Context-Aware Synthetic Data Generation for Object Detection Using Alpha Blend

Ponlawat Chophuk
Faculty of Informatics
Burapha University
Chonburi, Thailand 20131
ponlawat.ch@go.buu.ac.th

Pinphong Ruangraweenukit
Faculty of Informatics
Burapha University
Chonburi, Thailand 20131
64160205@go.buu.ac.th

Waranrach Viriyavit
Faculty of Informatics
Burapha University
Chonburi, Thailand 20131
waranrach.vi@go.buu.ac.th

Thatsanee Charoenporn
Faculty of Data Science
Musashino University
Tokyo, Japan
thatsanee@ds.musashino-u.ac.jp

Virach Sornlertlamvanich*
Faculty of Data Science
Musashino University
Tokyo, Japan
virach@gmail.com

Abstract—**Object detection for waste management is limited by the scarcity of large, labeled real-world datasets. To overcome this, we developed a synthetic dataset generation method that enhances background diversity using depth estimation, angle adjustment, image feathering, and alpha blending. These techniques seamlessly integrate foreground waste objects into varied environments—ground, urban, and field—thereby increasing scene variety and complementing existing data augmentation strategies. We evaluated the synthetic dataset through two test cases using YOLOv8L models. In Test Case 1, models trained with synthetic data achieved a precision of 0.611 and an mAP50 of 0.453, outperforming those trained solely on real data (precision: 0.529; mAP50: 0.399). Test Case 2 examined different real-synthetic data ratios, where the 75% Real / 25% Synthetic model achieved the highest mAP50 of 0.884, surpassing the 100% Real model's mAP50 of 0.823. These results demonstrate that our synthetic dataset effectively mitigates data scarcity and enhances object detection performance. Additionally, incorporating real image features is crucial for preventing misclassifications and partial detections, highlighting the importance of balancing real and synthetic data. Our approach offers a scalable and cost-effective solution for augmenting training data with diverse environmental contexts, leading to more reliable and accurate object detection models in challenging real-world settings.**

Keywords—Synthetic Data Generation, Object Detection, Data augmentation, Background diversity, Image Blending, Environmental Variability

I. INTRODUCTION

The growing demand for robust object detection models across various domains highlights the challenges of collecting and annotating large-scale datasets, especially in waste detection where objects appear in diverse and unpredictable environments. Acquiring real-world data is both expensive and time-consuming, limiting the availability of comprehensive datasets necessary for training high-performing models. To overcome these limitations, synthetic data generation has emerged as a promising solution [1]. Techniques such as image blending offer scalable and cost-effective alternatives that maintain the diversity and complexity essential for effective model training [2]. By artificially generating data, researchers can address issues related to data scarcity, privacy concerns, and high annotation costs, facilitating the development of more accurate and generalizable object detection systems.

Synthetic data not only enhances model performance but also enables domain adaptation, as demonstrated by various studies. Shorten and Khoshgoftaar [1] highlighted techniques like image augmentation, inpainting, and generative models that effectively supplement or replace real-world datasets. Frid-Adar et al. [7] showcased the potential of Generative Adversarial Networks (GANs) in creating synthetic medical images with performance comparable to models trained on real data. Similarly, Richter et al. [8] utilized synthetic data from video games for autonomous driving, and Wang et al. [9] applied it to crowd counting, enhancing accuracy in varying densities. Despite these advancements, challenges such as artifact boundaries—visible seams and inconsistencies—reduce image realism [5]. Advanced blending techniques like Poisson Image Editing [10], Multi-layer Alpha Blending [11], and Multiblending [12] have been developed to address these issues, ensuring seamless integration of foreground objects into diverse backgrounds. Building on these methods.

This research explores generating synthetic datasets using inpainting techniques for waste detection. By integrating alpha blending and image feathering, the study aims to produce more realistic and contextually accurate synthetic images, enhancing dataset diversity and quality. The generated dataset will be evaluated by training YOLO [18] models and testing on real-world datasets such as COCO and TACO, anticipating improved object detection performance in waste management tasks. This study contributes to understanding how synthetic data can support real-world applications, particularly where real data is limited or difficult to acquire and lays the groundwork for more effective data generation methodologies in computer vision research. The paper is organized as follows: Problem Analysis, Methodology, Results, and Conclusion.

II. Problem Analysis

Effective training of object detection models for waste management requires datasets that encompass diverse environments such as urban, field, and muddy settings. However, existing public datasets present significant limitations in environmental diversity and data volume.

As illustrated in Table I, the COCO 2017 dataset offers a vast number of images with diverse object classes but lacks sufficient environmental diversity, making it less ideal for specialized tasks like waste detection. The TACO dataset, while focused on trash annotations, is limited by its small size and a high proportion of tiny objects, which can negatively impact model training and performance.

TABLE I. COMPARISON OF DATASETS BY ENVIRONMENTAL RATIOS

Dataset	**Total Image**	**Environment (%)**				**Limitation**
		Urban	***Field***	***Muddy***	***Other***	
COCO 2017	164K	~50	~30	~10	~10	General object classes, limited environmental diversity
TACO	1500	~45	~30	~20	~5	Small dataset size, many tiny objects
Roboflow [19]	7600	~56	~20	~10	~10	Limited environmental diversity
Roboflow [20]	2100	~60	~15	~15	~10	Limited environmental diversity
Our	3000	33	33	33	1	Balanced environments

The Roboflow datasets ([19], [20]) provide a moderate number of images with a predominant focus on urban environments. Although they offer more targeted data compared to COCO, they still suffer from limited environmental diversity, restricting the model's ability to generalize across different settings.

In contrast, our Synthetic Dataset addresses these limitations by ensuring a balanced representation across urban, field, and muddy environments. With 3,000 images evenly distributed among these contexts, our dataset enhances environmental diversity and contextual accuracy. Additionally, the use of synthetic generation techniques mitigates issues related to data scarcity and high annotation costs, providing a scalable and cost-effective solution for training more reliable and accurate object detection models in waste management.

III. Methodology

The primary objective of this research is to develop a methodology for generating a customized synthetic dataset aimed at improving object recognition accuracy in waste detection tasks. This involves automatically separating foreground objects from their backgrounds, generating new background images, reconstructing composite images, and validating the effectiveness of these synthetic images. The methodology comprises can be illustrated in Figure 1.

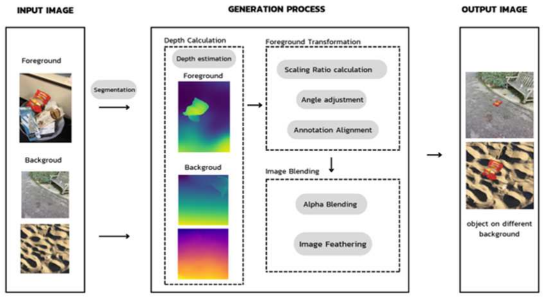

Fig. 1. A workflow of a proposed methodology.

A. Foreground Object Separation

For the extraction of foreground objects, the Trash Annotations in Context (TACO) dataset [16] was utilized. The TACO dataset provides annotated images of litter and waste across various environments, making it an ideal source for identifying relevant objects such as plastic bottles and cans. The following steps were undertaken to prepare the foreground objects:

1) Annotation-Based Segmentation

The annotations provided in the TACO dataset were used to identify pixel regions corresponding to foreground objects. This involved isolating objects like plastic bottles and cans from their backgrounds.

2) Binary Mask Generation

Binary masks were created where foreground objects were represented in black and the background in white. Image processing libraries such as OpenCV and PIL were employed to generate precise masks that accurately align with object boundaries.

3) Quality Assurance

Visual inspections were conducted to ensure that the masks accurately represented the objects without including any background elements. This step was crucial to maintaining the integrity of the synthetic images. As shown in Figure 2.

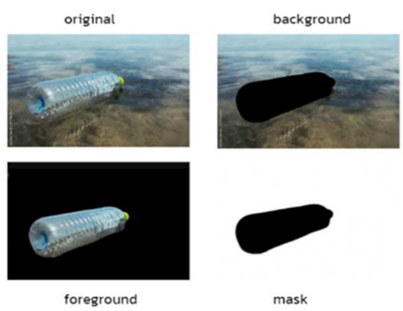

Fig. 2. Segmented images and mask for foreground object.

B. Background Image Selection

The selection of background images was critical to ensure environmental diversity in the synthetic dataset. Three distinct background environments were chosen: field, muddy, and urban. A total of 90 background images were curated, with 30 images representing each environment type, presented in Table II. The preparation process included the following steps:

1) Environment Categorization

Background images were categorized into three environments—field, muddy, and urban—to capture a wide range of real-world scenarios where waste is commonly found.

2) Background Selection

From each category, 30 high-quality background images were selected. To prevent data overfitting, three images from each environment type were randomly chosen for each synthetic image generation cycle.

3) Background Suitability

We selected object-free backgrounds with angles between 20° and 75° to prevent interference with foreground waste objects and ensure diverse perspectives, enhancing the synthetic dataset's variety.

TABLE II. COMPARISON OF DATASETS BY ENVIRONMENTAL RATIOS

Environment	Quantity	Image Example
Urban	30	
Field	30	
Muddy	30	

C. Synthetic Image Generation

This study generates synthetic images by integrating five foreground objects with randomly selected backgrounds from 90 urban, field, and muddy environments. Using depth estimation, angle adjustment, feathering, and alpha blending, this method enhances dataset diversity and realism, improving object detection in waste management. Figure 3 illustrates the workflow.

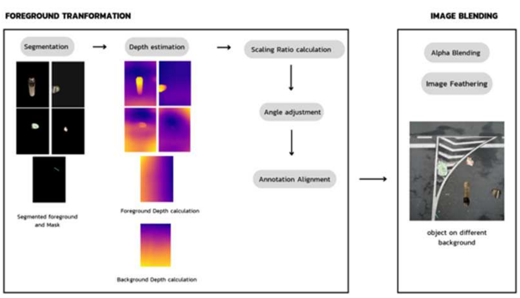

Fig. 3. Workflow of Synthetic Image Generation.

1) Environment Categorization

a) Depth Estimation

The MiDas model was utilized to determine the depth of both foreground and background elements. MiDas [22] estimates depth based on visual cues such as object shapes, shading, occlusion, perspective, and texture gradients, providing a spatial understanding necessary for realistic object placement.

b) Scaling Ratio Calculation

By analyzing the average depth of foreground and background elements, the appropriate scaling ratio was calculated using the following formula:

$$\text{Scaling Ratio} = \frac{\bar{d}_{\text{foreground}}}{\bar{d}_{\text{background}}} \tag{1}$$

Here, $\bar{d}_{\text{foreground}}$ represents the average depth of the foreground, and $\bar{d}_{\text{background}}$ represents the average depth of the background. The bar over the $\bar{d}$ indicates the average (mean) of those depths. This ratio from (1) ensures that the foreground objects are proportionally scaled to fit naturally within the background environments.

c) Annotation Alignment

Alongside scaling, annotations were adjusted to align accurately with the transformed foreground objects, maintaining the integrity of object boundaries.

d) Angle Adjustment

The foreground objects were rotated and scaled using the Scale-Invariant Feature Transform (SIFT) algorithm. SIFT [17] calculates keypoints within an image and compares their angles to align the foreground with the background, ensuring that the objects match in terms of rotation and orientation.

2) Image Blending

To achieve seamless integration of foreground objects into diverse backgrounds and address artifact boundaries, advanced image blending techniques were employed

a) Alpha Blending

In ground-based environments, alpha blending is utilized to seamlessly merge foreground objects with backgrounds using a process known as Equation (2). Initially, a binary mask is converted to a three-channel format to align with the RGB channels of the images. The edges of the mask are then feathered to smooth transitions between the foreground and background, effectively reducing visible seams and artifacts. After overlaying the smooth mask onto the background image, the mask is normalized to ensure consistent blending across different images. The final composite image is created using the formula:

$$\text{Composite Image} = \alpha \cdot F + (1 - \alpha) \cdot B \tag{2}$$

Where α is the normalized, blurred mask that dictates the transparency of each pixel, ensuring a seamless integration of the Foreground (F) and Background (B) layers.

b) Image Feathering

To enhance the realism of composite images, we used image feathering and alpha blending to smooth foreground edges and create seamless transitions. After blending, we adjusted the brightness and contrast to match the background's lighting and corrected the color balance to harmonize color tones. These steps ensured visual consistency and minimized artifact boundaries, resulting in more realistic composite visuals.

D. Model Training and Evaluation

To evaluate the effectiveness of the synthetic dataset in enhancing object detection performance, two test cases were conducted:

1) Test Case 1: Synthetic Dataset vs. Real Dataset Training

In the first test case, two YOLOv8L object detection models were trained separately for 200 epochs with a batch size of 16. One model utilized the synthetic dataset combined with 10% of the real Trash Annotations in Context (TACO) dataset [16] to help the model learn real image features and prevent overfitting. The other model was trained solely on the real TACO dataset. Both models were then evaluated on a common validation set from the TACO dataset to assess their generalization capabilities. The primary evaluation metric was mean Average Precision (mAP), supplemented by precision and recall scores. This comparison aimed to determine whether the synthetic dataset could match or surpass the performance of models trained exclusively on real data.

2) Test Case 2: Impact of Real and Synthetic Data Ratios on Model Performance

To evaluate the influence of varying real and synthetic data ratios on object detection performance, five YOLOv8L models were trained with different data compositions ratios in Table IV.

Each configuration was trained for 200 epochs with a batch size of 16, applying consistent augmentation techniques to ensure comparability. The trained models were then evaluated on a custom test set of 1,150 images encompassing three distinct environments—urban, field, and muddy—to assess their performance in diverse real-world scenarios. Evaluation metrics included Precision, Recall, mean Average Precision (mAP), This structured approach aims to identify the optimal balance between real and synthetic data that maximizes detection accuracy and generalization, thereby validating the effectiveness of synthetic data in enhancing object detection models for waste management tasks.

IV. Result

A. Test Case 1: Synthetic Dataset vs. Real Dataset Training

In the first test case, the YOLOv8L model trained on the synthetic dataset outperformed the model trained solely on the real-world TACO dataset, achieving a higher precision of 0.611 and an mAP50 of 0.453, compared to 0.529 precision and 0.399 mAP50 for the real dataset model as shown in Table III. However, the synthetic model exhibited a slightly lower recall (0.307 vs. 0.320), indicating that while it was more accurate in its detections, it missed a few relevant objects. This improvement in precision and mAP50 suggests that the synthetic dataset effectively enhanced the model's ability to correctly identify objects without increasing false positives.

TABLE III. Model Performance Metrics on Synthetic vs. Original Dataset

Metric	mAP50	Precision	Recall
Real dataset model [16]	0.529	0.320	0.399
Synthetic dataset model	0.611	0.307	0.453

Despite these gains, the synthetic model faced challenges in diverse environments, as illustrated in Figure 4, where it sometimes misclassified objects or only partially detected them (e.g., detecting 50% of an actual object). These issues stem from the variability in backgrounds and the complexity of real-world scenes, which may not be fully captured by the synthetic data. Additionally, certain classes may have overlapping features with backgrounds, leading to incorrect predictions. To address these limitations, further refinement of the synthetic data generation process is necessary, potentially by increasing the diversity and complexity of the synthetic backgrounds and incorporating more varied object orientations and occlusions. Overall, the results demonstrate that while synthetic data can significantly enhance object detection performance in waste management tasks, ongoing improvements are required to achieve comprehensive detection across highly varied and complex real-world environments.

Fig. 4. Example of detected image in Test Case 1.

B. Test Case 2: Impact of Real and Synthetic Data Ratios on Model Performance

To assess the influence of varying real and synthetic data ratios on object detection performance, five YOLOv8L models were trained with different data compositions. The performance metrics for each configuration are presented in Table IV.

TABLE IV. MODEL PERFORMANCE METRICS ON SYNTHETIC VS. ORIGINAL DATASET

Data composition		Precision	Recall	mAP50
Real (%)	*Synthetic (%)*			
0	100	0.461	0.079	0.262
25	75	0.918	0.494	0.714
50	50	0.959	0.726	0.854
75	25	0.988	0.772	0.884
100	0	0.948	0.753	0.823

The results demonstrate that incorporating synthetic data enhances model performance up to a certain threshold. Notably, the 75% Real / 25% Synthetic model achieved the highest mAP50 of 0.884, surpassing the 100% Real model's mAP50 of 0.823. This improvement suggests that synthetic data introduces beneficial diversity, enabling the model to generalize better across varied environments.

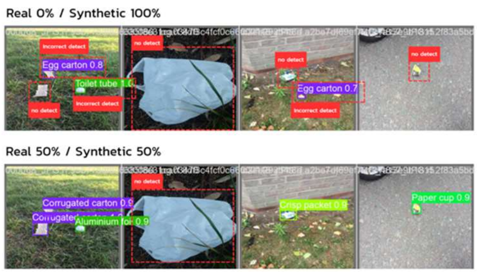

Fig. 5. Example of detected image in Test Case 2.

However, as illustrated in Figure 5, the 0% Real / 100% Synthetic model struggled with misclassifications and partial object detections, likely due to insufficient learning of real image features and the presence of artifact boundaries. Without exposure to real-world image characteristics, the model cannot accurately interpret and detect objects in complex settings. In contrast, the 50% Real / 50% Synthetic model demonstrated significantly better accuracy and reliability by balancing real image features with synthetic diversity. Testing on datasets encompassing three different environments—urban, field, and muddy—highlighted that models trained with a mix of real and synthetic data could better handle environmental variability. These findings underscore the importance of integrating real images to provide foundational visual cues that synthetic data alone cannot fully replicate. Additionally, addressing artifact boundaries through techniques like alpha blending and image feathering, along with rigorous quality assurance measures—including visual inspections and quantitative metrics was crucial in maintaining the realism and integrity of synthetic images. This balance ensures that synthetic data effectively complements real data, leading to more accurate and robust object detection models in waste management tasks.

V. CONCLUSION

This study developed a methodology for creating a customized synthetic dataset to enhance object recognition accuracy in waste detection using the YOLOv8L model. By applying advanced image processing techniques—depth estimation, angle adjustment, alpha blending, and image feathering—foreground waste objects were seamlessly integrated into diverse backgrounds, resulting in realistic composite images. The synthetic dataset, combined with varying proportions of real-world data, effectively improved model performance.

In Test Case 1, models trained with synthetic data achieved a precision of 0.611 and an mAP50 of 0.453, outperforming those trained solely on real data (precision: 0.529; mAP50: 0.399). Test Case 2 further demonstrated the advantage of synthetic data, with the 75% Real / 25% Synthetic model achieving a higher mAP50 (0.884) compared to the 100% Real model (mAP50: 0.823). However, models trained exclusively on synthetic data showed lower precision and struggled with misclassifications and partial detections, highlighting the critical role of real image features. Without real images demonstrating how objects appear in specific environments, the model lacks contextual understanding to accurately identify or predict object-environment interactions, resulting in missed detections or incorrect classifications. Exposure to real images enables the model to discern whether waste objects and environmental conditions are realistic, thereby enhancing detection accuracy and reliability.

Future work should focus on optimizing object placement, incorporating metadata such as object angles and viewpoints (e.g., top, front, bottom), and ensuring the contextual appropriateness of objects within specific environments (e.g., determining if a clean bottle realistically fits in a muddy setting). Additionally, integrating more sophisticated generative models and expanding dataset diversity will further enhance model robustness and generalization capabilities. These improvements will contribute to developing more reliable and accurate object detection systems in varied and challenging real-world environments.

ACKNOWLEDGMENT

This work was supported by the Faculty of Informatics, Burapha University, Chonburi, Thailand, and the Faculty of Data Science, Musashino University, Tokyo, Japan.

REFERENCES

[1] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[2] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, … and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2018, pp. 969–977.

[3] D. Dwibedi, I. Misra, and M. Hebert, "Cut, paste and learn: Surprisingly easy synthesis for instance detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 1301–1310.

[4] G. Georgakis, A. Mousavian, A. C. Berg, and J. Kosecka, "Synthesizing training data for object detection in indoor scenes," *arXiv preprint arXiv:1702.07836*, 2017.

[5] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, … and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 740–755.

[6] B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, … and D. Amodei, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

[7] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[8] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, Netherlands, Oct. 2016, pp. 102–118.

[9] Q. Wang, J. Gao, W. Lin, and Y. Yuan, "Learning from synthetic data for crowd counting in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 8198–8207.

[10] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," in *ACM Trans. Graph. (TOG)*, vol. 22, no. 3, pp. 313–318, 2003.

[11] M. Salvi and K. Vaidyanathan, "Multi-layer alpha blending," in *Proc. ACM SIGGRAPH Symp. Interact. 3D Graph. Games*, Mar. 2014, pp. 151–158.

[12] P. Baudisch and C. Gutwin, "Multiblending: Displaying overlapping windows simultaneously without the drawbacks of alpha blending," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2004, pp. 367–374.

[13] C. Allene, J. P. Pons, and R. Keriven, "Seamless image-based texture atlases using multi-band blending," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–4.

[14] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, 2010.

[15] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, … and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2018, pp. 969–977.

[16] P. F. Proença and P. Simões, "TACO: Trash annotations in context for litter detection," *arXiv preprint arXiv:2003.06975*, 2020.

[17] M. Chen, Z. Shao, D. Li, and J. Liu, "Invariant matching method for different viewpoint angle images," *Appl. Opt.*, vol. 52, no. 1, pp. 96–104, 2013.

[18] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of YOLO algorithm developments," in *Proc. Comput. Sci.*, vol. 199, pp. 1066–1073, 2022.

[19] Garbage, "Garbage detection dataset [Open source dataset]," *Roboflow Universe*, May 2024. [Online]. Available: https://universe.roboflow.com/garbage-uo5xu/garbage-detection-vixig

[20] FYP, "YOLOv8-trash-detections dataset [Data set]," *Roboflow*, 2023. [Online]. Available: https://universe.roboflow.com/fyp-bfx3h/yolov8-trash-detections

[21] R. Birkl, D. Wofk, and M. Müller, "MiDaS v3.1—A model zoo for robust monocular relative depth estimation," *arXiv preprint arXiv:2307.14460*, 2023.