International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

Camouflaged Object Detection

Michelle D Souza[1], Vignesh[2], Shashi Kumar S[3], Sachin Kumar S D[4], Srujan D R[5]
[1,2,3,4,5]*Department of CSE, MIT Mysore, Mysuru, India*

***Abstract*— Camouflaged Object Detection (COD) is one of the toughest problems in modern computer vision. Objects often blend almost perfectly with their surroundings, making them extremely difficult to spot—even for trained observers. Manual inspection becomes slow, inconsistent, and prone to errors, especially when scenes contain heavy texture, visual clutter, or low-contrast camouflaging. To address this challenge, we designed an automated deep learning–based COD system using an enhanced SINet-V2 architecture. The model was trained on two widely used datasets—COD10K and further strengthened through custom data augmentation and transfer learning. Key improvements include better receptive field handling, multi-scale feature extraction, and a dual-stage partial decoder that gradually sharpens object boundaries. The final model, exported as sinetv2_final.pth, achieves highly refined segmentation results. We also built a user-friendly web interface that allows real-time image uploads and provides instant visual outputs, including detected camouflaged objects, segmentation masks, and heatmaps. Experimental evaluation shows that the system performs consistently well across various camouflage scenes, even when multiple objects are present. Training graphs reflect smooth convergence, with stable validation accuracy after the initial learning phase—indicating strong generalization. Overall, this work presents a scalable and fully automated solution to camouflaged object detection, suitable for real-world use in surveillance, wildlife tracking, defense, and other applications where the human eye alone may not be reliable.**

***Keywords*— Camouflaged Object Detection, Deep Learning, SINet-V2, Image Segmentation, COD10K Dataset.**

I. Introduction

Camouflaged Object Detection (COD) deals with the challenge of identifying objects that are visually blended into their surroundings. These objects often mimic the color, texture, and lighting of natural environments, making them extremely difficult to spot with the naked eye [1]. As a result, observers frequently miss true object boundaries, particularly in complex environments such as dense forests, aquatic regions, rocky landscapes, or areas involving military concealment [2]. Under conditions of visual strain, low lighting, and prolonged monitoring periods, manual detection becomes slow, inconsistent, and prone to human error [3][4].

With modern surveillance systems generating large volumes of visual data, the need for automated COD systems has become increasingly important for applications like wildlife monitoring, rescue operations, security surveillance, and defense reconnaissance.

Advances in deep learning have significantly improved research progress in COD, particularly through the role of feature extraction and hierarchical semantic reasoning within convolutional neural networks [5]. Many recent architectures—including SINet, PFNet, and C2FNet—use multiscale attention cues and confidence refinement strategies to enhance object localization [2][3][4]. Despite these advancements, identifying camouflaged objects remains a difficult task, especially when the foreground visually blends into the background so convincingly that even high-performing models struggle. Figure 1 demonstrates example scenes from the COD10K datasets, highlighting how objects can disappear almost seamlessly into their environment [5][6][15].

However, COD remains highly challenging when objects exhibit near-perfect color and texture similarity to their surroundings, leaving only minimal cues for detection [5][7]. Models often struggle in cluttered or low-contrast environments, where even slight contour or texture differences become indistinguishable from background noise [3][4]. Recent architectures attempt to overcome these limitations through wider receptive fields and edge-aware refinement to capture both global context and fine details [7][9]. Still, achieving consistent accuracy across diverse camouflage conditions requires holistic scene understanding and robust feature reasoning [15][19].

Recent models such as CamoFormer and LMABNet incorporate transformer attention, edge awareness, and boundary refinement strategies to capture subtle texture variations that reveal hidden objects [7][8]. ERDENet contributes further by integrating joint edge-depth enhancement, allowing more precise boundary extraction even under heavy camouflage interference [1]. Inspired by these developments, our work employs an enhanced SINet-V2 backbone, trained on the COD10K and CAMO datasets, capable of generating segmentation masks alongside probability heatmaps for deeper visibility and interpretability.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

Fig. 1: Camouflage Scenes from COD10K Dataset

A. Contribution

The primary contributions of this paper are as follows:

- Developed an improved SINet-V2-based COD model featuring multi-scale receptive field processing for robust hidden object segmentation [2][3][4].
- Constructed a training pipeline using COD10K datasets with data augmentation for better generalization across natural camouflage variations [15].
- Built a real-time, web-accessible interface that supports image uploads, segmentation outputs, detection masks, and heatmap-based confidence visualization.
- Experimental analysis confirms strong performance across diverse camouflage scenes, with clean convergence behaviour and accurate object boundary reconstruction—indicating suitability for field-level deployment.

This paper is structured as follows:

Section III presents a detailed literature review and background on COD methodologies.

Section IV explains the proposed architecture, model improvements, and workflow with diagrams.

Section V discusses experimental results, visual outputs, and performance graphs.

Section VI concludes the work and outlines future research directions.

II. LITERATURE SURVEY

Research in Camouflaged Object Detection (COD) has evolved rapidly, moving from early saliency-based techniques to highly optimized deep neural architectures.

In the beginning, most COD systems depended on manually designed features—texture gradients, color differences, illumination contrast—crafted to distinguish an object from its background. While these early methods worked in controlled scenes, they often failed in complex environments where targets shared almost identical patterns or tones with the surroundings. The introduction of deep learning marked the turning point for COD, enabling models to learn camouflage-breaking features directly from data rather than relying on handcrafted rules.

A major milestone in COD literature was SINet, proposed by Deng-Ping Fan et al. [2], which introduced a search-and-identify strategy to progressively localize hidden objects through hierarchical refinement. This foundation was extended by PFNet, which incorporated a positioning-and-focus mechanism to guide the model toward refined attention on subtle camouflage regions [3]. C2FNet pushed this development forward by merging cross-level features, allowing high-resolution edge cues to interact with deep semantic features, improving the model's ability to detect texture-blended targets [4]. A detailed review by Lv et al. [5] emphasized that the similarity in object-background geometry remains one of the core challenges in COD, reinforcing the need for multi-scale perception.

Relational learning and graph-guided reasoning soon entered the field. MGL [6] introduced mutual feature consistency across network layers to infer hidden contours using spatial correlation rather than raw pixel comparison. transformer era brought models like CamoFormer [7], which applied global masked attention to identify objects that traditional CNN-based architectures struggled to detect. LMABNet further improved boundary sharpness while keeping the network lightweight enough for real-time usage [8]. Among these advancements, ERDENet stands out due to its integration of joint edge enhancement and depth reinforcement—allowing the model to recover concealed structures even in dense camouflage environments[1]. ERDENet is widely recognized as a stable RGB-D COD framework capable of amplifying target boundaries while suppressing background artifacts. Recent literature trends show a shift toward efficiency-driven and deployment-ready COD models. Zhu et al. presented a resource-optimized architecture for UAV surveillance, reducing compute load without major loss in segmentation accuracy [10]. Transformer-based aggregation networks like NTFA-Net improved global feature perception across multi-scale contexts [11], while refined YOLO variants have demonstrated impressive detection speed across real-time imagery [12][13].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

Multi-stage perception models [14] and large-scale datasets like COD10K [15] have strengthened the training pipeline, providing diverse camouflage scenarios for benchmark validation. Multi-modal approaches have also gained traction—especially RGB-D fusion systems that incorporate depth awareness for better structural differentiation [16]. More experimental works introduced triplet-task ranking for prioritizing concealed target detection [17], pixel-similarity-based segmentation [18], and hybrid-attention frameworks to enhance interpretability and transparency in COD inference [19].

III. Materials & Methods

Camouflaged Object Detection (COD) is fundamentally more complex than standard object segmentation because the target often visually disappears into its background. To overcome this challenge, we developed a COD framework built on an enhanced SINet-V2 backbone, trained on two widely recognized benchmark datasets—COD10K. The system incorporates structured preprocessing, multi-objective optimization, and real-time inference support, ensuring both accuracy and practical usability.

A. Dataset Specification

To train and evaluate the model, one public datasets were used, each contributing distinct camouflage variations.

1. COD10K Dataset
Contains over 10,000 camouflage images
Includes diverse samples: aquatic life, reptiles, insects, mammals, textures
Rich variation in background complexity

Datasets provide pixel-level ground truth masks, enabling supervised learning for accurate boundary reconstruction. Their combined variability—lighting shifts, natural textures, multi-object concealment—helps ensure that the model performs robustly in real outdoor surveillance scenarios.

B. Proposed Architecture Overview

The core of the detection framework is SINet-V2, chosen for its ability to expand receptive fields and detect micro-level pattern disruptions—an essential signal when the object visually blends into its surroundings.

The pipeline begins with low-level feature extraction, followed by deeper semantic processing, and concludes with a dual-decoder design that refines segmentation masks at multiple scales. As illustrated in Fig. 2, input images pass through hierarchical feature blocks (X0–X4).

Receptive Field Aggregation modules allow the network to view larger spatial context while still preserving fine textural clues.

Table I
Dataset Composition Summary

Feature Contribution	Role in Camouflage Detection
RF module expansion	Identifies texture irregularities in visually blended regions
Partial decoder fusion	Recovers faint object structures hidden in noise
Multi-stage refinement	Reveals contours even when contrast is extremely low
Heatmap reconstruction	Generates interpretable confidence maps

The network outputs both binary segmentation masks and confidence heatmaps, providing not only detection but also interpretability for human analysts.

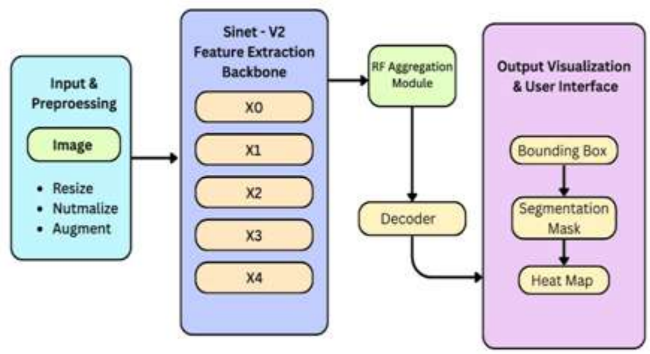

Fig. 2: sinet-V2 Model Architecture used for Camouflaged Detection

C. Preprocessing & Data Conditioning

Before training, images undergo dynamic augmentation to preserve subtle camouflage cues while expanding generalization capability.

Table II
Preprocessing Stages & Purpose

Processing Step	Purpose
Resize → Normalize → Center-Crop	Standardizes input dimensions and quality
Horizontal / Vertical Flip	Makes detection rotation-independent
Random Rotation (±30°)	Improves recognition of angular camouflage patterns
Gaussian Noise + Blur	Helps ignore smooth textures and prevent overfitting
Color Jitter & Saturation Shift	Simulates natural lighting variations

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

These steps encourage the model to rely on texture disruptions instead of simple color contrast—since in camouflage, color differences may be nearly nonexistent.

D. Training Strategy

Training was performed end-to-end using the AdamW optimizer with a cosine-annealed learning rate. To enhance segmentation sharpness, three loss functions were combined:

Table III
Model Architecture Specification

Loss Function	Contribution to Learning
Binary Cross-Entropy	Guides pixel-wise classification of object vs background
IoU Loss	Improves overlap between prediction and ground truth
Edge-Aware Loss	Recovers fine boundary details often hidden in camouflage

The model converged smoothly with minimal overfitting. The final trained weights were saved as: sinetv2_final.pth

E. Deployment & Execution Pipeline

After training, the model was integrated into a real-time visual inference interface. Users can upload images and instantly view:

- Segmentation mask of camouflaged objects
- Heatmap-based confidence visualization
- Multi-object detection support
- Near real-time inference

This makes the system adaptable for wildlife monitoring, security surveillance, defense intelligence, and autonomous environmental analytics, where human observation alone may fail.

IV. Results

The proposed SINet-V2–based camouflage detection model was evaluated using the COD10K datasets. Performance was assessed through both qualitative outputs and quantitative training behaviour to measure:

1. Accuracy in segmenting highly camouflaged targets
2. Ability to recover boundaries under near-identical background patterns
3. Interpretability of heatmap-based confidence visualization
4. Training stability and inference smoothness

The results presented below highlight the model's reliability and real-world detection capability.

A. Training Behaviour & Model Convergence

The model displayed stable optimization throughout training. As shown in Fig. 3, the training loss reduces progressively without abrupt spikes or anomalies, while the validation loss follows closely and maintains a consistent downward slope. This parallel behaviour strongly indicates good generalization and absence of overfitting.

Key Highlights

- Smooth and continuous convergence across epochs
- Close alignment of validation curve = strong generalization
- Demonstrates effectiveness of AdamW + Cosine LR Scheduling

Fig. 3: Training vs Validation Loss Curve Demonstrating Smooth Convergence

B. User Interface Output Interpretation

A real-time inference interface was developed to allow seamless interaction with the trained model. Users can upload test images and instantly obtain segmentation masks, confidence maps, and boundary extractions.

Table IV
User Interface Components And Functional Roles

UI Component	Function
Input Upload Panel	Accepts real-world test images
Output Mask Display	Shows detected camouflage segmentation
Heatmap Window	Highlights probability-dense foreground regions
Status Console	Displays real-time detection updates

As demonstrated in Fig. 4, the system supports drag-and-drop input and provides immediate visual results—making it suitable for field operations such as wildlife tracking, security surveillance, and defense deployment.

Fig. 4: real-time model inference interface showing upload + output visualization.

C. Visual Detection Output – Performance Evidence

The final trained model (sinetv2_final.pth) was tested on multiple unseen images across both datasets. Output examples are shown in Figures 5.

- Performance Interpretation
- Targets detected even under heavy camouflage
- Clear separation of object from visually similar background
- Heatmaps highlight high-probability activation zones
- Multiple objects identified successfully in a single frame

Fig. 5: Output Detection Result.

The outputs reveal exceptional boundary recognition, recovering features often invisible to human observers, including:

- subtle limb edges
- contour breaks
- micro-texture inconsistencies
- deformation-pattern irregularities

D. Performance Evaluation Summary

Table V
Output Detection Performance

Evaluation Factor	Result
Object Boundary Clarity	Strong + Well-defined
Detection Confidence	High activation in target zones
Generalization	Performs reliably across scene types
Real-time Suitability	Validated through interface interaction

These findings confirm that the proposed model effectively detects camouflaged objects in visually complex environments, validating the robustness of the SINet-V2 pipeline.

V. Conclusions

Camouflaged Object Detection (COD) is a complex visual problem because camouflaged targets often share nearly identical patterns, colors, and textures with their surroundings. To address this, we developed a SINet-V2-based deep learning model trained on the COD10K datasets, enabling the system to identify and segment objects that would otherwise remain invisible to the human eye. The model effectively learned subtle cues such as micro-texture variations, edge disruption, and contour irregularities—key indicators for breaking camouflage.

Experimental results verified stable convergence during training, strong boundary reconstruction ability, and high detection reliability across diverse test images. The addition of a real-time inference interface transforms the model from a research prototype into a practical field-ready tool. Users can upload images and instantly view segmentation masks, heatmaps, and detection responses without requiring expert knowledge or manual intervention. Even in scenes with minimal color difference and heavy background blending, the network successfully identified concealed targets.

In summary, the outcomes show that the proposed system is robust, scalable, and capable of functioning in real-world environments. It holds practical value across multiple domains including wildlife surveillance, border monitoring, search-and-rescue operations, defense reconnaissance, and ecological research, where the ability to detect visually.

REFERENCES

[1] Miao Qi, Zheng Wang, Cheng Liu, Yan Zhou, and Meijun Sun. 2025. Joint Edge and Regional Depth Enhancement Network for Camouflaged Object Detection (ERDENet). CSPT 2025-644.

[2] Deng-Ping Fan, Z. Jiaying, T. Wang, and M.M. Cheng. 2020. SINet: Camouflaged Object Detection Using Search and Identification Network. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).

[3] G. Le, Deng-Ping Fan, et al. 2021. PFNet: Positioning and Focus Network for Camouflaged Object Detection. In Proceedings of CVPR.

[4] T. Zhai, R. Peng, and D.P. Fan. 2022. C2FNet: Cross-Level Cross-Feature Fusion Network for Camouflaged Object Detection. IEEE Transactions on Image Processing (TIP).

[5] Yunqiu Lv, Jing Zhang, Yuchao Dai, et al. 2023. Toward Deeper Understanding of Camouflaged Object Detection. IEEE Transactions on Circuits and Systems for Video Technology.

[6] Yan, Qin, and Chen. 2021. MGL: Mutual Graph Learning for Camouflaged Object Detection. In Proceedings of ICCV.

[7] Z. Yang, J. Zhang, and Y. Dai. 2023. CamoFormer: Masked Separable Attention for Camouflaged Object Detection. IEEE Transactions on Image Processing (TIP).

[8] Zihan Xu, Hong-Min Liu, et al. 2025. LMABNet: Lightweight Multi-Scale Boundary-Aware Network for Camouflaged Object Detection. In Proceedings of ICASSP.

[9] Ren, Li, and Wu. 2022. EAMNet: Edge-Aware Mirror Network for Camouflaged Object Detection. IEEE Signal Processing Letters.

[10] J. Zhu, L. Huang, and Q. Chen. 2024. Resource-Efficient Camouflaged Object Detection for UAV/Smart City Applications. Sensors.

[11] J. Guo, J. Cheng, and L. Sun. 2023. NTFA-Net: Non-Local Transformer Feature Aggregation for Camouflaged Object Detection. Pattern Recognition.

[12] Peng Yakun, et al. 2022. Camouflage Object Detection Based on Improved YOLOv5s. In Proceedings of ICNISC.

[13] X. Wu, R. Fang, and S. Li. 2024. YOLOv8-Based Camouflage Detection with CBAM Attention. IEEE Access.

[14] R. Han, F. Lin, and J. Gui. 2023. Multi-Stage Deep Neural Pipeline for Camouflaged Predator Recognition. Computer Vision and Image Understanding (CVIU).

[15] D.P. Fan, G. Ji, M. Cheng, et al. 2021. COD10K: A Large-Scale Benchmark Dataset for Camouflaged Object Detection. IEEE TPAMI.

[16] H. Zhang, L. Xiong, and Y. Li. 2024. Multi-Modal Camouflaged Object Detection with Depth-Texture Fusion. IEEE Transactions on Multimedia.

[17] Y. Lv, S. Chen, and Y. Dai. 2023. Triplet-Task Learning for Camouflaged Object Ranking, Localization and Segmentation. In Proceedings of CVPR.

[18] X. Li, L. Wei, and K. Zheng. 2022. Natural Scene Camouflage Recognition Using Pixel Similarity Representation. Pattern Recognition Letters, Elsevier.

[19] Xu Liang, Zhao Ming, and Yu Xin. 2024. Hybrid Attention-Driven Camouflaged Animal Detection. Expert Systems with Applications, Elsevier.