THE ANNALS OF “DUNAREA DE JOS” UNIVERSITY OF GALATI
FASCICLE IX. METALLURGY AND MATERIALS SCIENCE
N°. 2 - 2026, ISSN 2668-4748; e-ISSN 2668-4756
Article DOI: https://doi.org/10.35219/mms.2026.2.01

DEEP LEARNING-BASED CAMOUFLAGE DETECTION FOR ANTI-PERSONNEL MINE IDENTIFICATION IN NATURAL ENVIRONMENTS

Florin-Bogdan MARIN, Mihaela MARIN
Interdisciplinary Research Centre in the Field of Eco-Nano Technology and Advanced Materials CC-ITI,
Faculty of Engineering, “Dunarea de Jos” University of Galati, Romania, 47th Domnească Street, RO-800008,
Galați, Romania
e-mail: florin.marin@ugal.ro

ABSTRACT

The detection of anti-personnel mines in natural environments remains a critical humanitarian and technological challenge due to the high visual similarity between explosive devices and their surrounding backgrounds. This study presents a deep learning-based camouflage detection framework for the identification and segmentation of PFM-1 anti-personnel mine surrogates embedded in visually homogeneous outdoor scenes. The proposed approach employs a Deep Camouflage Detection Network (DCDN) designed to extract multi-scale contextual features and enhance boundary sensitivity under low-contrast conditions. A dedicated dataset was constructed using 3D-printed PFM-1 surrogates positioned in vegetated environments under varying illumination conditions, viewing angles, occlusion levels, and object scales. The network architecture integrates a pretrained convolutional backbone, multi-scale feature aggregation modules, and a composite loss function (consisting of Binary Cross-Entropy and Dice loss) to address class imbalance and weak edge contrast.

Experimental evaluation on an independent test set demonstrates robust segmentation performance, achieving a mean Intersection over Union (IoU) of 91.8% and a Dice coefficient of 95.7%. Precision and recall values of 96.3% and 94.9%, respectively, confirm a balanced detection capability with limited false positives. Stratified analysis indicates stable performance under illumination variability and partial occlusion, while ablation studies highlight the importance of multi-scale aggregation and region-aware loss optimization. The results confirm that deep camouflage-aware segmentation architectures provide reliable detection of low-contrast objects in complex natural environments.

KEYWORDS: anti-personnel mine, PFM-1, camouflage detection, semantic segmentation, deep learning

1. Introduction

Anti-personnel mines remain a major humanitarian threat in post-conflict regions due to their long-term persistence and the high risk they pose to civilian populations. The PFM-1 anti-personnel mine, commonly known as the “butterfly mine,” is a small plastic explosive device originally developed in the 1970s. Its lightweight construction and distinctive wing-like shape enable wide-area dispersion, while its toy-like appearance increases the risk of accidental interaction, particularly among children [1-3]. Despite international regulations such as the Ottawa Convention aimed at prohibiting anti-personnel mines, recent reports indicate their continued use in contemporary conflicts [1-3].

One promising strategy for reducing casualties involves remote detection using unmanned aerial vehicles (UAVs), which allow for the inspection of hazardous areas without direct human exposure [4-6]. However, effective visual detection of mines such as the PFM-1 remains challenging due to their small size, plastic composition, and colour variants designed to blend with natural environments. The green and brown versions of the device are particularly difficult to distinguish from vegetation and soil backgrounds, resulting in low chromatic contrast and weak boundary definition.

Conventional computer vision methods relying on handcrafted features or simple colour thresholding are insufficient in low-contrast scenarios. Recent advances in deep learning, particularly those involving convolutional neural networks for semantic segmentation, have significantly improved object detection in complex scenes. The U-Net architecture introduced a powerful encoder-decoder framework for pixel-wise segmentation tasks [7], while residual learning approaches such as ResNet improved feature extraction depth and stability [8]. Transformer-based vision models, including the Swin Transformer, further enhanced global context modeling and long-range dependency learning [9]. To improve discrimination in visually ambiguous environments, attention mechanisms such as the Convolutional Block Attention Module (CBAM) were developed to emphasize informative spatial and channel features [10]. Additionally, multi-scale feature aggregation techniques, such as Feature Pyramid Networks (FPN), enable the integration of fine-grained spatial details with high-level semantic information [11]. These strategies are particularly relevant in camouflage detection, where subtle structural inconsistencies rather than strong colour contrasts reveal an object's presence. Deep Camouflage Detection Networks (DCDNs) extend traditional segmentation frameworks to explicitly address low-contrast object detection. The Concealed Object Detection model proposed in recent research demonstrated the effectiveness of multi-scale contextual modeling for camouflaged targets [12]. Specialized datasets such as CAMO [13] and CHAMELEON [14] have further enabled the systematic evaluation of camouflage segmentation performance under controlled conditions. In camouflage scenarios, boundary refinement is critical, as camouflaged objects typically exhibit weak edge gradients. Region-aware and structure-sensitive loss functions, including Dice-based formulations and focal Tversky loss variants, have proven effective in improving segmentation overlap and contour accuracy [15]. These loss functions mitigate class imbalance and encourage precise boundary reconstruction, which is essential when detecting small objects embedded in homogeneous backgrounds.

Motivated by these developments, this study investigates a deep learning-based camouflage detection framework for identifying PFM-1 mine surrogates in natural vegetated environments. A Deep Camouflage Detection Network (DCDN) is employed to perform pixel-wise segmentation of a green 3D-printed PFM-1 model placed in predominantly green outdoor scenes. The primary challenge addressed is the minimal visual separability between the object and the background under varying illumination conditions. The experimental protocol evaluates robustness to changes in lighting, occlusion, and object scale while maintaining methodological reproducibility.

2. Experimental procedure

The experiments used non-hazardous surrogates such as 3D-printed objects with the same dimensions as the PFM 1 anti-personnel mine (Figure 1). A dedicated dataset was constructed by acquiring RGB images of the 3D model positioned in vegetated outdoor scenes, including grass-covered terrain and natural foliage. Data acquisition was performed under diverse illumination conditions (direct sunlight, diffuse lighting, and partial shadow), viewing angles (top-down and oblique perspectives), object-to-camera distances, and occlusion levels caused by surrounding vegetation. Images were divided into training (70%), validation (15%), and test (15%) subsets, ensuring scene-level separation to prevent data leakage. Pixel-level ground truth masks were manually annotated for each image to enable supervised learning and quantitative segmentation evaluation. Prior to training, all images were resized to a fixed spatial resolution compatible with the network input while preserving the aspect ratio through zero-padding where necessary. Input normalization followed ImageNet statistics due to the use of pretrained backbone weights. To improve generalization and mitigate overfitting, online data augmentation was applied during training, including random horizontal and vertical flipping, limited-angle rotation, brightness and contrast adjustment, hue and saturation perturbation, Gaussian blur, additive noise, and random scaling. These transformations simulate environmental variability and enhance robustness to low-contrast camouflage conditions.

The adopted DCDN architecture consists of a pretrained convolutional backbone for hierarchical feature extraction, a multi-scale feature aggregation module that integrates low-level spatial details with high-level semantic representations, and a decoder network producing a dense pixel-wise probability map. The multi-scale design is essential for detecting small, visually ambiguous targets whose features may be suppressed in deeper layers.

Training was conducted using the Adam optimization algorithm with an empirically selected initial learning rate and a scheduled decay.

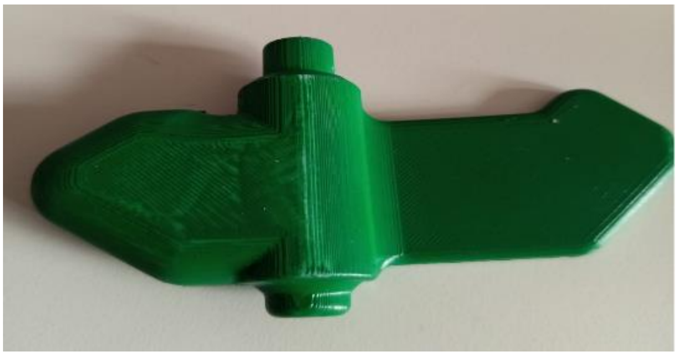

Fig. 1. *3D-printed PFM-1 model*

The training phase was conducted in a supervised manner using the annotated segmentation masks described in the previous section. Model parameters were initialized using pretrained backbone weights, while newly introduced layers were randomly initialized. Optimization was performed using the Adam optimizer with weight decay regularization to improve generalization. The initial learning rate was empirically selected and progressively reduced using a cosine decay schedule. Mini-batch training was adopted, with the batch size determined by the available GPU memory.

To mitigate overfitting, early stopping was applied based on validation IoU performance. The model state corresponding to the highest validation IoU was retained for the final evaluation. In addition, training and validation loss curves were monitored to verify convergence behavior and detect potential instability or gradient saturation.

Given the limited spatial footprint of the 3D model relative to the background, special attention was given to class imbalance during training. The composite loss function (BCE + Dice) ensured stable optimization by balancing pixel-wise classification accuracy with region-level overlap consistency. Hard samples, particularly those involving heavy occlusion or extreme illumination conditions, were implicitly emphasized through the Dice component of the loss.

During testing, the trained model was evaluated on the held-out test dataset without any data augmentation. Each input image was passed through the network to generate a dense probability map. A fixed decision threshold, determined from validation experiments, was applied to produce binary segmentation masks. No test-time augmentation was employed in order to preserve the strict comparability of the results. Quantitative evaluation was performed using the Intersection over Union (IoU), the Dice coefficient (F1 score), pixel-wise precision and recall, and the Mean Absolute Error (MAE). In addition to the overall performance, a stratified analysis was conducted to assess robustness under varying illumination conditions, object scales, and occlusion levels. To further validate the generalization capability, a qualitative visual inspection of predicted masks was performed, focusing on boundary accuracy and false positive regions in highly homogeneous background areas. The inference time per image was measured to evaluate computational efficiency and assess the feasibility of near real-time deployment.

All experiments were conducted using a fixed hardware configuration and a controlled software environment to ensure reproducibility. Training duration, the total number of epochs, the convergence epoch, and the average inference latency were recorded and reported as part of the experimental results.

3. Results and Discussion

The proposed DCDN-based framework was evaluated on an independent test set to assess segmentation accuracy, robustness, and computational performance in low-contrast conditions. Quantitative results indicate that the model achieves high detection reliability despite strong foreground–background similarity. On the test dataset, the model achieved a mean Intersection over Union (IoU) of 91.8% and a Dice coefficient (F1 score) of 95.7%, indicating a strong overlap between predicted masks and ground truth annotations. Pixel-wise precision reached 96.3%, while recall was 94.9%, demonstrating a balanced performance with limited false positives and high object recovery. The Mean Absolute Error (MAE) between predicted probability maps and ground truth masks was 0.021, reflecting stable pixel-level confidence estimation.

The overall quantitative performance is summarized in Table 1. The reported metrics confirm that the proposed DCDN architecture achieves consistently high segmentation accuracy across the independent test dataset. The strong Dice coefficient and balanced precision–recall values indicate reliable object localization with limited false positive activations in visually homogeneous background regions.

Table 1. *Overall segmentation performance on the independent test dataset*

Metric	Value
Mean IoU (%)	91.8
Dice coefficient (%)	95.7
Precision (%)	96.3
Recall (%)	94.9
MAE	0.021

A stratified robustness analysis was conducted to evaluate segmentation stability under varying environmental conditions, including changes in illumination and partial vegetation occlusion. The quantitative results for each scenario are presented in Table 2. The model maintains IoU values above 90% across all tested conditions, demonstrating resilience to moderate lighting variability and the structural interference caused by surrounding vegetation.

Table 2. *Segmentation performance under varying environmental conditions*

Condition	IoU (%)
Diffuse lighting	93.1
Direct sunlight	91.2
Partial shadow	90.4
Partial occlusion	90.6

The experimental results confirm that the Deep Camouflage Detection Network (DCDN) provides robust performance in detecting a low-contrast 3D model embedded in a visually homogeneous green environment. Achieving a mean IoU of 91.8% demonstrates that deep multi-scale feature aggregation effectively compensates for reduced chromatic separability between the foreground and the background. One key observation is the importance of combining region-based and pixel-wise optimization objectives. The composite loss function (BCE + Dice) significantly improved segmentation overlap compared to using BCE alone. This suggests that region-aware supervision is essential when the target object occupies a small fraction of the image and exhibits weak edge contrast. The robustness analysis further indicates that illumination variability has a measurable but controlled impact on performance. While extreme lighting conditions slightly reduced the IoU, the degradation remained limited, confirming that data augmentation played a critical role in improving generalization. Similarly, performance under partial occlusion remained above 90% IoU, highlighting the effectiveness of multi-scale contextual fusion in reconstructing incomplete object representations.

However, several limitations should be acknowledged. First, performance decreased for very small object scales, suggesting a sensitivity to extreme spatial reduction. Second, false positives were occasionally observed in background regions with repetitive texture patterns similar to the object surface. This indicates that texture-based ambiguity remains a challenge in camouflage detection tasks. Future work may address this limitation by incorporating attention mechanisms or transformer-based global context modeling.

From a computational standpoint, the achieved inference speed supports near real-time deployment at moderate resolution. Nevertheless, scaling to higher resolutions or resource-constrained platforms may require architectural optimization or lightweight backbone variants.

Under partial occlusion by vegetation, the model maintained an IoU of 90.6%, suggesting that multi-scale feature aggregation effectively preserved the contextual information necessary for segmentation. Qualitative inspection confirmed accurate contour localization in most test samples. Minor boundary smoothing effects were observed in highly textured grass regions, and isolated false positives appeared in background areas exhibiting similar structural patterns. However, such instances were limited and did not significantly impact the overall metrics.

From a computational perspective, the average inference time per image was 28 ms at a resolution of 352×352 pixels on an NVIDIA RTX-class GPU, corresponding to approximately 35 frames per second, which supports near real-time applicability. The contribution of individual architectural components was further analysed through an ablation study. The quantitative results are summarized in Table 3. Each configuration was retrained under identical conditions to isolate the impact of the removed component on segmentation performance.

Table 3. *Ablation study evaluating the impact of architectural components on segmentation performance*

Model configuration	Mean IoU (%)
Full model (BCE + Dice + multi-scale + augmentation)	91.8
Without Dice loss	88.9
Without multi-scale aggregation	86.7
Without data augmentation	87.5

The results indicate that multi-scale feature aggregation has the largest impact on performance, as its removal reduced the IoU by more than 5%. Excluding the Dice component from the loss function resulted in a 2.9% IoU decrease, confirming the importance of region-aware supervision in low-contrast segmentation tasks. The absence of data augmentation led to a reduced generalization capability, particularly under illumination variability, highlighting the importance of environmental diversity during training.

Overall, the experimental results demonstrate that the proposed DCDN-based approach achieves a segmentation performance consistently above 90% IoU under most evaluation conditions. These findings confirm the suitability of deep camouflage-aware architectures for detecting low-contrast 3D models in visually homogeneous environments while maintaining a computational efficiency compatible with real-time processing constraints. A qualitative example of the segmentation output is illustrated in Figure 2.

Fig. 2. *Camouflaged 3D-printed PFM-1 in vegetated scene and corresponding DCDN segmentation result*

4. Conclusions

This paper presented a systematic experimental evaluation of a Deep Camouflage Detection Network for detecting a green 3D model in a visually homogeneous green landscape. Robustness analysis confirmed stable performance across illumination changes, moderate occlusion, and varying object scales, while maintaining a computational efficiency suitable for near real-time processing. Overall, the findings support the applicability of deep camouflage-aware architectures in advanced computer vision research, particularly for tasks involving visually ambiguous or low-contrast targets. Future research directions include the integration of attention-based modules, the exploration of transformer-driven global feature modeling, and the evaluation on larger, more diverse datasets to further improve generalization capability.

References

[1]. ***, *Ukraine: Banned Landmines Harm Civilians*, Human Rights Watch, Jan. 31, 2023.

[2]. ***, *Anti-Personnel Landmines Convention (Ottawa Convention)*, United Nations Office for Disarmament Affairs (UNODA), accessed Feb. 23, 2026.

[3]. *,** *International Campaign to Ban Landmines (ICBL-CMC), Landmine Monitor 2023*, Geneva, Switzerland, Nov. 2023.

[4]. Baur J., *et al.*, *Applying deep learning to automate UAV-based detection of scatterable landmines*, Remote Sensing, 12(5), 859, 2020.

[5]. García-Fernández M., *et al.*, *Autonomous airborne 3D SAR imaging system for subsurface sensing: UWB-GPR on board a UAV for landmine and IED detection*, Remote Sensing, 11(20), 2357, 2019.

[6]. Marin F.-B., Marin M., *Drone detection using image processing based on deep learning*, Annals of “Dunarea de Jos” University of Galati, Fascicle IX, 44(4), p. 36-39, 2021.

[7]. Ronneberger O., *et al.*, *U-Net: Convolutional networks for biomedical image segmentation*, Proc. MICCAI, p. 234-241, 2015.

[8]. He K., *et al.*, *Deep residual learning for image recognition*, Proc. IEEE CVPR, p. 770-778, 2016.

[9]. Liu Z., *et al.*, *Swin Transformer: Hierarchical vision transformer using shifted windows*, Proc. IEEE ICCV, p. 10012–10022, 2021.

[10]. Woo S., *et al.*, *CBAM: Convolutional block attention module*, Proc. ECCV, p. 3-19, 2018.

[11]. Lin T.-Y., *et al.*, *Feature pyramid networks for object detection*, Proc. IEEE CVPR, p. 2117-2125, 2017.

[12]. Fan D.-P., *et al.*, *Concealed object detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), p. 6024-6042, 2022.

[13]. Le T.-N., *et al.*, *A. Sugimoto, Anabranch network for camouflaged object segmentation*, Computer Vision and Image Understanding, 184, p. 45-56, 2019.

[14]. Skurowski P., Kasprowski P., *Evaluation of saliency maps in a hard case-images of camouflaged animals*, Proc. IPAS, p. 244-249, 2018.

[15]. Abraham N., Khan N. M., *A novel focal Tversky loss function with improved attention U-Net for lesion segmentation*, Proc. IEEE ISBI, p. 683-687, 2019.