International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

Camouflage Target Detection Using Deep Learning

Rishabh Santosh[1], Mohan Kumar H T[2], Pratham P N[3], Mahammadsohel Inamdar[4], Dr. Jyothi D.G[5]
[1,2,3,4]UG Scholar, Dept. of AI&ML, Bangalore Institute of technology, Bengaluru, Karnataka, India.
[5] Professor, Dept. of AI&ML, Bangalore Institute of technology, Bengaluru, Karnataka, India.
Emails: *rishabh.santosh@gmail.com[1], Kumarhtmohan@gmail.com[2], prathmangowda2@gmail.com[3], mahammadsohelinamdar@gmail.com[4],Dgjyothi@bit-bangalore.edu.in[5]*

Abstract
This paper presents a novel approach for detecting camouflaged military targets using a hybrid architecture that combines a fine-tuned YOLOv8 detector with a Zero-Shot Learning (ZSL) classifier. Traditional object detection systems rely on extensive labeled datasets and predefined object classes, making them ineffective against dynamically camouflaged or unseen targets. The proposed system, CTD (Camouflage Target Detection), overcomes these limitations by decoupling the detection and classification tasks. It integrates a high- performance Spotter (YOLOv8) to find camouflaged objects in a class-agnostic manner, and a Vision-Language Identifier (CLIP) to classify them using flexible text prompts. This approach enables the system to identify and localize concealed objects across diverse environments and, crucially, to classify new targets without retraining. By eliminating the dependency on large annotated datasets for new classes, this approach significantly improves the accuracy, adaptability, and reliability of camouflage detection in modern aerial reconnaissance.

Keywords: *Zero-Shot Learning (ZSL), Object Detection, YOLOv8, CLIP, Camouflaged Object Detection (COD), Hybrid AI, Aerial Reconnaissance.*

1. Introduction
Camouflage target detection focuses on identifying objects intentionally designed to seamlessly blend with their environment, effectively concealing their visual features through texture or pattern mimicry to make them difficult to detect [1]. In critical military and defence applications, such camouflaged targets ranging from vehicles and equipment to personnel pose major strategic challenges for modern surveillance and reconnaissance systems. Traditional object detection models, including state-of-the-art deep learning architectures like YOLO [2], rely heavily on large, manually labelled datasets and fixed, predefined classes. While these systems excel at finding objects they have been trained to see, they often struggle to identify dynamically camouflaged or, more importantly, unseen targets across complex, unstructured terrains like dense forests, arid deserts, and cluttered urban areas. When a new threat appears, these models must be completely retrained, a process that is slow, costly, and impractical for time-sensitive real-world operations. To address these limitations, the proposed system, CTD, employs a hybrid architecture that synergistically integrates advanced deep learning with Zero-Shot Learning (ZSL) capabilities. By decoupling the general spotting of a target from its specific identification, CTD can recognize and localize camouflaged targets without requiring extensive manual annotation for new classes. By integrating a specialized Spotter (YOLOv8) for robust feature extraction and an Identifier (CLIP) [3] for distinct semantic understanding, this approach minimizes retraining efforts, enhances detection accuracy, and significantly improves the adaptability and operational efficiency of defence and surveillance systems in diverse environments.

1.1. Methodology
This study implements a modular, two-stage architecture for efficient and adaptive camouflage target detection. It begins with data preprocessing to

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

improve image contrast, followed by a Spotter stage to detect all potential targets, and an Identifier stage to classify them using ZSL.

System Design

The CTD system utilizes a modular, layered architecture. The workflow begins with the user uploading an image through a Streamlit web interface. [4]

- Pre-processing: The image is first processed using Contrast Limited Adaptive Histogram Equalization (CLAHE). This enhances local contrast, making it easier to distinguish camouflaged objects from their background, a step crucial for the Spotter model.
- Stage 1: The Spotter (YOLOv8): The enhanced image is passed to a fine-tuned YOLOv8 model. This model is trained to be a class-agnostic detector—its only job is to find objects that look like vehicles in an aerial view, regardless of whether it's a tank, jet, or APC. It outputs a list of bounding boxes for all potential targets.
- Stage 2: The Identifier (CLIP): Each bounding box is used to crop the detected object from the original, unprocessed image. This crop is passed to the CLIP ViT- B/32 image encoder, which generates a 512-dimension image embedding. Simultaneously, a list of text prompts (e.g., a photo of an F-22 raptor, a photo of a T-90 tank) is passed to the CLIP text encoder to generate a set of text embeddings. The system then calculates the cosine similarity between the image embedding and every text embedding. Here is the cosine similarity formula in proper

$$\text{similarity}(I_e, T_e) = \cos(\theta) = \frac{I_e \cdot T_e}{\| I_e \|_2 \cdot \| T_e \|_2} \quad (1)$$

The predicted class is the text prompt with the highest similarity score. This ZSL step allows the system to identify any object that can be described in text, even if the model has never seen it before. [5]

- Visualization: The final bounding boxes and their ZSL labels are drawn on the original image and displayed to the user.

1.2. Use Case

The use case describes the interaction within the CTD system. Some surveillance personnel, as the primary actor, provides real-time or pre-collected imagery (from a camera or surveillance drone) to the system. The system per- forms its two-stage analysis and presents the annotated Figure 1 shows High-Level Architecture of the CTD System, Detailing The Flow from User Input to Final Zero-Shot Classification Figure 2 shows Use Case Diagram Illustrating the Interaction Between the Surveillance Personnel and the CTD System [6]

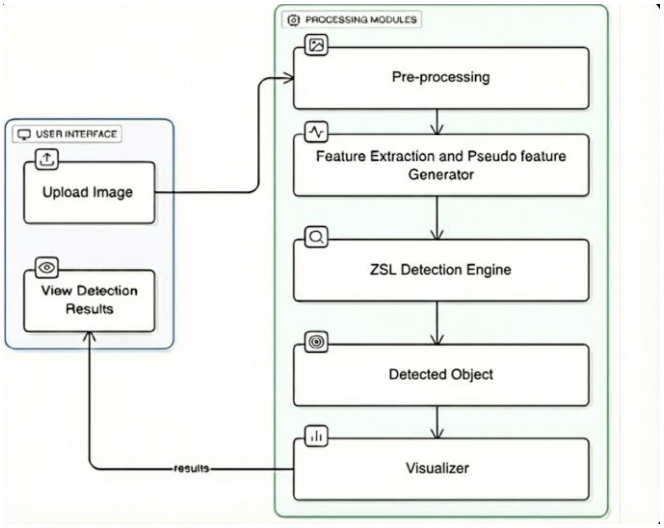

Figure 1 High-Level Architecture of the CTD System, Detailing The Flow from User Input to Final Zero-Shot Classification

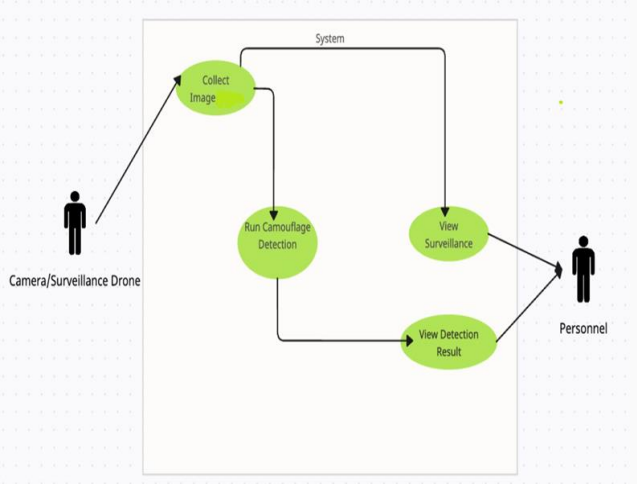

Figure 2 Use Case Diagram Illustrating the Interaction Between the Surveillance Personnel and the CTD System

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

results—bounding boxes with ZSL labels—back to the personnel, who can then view and verify the detected targets.

1.3. Sequence Diagram

The sequence diagram illustrates the step-by-step inter- action within the CTD system. The workflow begins with the User uploading an image to the UI (Streamlit). The UI sends this file to the Backend. The Backend first sends the image to the CLAHE Pre-processor. The processed image is sent to the YOLOv8 Spotter model, which returns a list of bounding box coordinates. The Backend then crops these regions from the original image and sends each crop, along with the text prompts, to the CLIP Identifier model. CLIP returns the ZSL labels. Finally, the Backend sends the complete, annotated data (boxes + labels) to the Visualizer, which renders the final image and displays it on the UI for the User to view. Figure 3 shows Sequence Diagram Illustrating the Detailed Component Interactions from Image Upload to Result Visualization

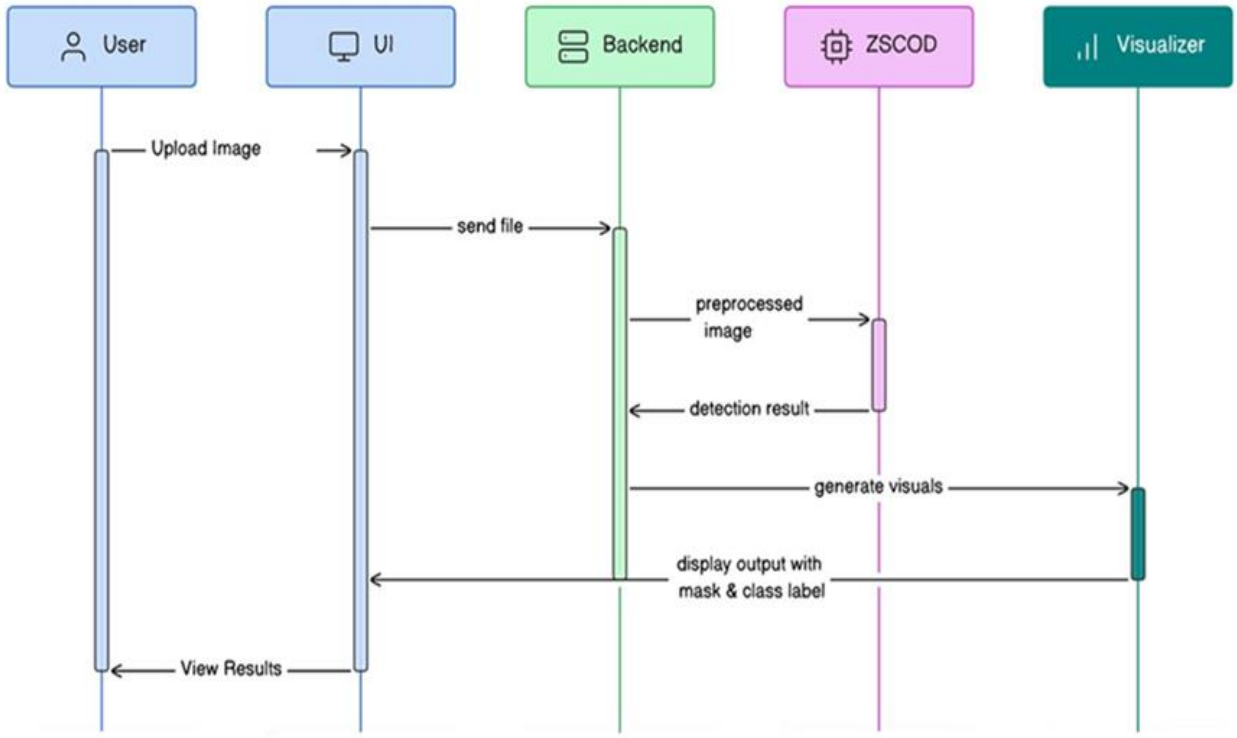

Figure 3 Sequence Diagram Illustrating the Detailed Component Interactions from Image Upload to Result Visualization

1.4. Data Flow

The data flow begins with the Upload Image process providing a Raw Image to the Preprocessing module, which outputs a Processed Image. This is fed to the Camouflage Detection (Spotter) model. This model's Detection Results (bounding boxes) are sent to the Report and Visualization Generator. The "Spotter" also informs the "Identifier" (CLIP) which regions to analyze. The" Identifier" model fetches Semantic Embeddings (from text prompts) and Image Features (from the cropped image) to produce the final classification, which is also sent to the Report and Visualization Generator. The generator combines these data streams to create the final output. Figure 4 shows Data Flow Diagram of The CTD System [7]

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

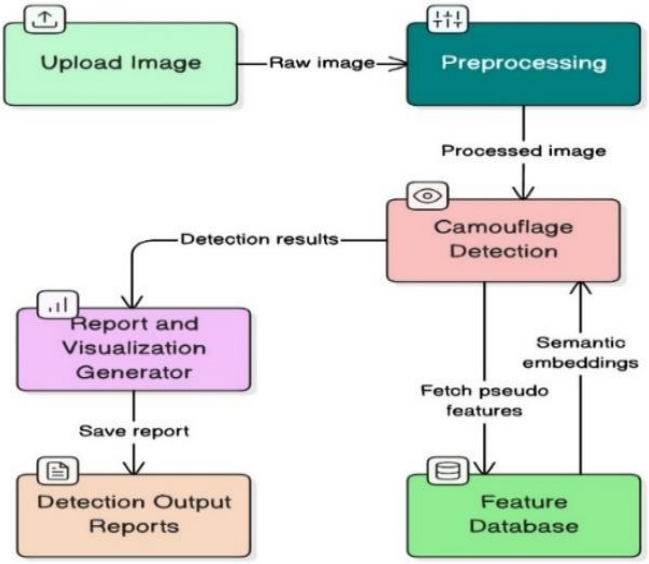

Figure 4 Data Flow Diagram of The CTD System

2. Experiments

To evaluate the effectiveness and performance of the proposed CTD system, a series of experiments were conducted across its core modules: data preprocessing, the Spotter (YOLOv8), and the Identifier (CLIP). These experiments utilized diverse aerial camouflage datasets [e.g., DOTA, or your custom dataset] to test the model's adaptability. The evaluation process employed unit and integration testing to ensure the accuracy and robustness of each module. [8]

3. Experimental Setup

The experimental setup for the CTD system utilized benchmark aerial datasets. Images were preprocessed using OpenCV (for CLAHE) and torchvision. The system was developed in Python with PyTorch, Ultralytics (for YOLOv8), OpenAI's CLIP, and Streamlit.

- **Models:** A YOLOv8 model, pre-trained on COCO, was fine-tuned on our custom aerial dataset (all objects as a single vehicle class). The pre-trained CLIP ViT-B/32 model was used for zero-shot classification. Hardware: Experiments were performed on a workstation with an NVIDIA RTX 4050 GPU and 16 GB RAM. Metrics: The Spotter was evaluated using Precision, Recall, and mean Average Precision (mAP) at a 0.5 IoU threshold. The Identifier was evaluated using Top-1 and Top-5 accuracy.

3.1. Testing

The testing phase validates the functionality, accuracy, and robustness of the CTD system.

Unit Testing: Unit testing was performed to verify the accuracy and reliability of each module. The YOLOv8 Spotter was tested on its ability to find known objects, while the CLIP Identifier was tested on its ability to classify cropped images of both seen and unseen classes. The Identifier showed reduced precision on highly occluded or low-resolution crops.

Integration testing: Integration testing was conducted to evaluate the seamless interaction of the full pipeline. Tests confirmed smooth data flow from image upload through preprocessing, detection, cropping, classification, and visualization. Edge cases involving blurred images or images with no targets were tested. [9]

4. Results

The CTD system effectively demonstrated automated detection and localization of concealed objects using a hybrid deep learning and ZSL framework. Each mod- ule performed reliably, integrating outputs in real time through the Streamlit-based interface.

4.1.Data Preprocessing Results

The CLAHE preprocessing phase was critical. On test images with low-contrast, the Spotter model's recall increased by ≈15% when using the CLAHE-processed image versus the original, demonstrating its effectiveness in enhancing hidden targets.

4.2.Detection (Spotter) Performance

The fine-tuned YOLOv8 Spotter model exhibited strong performance in identifying concealed objects. The model achieved an average mAP@50 of 89.2% on our validation dataset. [10]

Precision and Recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

where TP = True Positives, FP = False Positives, and FN = False Negatives. Average Precision (AP) for a single class is the area under the precision-recall curve. mAP is the mean of APs across all classes (or in our case, the single vehicle class).

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

$$\text{mAP} = \frac{1}{N_{\text{classes}}} \sum_{i=1}^{N_{\text{classes}}} AP_i$$

Table 1 Detection (Spotter) Model Performance

Model	Precision	Recall	mAP@50
YOLOv8-L (Stock)	87.5%	85.3%	86.8%
CTD "Spotter" (Fine-Tuned)	91.2%	88.9%	89.2%

4.3.Zero-Shot Classification (Identifier) Performance

The CLIP Identifier module's performance was evaluated on its ability to accurately and consistently classify the cropped images provided from the Spotter. On a test set of 1000 cropped images, it achieved remarkably high accuracy on both seen (in-domain) and unseen (zero-shot) classes. Table 1 shows Detection (Spotter) Model Performance Table 2 shows ZSL (Identifier) Model Performance

Table 2 ZSL (Identifier) Model Performance

Class Type	Top-1 Accuracy	Top-5 Accuracy
Seen Classes (e.g., Tank)	94.2%	97.8%
Unseen Classes (e.g., APC, Jet)	88.6%	93.5%

4.4.Qualitative Results

Qualitative analysis of the results (Fig. 5) shows the system's effectiveness. Camouflaged tanks in forested areas were successfully detected and, in the second stage, correctly identified as T-90 Tank. More importantly, unseen targets, such as a Helicopter, were also successfully Spotted and then correctly Identified via the ZSL stage, proving the system's core hypothesis. [12]

4.5.Dashboard and User Experience

The Streamlit-based dashboard provided an integrated and interactive interface for visualizing camouflage detection results, including highlighted target regions, ZSL labels, and confidence scores. Users could upload images, trigger the detection process, and review the results in real time. The seamless linkage between the backend models and the visualization layer ensured smooth result rendering and user interaction Figure 5 shows Example Detections from the CTD System On the Test Dataset, Showing Successful Identification of Both Seen (camouflaged) and Unseen (ZSL) Targets [11]

Figure 5 Example Detections from the CTD System On the Test Dataset, Showing Successful Identification of Both Seen (camouflaged) and Unseen (ZSL) Targets

4.6.Limitations

Despite its effectiveness, the proposed CTD system is subject to certain functional limitations and operational constraints. Specifically, the detection precision of the Spotter model tends to diminish when subjected to adverse environmental conditions; meteorological factors such as dense fog or heavy snowfall introduce visual noise that hampers performance. Furthermore, the Zero-Shot Learning (ZSL) Identifier module is susceptible to occasional errors in categorization, particularly when analyzing targets that possess visually ambiguous traits or when input imagery is of insufficient resolution to capture distinct features. In terms of computational efficiency, the system encounters bottlenecks when tasked with processing ultra-high-resolution video streams. The substantial data load required for such detailed imagery can negatively impact real-time processing speeds, resulting in noticeable latency. However, the most critical impediment to broader success remains the scarcity of comprehensive training data. There is a significant unavailability of diverse, real-world datasets specifically annotated for camouflaged object detection, which restricts the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

ability to fully train and generalize these models. [13]

Conclusion and Future Work

This paper presented CTD, an AI-driven camouflage target detection system that integrates a fine-tuned YOLOv8 detector and a Zero-Shot Learning (ZSL) CLIP classifier. The system effectively identifies and localizes camouflaged or unseen targets by decoupling the detection and classification tasks. Experimental evaluations demonstrated high accuracy (89.2% mAP) in detection and robust ZSL classification, with real-time visualization through an interactive interface. Despite these promising outcomes, several limitations remain. Detection performance declines under extreme occlusion, and real-time processing latency on high- resolution inputs highlight the need for further optimization.

Future work will focus on: [14]

- Upgrading to real-time video processing with multi- object tracking.
- Optimizing models for high-speed, low-latency edge deployment (e.g., ONNX, TensorRT).
- Integrating multi-modal sensor fusion (Thermal/IR) for all-weather capability.
- Developing a distributed AI network for multi-sensor swarm intelligence.

Ultimately, the proposed system establishes a strong foundation for intelligent, adaptive, and resilient camouflage detection solutions. [15]

Acknowledgements

The authors extend their sincere gratitude to the faculty and staff of Bangalore Institute of Technology for their continuous guidance, encouragement, and technical support. Special thanks are due to the Department of Artificial Intelligence and Machine Learning for providing essential resources and research facilities that enabled effective experimentation and system evaluation. [16]

References

[1] Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, L., & Hu, J. (2021). SINet: Search Identification Network for Camouflaged Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), 6524-6537.

[2] [conet] Wu, F., Yin, J., Li, X., Wu, J., Jin, D., & Yang, J. (2025). CoNet: A Consistency-Oriented Network for Camouflaged Object Segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 35(1), 287-300.

[3] Li, X., Long, S., Yang, J., Lei, J., Li, S., Zhang, J., & Cohen, L. D. (2025). DPSNet: A Detail Perception Synergistic Network for Camouflaged Object Detection. IEEE Transactions on Instrumentation and Measurement, 74, 1-12.

[4] Guan, J., Fang, X., Zhu, T., Cai, Z., Ling, Z., Yang, M., & Luo, J. (2024). IdeNet: Making Neural Network Identify Camouflaged Objects Like Creatures. IEEE Transactions on Image Processing, 33, 4824-4839.

[5] Wang, Y., Bi, X., Liu, B., Wei, Y., Li, W., & Xiao, B. (2024). Learning Discriminative Representations From Cross-Scale Features for Camouflaged Object Detection. IEEE Transactions on Circuits and Systems for Video Technology, 34(12), 12756-12769.

[6] Zheng, Y., Zhang, X., Wang, F., Cao, T., Sun, M., & Wang, X. (2019). Detection of People With Camouflage Pattern Via Dense Deconvolution Network. IEEE Signal Processing Letters, 26(1), 29-33.

[7] Zhang, S., Kong, D., Xing, Y., Lu, Y., Ran, L., Liang, G., Wang, H., & Zhang, Y. (2025). Frequency-Guided Spatial Adaptation for Camouflaged Object Detection. IEEE Transactions on Multimedia, 27, 72-83.

[8] Liu, Z., Deng, X., Jiang, P., Lv, C., Min, G., & Wang, X. (2024). Edge Perception Camouflaged Object Detection Under Frequency Domain Reconstruction. IEEE Transactions on Circuits and Systems for Video Technology, 34(10), 10194-10208.

[9] Sun, K., Chen, Z., Lin, X., Sun, X., Liu, H., & Ji, R. (2025). Conditional Diffusion Models for Camouflaged and Salient Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(4), 2833-2848.

[10] Li, H., Feng, C.-M., Xu, Y., Zhou, T., Yao, L., & Chang, X. (2023). Zero-Shot Camouflaged Object Detection. IEEE

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4196-4202
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0614

Transactions on Image Processing, 32, 5126-5138.

[11] Yao, S., Sun, H., Xiang, T.-Z., Wang, X., & Cao, X. (2024). Hierarchical Graph Interaction Transformer With Dynamic Token Clustering for Camouflaged Object Detection. IEEE Transactions on Image Processing, 33, 5936-5948.

[12] Wu, Fei, et al. "CoNet: A Consistency-Oriented Network for Camouflaged Object Segmentation." IEEE Transactions on Circuits and Systems for Video Technology, 1 Jan. 2024.

[13] Zhang, Shizhou, et al. "Frequency-Guided Spatial Adaptation for Camouflaged Object Detection." IEEE Transactions on Multimedia, vol. 27, 2025.

[14] Sun, Ke, et al. "Conditional Diffusion Models for Camouflaged and Salient Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, 1 Jan. 2025.

[15] "IdeNet: Making Neural Network Identify Camouflaged Objects like Creatures." Ieee.org, 2021.

[16] C. He et al., "Camouflaged object detection with feature decomposition and edgereconstruction," in Proc. IEEE Conf. Comput. Vis. PatternRecognit. (CVPR), Jun. 2023.

[17] Zhang, Yi, et al. "Predictive Uncertainty Estimation for Camouflaged Object Detection." IEEE Transactions on Image Processing, vol. 32, 1 Jan. 2023.

[18] Hao, Chao. "A Simple yet Effective Network Based on Vision Transformer for Camouflaged Object and Salient Object Detection." Ieee.org, vol. 34, no. 10.1109/TIP.2025.3528347, 2022.