

Detecting Camouflaged Object in Frequency Domain

Yijie Zhong¹ Bo Li² Lv Tang Senyun Kuang³ Shuang Wu² Shouhong Ding²
¹ Tongji University, ² Youtu Lab, Tencent, ³ Southwest Jiaotong University
dun.hasaki@gmail.com, libraboli@tencent.com, luckybird1994@gmail.com
syKuang@my.swjtu.edu.cn, calvinwu@tencent.com, ericshding@tencent.com

1. Introduction

This supplementary materials contains three parts:

- Section 2 provides more comprehensive analyses of the proposed network.
- Section 3 provides an analysis of the failure samples to show the limitation of our proposed model.
- Section 4 explores possible application scenarios for the COD task.

2. Function of frequency information

In this part we discuss the importance of using both RGB and frequency domain information. First, we change the input of the vllina U-Net to the frequency signals $\mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 192}$ after the input enhancement, as shown in Figure 1 (b). Due to the low resolution, the network is only able to localize the camouflaged objects. Also, our adjustment of the first layer of the backbone can greatly affect the powerful prior of the pre-trained model trained on ImageNet. Our proposed network can get the best performance (show in Figure 1 (c)) when RGB and frequency domain information are used at the same time and combined by FA. While using frequency domain information can help locate the targets, this is not enough for COD task. The RGB information also contains some clues that are not presented in the frequency signals.

3. Failure case study and limitation

Firstly, we only consider the DCT transform. There are other different transformations (*e.g.*, Wavelet, FFT). This will serve as our future work. Also, some edges and details of the objects cannot be accurately reproduced, such as spider feet, which is not concern in this paper.

Secondly, we introduce the frequency domain information to help the network locate the camouflaged objects. However, when the original JPEG image has a large compression rate, the image does not change dramatically under human eye observation, but it leads to a large amount

of information loss. As shown in Figure 2, the coefficients extracted from the original JPEG images cannot help us to find the camouflaged objects. We can find that there is almost no difference to help us find the camouflaged objects, as shown in Figure 2 (d). The high-frequency information already lost (show in Figure 2 (e)). Most of the energy is concentrated in the low frequency spectrum, so it resulted in a high frequency signal that is no longer discriminative. This also leads to a failure of locating the camouflaged objects even if we use the frequency domain information as an aid, because the camouflaged regions have similar frequency signals to other regions.

4. Potential application

Concealed object detection systems have various downstream applications in fields such as medicine, art, and agriculture. Here, we envision some potential uses.

Medicine. Early diagnosis through medical imaging plays a key role in the treatment of diseases. However, the early disease area usually have a high degree of homogeneity with the surrounding tissues. Similar to camouflaged object detection, polyp segmentation (show in Figure. 3) faces several challenges, such as variation in appearance.

Agriculture. Our method may be used for the pest detection. Since early 2020, plagues of desert locusts have invaded the world. Large numbers of locusts gnaw on fields and completely destroy agricultural products, causing serious financial losses and famine due to food shortages. As shown in Figure 4, introducing AI-based techniques to provide scientific monitoring is feasible for achieving sustainable regulation/containment by governments.

Art. Background warping to conceal objects is a fascinating technique in the SIGGRAPH community. Some examples can be seen in [1]. We think that this technique will provide more training data for existing data-hungry deep learning models, and thus it is of value to explore the underline mechanism behind the feature search and conjunction search theory.

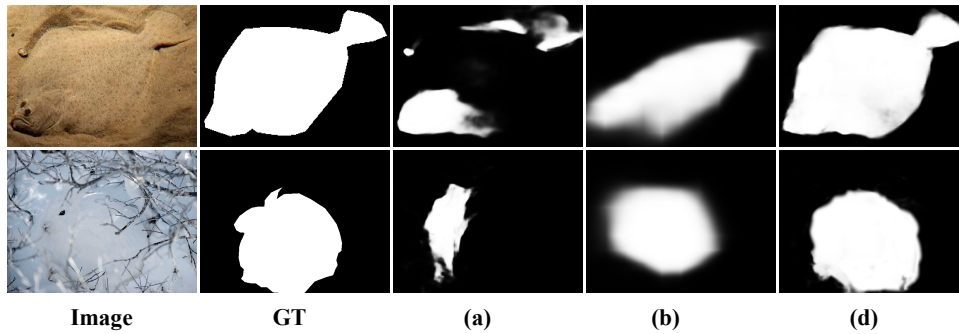


Figure 1. (a): model only using RGB information; (b): model only using frequency domain information; (c): model using both RGB and frequency information.

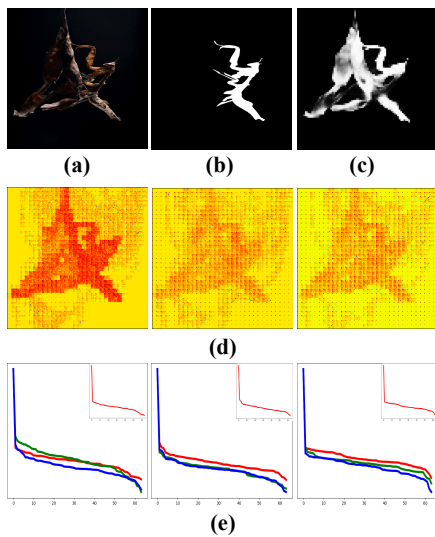


Figure 2. (a) denotes the input JPEG images which have a high compression ratio ($< 4\%$). (b) denotes the ground truth. (c) denotes our predictions. (d) denotes the DCT coefficients extracted from the JPEG images in different channels. (e) denotes the frequency energy from the low-frequency to the high-frequency. The red line represents the energy from the whole image. And the green and blue lines represent the energy extracted from the camouflaged region and the other region, respectively.

References

- [1] Hung-Kuo Chu, Wei-Hsin Hsu, Niloy J. Mitra, Daniel Cohen-Or, Tien-Tsin Wong, and Tong-Yee Lee. Camouflage images. *TOG*, 29(4):51:1–51:8, 2010.

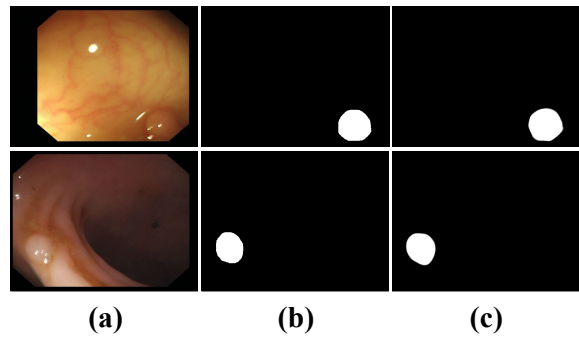


Figure 3. Polyp segmentation. (a) is input polyp images. (b) is corresponding ground truth. (c) is predictions from our method.

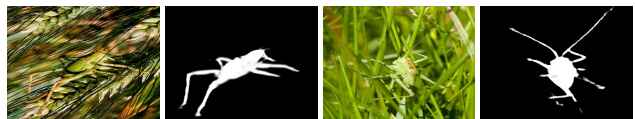


Figure 4. Some examples and our predictions.