Data Augmentation Techniques for Camouflaged Object Detection

Vinitra Muralikrishnan
UMass Amherst
vmuralikrish@umass.edu

Rahul Saxena
UMass Amherst
rahulsaxena@umass.edu

Mustafa Chasmai
UMass Amherst
mchasmai@umass.edu

1. Overview

The essence of image segmentation and detection is to find objects that stand-out or are different from their background, which is also how humans typically imagine "objects". But what if an object does not want to be found? Camouflaged objects tend to blend-in with their surroundings by using similar colors and textures. Natural selection has forced many species to evolve and use sophisticated camouflage mechanisms to avoid detection by predators. Soldiers use camouflage to move covertly and avoid detection by enemies. We address the problem of detecting these objects that expressly wish to avoid detection. Camouflaged objects are abundant in nature and can be extremely challenging to detect, even for humans. Having the capability to identify objects concealed in plain sight holds significant implications for search and rescue operations, tracking aquatic species, defect detection, and more. However, unlike generic object detection datasets, benchmark datasets for Camouflaged Object Detection (COD) are relatively smaller and lack diversity in labels. In this project, we aim to explore various methods to alleviate this limitation and catalyze future research in this area.

2. Related Work

Research into camouflaged object detection has a rich history in biology and the potential impact of tackling this is tremendous. However, within the domain of computer vision, this area remains relatively under-explored compared to other forms of object detection, such as Generic Object Detection [2, 6] and Salient Object Detection [1]. This discrepancy can be partly attributed to the inherent complexity of the task and the scarcity of well-labelled annotated data. Interest in this space was kickstarted with the contributions of CamouflagedAnimals [7] and CAMO [5], and found momentum with the contribution of the COD10K dataset [3]. CamouflagedAnimals [7] includes videos of camouflaged animals, which immediately pop-out as they start to move. CAMO [5] consists of 1250 images and was the first work dedicated primarily to camouflaged objects. MirrorNet [12] improved over existing approaches on CAMO [5] by fusing predictions from mirrored data. COD10K [3] greatly increased the scale of this problem by constructing a dataset $\sim$ 10 times bigger. They also propose SiNet, which mimics the human receptive fields and the "search and identification" stages of predators. Many animals are adept at using the environment to blend and conceal themselves to avoid detection. The role of background and textures in this concealment has been studied and explored by Ren el al., 2021 [8]. Xiang el al., 2022 [11] explore the contribution of depth and attempts to use them for detection.

3. Dataset Description

For this project we will be using 2 benchmark datasets to test our hypotheses: CAMO [5] and COD10K [3]. CAMO [5] consists of 1250 images with naturally camouflaged objects like fish, chameleons and insects as well as artificially camouflaged objects like soldiers and body painting. Each image has at least one camouflaged object and corresponding manually-annotated pixel-wise segmentation masks. COD10K [3] consists of 10,000 images containing terrestial, aquating and flying animals as well as amphibians and body art. Most images were downloaded from Flicker by filtering with appropriate key-words. Each image is accompanied with rich hierarchical annotations including bounding-box, object-level, and instance-level labels. We will be working on a semantic segmentation task and will use only the corresponding labels in COD10K [3].

4. Proposed Plan

Our work would be focused towards exploring different ways to improve performance of a baseline segmentation model with a Resnet-18 [4] backbone. We also plan to conduct a study on leveraging generative models and explore whether it can be used to reduce the gap between the COD datasets and generic image segmentation datasets.

4.1. Style-Transfer as Augmentation

The existing benchmark datasets like COD10K [3] has a great label and object diversity but still is quite imbalanced at the fine-grained level. To further improve the mod-

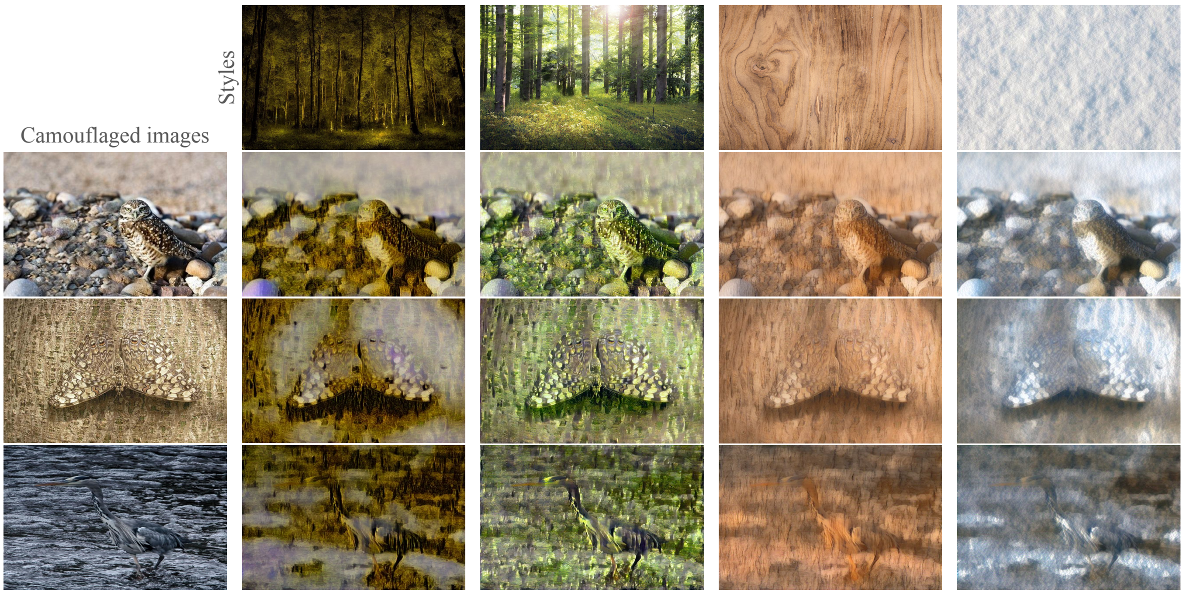

Figure 1. **Style-Transfer as Augmentation.** We consider a few images from the COD10K dataset, and apply style transfer on them with a few natural-like textures. We use a VGG-19 model pre-trained on ImageNet for neural style transfer.

els become more robust to this task, we propose a data augmentation pipeline by generating images using Neural Style Transfer. This would involve using images from COD10K transformed into different nature-like backgrounds and textures. This pipeline can be applied on other datasets of concealed object detection, further improving the robustness towards detection in diverse backgrounds and environments. Fig 1 presents some samples taken from COD10K transformed so far.

4.2. Finding New Camouflaged Images

Camouflaged objects are abundant in nature, and a good proportion of the data in COD10K [3] consists of species. However, the scale of this datasets is still much smaller than species detection datasets like the iNaturalist [10] dataset, with over 450K images with bounding box annotations. We expect some proportion of these images (for e.g. butterflies and chameleons) to already have camouflaged objects. The internet has a significantly larger collection of natural images, and we could use existing detectors to pseudo-label bounding boxes for them as well. We propose to use a classification model to filter such images and add them to the COD10K [3]. We could train a simple image classifier with camouflaged and non-camouflaged objects in COD10K. We could also compare the pixels inside and around the bounding boxes and use some heuristics to find objects that are well blended with their surroundings, which is essentially what a camouflaged object would look like. Even with aggressive filtering, if we can get about 1% of the data in iNaturalist, that would increase the size of COD10K by 50%, which would be a good contribution to this field. If we consider the entire set of internet images, this could go an order of magnitude higher.

4.3. Exploring Benefits of Synthetic Data

On a low priority, but we would like to explore recent advances in image generation to generate synthetic examples of camouflaged objects. Similar to the style-transfer based augmentation idea, we can generate new backgrounds for objects in existing data (same object in a different background). With the famed power of stable-diffusion [9], it might even be possible to generate camouflaged images directly from simple text prompts like “butterfly camouflaged in a wood background”. Although the quality and benefits of AI-generated images to train new AI is often debatable, this is an interesting direction to explore. Considering the relatively smaller size of the datasets here and the difficulty in annotation, generating synthetic data, if helpful, could bridge the gap between camouflaged and generic image segmentation datasets like COCO [6] and Pascal-VOC [2].

References

[1] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. Salient object detection: A survey. *Computational visual media*, 5:117–150, 2019. 1

[2] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111:98–136, 2015. 1, 2

[3] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, 2020. 1, 2

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[5] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Computer vision and image understanding*, 184:45–56, 2019. 1

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 1, 2

[7] Erik Learned-Miller Pia Bideau. It's moving! a probabilistic model for causal motion segmentation in moving camera videos. In *European Conference on Computer Vision (ECCV)*, 2016. 1

[8] Jingjing Ren, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Yangyang Xu, Weiming Wang, Zijun Deng, and Pheng-Ann Heng. Deep texture-aware features for camouflaged object detection, 2021. 1

[9] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[10] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018. 2

[11] Mochu Xiang, Jing Zhang, Yunqiu Lv, Aixuan Li, Yiran Zhong, and Yuchao Dai. Exploring depth contribution for camouflaged object detection, 2022. 1

[12] Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, and Tam V. Nguyen. Mirrornet: Bio-inspired camouflaged object segmentation. *IEEE Access*, 9:43290–43300, 2021. 1