

COMPARISON OF AUTHENTIC MEDIA AND SYNTHETIC MEDIA IN COMPUTER VISION TRAINING

Jonathan Adams
Florida State University
jladams@fsu.edu

Odin Moja
Florida State University
omoja@fsu.edu

Christopher Ko
Florida State University
cyk21@fsu.edu

ABSTRACT

Air-sea rescue is a highly time sensitive endeavor. Victim submersion time strongly influences survival time. The problem is exasperated by the vast search spaces presented by the ocean. Finding individuals quickly in such spaces can be challenging. Drones using computer vision can be a great boon in these situations. Computer vision models require vast amounts of labelled data. The acquisition and labelling of said data can make the use of computer vision models infeasible. Synthetic data represents a possible solution. Generated images may present a sufficient facsimile of authentic images to be used in their stead. Our study used the combination of several tools to produce an automated synthetic data pipeline which both generated and labelled images. Our results showed the model performed well even in comparison to studies using purely authentic data.

Keywords

Machine Learning, Computer Vision, Synthetic Media

INTRODUCTION

Air-sea search and rescue operations are a two-part endeavor, where rescue is the culmination of the search efforts to locate those who need assistance, the environmental conditions, and the length of time the individuals have been in the water (Futch & Allen, 2019). Despite current technologies, there are occasions when the location of people in distress is not known, triggering a visual search from an aerial platform. This may require spotters to scan vast, undulating areas of ocean for something as small as a lifejacket, or a person treading water. The mechanical and human resources involved in aerial search operations are daunting. Search aircraft are expensive to man and operate, Coast Guard helicopters are limited in number, and in their hours of operation. Aircraft are limited in number, and spotters are prone to fatigue, which may compromise efforts.

In today's environment, the concept of using drones, equipped with computer vision, to enhance search efforts makes a great deal of sense. Deep learning models are sufficient to allow for autonomous home deliveries, some waitstaff functions, and automobiles (Gao, 2022). There are challenges to training such resources however, most notably finding and labeling data (Jaiswal et al, 2024). The current research proposes a method of enhancing the efficiency and quality of computer vision without manually labeling thousands of training data images, creating new opportunities to include autonomous spotters in air-sea search operations.

LITERATURE REVIEW

Synthetic media is also known as "deep fakes," or representations of known objects that have been generated by an artificial intelligence algorithm, such as DALL-E. However, synthetic media is the term generally used to describe these representations when employed for training artificial intelligence algorithms to identify (i.e. classify) what it sees. It has been noted that computer vision is, in many ways, superior to human vision (Gao, 2022) if one has the proper amount of data. The use case for synthetic training data has been consistent over the past 7 years, which emphasizes the savings in time, and cost but also has a meaningful effect on data procurement, and management. These characteristics are shown in Table 1.

Key rationale for using synthetic media

Insufficient volume of specific training images available.
The cost of acquiring data is prohibitive.
Data labeling is time-consuming and error prone.

Privacy regulations inhibit data acquisition.
Production values can be inconsistent, nominal image qualities are not clearly defined.

Table 1. List of challenges associated with using authentic images.

The current research aims to compare the performance of computer vision training using photographic images of real individuals in maritime settings with a dataset comprising synthetic media (i.e., ‘deep fakes’) combined with a small percentage (0.1%–8%) of authentic images. Previous studies have shown that including a small proportion of authentic images improves efficiency, particularly when datasets feature varied backgrounds, lighting, and textures (Zhang et al., 2020). Smutny and Jones (2024) reported a significant performance increase when 10% of the dataset consisted of authentic images. Additionally, experiments that tested different combinations of data types and image qualities—such as texture, lighting, and composition—found that computer vision algorithms performed better with mixed datasets (synthetic and authentic images) compared to datasets composed solely of synthetic or authentic images, particularly for classifying individuals on land (Yu et al., 2020). Across all studies utilizing synthetic media, model performance improvements were consistently observed (Adams, 2020).

The use of synthetic media in training datasets offers significant advantages while addressing challenges in data collection, ethical concerns, and model robustness. Specifically, synthetic media aids in model generalization, which is crucial for complex tasks like drowning detection, where environmental factors such as waves, lighting, and depth variations can affect accuracy (Schwoch et al., 2024). Unlike authentic images, synthetic media allows for precise control and customization of image quality, camera positions, backgrounds, and scene composition to adapt to diverse scenarios (Man & Chahl, 2022). Authentic images often lack comprehensive coverage of subjects from multiple angles or diverse poses. Synthetic media overcomes this limitation by mathematically generating random camera positions on an inverted hemisphere, ensuring full coverage of the object (Deserno, 2004). Each image is then superimposed on a randomly selected background from a curated dataset of prepared backgrounds.

The most significant cost-saving aspect of synthetic media lies in data labeling. Synthetic data generation enables the automatic creation of ground truth data, including image-level labels and bounding boxes, thereby reducing the need for manual annotation, minimizing human error, and accelerating the training process (Delussu et al., 2024; Afonso & Giufreda, 2023). For instance, an RTX 3000 series GPU can generate 15,000 labeled images with bounding boxes and varied backgrounds per hour, without errors. Additionally, synthetic media offers substantial privacy benefits by enabling the sharing, pooling, or transferring of datasets without privacy or consent-related concerns (Nikolenko, 2019; Goyal & Mahmoud, 2024). Leveraging these advantages, researchers can develop more efficient and accurate models for detecting drowning incidents, contributing to improved safety in aquatic environments.

This study employs a feasibility design to evaluate the effectiveness of using video feeds from an aerial drone and a computer vision neural network (YOLO11) trained to detect individuals in water settings (e.g., drowning victims). The research aims to compare the performance of a computer vision model trained on synthetic media with results from a model trained exclusively on authentic data. The objective is to determine whether a combination of synthetic and authentic training data can outperform models trained solely with authentic media in classification tasks.

Additionally, this study seeks to outline a process for producing synthetic training media and to demonstrate the feasibility of developing a high-performing model within weeks. Here, “synthetic data” refers to multiple components of a pipeline, including the models, the scenes where they are integrated, and the library that combines them. For comparison, Claesson et al. (2020) conducted a study in which trained lifeguards simulated drowning motions under controlled conditions at selected beaches in Sweden. These simulations were captured by a drone flying overhead and served as authentic training data for their model.

METHODOLOGY

In 2020, Claesson et al. conducted a study to evaluate the feasibility of using computer vision in air-sea operations. Their dataset consisted of 593 authentic photographs of swimmers in various poses, divided into a training set (393 images) and a testing set (200 images). All photos were captured during the day in clear sea water under favorable conditions. The Artificial Intelligence component was powered by Microsoft Azure Cognitive Services (customvisiononline.ai). However, due to the proprietary nature of the technology, detailed information regarding the hardware setup, software configurations, training parameters (e.g., number of epochs, model hyperparameters, and pre-training settings) was unavailable, limiting the possibility of exact replication.

For generating synthetic data, several approaches were considered, including sourcing models from open-source archives, commissioning human artists to create models, and using generative tools such as Meshcapade (Meshcapade—The Digital Human Company, n.d.) and Meshy (Meshy, 2025). These tools enabled the researchers to produce male and female models

with significant variation in skin tone and body shape. Through iterative refinement of prompts, the team consistently generated high-quality models that incorporated diverse physical characteristics. These models were then used to create boxed and labeled images, ready for training. The synthetic dataset was subsequently employed to replicate the testing methods from the original study.

Three-dimensional models were processed into two-dimensional images using Blender, an open-source modeling application. Each image is saved with a companion text document that lists the class of the various boxed objects. Half of the models were styled in a water treading pose with shoulders and head prominent and the lower half of the body underwater. Half of the models were flat splayed out in the same poses as seen in Claesson et al. (2020). Half of the flat models were integrated into scenes face down and the other face up. All the models were integrated into scenes such that the object in frame occupied a large relative proportion of the image. In many cases objects took up approximately 60% of the image. This was done to aid in the definition of object edges.

STUDY DESIGN

A batch of 4,000 synthetic images was generated and randomly sorted into training, validation, and testing datasets using a script. The data was split into 80% for training, 10% for validation, and 10% for testing, resulting in 3,200 training images. To incorporate authentic data, 432 synthetic images were replaced with real images, increasing the proportion of authentic data to 13.5%. The proportion of authentic to synthetic imagery is based on a review of research literature, which suggests synthetic media can be used successfully in proportions ranging from .01 to 10 percent, depending on the complexity of the environment and the features of the object to be classified (Tsai, et al, 2017; Nabati, [et.al.](#), 2020). Of these, 272 images contained people, while 160 were blank or null images. This adjustment yielded 8.5% natural or authentic data and an additional 5% authentic background data. Blank or background images were included following the recommendations of the Ultralytics dataset guidelines to reduce false positives (Ultralytics, n.d.b).



Figure 1. Example of synthetic and authentic training media

For additional testing after initial training, a subset of the publicly available Man Overboard Drone (MOBDrone) dataset was used. This dataset is designed to assist in training AI models for detecting individuals who have fallen overboard during ocean search and rescue operations (Cafarelli et al., 2022). The study utilized YOLOv11m, a validated object detection model (Alif & Hussain, 2024; Ultralytics, 2025c), and trained it for 8,000 epochs. Initial training was conducted on an Alienware R15 desktop running Ubuntu 20.04 LTS, equipped with a Nvidia RTX 4090 graphics processing unit.

The model was initialized with a pre-trained .pt file based on the COCO8 dataset. This dataset includes 80 default object classes, one of which is "person" (Ultralytics, n.d.-a). Instead of creating a distinct "swimmer" class, the researchers opted to further train the "person" class. The decision was based on the fact that the COCO dataset primarily depicts people in terrestrial and unobstructed contexts, such as playing sports, standing in front of stores, or sitting at a table (Lin et al., 2014). These scenarios differ significantly from the challenging visual conditions encountered in ocean rescue. The goal was not only to recognize a person but also to adapt the model to identify individuals in complex aquatic environments.

RESULTS

To ensure parity with the original study, testing was designed to approximate its conditions. The MOBDrone dataset was used as it features clear, near-coastal, daytime ocean scenes captured from heights of 10–60 meters, closely resembling the conditions described by Claesson et al. (2020). However, some differences were noted: the MOBDrone test set occasionally included two victims per image, while the Claesson dataset consistently featured only one. Additionally, the MOBDrone dataset included objects such as boats in the frame. These factors made the MOBDrone dataset a comparable yet more challenging test set for the model.

Two critical metrics influenced testing performance: Intersection over Union (IoU) and confidence (conf). IoU measures the overlap between the predicted bounding box and the ground-truth region, while confidence reflects the model's likelihood of correctly detecting and localizing an object. For a detection to qualify as a hit, it must meet the thresholds for both IoU and confidence. YOLO allows users to adjust these thresholds, which can affect performance metrics such as precision, recall, and mean Average Precision (mAP). As of this writing, Ultralytics lists default IoU and confidence values as 0.70 and 0.001, respectively.

Initial testing was conducted using these default IoU and confidence values, yielding Precision: 74.1%, Recall: 76.1%, and mAP@50: 75%. Figure 4 displays the normalized confusion matrix, where the model achieved 80% true positives in detecting instances of drowning individuals. When interpreting the confusion matrix, it is important to consider YOLO's mechanics. Each column total must sum to 1.0. The "blank" sector in the bottom-right represents correctly identified instances of "background." YOLO prioritizes identifying targets rather than non-targets. For background hits to occur, the model must predict "background" and match ground truths for all background elements. This can result in a confusion matrix dominated by unnecessary "correct" background predictions.

Another key observation is the value of 1.00 in the top-right quadrant of the confusion matrix, which represents false positives. Since the matrix is normalized and the columns must sum to 1.0, this indicates that 100% of false positives were associated with the single victim class.

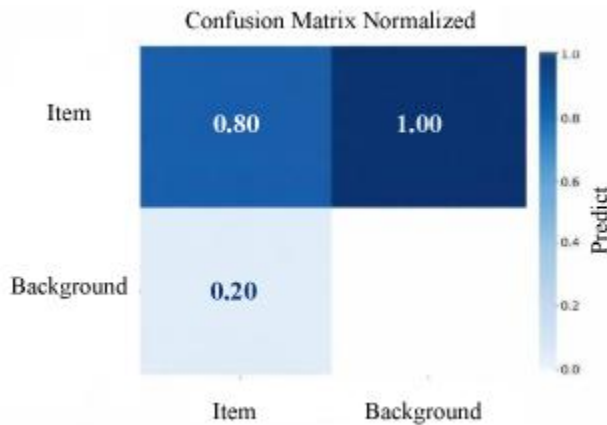


Figure 4. Normalized confusion matrix for Yolo 11m at IoU 0.7 and conf 0.001

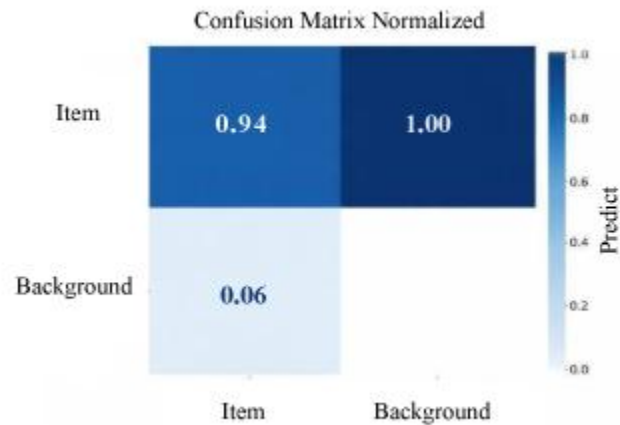


Figure 5. Normalized confusion matrix for Yolo 11m at IoU 0.01 and conf 0.0001

To test the extremes of detection possibility, the values of IoU and conf were then set to 0.01 and 0.0001 respectively. The model was tested again and produced the following metrics; precision 74.2%, recall 76.4%, and mAP@50 74.9%. Figure 5 illustrates the model's normalized confusion matrix wherein it achieved 94% true positives for the victim class.

DISCUSSION

While the overall results were satisfactory, the most surprising finding was the stability of many metrics, even as IoU and confidence (conf) thresholds were adjusted. The true positive rate increased, and advanced metrics, such as precision-recall curves, remained stable. The most unexpected result was the minimal decrease in precision, despite reductions in IoU. Lowering IoU typically reduces the stringency for a detection to be considered a hit, as it requires less overlap with the ground truth, which should theoretically lower precision. However, across the tested range of conf values (0.0001 to 0.8), the lowest precision observed was 74.1%, while the highest was 88.2% at conf = 0.8, with similar stability across varying IoU values. Only at extreme thresholds (e.g., IoU or conf ≥ 0.9) did precision and other metrics drop to zero.

Precision measures how often the model is correct when it claims to have detected an object, while recall measures how many actual objects the model successfully identifies. Given the critical nature of detecting drowning individuals, recall (minimizing missed detections) is arguably more important than precision (minimizing false positives). Encouragingly, the model demonstrated robustness to reduced confidence thresholds, enabling higher recall without a significant drop in precision. Additionally, the model's resilience under high IoU values suggests strong localization capabilities, which are particularly valuable for this study's context.

The results highlight the effectiveness of the latest YOLO architecture when trained on synthetic data and tested on natural data, supporting the feasibility of using synthetic data as a substitute for natural data. Future research could explore the potential of using smaller proportions of authentic data and optimizing training hyperparameters to further enhance model performance.

FUTURE WORK

It is important to bear in mind that experiments using synthetic imagery have been ongoing for nearly 10 years. Over time it has become apparent that each classification context has different requirements that may change the volume of data, require stringent feature sets, or include complex environment features such as one might expect to find in ocean scenes due to changing features of water in motion. In addition, current hardware capabilities include significant opportunities for improvement in the future. For example, obtaining a license for transmission in reserved or restricted bandwidth frequencies, will increase the transmission range to the horizon (about 10 miles). Another consideration is the compute required by state-of-the-art computer vision systems, which are currently undergoing rapid development because of the increasing research with humanoid robotics. It is also important to note that the current study is intended as proof-of-concept, the development of a product intended for real-world circumstances would require greater diversification of the data to include numerous meteorological conditions, and levels of ocean turbulence.

LIMITATIONS

The study was approved by the Swedish ethical review authority, ensuring compliance with ethical standards in research involving human participants.

Some qualities used in the prior study were not noted, for example specifics about the training parameters, or qualities of the training media were not discussed. The original study reported metrics of sensitivity, specificity and no mAP. Granted in detection tasks sensitivity is essentially the same as recall, but the lack of other industry standard metrics reported such as precision and mAP@50 greatly reduced the ability to do a full comparison.

REFERENCES

1. Adams, J., Murphy, E., Sutor, J. (2021). Assessing the Qualities of Synthetic Visual Data Production, *2021 9th International Conference on Information and Education Technology (ICIET)*, Okayama, Japan, 2021, pp. 452455, doi: 10.1109/ICIET51873.2021.9419586.
2. Afonso, F., & Giuffrida, N. (2023). Editorial: Synthetic data for computer vision in agriculture. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2023.1277073>
3. Belanger Smutny, J. W., & Jones, C. (2024). *The effects of using created synthetic images in computer vision training*. Virginia Tech.
4. Claesson, A., Schierbeck, S., Hollenberg, J., Forsberg, S., Nordberg P., Ringh, M., Olausson, M., Jansson, A., Nord, A. (2020). The use of drones and a machine-learning model for recognition of simulated drowning victims—A feasibility study. *Resuscitation*, 152, 30-36. <https://doi.org/10.1016/j.resuscitation.2020.09.022>
5. Deserno, M. (2004). How to generate equidistributed points on the surface of a sphere. Unpublished. Sept. 28 2004, Retrieved on Jan. 18, 2021 from https://www.cmu.edu/biolphys/deserno/pdf/sphere_equi.pdf.
6. Delussu, A., Putzu, L. Fumera, G. (2024). Synthetic data for video surveillance applications of computer vision: A review. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-024-02102-x>
7. Futch V and Allen A (2019) Search and Rescue Applications: On the Need to Improve Ocean Observing Data Systems in Offshore or Remote Locations. *Front. Mar. Sci.* 6:301. doi: 10.3389/fmars.2019.00301
8. Gao, Y. (2022). Research on the application of artificial intelligence technology in the development of computer vision. *Highlights in Science, Engineering and Technology*, 9, 80-84.
9. Goyal, M., & Mahmoud, Q. (2024). A systematic review of synthetic data generation techniques using generative AI. *Electronics*, 13(17), 3509. <https://doi.org/10.3390/electronics13173509>
10. Jaiswal, A., Shah, A., Harjadi, C., Windgassen, E., & Washington, P. (2024). Ethics of the Use of Social Media as Training Data for AI Models Used for Digital Phenotyping. *JMIR formative research*, 8, e59794. <https://doi.org/10.2196/59794>
11. Kar, A., Prakash, A., Liu, M. Y., Cameracci, E., Yuan, J., Rusiniak, M., ... & Fidler, S. (2019). Meta-sim: Learning to generate synthetic datasets. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4551-4560). <https://doi.org/10.48550/arxiv.1904.11621>
12. Lin, F., Hou, T., Jin, Q., & You, A. (2021). Improved YOLO Based Detection Algorithm for Floating Debris in Waterway. *Entropy*, 23(9), 1111. <https://doi.org/10.3390/e23091111>
13. Man, K., & Chahl, J. (2022). A Review of Synthetic Image Data and Its Use in Computer Vision. *Journal of Imaging*, 8(11), 310. <https://doi.org/10.3390/jimaging8110310>
14. Nabati, M., Navidan, H., Shahbazian, R., Ghorashi, S. A. and Windridge, D., (2020) "Using Synthetic Data to Enhance the Accuracy of Fingerprint-Based Localization: A Deep Learning Approach," in *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1-4, doi: 10.1109/LSSENS.2020.2971555.

15. Nikolenko, S. I. (2019). Synthetic data for deep learning. *arXiv preprint arXiv:1909.11512*. <https://doi.org/10.48550/arxiv.1909.11512>
16. Schwoch, S., Dammann, M. P., Bartl, J. G., Kretschmar, M., Saske, B., & Paetzold-Byhain, K. (2024). Towards a process for the creation of synthetic training data for AI-computer vision models utilizing engineering data. *Proceedings of the Design Society*, 4, 2237–2246. doi:10.1017/pds.2024.226
17. Tremblay, J., Prakash, 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. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 969-977). <https://doi.org/10.48550/arxiv.1804.06516>
18. Tsai, C., Tsai, Y., Hsu, S., & Wu, Y. (2017). Synthetic Training of Deep CNN for 3D Hand Gesture Identification. *2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, 165-170.
19. T Varga, L. A., Kiefer, B., Messmer, M., & Zell, A. (2022). Seadronessee: A maritime benchmark for detecting humans in open water. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2260-2270). <https://doi.org/10.48550/arxiv.2105.01922>
20. Yeong, S. P., King, L. M., & Dol, S. S. (2015). A review on marine search and rescue operations using unmanned aerial vehicles. *International Journal of Marine and Environmental Sciences*, 9(2), 396-399. ISNI:0000000091950263
21. Yu, J., Farin, D., Krüger, C., Schiele, B. (2020) “Improving person detection using synthetic training data” in *2010 IEEE International Conference on Image Processing*, Sept. 2010. Accessed on May 29, 2020. Available: IEEE Xplore <https://doi.org/10.1109/ICIP.2010.5650143>
22. Zhang, L., Gonzalez-Garcia, A. et al. (2020) “Synthetic data generation for end-to-end thermal infrared tracking.” in *Transitions On Image Processing*, Vol. 28 (4). Accessed on May 13, 2020. Available: IEEE Xplore.