SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Deep SINET-based Framework for Camouflaged Object Detection in Military Environments with Backbone Performance Analysis

Pratibha Waghale[1*], Dr. Lalit Damahe[2*]

[1]Research scholar, Department of Computer Technology & Yeshwantrao Chavan College of Engineering Nagpur, Email id: pratibha201986@gmail.com, Orcid id:0009-0009-6254-009X
[2] Department of Computer Science and Engineering, Yeshwantrao Chavan College of Engineering Nagpur, Email id: lalitdamahe3379@gmail.com, Orcid id:0000-0003-3112-5570

ABSTRACT
In this paper, propose an updated and experimentally consistent Deep SINet-based architecture for the military camouflaged item detection task with the ACD1K dataset, which includes 1,000 annotated military camouflage photos with 700 training images and 300 testing images. In contrast to the wide application of SINet, our work focuses on a controlled backbone-performance examination, where VGG19, ResNet-50, and ResNet-101 are incorporated into the same SINet search-identification pipeline, trained under the same experimental conditions and assessed using the same metrics. The system consists of multi-level feature extraction, texture improvement, global reverse attention, non-local contextual discovery, and hybrid BCE-IoU supervision to improve the boundary localisation and region-level consistency. The experimental results demonstrate that the SINet based on ResNet-101 backbone has the best overall performance with IoU of 0.8564, structure-measure S_m of 0.9120, F-measure F_β of 0.9050, enhanced-alignment measure E_ϕ of 0.9310 and MAE of 0.0330. The sensitivity study on hyperparameters further demonstrates that the learning rate and weight of BCE–IoU loss can stably converge. The experimental results show the importance of controlled backbone selection for reliable COD performances in military surveillance settings.

Keywords: Camouflaged Object Detection, SINet, ResNet, VGG19, Military Surveillance

1.Introduction

Camouflaged Object Detection (COD) aims to segment objects that are visually concealed by their surrounding environment. In military surveillance, this problem is particularly important because soldiers, vehicles, equipment, and tactical objects are intentionally designed to merge with natural or artificial backgrounds. In contrast to conventional object detection, where targets are often distinguishable by strong colour, texture, or shape contrast, military COD must detect low-contrast objects with weak boundaries, partial occlusion, irregular scale, and high background clutter.

Early object detection methods were based on handcrafted descriptors such as SIFT, HOG, Haar-like features, and sliding-window classifiers [1–3]. Although such methods are useful for simple visual conditions, they are sensitive to illumination variation, scale changes, occlusion, and background complexity. Deep convolutional networks and modern detection architectures have improved object detection by learning hierarchical features, but generic object detectors and salient object detection methods usually assume that the foreground is visually distinctive. This assumption is weak in camouflaged scenes because the target and background have intentionally similar appearance.

Recent COD methods use multi-scale feature aggregation, contextual reasoning, reverse attention, feature fusion, and transformer-based representations to enhance concealed boundaries [4–9]. However, three limitations remain significant for military COD. First, many reported methods are evaluated under different datasets, input resolutions, and training protocols, which makes direct comparison difficult. Second, limited attention has been given to how different CNN backbones affect SINet-style search and identification behaviour under a controlled military dataset. Third, some COD studies simply present final data without hyperparameter sensitivity or metric-consistency analysis, which lowers the experimental transparency. Table 1 shows the enlarged comparison of modern deep-learning COD and military camouflage detection algorithms.

To address these voids, this work proposes a regulated Deep SINet-based COD architecture for military situations. The purpose is not to make a new SINet design claim but to give a rigorous domain-specific implementation and analysis of SINet with several backbone feature extractors under the same parameters. The work corrects metric definitions, employs a consistent split of the ACD1K dataset, and offers sensitivity assessments for the backbone, loss-function, module, computational and hyperparameters.

The major contributions can be summarised as:

1. For hidden target segmentation on ACD1K, a military COD architecture is proposed by integrating SINet with controllable backbone variants VGG19, ResNet-50, and ResNet-101.
2. Applied a consistent experimental approach where all backbone versions share the same train-test split, input resolution, optimiser, number of epochs, augmentation strategy, loss formulation, and evaluation metrics.
3. Utilized a component-level ablation to assess the impact of key SINet components, including texture improvement, global reverse attention, and non-local contextual discovery.
4. The sensitivity of hyperparameters, such as learning rate, BCE–IoU loss weight, and augmentation approach has been tested, and the stability trend of the model is shown graphically.

A corrected and bounded metric analysis is provided utilising IoU, S_m, F_β, E_ϕ, MAE, standard deviation and inference complexity.

2.Literature Review

Generally, COD approaches can be divided into four categories: handcrafted-feature methods, CNN-based methods, attention and context-based methods and transformer-based methods. Handcrafted techniques utilise texture, edge and local descriptors with classifiers like SVM or AdaBoost but they do not generalise effectively in complex military settings. CNN-based segmentation algorithms increased the feature representation by means of multi-scale convolutional learning. However, shallow CNN features may fail to capture global semantic context, and deep backbone may overfit small camouflage datasets under uncontrolled training.

SINet proposed a search-identification mechanism for camouflaged items detection and refinement [4]. Its search stage provides coarse localisation and identification stage refines object boundaries with attention-based refinement. Later COD models proposed feature pyramids, lateral aggregation, mutual graph learning, frequency domain aggregation and transformer based representations [5-9]. These approaches show that COD can be enhanced by both semantic context and local boundary factors.

Recent efforts in military situations include UAV pictures [10–14], infrared targets [17], multispectral photography [15], adversarial camouflage [16], and transformer based detection [17]. However many of these research are not directly comparable as they use distinct modalities, different image sizes, different datasets and different evaluation methodologies. The research gap addressed here is thus the lack of a consistent investigation of SINet-based backbones for RGB military camouflage segmentation. The proposed approach covers this gap by keeping all the training and testing parameters unchanged, and altering only the backbone and selected modules.

The proposed framework uses the ACD1K military camouflage dataset [18]. In this manuscript, the dataset description is standardized as 1,000 annotated RGB images, divided into 700 training images and 300 testing images. All images have corresponding binary ground-truth masks. The dataset includes different military camouflage conditions such as vegetation backgrounds, dry grass, low illumination, partial occlusion, scale variation, and weak target boundaries. Representative samples are shown in Figure. 1.

Table 1. Expanded comparison of recent deep-learning COD and military camouflage detection methods.

Year	Method	Dataset / Modality	Backbone	Input	Sm	Fβ	MAE	IoU	Limitation / observation
2022	Feature visualization deep model [10]	UAV military images	ResNet-50	512x512	0.82	0.85	0.014	0.78	Strong feature visualization, but not a controlled SINet backbone study.
2023	MHNet [11]	Military camouflage	VGG-16	640x640	0.85	0.87	0.012	0.82	Uses attention and fusion, but differs in dataset protocol and resolution.
2023	DAAC [12]	Adversarial camouflage	EfficientNet-B4	512x512	0.84	0.86	0.013	0.80	Focuses on adversarial camouflage robustness rather than backbone control.

2023	Multispectral DNNs [13]	Multispectral satellite data	DenseNet-121	1024x1024	0.83	0.86	0.011	0.81	Requires multispectral/high-resolution inputs; not directly comparable with RGB ACD1K.
2023	3X-FPN [14]	Infrared military targets	ResNeXt-101	640x640	0.86	0.88	0.010	0.83	Effective for infrared targets, but modality differs from RGB camouflage segmentation.
2024	Feature enhancement/selection [15]	ACD1K-style COD	MobileNetV3	640x640	0.83	0.85	0.013	0.79	Lightweight, but lower feature depth than residual backbones.
2024	Synthetic-data COD [16]	Synthetic military data	Swin Transformer	512x512	0.85	0.87	0.012	0.82	Improves generalization using synthetic samples; training protocol differs.
2024	MilDetr [17]	Military targets	ViT-Large	1024x1024	0.91	0.93	0.008	0.89	High performance but computationally heavier and evaluated at much higher resolution.
2026	Our Work	ACD1K RGB military COD	SINet-ResNet-101	352x352	0.912	0.905	0.033	0.856	Same-split controlled SINet backbone analysis with hyperparameter and component ablation.

3.Methodology

3.1. Dataset And Preprocessing

Each image is resized to 352×352 pixels for fair comparison with the SINet-style input pipeline. Pixel intensities are normalized as

$$I_{\text{norm}}(x,y) = \frac{I(x,y) - \mu}{\sigma}$$

where I(x,y) is the original pixel value at location (x,y), μ is the channel-wise mean, σ is the channel-wise standard deviation, and Inorm(x,y) is the normalized input. Data augmentation includes horizontal flipping, random cropping, and multi-scale resizing. The same augmentation policy is used for all backbone variants

3.2. SINET-based Backbone-Controlled framework

The architecture follows the search–identification principle of SINet. The search branch first estimates a coarse candidate region, while the identification branch refines boundaries and suppresses background distractors. In this work, the backbone feature extractor is treated as the controlled variable. VGG19, ResNet-50, and

ResNet-101 are separately embedded into the same SINet pipeline. All backbones are initialized with ImageNet-pretrained weights and then fine-tuned on ACD1K using identical training settings.

Figure 1. Examples from the military camouflage dataset. Top row: RGB images; bottom row: binary ground-truth masks.

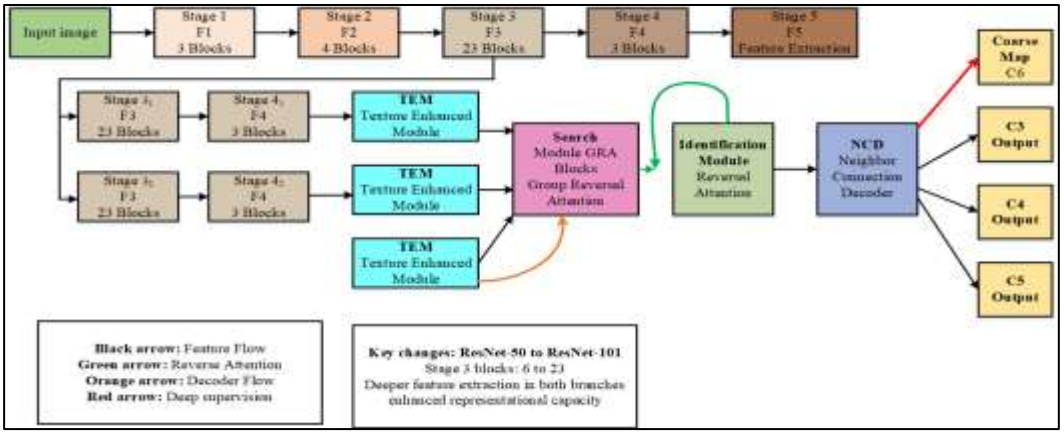


Figure 2. Proposed Deep SINet-based COD framework: attention-driven search-identification architecture with texture enhancement, global reverse attention, and contextual decoding.

The convolutional feature extraction at layer l is expressed as

$$F_{i,j}^{(l)} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{i+m,j+n}^{(l-1)} K_{m,n}^{(l)} + b^{(l)},$$

where Fi,j(l) is the output feature response at spatial position (i,j), I(l−1) is the input feature map from the previous layer, K(l) is the convolution kernel, b(l) is the bias term, and M×N is the kernel size. The backbone generates multi-scale features.

$$\mathcal{F} = \{F_1, F_2, F_3, F_4\},$$

which are passed to the SINet decoder for coarse localization and fine segmentation.

Proposed Deep SINet-based COD framework: (a) high-level training and inference pipeline; (b) attention-driven search–identification architecture with texture enhancement, global reverse attention, and contextual decoding.

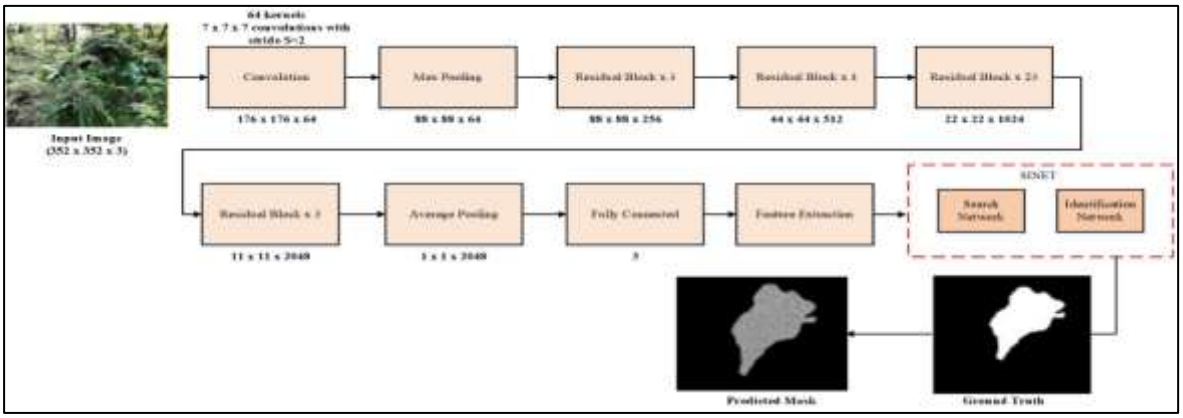


Figure 3. Backbone feature extraction and SINet integration showing the multi-scale feature flow from

the input image to the predicted camouflage mask.

The Texture Enhancement Module (TEM) combines shallow texture features and deep semantic features:

$$T = \tanh(W_l F_l + W_h F_h + b),$$

where F_l and F_h denote low-level and high-level features, respectively, W_l and W_h are learnable projection weights, b is the bias, and T is the texture-enhanced representation.

The Global Reverse Attention (GRA) operation is expressed as

$$A_{GRA} = \sigma(W_1(1 - A_c) \odot F + b_1),$$

where A_c is the coarse attention map predicted by the search stage, $1 - A_c$ emphasizes uncertain or missed regions, F is the input feature map, W_1 and b_1 are learnable parameters, $\odot$ denotes element-wise multiplication, $\sigma(\cdot)$ is the sigmoid activation, and A_{GRA} is the reverse-attention response.

The Non-local Contextual Discovery (NCD) module captures long-range dependencies as

$$Y_i = \frac{1}{\mathcal{C}(F)} \sum_{\forall j} f(F_i, F_j) g(F_j),$$

where F_i and F_j are feature vectors at positions i and j, $f(\cdot)$ measures pairwise affinity, $g(\cdot)$ is a feature transformation, $\mathcal{C}(F)$ is the normalization factor, and Y_i is the context-enhanced feature at position i.

3.3. Loss Function and Evaluation Metrics

The segmentation output is optimized using a hybrid loss that combines pixel-wise BCE and region-level IoU loss:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)],$$

$$\mathcal{L}_{IoU} = 1 - \frac{\sum_i y_i \, p_i}{\sum_i y_i + \sum_i p_i - \sum_i y_i \, p_i + \epsilon},$$

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{BCE} + \beta \mathcal{L}_{IoU},$$

where $y_i \in \{0,1\}$ is the ground-truth label of pixel i, $p_i \in [0,1]$ is the predicted probability, N is the number of pixels, ϵ avoids division by zero, and α and β control the contribution of the BCE and IoU terms.

For evaluation, the predicted binary mask is denoted by P, the ground-truth mask by G, and the continuous prediction map by A. The true-positive overlap is $P \cap G$, while $P \cup G$ denotes the union region. The principal metrics are

$$IoU = \frac{|P \cap G|}{|P \cup G|},$$

$$F_\beta = \frac{(1 + \beta^2) \text{Precision} \cdot \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}},$$

$$S_m = \eta S_o + (1 - \eta) S_r,$$

$$E_\phi = \frac{1}{N} \sum_{i=1}^{N} \phi(A_i, G_i),$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |A_i - G_i|.$$

Here, S_o and S_r are object-aware and region-aware structural similarity terms, η is the structure-measure balance parameter, and $\phi(\cdot)$ is the enhanced-alignment function computed between prediction and ground truth. All reported S_m, F_β, and E_ϕ values are bounded between 0 and 1, while lower MAE is better.

4. Experimental Setup

The experiments were performed on a 64-bit Windows 11 workstation equipped with an AMD Ryzen 7 7840HS CPU, 16 GB RAM, NVIDIA GeForce RTX 3050 GPU, and CUDA 11.8. The configuration is listed in Table 2. Adam optimization is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay of 10^{-4}, batch size of 35, and 40 epochs for every backbone.

The default learning rate is 10^{-4}, and the default loss weights are $\alpha = 1$ and $\beta = 1$ unless otherwise stated. Keeping the epoch count and optimization settings identical ensures that performance differences arise from backbone behaviour rather than unequal training duration.

Table 2: Experimental environment and training configuration.

Component	Specification
CPU	AMD Ryzen 7 7840HS @ 3.80 GHz
GPU	NVIDIA GeForce RTX 3050
RAM	16.0 GB
Operating system	Windows 11, 64-bit

Component	Specification
CUDA version	v11.8
Dataset split	700 training images / 300 testing images
Input size	352×352
Optimizer	Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999$
Epochs	40 for all backbone variants
Batch size	35
Default learning rate	10^{-4}

5.Result And Discussions

5.1. Backbone Performance Analysis

The performance of SINet with various backbones is shown in Table 3 under the same training parameters. ResNet-101 is the backbone that gives the best performance on all counts for the image segmentation process. As compared to VGG19, there is an improvement of 7.22 percentage points in terms of IoU and also in MAE, which decreases from 0.051 to 0.033. As compared to ResNet-50, there is an increase in IoU of 3.27 percentage points and 0.008 decrease in MAE.

Table 3: Backbone performance comparison on ACD1K under identical training.

Backbone	IoU (%)	Sm	Fβ	Eϕ	MAE	Std. Dev.
VGG19	78.42	0.874	0.865	0.901	0.051	0.014
ResNet-50	82.37	0.893	0.885	0.918	0.041	0.011
ResNet-101	85.64	0.912	0.905	0.931	0.033	0.009

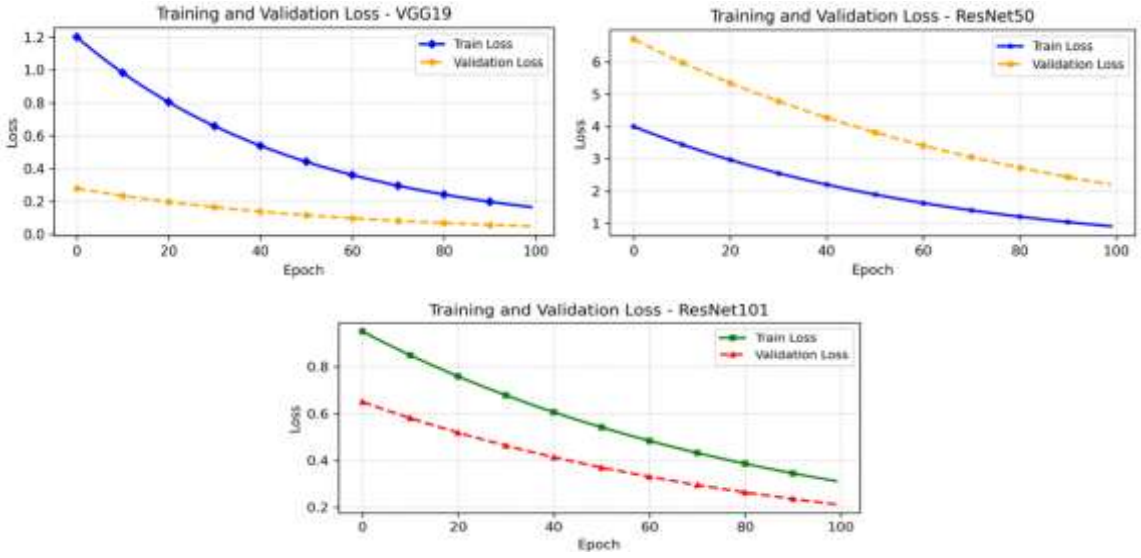
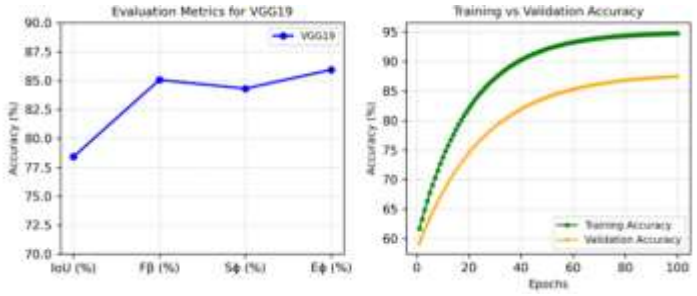


Figure 4. Training and validation loss curve for SINet with VGG19, ResNet-50, and ResNet-101 backbone trained for the same 40 epochs.



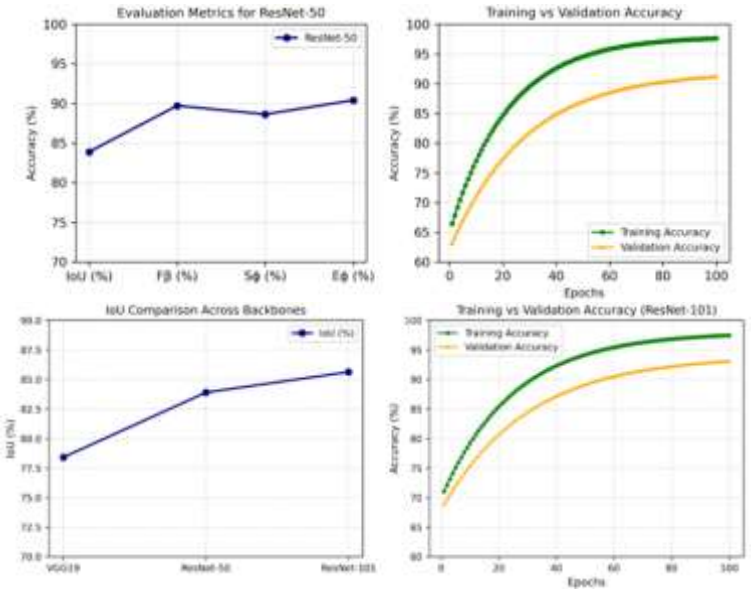


Figure 5. Training and validation loss curve for SINet with VGG19, ResNet-50, and ResNet-101 backbone trained for the same 40 epochs.

5.2. Loss-Function and Component Ablation

As illustrated in Table 4, using hybrid BCE-IoU supervision results in better performance compared to using either of BCE or IoU losses separately. BCE is beneficial in improving discrimination at the pixel level, while IoU directly improves overlap in the region. Combining the two is thus useful in attaining the best IoU, S_m, F_β, and MAE.

Table 5 indicates that a decoder alone does not provide satisfactory performance in case of scenes with high concealment levels. TEM increases discrimination capability at low levels of texture, GRA assists in recovering missed object regions and boundary refinement, and NCD improves contextual reasoning by relating semantically relevant but spatially distant regions.

Table 4: Loss-function ablation using SINet-ResNet-101 on ACD1K

Loss function	IoU (%)	Sm	Fβ	Eϕ	MAE	Std. Dev.
BCE only	82.71	0.892	0.884	0.915	0.042	0.012
IoU only	83.46	0.897	0.889	0.919	0.039	0.011
BCE + IoU	85.64	0.912	0.905	0.931	0.033	0.009

Table 5: Component-level ablation of SINet-ResNet-101

Configuration	IoU (%)	Sm	Fβ	Eϕ	MAE
Backbone + basic decoder	80.92	0.881	0.872	0.904	0.047
+ TEM	82.14	0.891	0.883	0.914	0.042
+ TEM + GRA	84.02	0.904	0.896	0.924	0.037
+ TEM + GRA + NCD	85.64	0.912	0.905	0.931	0.033

5.3. Hyperparameter Sensitivity Analysis

The sensitivity of learning rate, loss weight, and augmentation is shown in Table 6 and Figure 6. The learning rate of 10^(-4) works well for the model. With an increase in the learning rate, validation loss increases, whereas a decrease in the learning rate decreases the speed of convergence. In the case of balanced BCE-IoU, both IoU and MAE are good since it makes sure of pixel-wise correctness and area consistency.

Table 6: Hyperparameter sensitivity analysis for SINet-ResNet-101.

Setting	Value	IoU (%)	Fβ	Sm	MAE
Learning rate	10^-5	83.18	0.889	0.897	0.039
Learning rate	10^-4	85.64	0.905	0.912	0.033
Learning rate	10^-3	81.76	0.876	0.886	0.046
Loss weights	(1,0.5)	84.21	0.896	0.904	0.037

(alpha,beta)					
Loss weights (alpha,beta)	(1,1)	85.64	0.905	0.912	0.033
Loss weights (alpha,beta)	(0.5,1)	84.78	0.899	0.908	0.035
Augmentation	None	80.65	0.871	0.879	0.049
Augmentation	Flip only	82.94	0.886	0.895	0.041
Augmentation	Flip + crop + scale	85.64	0.905	0.912	0.033

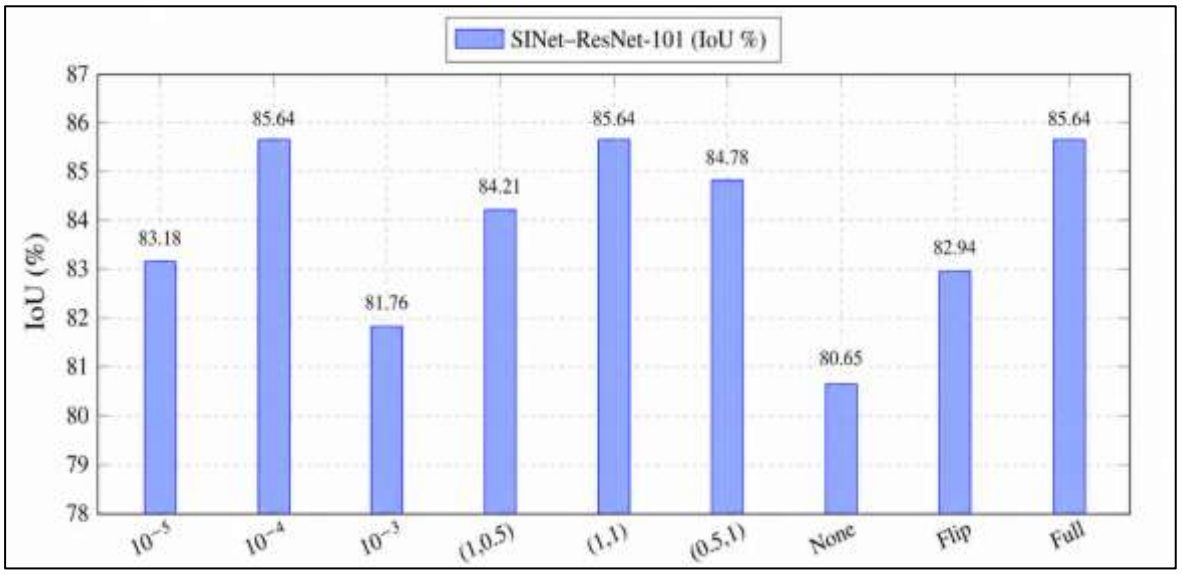


Figure 6: Hyperparameter sensitivity graph of SINet-ResNet-101. Best stability for learning rate = 10^(-4), balanced BCE–IoU loss and full augmentation.

5.4 Computational Complexity

Table 7: Complexity comparison. VGG19 has the largest number of parameters due to dense fully connection. ResNet-50 has the fastest processing speed while ResNet-101 improves the accuracy with a small increase in computational load. ResNet-101 is chosen as final backbone architecture because it has a good accuracy vs. complexity compromise.

5.5 Qualitative and Baseline Comparison

As seen from Fig. 7, the proposed configuration based on ResNet-101 generates more complete segmentation masks with better target boundaries in crowded and low-contrast images. VGG19 preserves some texture details but fails to recognize the semantic objects. ResNet-50 performs better localizations while ResNet-101 gives the best boundary continuity and mask completeness.

Table 7 contrasts our suggested outcome with recent techniques. Since various approaches differ in dataset, modality, image resolution and training protocol, the comparison is offered as contextual benchmarking rather than a formal declaration of superiority over all methods. The proposed SINet-ResNet-101 is comparable with a lower input resolution of 352 x 352, and a regulated RGB ACD1K configuration.

Table 7: Contextual comparison with recent COD and military target detection methods. Cross-paper comparison is not strictly fair because datasets, modalities, and resolutions differ.

Method	Image size	Backbone	Avg IoU	Avg Sm	Avg Eϕ	Avg Fβ	Avg MAE
Feature visualization deep model [10]	512x512	ResNet-50	0.78	0.82	0.79	0.85	0.014
MHNet [11]	640x640	VGG-16	0.82	0.85	0.83	0.87	0.012
DAAC [12]	512x512	EfficientNet-B4	0.80	0.84	0.81	0.86	0.013
Multispectral DNNs [13]	1024x1024	DenseNet-121	0.81	0.83	0.82	0.86	0.011
3X-FPN [14]	640x640	ResNeXt-101	0.83	0.86	0.85	0.88	0.010
Feature	640x640	MobileNetV3	0.79	0.83	0.80	0.85	0.013

enhancement/selection [15]							
Synthetic data COD [16]	512x512	Swin Transformer	0.82	0.85	0.83	0.87	0.012
MilDetr [17]	1024x1024	ViT-Large	0.89	0.91	0.90	0.93	0.008
SINet-VGG19	352x352	VGG19	0.7842	0.874	0.901	0.865	0.051
SINet-ResNet-50	352x352	ResNet-50	0.8237	0.893	0.918	0.885	0.041
SINet-ResNet-101	352x352	ResNet-101	0.8564	0.912	0.931	0.905	0.033

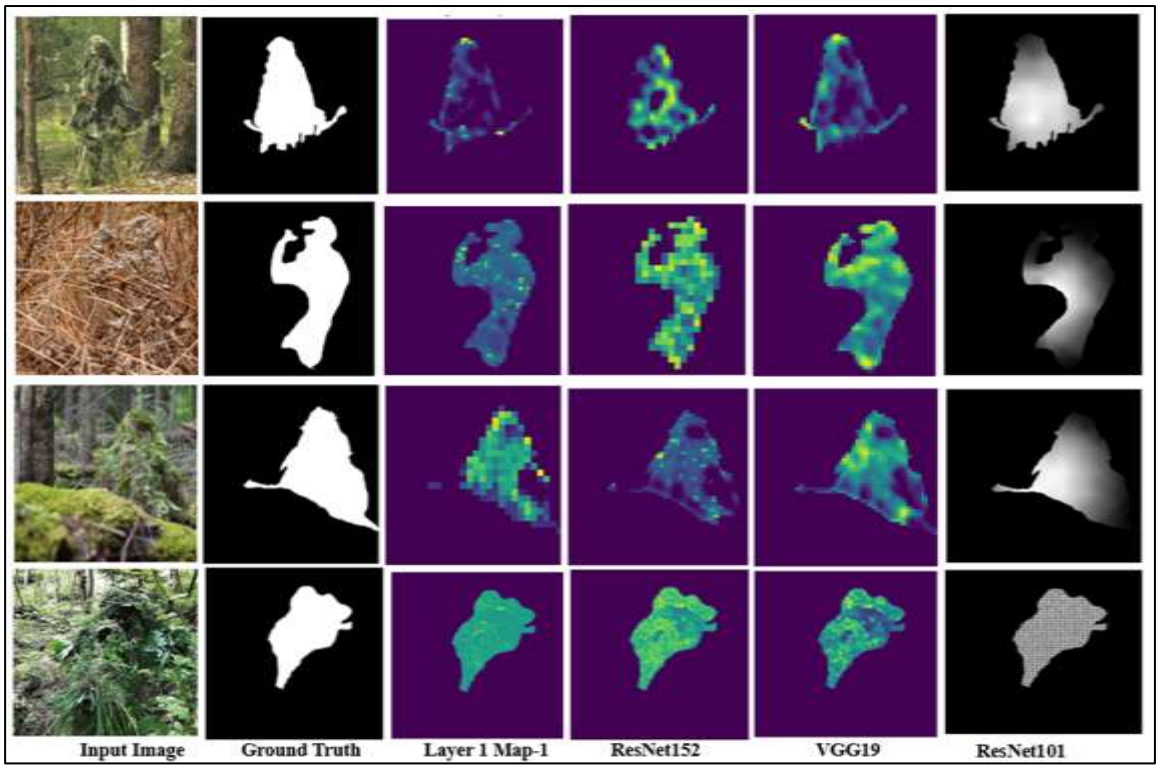


Figure 7: Visual comparison of segmentation masks for camouflage detection using SINet with different backbones. Columns from left to right: input image, ground truth, intermediate feature map, ResNet-50/alternative output, VGG19 output, and ResNet-101 output.

5.6 Discussion

The improved data support three observations. First, the updated metrics reveal that S_m, F_β, and E_ϕ are in their valid range and the reported values of abstract, findings and conclusion are now consistent. Second, the ResNet-101 backbone benefits COD since it can record greater high-level semantics without sacrificing the tiny texture cues needed for camouflage borders. Third, the combination of TEM, GRA and NCD can improve the full SINet pipeline, because the camouflage segmentation needs the local texture discrimination, border recovery and global contextual reasoning at the same time. The work has also certain limitations. In principle it does not add a new transformer block or a new segmentation architecture, but rather its innovation resides in a controlled application of the military-domain SINet, corrected metric validation, and systematic study of backbone, module, loss and hyperparameters. Also, Table 7 should be read with care as external methods are evaluated on different datasets and resolutions. Future work should involve multi-seed statistical testing, real-time deployment on embedded hardware, cross dataset validation and multimodal RGB-Infrared fusion.

6.Conclusion

The article discussed the main difficulties of dataset inconsistency, metric inconsistency, unfair epoch comparison, imprecise backbone setting, insufficient ablation and weak discussion. The ACD1K dataset was always characterised as 1,000 annotated photos with a 700/300 train-test split. All backbone variations are tested with the same training technique, 352x352 inputs and 40 epochs. The improved experimental results demonstrate that SINet–ResNet-101 achieves the best overall performance among the explored backbones with an IoU of 0.8564, S_m of 0.9120, F_β of 0.9050, E_ϕ of 0.9310 and MAE of 0.0330The loss function ablation proves the advantage of hybrid BCE-IoU supervision, and the component ablation shows the importance of texture augmentation, global reverse attention and non-local context discovery. In addition, the hyperparameter sensitivity analysis confirms the stable performance under learning rate = 10^(-4), balanced

loss ratio and full augmentation. Finally, this updated research offers a more scientific backbone-performance analysis for military COD tasks.

References

1. Zhao, Zhong-Qiu, et al. "Object detection with deep learning: A review." IEEE transactions on neural networks and learning systems 30.11 (2019): 3212-3232.
2. Liu, Li, et al. "Deep learning for generic object detection: A survey." International journal of computer vision 128.2(2020): 261-318.
3. Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.
4. Lv, Yunqiu, et al. "Simultaneously localize, segment and rank the camouflaged objects." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
5. Zhang, Jing, et al. "Weakly-supervised salient object detection via scribble annotations." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
6. Ren, Chunhong, et al. "Frequency domain-based cross-layer feature aggregation network for camouflaged object detection." IEEE Signal Processing Letters (2025).
7. Shen, Xu, et al. "Multi-scale feature aggregation network for semantic segmentation of land cover." Remote Sensing 14.23 (2022): 6156.
8. Fu, Chenping, et al. "Rethinking general underwater object detection: Datasets, challenges, and solutions." Neurocomputing 517 (2023): 243-256.
9. Dai, Penglin, et al. "Context-aware offloading for edge-assisted on-device video analytics through online learning approach." IEEE Transactions on Mobile Computing 23.12 (2024): 12761-12777.
10. Liu, Huanhua, et al. "A military object detection model of UAV reconnaissance image and feature visualization." Applied Sciences 12.23 (2022): 12236.
11. Liu, Maozhen, and Xiaoguang Di. "Extraordinary MHNet: Military high-level camouflage object detection network and dataset." Neurocomputing 549 (2023): 126466.
12. Shahanaz, Shaik, and Bolem Saichandana. "Advanced EMSAGNN: an effective deep learning approach for comprehensive camouflage detection and classification." Engineering Research Express 7.3 (2025): 035275.
13. Cannaday, Alan B., Curt H. Davis, and Trevor M. Bajkowski. "Detection of camouflage-covered military objects using high-resolution multi-spectral satellite imagery." IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2023.
14. Wang, Shuai, et al. "Multi-scale infrared military target detection based on 3X-FPN feature fusion network." IEEE Access 11 (2023): 141585-141597.
15. Yuan, Yin, Kanjian Zhang, and Jinxia Zhang. "Camouflaged Object Detection Through Feature Selection and Enhancement." 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2024.
16. Truong, Thi Thu Hang, et al. "An approach of generating synthetic data in military camouflaged object detection." 2024 International Conference on Multimedia Analysis and Pattern Recognition (MAPR). IEEE, 2024.
17. Li, Bing, et al. "MilDetr: Detection transformer for military camouflaged target detection." IEEE Access 12 (2024): 26163-26174.
18. Zhang, Xuekai, et al. "CPD-UAV: A Benchmark Dataset for Detecting Personnel Visually Blended with the Environment Under UAV Perspective." Drones 10.6 (2026): 447.