

Distribution-Aware Synthetic Data Refinement for Lightweight Military Camouflaged Object Detection

Thi-Thu-Hang Truong* and Trung-Kien Tran*[†] Hai-Hong Phan[‡]

* Institute of Information Technology and Electronics, AMST, Hanoi, Vietnam

[‡] Institute of Information and Communications Technology, Le Quy Don Technical University, Hanoi, Vietnam

[†] Corresponding author (t2kien@gmail.com)

Abstract—Military camouflaged object detection remains a challenging task due to the deliberate concealment of targets and the scarcity of annotated training data. In this paper, we propose an object-level distribution-aware synthetic data refinement framework that selects 3D-generated camouflage samples based on feature-space alignment with real and style-transferred images, measured by Mahalanobis distance in the YOLO embedding space. The refined synthetic data is combined with real-image augmentation to form a series of training configurations evaluated across six lightweight YOLO variants targeting edge deployment on resource-constrained platforms.

Experiments show that feature-based selection consistently outperforms using the full synthetic set, confirming that alignment quality matters more than quantity. Real-image style transfer drives the strongest per-class gains for visually distinctive but underrepresented classes, while selected 3D data contributes complementary geometric diversity. The combined strategy achieves the highest overall detection performance, with augmented configurations also converging faster and generalizing better to real-image validation data. Trade-off analysis across models reveals practical recommendations for balancing accuracy and computational cost under edge deployment constraints. Severely underrepresented categories remain a persistent bottleneck, highlighting the need for targeted data strategies beyond uniform augmentation.

Index Terms—Military camouflaged object detection, YOLO, data augmentation, synthetic data, style transfer, benchmark, lightweight, edge deployment, distribution-aware data refinement.

I. INTRODUCTION

Military camouflaged object detection (MCO) remains a challenging problem in computer vision due to the deliberate design of targets to blend with their surrounding environments. Unlike naturally camouflaged animals, military objects such as tanks, warships, and aeroplanes are concealed using engineered patterns and materials that suppress key visual features such as edges and textures, making them particularly difficult for automated detection systems to identify [1]. The YOLO family of detectors offers a strong balance between accuracy and inference speed, with lightweight variants especially attractive for deployment on resource-constrained embedded platforms such as unmanned aerial vehicles (UAVs) and handheld devices. However, their application to MCO is hindered by the scarcity of annotated training data, significant class imbalance, and the visual complexity introduced by camouflage.

To address data scarcity, a series of prior works have progressively developed augmentation pipelines for the MHCD2022 benchmark [1], the only publicly available dataset

for this task. Truong et al. [2] demonstrated that 3D-rendered synthetic images can improve detection, and evaluated style transfer for bridging the synthetic-to-real domain gap. [3] extended this to all five classes with SSIM-based quality filtering, and [4] proposed the 3DSMCOS pipeline covering all classes with distractor-augmented rendering. While these works individually demonstrated the value of each strategy, two key gaps remain: the combined impact across multiple YOLO variants has not been systematically studied, and no prior work has addressed the quality of synthetic data - using the full 3D set regardless of how well individual samples align with the real-image distribution.

In this paper, we address both gaps through a systematic benchmark study on military camouflaged object detection, targeting edge deployment on resource-constrained platforms. Our main contributions are as follows:

- We propose an object-level distribution-aware synthetic data refinement framework that selects 3D-generated camouflage samples based on Mahalanobis distance in the YOLO feature space, improving alignment with real and style-transferred images without additional annotation cost.
- We provide the systematic benchmark of six modern lightweight YOLO variants on MHCD under seven augmentation configurations, covering mAP50 and mAP50-95 across all five object classes with explicit consideration of edge deployment constraints.
- We demonstrate that feature-based selection consistently outperforms using the full synthetic set, and provide practical recommendations for model and data strategy selection in resource-constrained military camouflage detection.

II. RELATED WORK

A. Military Camouflaged Object Detection

Military camouflaged object detection (MCO) focuses on identifying and localizing personnel and equipment deliberately concealed using camouflage techniques. Unlike general object detection, MCO must contend with engineered suppression of visual cues - edges, textures, and color contrast - making it significantly harder for standard detectors to generalize. Early deep learning research paid little attention to this domain, focusing instead on naturally camouflaged animals.

Liu et al. [1] addressed this gap by introducing MHCD2022, the first and, to the best of our knowledge, only publicly available benchmark for multi-class military camouflaged object detection, spanning five categories across 3,000 images.

Subsequent works have explored several directions to advance MCO. Peng et al. [5] proposed an attention mechanism-based network with pyramidal feature reduction for detecting camouflaged soldiers. Hwang et al. [6] investigated combining military and non-military datasets to improve detection robustness. Zhou et al. [7] integrated neural style transfer to enrich training data for military vehicle detection. Other works have explored synthetic data generation [4] and transformer-based architectures [8] to improve accuracy across diverse military environments.

Despite these advances, existing approaches still face notable limitations: detection performance remains modest on minority classes, object diversity across categories is limited, and the degree of camouflage realism in synthetic data varies widely. These gaps motivate the systematic augmentation benchmark presented in this paper.

B. Augmentation Pipelines for MHCD2022

Two lines of prior work have developed augmentation strategies specifically for the MHCD2022 dataset, differing fundamentally in their data sources and mechanisms.

The first line generates synthetic training images through 3D rendering. Truong et al. [4] proposed the 3DSMCOS pipeline, which renders 10 distinct 3D military models per class in BlenderProc across multiple terrain environments (jungle, desert, snow) under varied camera positions and lighting conditions. Camouflage textures are applied automatically based on background color similarity, and bounding box labels are derived from segmentation masks generated during rendering, eliminating manual annotation cost.

The second line augments real training images through style transfer. Truong et al. [3] extracted 366 images containing instances of the four minority classes from the MHCD training set and stylized them using 10 weather and seasonal reference images (sunshine, rain, snow). Multiple style transfer methods were evaluated with the best-performing method selected per class based on SSIM histograms. After filtering at an SSIM threshold of 0.75, 2,680 high-quality augmented images were retained.

C. YOLO Family of Detectors

The YOLO family [9] has become the dominant paradigm for real-time object detection, with each generation introducing architectural innovations while maintaining a favorable accuracy-efficiency trade-off. Each version provides a range of model scales - typically from nano (n) to extra-large (x) - allowing deployment across diverse computational budgets.

In this work, we focus on the smallest variant of each version - nano (n) for YOLOv8, YOLOv10, YOLO11, YOLO12, and YOLO26, and tiny (t) for YOLOv9 - as these best represent the computational constraints of edge deployment on some mobility devices such as UAV and other embedded platforms.

III. PROPOSED METHOD

A. Overview

In this work, we aim to improve military camouflage object detection performance by refining the synthetic 3D dataset according to the feature distribution of real and style-transferred images. Let \mathcal{R} , \mathcal{S} , and \mathcal{D} denote the real dataset, style-transferred dataset, and synthetic 3D dataset, respectively. The objective of the proposed method is to select an optimal subset $\mathcal{D}^* \subseteq \mathcal{D}$ whose feature distribution is statistically consistent with that of the real and style-transferred data in the embedding space of the YOLO detector. The selected dataset is then used to train the object detection model by minimizing the detection loss:

$$\theta^* = \arg \min_{\theta} \mathcal{L}_{det}(\mathcal{R} \cup \mathcal{S} \cup \mathcal{D}^*; \theta), \quad (1)$$

where θ denotes the parameters of the detector and \mathcal{L}_{det} represents the detection loss. To reduce the domain gap between synthetic and real data, the subset \mathcal{D}^* is selected by minimizing the Mahalanobis distance in the feature space:

$$\mathcal{D}_c^* = \arg \min_{\mathcal{D}'_c \subseteq \mathcal{D}_c} \sum_{x_i \in \mathcal{D}'_c} d_M(f(x_i), \mu_c, \Sigma_c), \quad (2)$$

where $f(x_i)$ denotes the feature embedding of the synthetic sample x_i , and μ_c and Σ_c are the mean vector and covariance matrix estimated from the real and style-transferred samples of class c , respectively. d_M is the Mahalanobis distance.

Figure 1 illustrates the proposed 2-phase-framework for training data preparation, which are Military Camouflaged Synthetic Data Generation (Phase 1) and Object-level Distribution-Aware Sample Selection (Phase 2). Given the MHCD baseline, a pool of 3D synthetic images, and a set of weather-stylized real images created with the two strategies [4] and [3], respectively, and summarized in Phase 1, the proposed framework focuses on Phase 2 to select the subset of synthetic samples whose appearance distribution is most aligned with the real-image domain. The selected synthetic data is then combined with the stylized real images to form the final augmented training set. Selection is performed independently per class to account for the different representational quality of 3D models across categories. The proposed approach's psuedo code is listed in Algorithm 1.

B. Object-Level Feature Extraction

For each training image, individual object instances are cropped according to their bounding box annotations and resized to 224×224 pixels. Features are extracted using a pretrained backbone (i.e, we utilized YOLO26x's backbone feature extraction), which serve as a fixed feature extractor. Global average pooling is applied to the resulting feature maps to produce a compact feature vector $\mathbf{f} \in \mathbb{R}^d$ per object instance. This object-level representation captures appearance and texture cues relevant to camouflage detection while remaining class-agnostic in the embedding space.

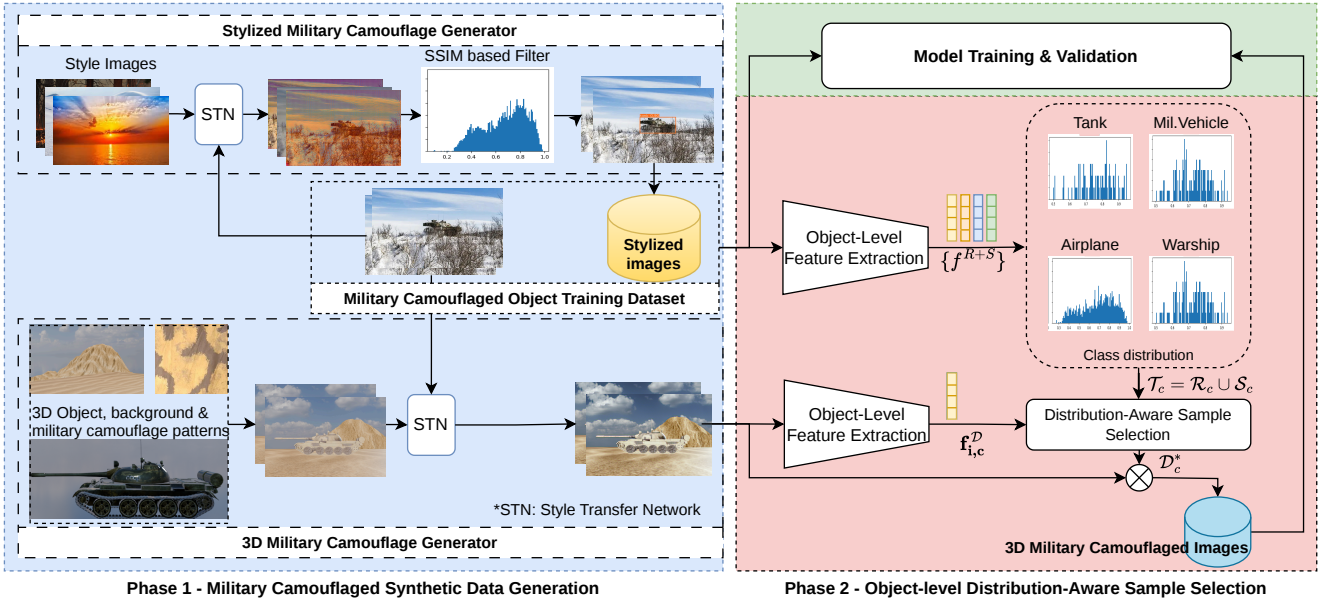


Fig. 1: The proposed Distribution-Aware Synthetic Data Refinement framework for the Military Camouflaged Object Detection.

C. Class Distribution Modeling

For each class c , a target feature distribution is constructed by aggregating object-level feature embeddings extracted from two complementary sources: real training images (\mathcal{R}_c) and stylized images (\mathcal{S}_c) generated by the Stylized Military Camouflage Generation pipeline in Phase 1 [3]. The combined feature set $\mathcal{T}_c = \mathcal{R}_c \cup \mathcal{S}_c$ is used to approximate the desired appearance manifold for class c , capturing both the semantic characteristics of real camouflage objects and the domain-adaptive visual variations introduced by style transfer. Based on \mathcal{T}_c , the class-wise feature distribution is modeled using an empirical multivariate Gaussian distribution parameterized by the mean vector μ_c and covariance matrix Σ_c . These statistical representations are subsequently employed to measure the distributional similarity between synthetic 3D samples and the target real-style feature manifold in the embedding space.

D. Distribution-Aware Sample Selection

We adopt the Mahalanobis distance [10] as the alignment metric. Mahalanobis distance accounts for the covariance structure of the target distribution, making it sensitive to the directions of greatest variance in the real-image feature space. This property is particularly important for camouflage detection, where discriminative features may occupy a low-dimensional subspace within the high-dimensional embedding.

Given a synthetic image, which is collected from the 3D Military Camouflage Generator in Phase 1, containing an instance of class c with feature vector $\mathbf{f}_{i,c}^D$, the Mahalanobis distance to the target distribution is computed as:

$$d_M(\mathbf{f}_{i,c}^D) = \sqrt{(\mathbf{f}_{i,c}^D - \mu_c)^\top \Sigma_c^{-1} (\mathbf{f}_{i,c}^D - \mu_c)} \quad (3)$$

A lower distance indicates higher alignment with the real-image distribution. All synthetic instances of class c are ranked by d_M , and the top- $k\%$ most aligned source images are retained, where $k \in \{25, 50, 75\}$.

Selection is performed at the source image level: an image is retained if any of its instances ranks in the top- $k(\%)$ for its respective class. The final augmented dataset combines the selected synthetic images with the full MHCD+Style set.

The full procedure is summarized in Algorithm 1.

IV. EXPERIMENTS

A. Datasets

All experiments use the MHCD2022 benchmark [1], the first publicly available dataset for military camouflaged object detection. It contains 3,000 images across five categories: person, military vehicle, tank, aeroplane, and warship, split into 2,400 training, 480 validation, and 600 test images. Figure 2 details the instance distribution per split, revealing severe class imbalance - person accounts for 76% of training instances while military vehicle (145) and warship (129) are critically underrepresented.

We evaluate seven training configurations organized into two groups. Base configurations use the full augmented datasets without selection:

- MHCD: the unmodified baseline using only the original 2,400 training images.
- MHCD+3D: augments the baseline with 3D synthetic images generated by the 3DSMCOS pipeline [4], covering all five classes across diverse terrain environments and lighting conditions.
- MHCD+Style: augments with weather-stylized real images produced by the pipeline in [3], retaining 2,680 high-

Algorithm 1 Distribution-Aware Synthetic Data Selection

Require: Image set $\mathcal{R}, \mathcal{S}, \mathcal{D}$ with labels; YOLO26x backbone ϕ (frozen); selection ratio k

Ensure: Selected synthetic feature set \mathcal{D}^*

```

1: // Step 1: Extract object-level features
2:  $\mathcal{R}_c, \mathcal{S}_c, \mathcal{D}_c \leftarrow \emptyset \quad \forall c$ 
3: for each source  $\mathcal{I} \in \{\mathcal{R}, \mathcal{S}, \mathcal{D}\}$  do
4:   for each image  $I \in \mathcal{I}$  with label  $(c, x, y, w, h)$  do
5:     Crop object region  $o \leftarrow \text{crop}(I, x, y, w, h)$ 
6:      $\mathbf{f} \leftarrow \text{GAP}(\phi(o)) \in \mathbb{R}^d$ 
7:     Append  $\mathbf{f}$  to  $\mathcal{R}_c, \mathcal{S}_c$ , or  $\mathcal{D}_c$  accordingly
8:   end for
9: end for
10: // Step 2: Model target distribution per class
11:  $\mathcal{D}^* \leftarrow \emptyset$ 
12: for each class  $c$  do
13:    $\mathcal{T}_c \leftarrow \mathcal{R}_c \cup \mathcal{S}_c$ 
14:   Estimate  $\mu_c, \Sigma_c^{-1}$  from  $\mathcal{T}_c$ 
15: // Step 3: Rank and select 3D synthetic samples
16: for each  $\mathbf{f} \in \mathcal{D}_c$  do
17:    $d_M(\mathbf{f}) \leftarrow \sqrt{(\mathbf{f} - \mu_c)^\top \Sigma_c^{-1} (\mathbf{f} - \mu_c)}$ 
18: end for
19: Sort  $\mathcal{D}_c$  by  $d_M$  ascending
20:  $n \leftarrow \lfloor k \cdot |\mathcal{D}_c| \rfloor$ 
21:  $\mathcal{D}^* \leftarrow \mathcal{D}^* \cup \{\mathbf{f}_i\}_{i=1}^n$ 
22: end for
23: return  $\mathcal{D}^*$ 

```

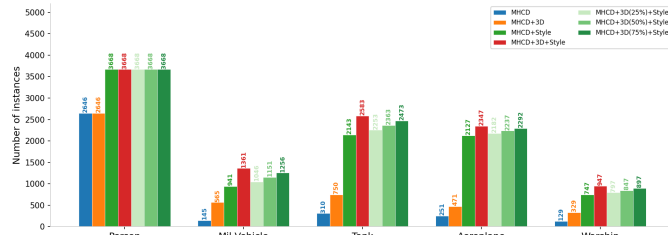


Fig. 2: Per-class instance counts in the training set across configurations. Both augmentation strategies reduce class imbalance, with MHCD+Style providing the largest absolute gains.

quality samples after SSIM-based filtering.

- **MHCD+3D+Style:** combines the full MHCD+3D and MHCD+Style sets.

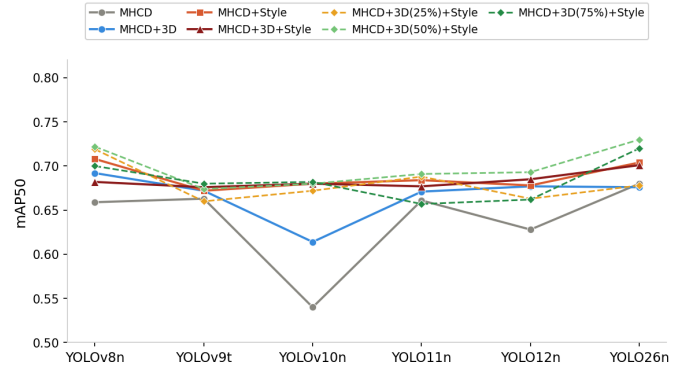
The second group evaluates the proposed selection-based augmentation strategy. The **MHCD+3D(k%)+Style** configuration retains the top k% of the most aligned images, where k is 25%, 50% and 75%, respectively.

B. Implementation Details

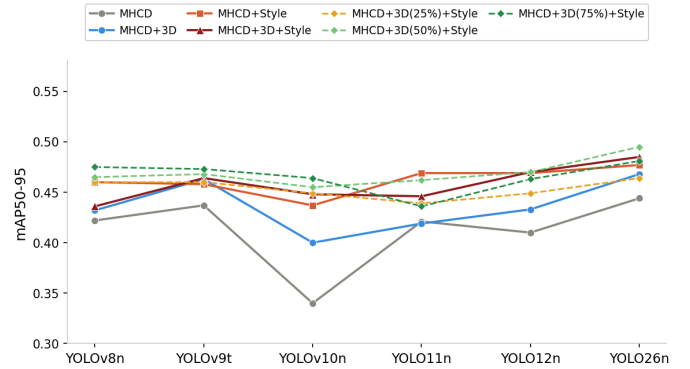
All models are trained using the Ultralytics framework with pretrained COCO weights, fine-tuned for 50 epochs at 640×640 resolution on a single NVIDIA GeForce RTX 5090 GPU with 4 data loading workers. The best checkpoint (highest validation mAP50) is selected for evaluation on the fixed test set of 600 images using the standard COCO evaluation

protocol. For the feature-based selection method, object crops are extracted using a pretrained YOLO26x backbone as a fixed feature extractor, and the empirical covariance is estimated per class from the combined real and style-transferred feature set.

C. Overall Results



(a) mAP50



(b) mAP50-95

Fig. 3: Overall detection performance of six YOLO variants across four augmentation configurations.

Figure 3 presents the overall mAP50 and mAP50-95 (averaged across all five classes) for each model-dataset combination. Table I gives the full per-class breakdown. Several observations stand out:

- **Every model improves over the MHCD baseline** when augmentation is applied - no model regresses on overall mAP50 under any augmented configuration, confirming the consistent benefit of data augmentation regardless of strategy.
- **Feature-based selection consistently outperforms full 3D augmentation.** The three selection-based configurations achieve higher overall mAP50 than MHCD+3D+Style for the majority of models, suggesting that quality of synthetic data matters more than quantity.
- **MHCD+3D(50%)+Style achieves the strongest results overall**, with YOLO26n reaching the highest mAP50 of 0.737 across the entire benchmark. This indicates a sweet spot where sufficient geometric diversity is retained while low-quality synthetic samples are filtered out.
- **The improvement is larger on mAP50-95 than on**

mAP50, suggesting that augmentation - particularly selection-based strategies - benefits localization quality in addition to detection recall.

The detailed per-class results in Table I reveal which classes drive these gains.

D. Distribution-Aware Synthetic Data Refinement

The selection-based configurations in Table I directly evaluate the impact of distribution-aware filtering on detection performance. The results confirm the following:

- **Alignment-based filtering outperforms naive combination.** MHCD+3D(50%)+Style achieves higher overall mAP50 than MHCD+3D+Style on all six evaluated models, demonstrating that domain-distant synthetic samples dilute rather than complement real-image augmentation.
- **The optimal ratio varies by model**, ranging from 25% to 75%, suggesting that model capacity influences how much synthetic diversity can be effectively absorbed.
- **50% selection is the most robust practical choice**, being the best or near-best configuration across all models and confirming that a balanced trade-off between quantity and alignment quality yields the most consistent gains.

These results validate the core premise of the proposed framework: filtering synthetic data by feature-space alignment leads to more effective augmentation than volume alone.

E. Efficiency-Accuracy Trade-off Analysis

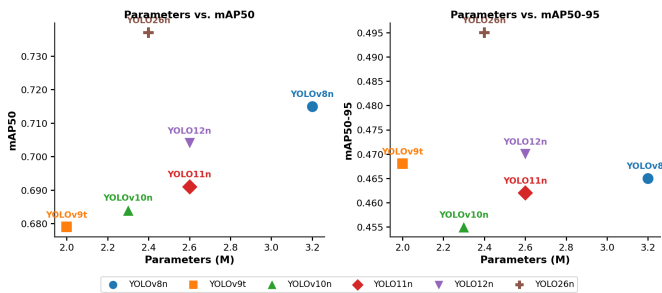


Fig. 4: Parameter efficiency of six lightweight YOLO variants trained on MHCD+3D(50%)+Style. Left: parameters vs. mAP50. Right: parameters vs. mAP50-95.

Figure 4 plots mAP50 and mAP50-95 against the number of parameters for all six models trained on MHCD+3D(50%)+Style. YOLO26n achieves the highest accuracy on both metrics with only 2.4M parameters – the fewest among all variants – demonstrating that its architecture extracts more discriminative features without increasing model size. YOLOv8n, despite having the most parameters (3.2M), ranks second on mAP50 but falls behind on mAP50-95, suggesting weaker localization quality relative to its model size. YOLOv10n consistently underperforms relative to its parameter count on both metrics, making it the least parameter-efficient choice in this benchmark.

F. Training Dynamics

Figure 5 shows the training loss, and validation loss curves of YOLO26n across seven configurations over 50 epochs.

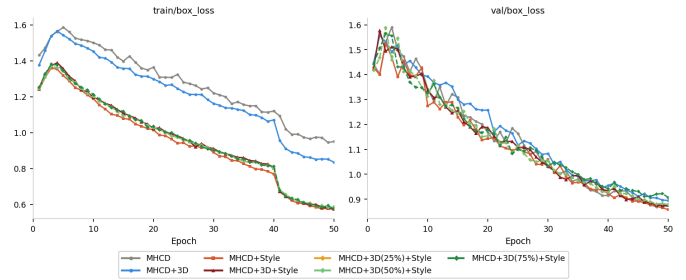


Fig. 5: Training dynamics of YOLO26n across seven augmentation configurations. From left to right: training box loss, validation box loss.

All configurations share the same fixed validation set of 480 images from the original MHCD split.

- **MHCD+3D alone exhibits the highest validation loss**, confirming a domain gap that real-image augmentation effectively compensates for.
- **Selection-based configurations follow convergence curves closely aligned with the full augmented sets**, confirming that feature-based filtering does not impair training dynamics.

V. CONCLUSION

We proposed a distribution-aware synthetic data refinement framework that selects 3D-generated military camouflage samples via Mahalanobis distance in the YOLO feature space, and evaluated it across six lightweight YOLO variants under multiple augmentation configurations targeting edge deployment. Feature-based selection consistently outperforms using the full synthetic set, confirming that alignment quality matters more than quantity. Real-image style transfer remains the primary driver of per-class gains for underrepresented classes, while selected synthetic data contributes complementary geometric diversity. Trade-off analysis across models provides practical guidance for balancing accuracy and computational cost under resource-constrained deployment.

A limitation of the current approach is that style transfer is applied uniformly across all classes, which proves beneficial for rare categories but introduces mild distribution shift for well-represented ones such as Person. Extending the selection framework to style-transferred images - applying distribution-aware filtering to retain only high-quality stylized samples per class - is a promising direction that could further improve data quality while mitigating the adverse effect on dominant categories.

REFERENCES

- [1] M. Z. Liu and X. Di, “Extraordinary mhnet: Military high-level camouflage object detection network and dataset,” *Neurocomputing*, vol. 549, p. 126 466, 2023.
- [2] T.-T.-H. Truong et al., “Impact of style transfer approaches on synthetic data for military camouflaged object detection,” in *Information and Communication Technology*, Springer Nature Singapore, 2025, pp. 27–41.

TABLE I: Comparison of mAP50 / mAP50-95 (mAP*) by different YOLO variants on MHCD and augmented datasets. **Bold** = best, *italic* = second best per model per class.

Model	Dataset	All		Person		Mil.Veh.		Tank		Aeroplane		Warship	
		mAP50	mAP*	mAP50	mAP*	mAP50	mAP*	mAP50	mAP*	mAP50	mAP*	mAP50	mAP*
YOLOv8n	MHCD	0.659	0.422	0.856	0.542	0.423	0.265	0.790	0.507	0.558	0.355	0.668	0.441
	+3D	0.692	0.432	0.845	0.531	0.350	0.195	<i>0.776</i>	0.527	0.630	0.347	0.861	0.561
	+Style	0.708	0.460	<i>0.850</i>	<i>0.533</i>	<i>0.478</i>	<i>0.284</i>	0.717	0.488	0.710	0.453	0.787	0.542
	+3D+Style	0.682	0.436	0.841	0.520	0.424	0.249	0.722	0.494	0.643	0.422	0.778	0.495
	+3D(25%)+Style	0.707	0.458	0.827	0.517	0.453	0.275	0.756	0.497	0.697	0.470	0.802	0.531
	+3D(50%)+Style	0.715	<i>0.470</i>	0.838	0.526	0.519	0.301	0.758	0.482	0.674	0.512	0.784	0.531
	+3D(75%)+Style	0.705	0.475	0.830	0.510	0.413	0.283	0.742	<i>0.515</i>	0.711	0.501	<i>0.827</i>	0.564
YOLOv9t	MHCD	0.663	0.437	0.864	0.547	0.347	0.211	0.798	<i>0.544</i>	0.594	0.408	0.714	0.472
	+3D	0.672	0.463	<i>0.851</i>	<i>0.530</i>	0.409	<i>0.294</i>	0.750	0.540	0.581	0.415	0.768	0.536
	+Style	0.672	0.458	0.841	0.517	0.444	0.269	0.723	0.522	0.640	0.477	0.710	0.504
	+3D+Style	0.676	0.464	0.848	0.514	0.395	0.270	0.766	0.524	0.666	0.500	0.706	0.514
	+3D(25%)+Style	0.666	0.465	0.846	0.523	<i>0.425</i>	0.294	0.664	0.484	0.710	0.558	0.685	0.466
	+3D(50%)+Style	0.679	0.468	0.842	0.525	0.372	0.256	0.706	0.504	0.642	0.475	0.832	0.580
	+3D(75%)+Style	0.682	0.473	0.830	0.517	0.382	0.275	<i>0.776</i>	0.575	<i>0.706</i>	0.477	0.715	0.520
YOLOv10n	MHCD	0.540	0.340	0.812	<i>0.494</i>	0.201	0.116	0.661	0.460	0.441	0.273	0.582	0.358
	+3D	0.614	0.400	0.812	0.488	0.325	0.215	0.656	0.457	0.520	0.327	0.759	0.512
	+Style	0.680	0.437	<i>0.824</i>	0.501	0.421	0.280	0.679	0.451	0.722	0.445	0.757	<i>0.507</i>
	+3D+Style	0.680	0.448	0.799	0.489	0.421	0.290	0.738	0.512	0.630	0.460	0.811	0.493
	+3D(25%)+Style	0.673	0.448	0.827	<i>0.494</i>	0.441	0.295	0.732	0.525	0.648	0.445	0.717	0.483
	+3D(50%)+Style	<i>0.684</i>	0.455	0.815	0.487	0.413	0.302	0.725	<i>0.520</i>	0.761	0.500	0.703	0.468
	+3D(75%)+Style	0.687	0.464	0.812	0.490	0.352	0.289	0.707	0.499	0.805	0.541	0.760	0.498
YOLO11n	MHCD	0.661	0.421	0.854	0.539	0.418	0.268	0.780	0.508	0.563	0.356	0.691	0.433
	+3D	0.671	0.419	0.852	0.523	0.402	0.268	0.768	0.538	0.575	0.325	0.757	0.443
	+Style	0.684	0.469	0.841	0.512	0.464	0.308	0.735	<i>0.527</i>	0.616	0.454	0.763	0.541
	+3D+Style	0.677	0.446	0.836	0.507	0.448	<i>0.281</i>	0.730	0.497	0.614	0.396	0.760	0.550
	+3D(25%)+Style	0.692	0.448	0.825	0.501	0.378	0.256	0.737	0.478	0.755	0.452	0.764	<i>0.554</i>
	+3D(50%)+Style	0.691	<i>0.462</i>	0.830	0.505	0.416	0.267	0.712	0.505	0.721	<i>0.486</i>	0.773	0.547
	+3D(75%)+Style	0.654	0.436	0.848	0.512	0.351	0.223	0.685	0.468	0.579	0.423	0.805	<i>0.554</i>
YOLO12n	MHCD	0.628	0.410	0.849	0.538	0.364	0.247	0.761	0.487	0.512	0.361	0.654	0.416
	+3D	0.677	0.433	0.854	<i>0.533</i>	0.393	0.246	0.745	0.519	0.610	0.365	0.784	0.503
	+Style	0.678	0.469	0.826	0.514	0.432	0.269	0.758	<i>0.537</i>	0.645	0.507	0.726	0.520
	+3D+Style	0.685	0.470	0.831	0.508	0.423	<i>0.279</i>	<i>0.765</i>	0.533	0.653	0.465	0.756	0.564
	+3D(25%)+Style	0.687	0.455	0.845	0.512	0.436	0.303	0.762	0.522	0.647	0.478	0.748	0.458
	+3D(50%)+Style	<i>0.704</i>	0.470	0.833	0.506	0.435	0.259	0.734	0.512	<i>0.739</i>	0.552	0.777	0.522
	+3D(75%)+Style	0.717	0.463	0.837	0.517	0.411	0.239	0.770	0.541	0.760	<i>0.537</i>	0.809	0.478
YOLO26n	MHCD	0.680	0.444	0.848	0.541	0.395	0.284	0.703	0.460	0.657	0.427	0.798	0.508
	+3D	0.676	0.468	<i>0.846</i>	<i>0.532</i>	0.362	0.256	0.740	0.529	0.634	0.451	0.796	0.572
	+Style	0.704	0.477	0.843	0.526	0.390	0.247	0.724	0.510	0.844	0.600	0.718	0.502
	+3D+Style	0.701	0.485	0.834	0.514	0.431	0.298	<i>0.746</i>	<i>0.541</i>	0.745	0.560	0.748	<i>0.512</i>
	+3D(25%)+Style	0.694	0.477	0.839	0.527	0.352	0.256	0.725	0.516	0.885	0.620	0.669	0.466
	+3D(50%)+Style	0.737	0.495	0.831	0.524	<i>0.430</i>	0.296	0.792	0.554	0.871	0.603	0.762	0.499
	+3D(75%)+Style	0.701	0.481	0.826	0.510	0.420	0.307	0.712	0.511	0.834	0.611	0.714	0.464

- [3] T. T. H. Truong and T. K. Tran, “A style transfer-based augmentation approach for detecting military camouflaged object,” *Journal of Military Science and Technology*, no. CSCE8, pp. 44–54, Dec. 2024.
- [4] T. T. H. Truong, T. K. Tran, N.-B. Hoang, T.-D. Nguyen, T. H. H. Phan, and C. T. Nguyen, “3dsmcos: A 3d model-based synthetic data pipeline for military camouflaged object segmentation with distractor-augmented realism,” in *2025 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 2025, pp. 1–6.
- [5] Y. Peng, J. Wang, Z. Yu, Y. You, and Y. Sun, “Camouflage soldier object detection network based on the attention mechanism and pyramidal feature shrinking,” *Multimedia Tools and Applications*, vol. 83, no. 33, pp. 79 917–79 938, Mar. 2024.
- [6] K.-S. Hwang and J. Ma, “Military camouflaged object detection with deep learning using dataset development and combination,” *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 2024.
- [7] Y. Zhou, J. Gong, W. Lu, N. Gu, K. Chong, and Z. Wang, “An object detection algorithm for military vehicles based on image style transfer and domain adversarial learning,” in *2022 IEEE International Conference on Unmanned Systems (ICUS)*, IEEE, 2022, pp. 1059–1064.
- [8] B. Li, R. Zhou, L. Yang, Q. Wang, and H. Chen, “Mildetr: Detection transformer for military camouflaged target detection,” *IEEE Access*, vol. 12, pp. 26 163–26 174, 2024.
- [9] Ultralytics, *YOLO by Ultralytics*, Accessed: 2025, 2025. [Online]. Available: <https://github.com/ultralytics/ultralytics>.
- [10] P. C. Mahalanobis, “On the generalized distance in statistics,” 1936.