.

3.4 VESSEL HEADINGS

Each vessel class was rendered at twenty different headings as depicted in Figure 2:

$$(0°, \pm5°, \pm15°, \pm30°, \pm45°, \pm60°, \pm90°, \pm120°, \pm135°, \pm160°, 180°)$$

Each image is labeled with the heading of the target vessel. In our experimentation we grouped the vessel headings into specific zones. These groupings are further explained in Section 4.
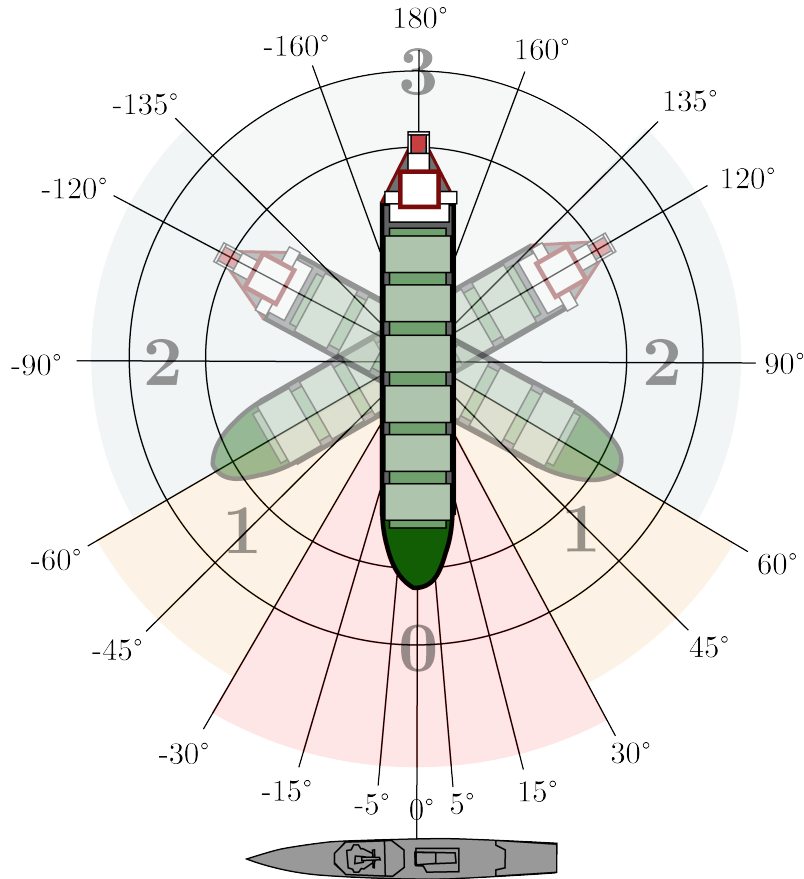
Figure 2. Target vessel heading angles.

3.5 WEATHER CONDITIONS AND SEA STATE

Environmental conditions at sea are highly variant. Leveraging the power of synthetic imagery, we attempted to capture this variance by training on a different combination of weather and sea conditions. NAVHAZ-Synthetic features five weather conditions: clear skies, dynamic clouds, dense dynamic clouds, overcast sky, and dense fog layer. Additionally, we selected five sea states (shown in Figure 3) that correlate to conditions described by the World Meteorological Organization Sea State codes [9]. Combined, these weather and sea state variations yield twenty-five environmental conditions across the training data.

3.5.1 Observers

In order to generate sufficient observational diversity, we virtualized ten observers in the modeling environment. It is common to have a human topside lookout in maritime evolutions, thus we included such an observer in our Unity environment. A topside lookout was modeled with a field-of-view that resembles the perspective a human observer with binoculars. The remaining nine observers represent a ship-mounted camera system augmented with degradations and effects inherent in real-world maritime conditions. The camera observers were modeled using common parameters from commercially available maritime cameras and placed at various points on ownship as depicted in Figure 4. We visualize the synthetic camera feeds in Figure 6 and show the effects and degradations associated with each observer in Table 1.

(a) Sea State: **2**
Wave height: 0.1m-0.5m
Smooth

(b) Sea State: **3**
Wave height:
0.5m-1.25m
Slight

(c) Sea State: **4**
Wave height:
1.25m-2.5m
Moderate

(d) Sea State: **5**
Wave height: 2.5m-4mm
Rough

(e) Sea State: **6**
Wave height: 4m-6m
Very rough

Figure 3. Sea states.

Forward Cameras

Mast Cameras
Topside Lookout

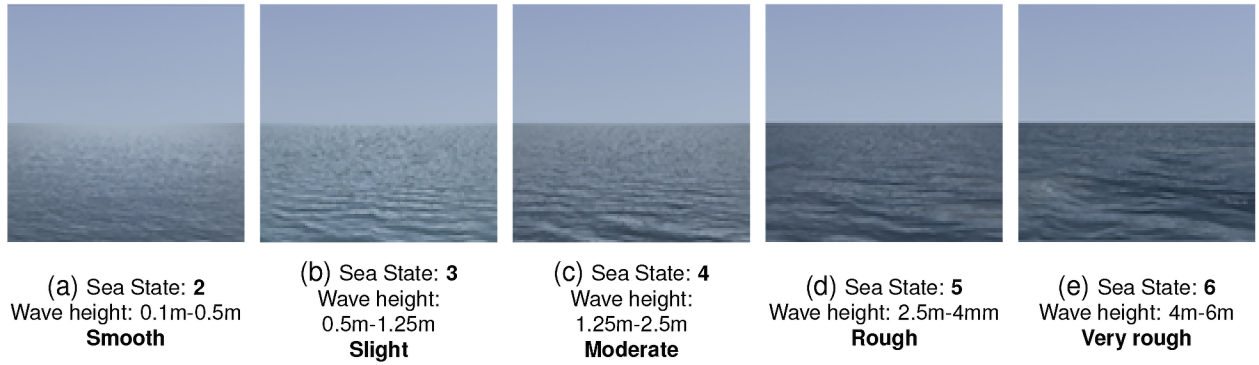Aft Cameras

Mast Cameras

Topside Lookout
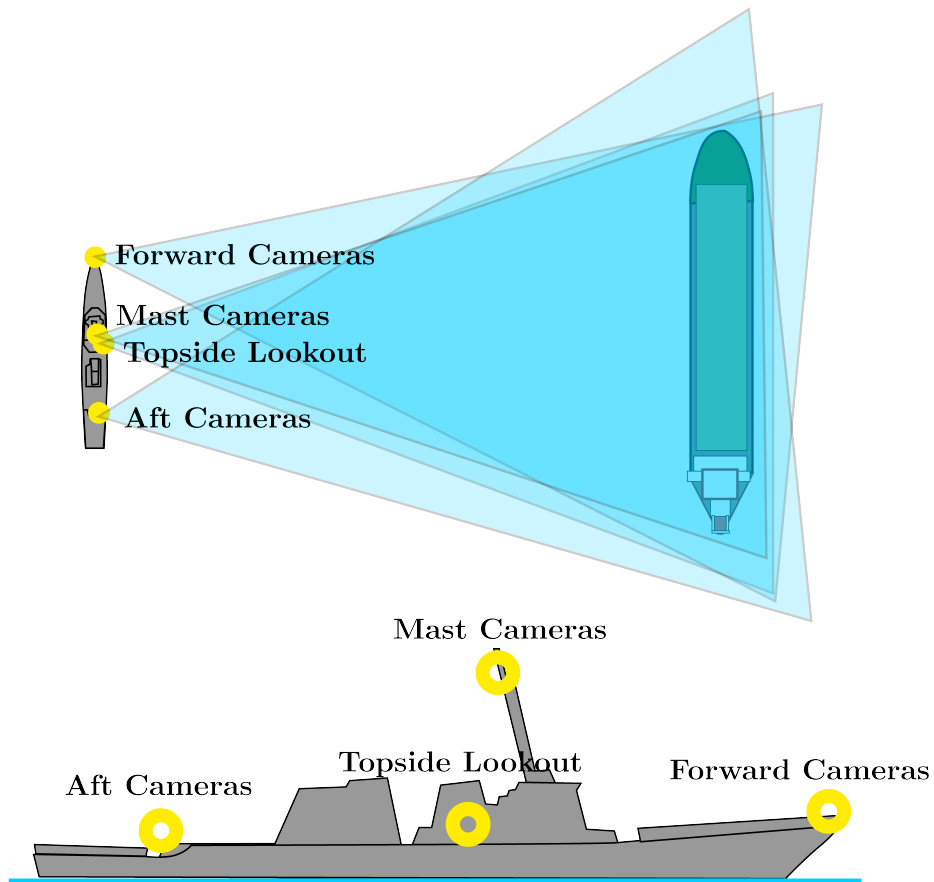
Aft Cameras

Forward Cameras

Figure 4. Observer Positions

Table 1. Observer Parameters

Observer Name	Field of View	Focal Length	Chromatic Aberration	Lens Flare	Lens Bloom	Lens Grit	Sensor Grain
Topside Lookout	31.0	54.7	-	-	-	-	-
Forward Camera 1	33.2	58.7	✓	-	-	-	✓
Forward Camera 2	33.2	58.7	✓	-	✓	✓	✓
Forward Camera 3	33.2	58.7	✓	✓	✓	✓✓	✓✓
Mast Camera 1	22.0	58.7	✓	-	-	-	✓
Mast Camera 2	22.0	58.7	✓	✓	✓	✓	✓
Mast Camera 3	22.0	58.7	✓	✓	✓	✓✓	✓✓
Aft Camera 1	53.1	35	✓	-	-	-	✓
Aft Camera 2	53.1	35	✓	✓	✓	✓	✓
Aft Camera 3	53.1	35	✓	✓	✓	✓✓	✓✓

4. EXPERIMENTATION

This section details our experimentation. We create a very large synthetic dataset (1-million images) and use it to train a deep neural network for a classification task: determining the heading of a commercial vessel. We evaluate our ability to train a large capacity model (VGG-19) [10] as well as the performance of the trained model on a small dataset of real imagery.

4.1 DATASETS

This section gives an overview of the image datasets used in our experimentation.

4.1.1 Training and validation data

In addition to using ImageNet weights [10, 11] for transfer learning, our training data consists of an 800,000 image class-balanced subset of NAVHAZ-Synthetic. We also curated a small subset (250 images) of commercial ships from the ImageNet dataset to use for domain adaptation. We used a 20% validation holdout in both training sets. We did not perform any training data augmentation in this experiment.

4.1.2 Test data

Our test data consists of a second small ImageNet subset (200 images) of commercial ships, balanced over four classes. Because it is difficult to infer the exact heading of the vessels in the ImageNet data, we apply weak labels, (general heading of the target vessel) to the test images. Our labeling schema is further described in 4.3.

4.2 ARCHITECTURE SPECIFICATION

For this experiment, we used the VGG-19 convolutional neural network described in Simonyan and Zisserman[10]. The input layer was modified to accommodate the NAVHAZ-Synthetic dimensions of 384x384 pixels. Following the VGG-19 stack of convolutional layers, the top of our model consists of three fully-connected layers: two with 4096 channels, followed by a 4-channel fully connected layer that corresponds to our four output classes. The model is finalized with a softmax output layer.

4.3 EXPERIMENT DESIGN

The problem of detection was bypassed in this experiment, however, maritime detection is a non-trivial problem. To alleviate detection-performance requirements for integration of our work, we train on images without tightly cropped bounding boxes, and targets generally centered in the frame.

Although NAVHAZ-Synthetic provides exact headings of target vessels, we weakened the labels to facilitate shaping this experiment as a classification problem, versus a more challenging regression problem. Weakening the labels also alleviates the challenge of labeling the ImageNet data with precise vessel headings. The data was grouped into four classes: Zone 0 (imminent collision), Zone 1 (collision danger), Zone 2 (safe maneuvering), and Zone 3 (no danger). These zones are visualized in Figure 2.

We trained VGG-19 in three discrete steps. First, we initialized all but the fully-connected layers of the model with ImageNet weights. We initialized the fully-connected layers with small random values chosen from a zero-mean Gaussian [12]. To establish a basis for performance comparison, we then attempted to fine-tune the model on our small ImageNet training subset of commercial ships.

In the second part of our experiment, we re-ininitialized the model with ImageNet weights and fine-tuned the network on our NAVHAZ-Synthetic training set. We used early stopping to achieve maximum validation accuracy without overfitting. We then tested our model on the ImageNet test set.

In the third part of our experimentation we performed domain adaptation by fine-tuning the model again on the ImageNet training set. We then retested our model on the ImageNet test set.

4.4 EXPERIMENTAL RESULTS

When training on just 250 real images we expected to see VGG-19 overfit, yielding a poorly-generalized model. However, dropout in the fully-connected layers is, by design, set to a rate of 50% to prevent such overfitting. We were unable to achieve convergence with such a small training set. As such, this base-performance comparison is omitted from our analysis.

When trained on fully-synthetic data we achieved a high validation accuracy (3% error), but poor test accuracy (70% error). We can see from the resulting confusion matrix (Figure 5a) that our model performs only slightly better than a random guess when presented with real test data.

After performing domain adaptation, we see a significant increase in accuracy on the test set, reducing the classification error from 70% to only 36.5% after a single epoch of fine-tuning on real training data. The confusion matrix for the domain adaptation results (Figure 5b) suggests that performance may improve even more when accounting for inter-class noise, noting that vessels classified with a Zone 0 heading are visually similar to those with a Zone 1 heading.
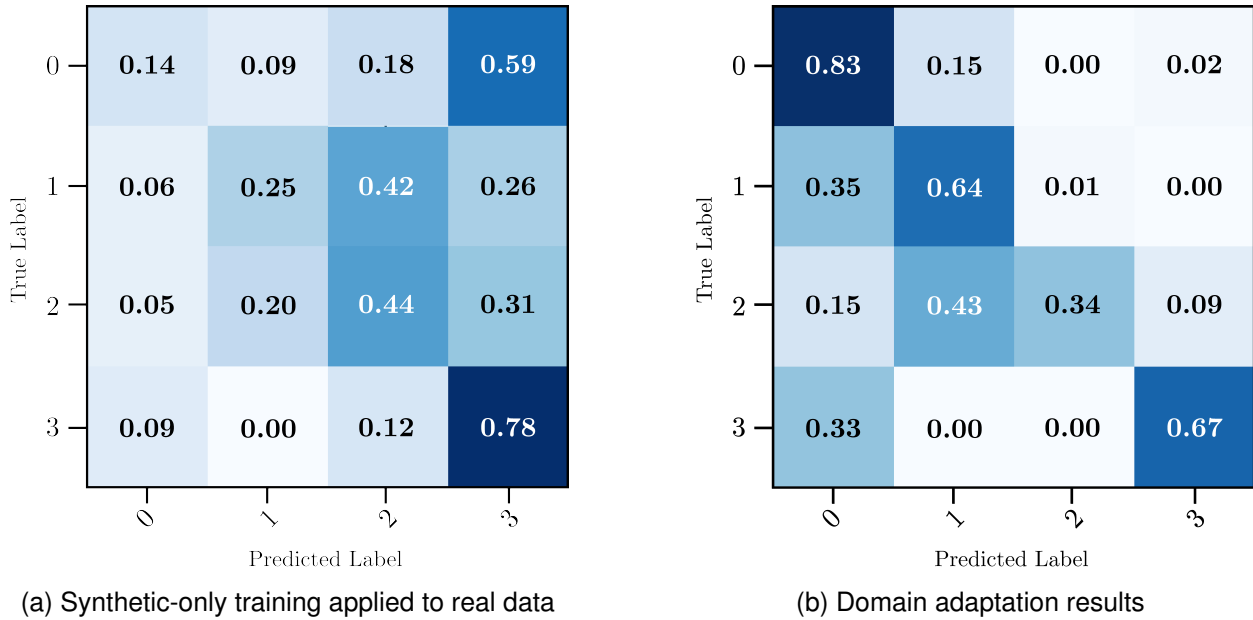
(a) Synthetic-only training applied to real data

(b) Domain adaptation results

Figure 5. Normalized confusion matrices

5. CONCLUSION AND FUTURE WORK

We presented NAVHAZ-Synthetic: a very large dataset of ten commercial vessel classes under varying lighting and environmental conditions. By applying this dataset, we demonstrated that synthetic data can be used to a train complex neural network architecture to convergence. We also show that domain adaptation can be accomplished with use of real imagery, yielding a well generalized model with very little target data. This is particularly significant for maritime and other Department of Defense (DoD) problems where data is often sparse and weakly labeled.

In leveraging synthetic imagery and constraining our problem to a simplified computer vision task, such as determining the general heading of an observed ship, we have demonstrated that current work in computer vision and machine learning can be adapted and applied to maritime problems like collision-avoidance. Alleviating the need for fielding a complex and expensive shipboard system, our proposed method is capable of running on a commercially available computing platform using video feeds from existing camera systems.

Statistical analysis of synthetic imagery versus real target data could help us understand how to further improve our results, although work in this domain seems to indicate that improving the "realism" of synthetic data yields a diminishing return [13]. Moreover, we have only implemented a very basic approach to domain adaptation, and will explore more sophisticated methods in order to understand and maximize the efficacy of synthetic imagery in training machine learning models.

Please contact the authors if you are interested in acquiring the NAVHAZ-Synthetic dataset, or our VGG-19 model weights.

7

(a) Forward camera 1

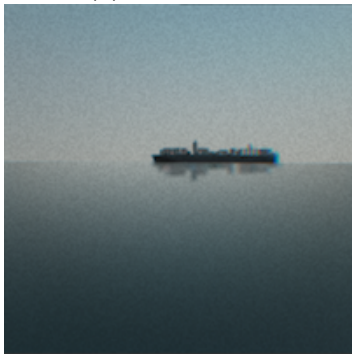(b) Forward camera 2

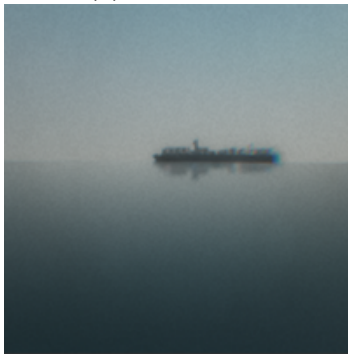(c) Forward camera 3

(d) Mast camera 1

(e) Mast camera 2

(f) Mast camera 3

(g) Aft camera 1

(h) Aft camera 2

(i) Aft camera 3

Figure 6. Scene Observers

REFERENCES

1. Gaidon, A., Wang, Q., Cabon, Y., and Vig, E., 2016. "Virtual worlds as proxy for multi-object tracking analysis," in *arXiv preprint arXiv:1605.06457*.

2. Ros, G., Sellart, L., Materzynska, J., Vazquez, D., and Lopez, A. M., 2016. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," pp. 3234–3243.

3. Handa, A., Patraucean, V., Badrinarayanan, V., Stent, S., and Cipolla, R., 2015. "Synthcam3d: Semantic understanding with synthetic indoor scenes," in *arXiv preprint arXiv:1505.00171*.

4. Hattori, H., Boddeti, V. N., Kitani, K., and Kanade, T., 2015. "Learning scene-specific pedestrian detectors without real data," in "Computer Vision and Pattern Recognition (CVPR), IEEE Conference on," IEEE, pp. 3819–3827.

5. Marin, J., Vázquez, D., Gerónimo, D., and López, A. M., 2010. "Learning appearance in virtual scenarios for pedestrian detection," in "Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on," IEEE, pp. 137–144.

6. Shafaei, A., Little, J. J., and Schmidt, M., 2016. "Play and learn: Using video games to train computer vision models," in *arXiv preprint arXiv:1608.01745*.

7. de Souza, C. R., Gaidon, A., Cabon, Y., and Pena, A. L., 2017. "Procedural generation of videos to train deep action recognition networks," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)," volume 2.

8. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., and Birchfield, S., 2018. "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization," in *arXiv preprint arXiv:1804.06516*.

9. Bales, S. L., 1983. "Designing ships to the natural environment," in *Naval Engineers Journal*, volume 95(2), pp. 31–40.

10. Simonyan, K. and Zisserman, A., 2014. "Very deep convolutional networks for large-scale image recognition," in *arXiv preprint arXiv:1409.1556*.

11. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., 2009. "Imagenet: A large-scale hierarchical image database," in "Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on," IEEE, pp. 248–255.

12. Hinton, G. E., 2012. "A practical guide to training restricted Boltzmann machines," in "Neural networks: Tricks of the trade," Springer, pp. 599–619.

13. Mayer, N., Ilg, E., Fischer, P., Hazirbas, C., Cremers, D., Dosovitskiy, A., and Brox, T., 2018. "What Makes Good Synthetic Training Data for Learning Disparity and Optical Flow Estimation?" in *International Journal of Computer Vision*, pp. 1–19.

INITIAL DISTRIBUTION

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-01-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
August 2018	Final	

4. TITLE AND SUBTITLE	5A. CONTRACT NUMBER
Leveraging Synthetic Imagery for Collision-at-sea Avoidance	5B. GRANT NUMBER
	5C. PROGRAM ELEMENT NUMBER

6. AUTHORS	5D. PROJECT NUMBER
Chris M. Ward	5E. TASK NUMBER
Josh Harguess	
Alexander G. Corelli	5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
SSC Pacific 53560 Hull Street San Diego, CA 92152–5001	TR 3149

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
SSC Pacific Naval Innovative Science and Engineering (NISE) Program 3560 Hull Street, San Diego, CA 92152-5001	NISE
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.

13. SUPPLEMENTARY NOTES

This is work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction.

14. ABSTRACT

Maritime collisions involving multiple ships are considered rare, but in 2017 several United States Navy vessels were involved in fatal at-sea collisions that resulted in the death of seventeen American Service members. The experimentation introduced in this paper is a direct response to these incidents. We propose a shipboard Collision-At-Sea avoidance system, based on video image processing, that will help ensure the safe stationing and navigation of maritime vessels. Our system leverages a convolutional neural network trained on synthetic maritime imagery in order to detect nearby vessels within a scene, perform heading analysis of detected vessels, and provide an alert in the presence of an inbound vessel. Additionally, we present the Navigational Hazards - Synthetic (NAVHAZ-Synthetic) dataset. This dataset, is comprised of one million annotated images of ten vessel classes observed from virtual vessel-mounted cameras, as well as a human "Topside Lookout" perspective. NAVHAZ-Synthetic includes imagery displaying varying sea-states, lighting conditions, and optical degradations such as fog, sea-spray, and salt-accumulation. We present our results on the use of synthetic imagery in a computer vision based collision-at-sea warning system with promising performance.

15. SUBJECT TERMS

Unity gaming engine; sea lighting positions; shipboard collisions; synthetic imagery, computer vision, machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19A. NAME OF RESPONSIBLE PERSON
A. REPORT	B. ABSTRACT	C. THIS PAGE			Chris Ward
U	U	U	U	16	19B. TELEPHONE NUMBER *(Include area code)* 619 553-2511

Standard Form 298 (Rev. 10/17)
Prescribed by ANSI Std. Z39.18

Approved for public release.

SSC Pacific
San Diego, CA 92152-5001