

Camouflaged Object Segmentation Using a U-Net Based Deep Learning Approach

Dharitree Devi¹, Arpita Prusty², Deepanjali Khadia³, Basudev Patra⁴, Jashasmita Pal⁵

¹UG Scholars, Dept. of CSE, RadhaKrishna Institute of Technology and Engineering, Khordha, Odisha, India

^{2,3,4,5}Associate Professor, Dept. of CSE, RadhaKrishna Institute of Technology and Engineering, Khordha, Odisha, India

Emails: d.dharitree30@gmail.com¹, arpitaprusty21@gmail.com², deepanjaliKhadia51@gmail.com³, basudevpatra131@gmail.com⁴, jashasmita_cse@riteindia.edu.in⁵

Abstract

The paper focuses on the computer vision task of camouflaged object detection that has various applications, including, but not limited to, military, security, biology, and industry. From the perspective of machine vision, camouflage is a state in which the target's texture, colour, and pattern are indistinguishable from the surrounding background, making it hard, if not impossible, to isolate those objects from the background. It implies that the COD problem statement extends far beyond the scope of ordinary object detection. To tackle the COD task, the authors used the 2D U-Net model that demonstrated its ability to address image segmentation problems through employing a contracting encoder and expansive decoder connected with skip links. In particular, the U-Net architecture was utilised due to its ability to maintain a high level of detail in an image, which is critical during camouflage detection, in contrast to typical bounding box estimation used in object detection tasks. The model was trained on MCSIK training sets and evaluated on MCSIK test sets, with an average accuracy of 70.14%. Overall, the U-Net architecture is promising for addressing camouflage object detection; however, there is a noticeable accuracy drop on real-world sensor data, which the paper tries to address further.

Keywords: COD, DeepLearning, 2DU-NetConvolutionalNeuralNetwork(CNN), MCSIK, Semantic Segmentation.

1. Introduction

Typically, object detection algorithms assume that the target is presented in the most contrasted manner on the background. The higher the contrast is between an object and its surrounding background, the easier it is for a bounding box to achieve the maximum level of localization accuracy. On the contrary, the camouflaged object detection (COD) paradigm puts forward an opposite requirement that the target should be presented in the least contrast manner possible

1.1.Problem Restatement

Annotated boxes often contain much background pixels, in this context, the use of bounding boxes for COD appears to lead to an increase in the relative amount of irrelevant information in the image. More specifically, the background around the target in COD is statistically close to the target itself, so there is a more significant challenge to distinguish the discontinuity. This necessitates a more involved segmentation engine, such as 2DU-Net convolutional

neural network described in the paper. As shown in Table 1 Traditional object detection versus camouflaged object detection.

Table 1 Traditional object detection versus camouflaged object detection

Aspect	Traditional Detection	Camouflaged Detection
Background relationship	Distinct and contrasting	Blended texture, colour and frequency
Boundary clarity	Sharp, continuous contours	Ambiguous, fuzzy, disrupted edges
Required technique	Bounding box (coarse)	Semantic segmentation (pixel-level)

1.2.Applications of Camouflaged Object Detection

Military & Security: Detection of ghillie suits and stealthy personnel is a challenging task due to the disruptive pattern painting and shadow concealment, which are actively used to bypass bounding-box detection.

Biological Research: The technology can be used to identify species based on the way they use camouflage in different biomes thus saving a lot of time on extensive biological research that requires expert knowledge on the targeted species.

Agriculture: Detecting agricultural pests like locusts or cotton bollworms that are about to destroy large portions of crops is significant in the agricultural industry.

Industrial Inspection: The inspection process benefits from COD's ability to separate subtle artifacts from large-scale patterns found in microstructure textures.

1.3.Contributions and Paper Organisation

There are three main contributions of this paper. First, we apply a 2D U-Net convolutional neural network to the task of camouflaged object segmentation on the MCSIK dataset and report a baseline average accuracy of 70.14%. Second, we evaluate the baseline model on the COD10K-C benchmark to characterise how segmentation accuracy depends on specific corruptions and their levels, quantifying to which extent such an accuracy drop is affected by the choice of image acquisition sensor. We posit that this kind of analysis is crucial for COD research to account for real-world conditions, seldom considered by COD studies that only employ clean benchmarks. Finally, we use auxiliary uncertainty maps to locate segmentation failures due to ambiguities inherent to camouflage, revealing patterns according to which such failures are distributed non-uniformly in the mask, and propose architectural improvements based on this analysis. The rest of this paper is organised as follows: Section 2 describes the methodology used for segmentation, including the structure of the segmentation engine, the U-Net encoder-decoder with skip-bindings, and the resources used for experimentation. Section 3 reports the baseline accuracy, corruption robustness analysis, and

provides a sanity check for the failure modes using uncertainty maps. Finally, Section 4 defines directions of future work to close the gap in robustness, and the paper concludes.

2. Methodology

2.1.The Segmentation Engine: Why 2D U-Net?

A box drawn around a camouflaged object will contain a substantial amount of background pixels, due to lack of local contrast. The need to discern a camouflage-blurred boundary requires the use of a semantic instance-level segmentation. This work utilizes the 2D U-Net convolutional neural network.

2.2.Network Architecture

The 2D U-Net is a symmetric encoder-decoder convolutional neural network, illustrated in Figure 1 below. The encoder extracts contextual patterns, downsampled by MAX-POOL blocks until reaching the lowest resolution. This resolution, referred to as the bottleneck, contains the global information while losing the local detail. The decoder upsamples the low-resolution global feature map generated by the globalpooling to recover the local image information through transpose convolution layers. Finally, a 1×1 convolution with softmax activation is applied to predict a class distribution over pixels.

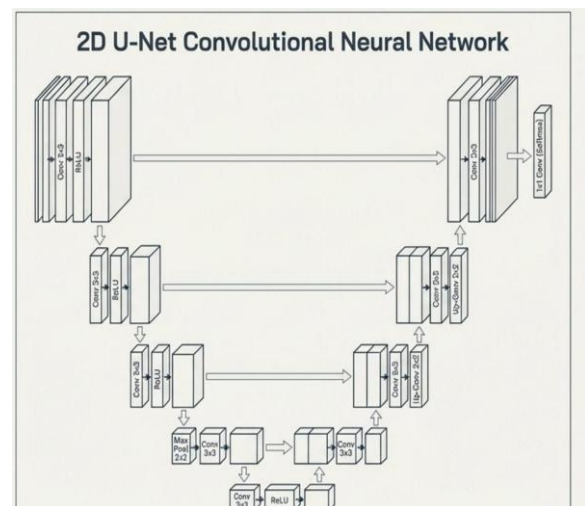


Figure 1 Convolution network architecture of the 2D U-Net

2.3.Skip Connections (Neglect Links)

U-Net: A Convolutional Network for Biomedical Image Segmentation Figure Essay.Camouflage

equals local contrast between an object and a background, resulting in reduced discontinuity information in skip connections (Figure 2). Given that skip connections are used in the encoder-bottleneck, the decoder lacks discontinuity data in favor of continuity. The U-Net design overcomes this challenge by introducing skip connections (also called neglect links), which enable the decoder to recover high-resolution spatial detail with a trade-off of complexity (Figure 2). Likewise, high-resolution skip connections enable the network to maintain spatial continuity, which is paramount for discontinuity localization in instances of camouflage.

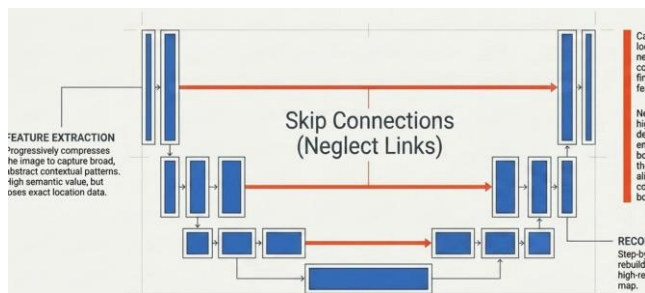


Figure 2 Skip connections (“neglect links”) of the 2D U-Net convolutional neural network

2.4.Dataset and Evaluation

The network was trained on MCSIK training set and benchmarked on the test set. Both contain images with hidden objects, which are challenging to locate. The efficacy of localization is measured by average pixel accuracy across the test set.

3. Results And Discussion

3.1.Baseline Evaluation

With 70.14% average pixel accuracy (Table 3), the 2D U-Net baseline demonstrates its capacity to encode fine spatial detail, making it a good candidate for network modifications. Specifically, for COD, accuracy typically decreases for objects with similar physical properties to the background. Therefore, the 2D U-Net baseline is a reasonable reference point, from which to explore further improvements for hidden object detection. With the addition of skip connections, the network can generalize fine spatial detail to other objects, thus enabling localization of ambiguous discontinuities.

3.2.Real-World Robustness

A baseline network that demonstrated excellent results in laboratory environment conditions may show poor accuracy in actual applications (Table 2). Besides favorable background/background conditions in laboratory settings, modern

Table 2 Corruption vulnerability matrix (COD10K-C benchmark) Corruption type Severity Dice coefficient
PhotometricLOW~0.97Geometric CRITICAL~-0.185

Corruption Type	Threat Level	Impact on Performance
Fog, brightness, contrast (Photometric)	LOW	~97% Dice retention; local texture gradients remain largely intact
Motion blur, Gaussian noise (Geometric)	CRITICAL	-18.5 Dice points; leading models lose over 25% performance

computer vision applications demand networks capable of operating reliably under partially severe weather conditions and objects moving at very high speeds. To estimate the network’s generalization performance, Table 2 below uses COD10K-C benchmark that characterizes corrupted images by two factors – photometric that impact the image’s brightness/color and geometric that impact object appearance.

3.3.Diagnostic Telemetry: Visualizing Uncertainty

In order to localize regions, where the network failed to encode fine spatial detail, one can utilize an auxiliary uncertainty logit. Typically, uncertainty is visualized as a narrow band along the boundary of a camouflaged object, rather than spanning the entire object mask (Figure 4). The same trend can be observed for the geometrically corrupted input: motion blur leads to inability to localize discontinuity. As such, visualization of uncertainty is

an effective method to identify failure modes of the network.

4. Towards Bridging the Robustness Gap

COD should be treated as a separate computer vision paradigm, requiring a separate set of tools. First, it is necessary to establish a baseline segmentation network, such as the 2D U-Net described earlier in this article, which manages to achieve 70.14% accuracy on the MCSIK test set (Table 3). Next, in order to achieve robustness to severe conditions, it is beneficial to equip the network with a dedicated high-frequency edge detection prior, such as the frequency-aware decoder blocks employed in several COD-specific publications. Finally, in to generalize to different domains, it is necessary to utilize generative augmentation to train the network on corrupted data, teaching it to handle domain-specific failure modes, such as motion blur. Altogether, these three principles can serve as a foundation for analyzing challenges of camouflaged object detection.

Conclusion

This paper has demonstrated that detecting objects in camouflage requires a fundamentally different approach, as compared to standard computer vision tasks. Specifically, COD benefits from being reframed as a semantic instance segmentation problem, rather than a classical bounding-box-based one. The introduction of skip connections allows the 2D U-Net convolutional neural network to serve as a valuable baseline for further network modifications. Meanwhile, the current work's corruption vulnerability analysis has identified motion blur and other geometric transformations as particularly detrimental to performance of current COD architectures. Future work will focus on incorporating the high-frequency edge detection priors and applying generative augmentation for generalization to different domains.

Acknowledgements

The authors would like to thank the department of Computer Science and Engineering for providing computational resources that made this research possible.

References

[1]. Ronneberger, O., P., & Brox, T. (2015). U-Net:

Convolutional network for biomedical image segmentation. In Medical image computing and computer-assisted intervention – MICCAI 2015 (pp. 234-241). Springer, Cham.

[2]. Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, J., & Shao, L. (2020). Camouflaged object detection: A benchmark and analysis. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34 pp. 9455-9463.) AAAI Press.

[3]. Zhu, Z., Liu, Z., Chen, C., Zhang, R., Fu, Q., Yang, X., ... & Zhang, Q. (2021). cod10k: A large-scale benchmark for camouflaged object detection [3] Skurowski, P., Abdulameer, H., Blaszczyk, J., Depta, T., Kornacki, A., & Kozieł, P. (2018). Animal camouflage analysis: CHAMELEON database. Unpublished manuscript.

[4]. R., Sun, M., S., Zhou, X., Zhang, Cong, W., & Zhao, Y. (2023). Frequency perception network for camouflaged object detection. In Proceedings of the 31 st ACM International Conference on Multimedia (pp. 1179-1189) .

[5]. K., Zaheer, M. Z., Islam, A., & Nandakumar, Mahmood, K. (2024). DiffuseMix: Label-preserving data augmentation with diffusion models. In Proceeded of the Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 27621-27630).