트랜스포머기반 유리 영역 검출방법

후샤오항 [1], 고서 [1], 양승준 [2], 조경은 [1]*
[1]동국대학교 멀티미디어공학과
[2]한국전자통신연구원
*e-mail: cke@dongguk.edu (교신저자)

Transformer-based glass area detection method

Xiaohang Hu[1], Rui Gao [1], Seung-Jun Yang[2], Kyungeun Cho[1]
[1]Dept of Multimedia Engineering, Dongguk University
[2]Electronics and Telecommunications Research Institute

Abstract

Glass is a common object in living environments, but even humans are sometimes unable to identify it. This study proposes a method for detecting glass area by learning edge information from images. The network structure of Transformer is used to accept the base features extracted by backbone and extract the boundary information of RGB images, and both features are used to learn the features of glass area and determine the glass area based on these boundary features. The experimental results show that our proposed method can detect glass area in images.

1. Introduction

Glass object detection refers to the detection of glass objects in natural scenes, such as glass doors, containers, and windows, which are common in living scenes. However, because glass has different properties in different environments, such as reflection, refraction, and varying colors, detection can be difficult, even for humans. Therefore, this study extracts features from RGB image to detect the presence of glass and determines the glass area based on these features.

2. Related works

Many studies and methods have attempted transparent object detection.

Mei et al. proposed a network that uses high-level and low-level contexts extracted from large fields to detect different sizes of glass in various scenes [1]. Cao et al. proposed a method to enhance the ability to distinguish boundaries. They also designed an adaptive Atrous spatial pyramid pooling module to capture features from multiple fields of view for semantic segmentation and boundary extraction [2].

Xie et al. proposed a transparent object segmentation dataset called Trans10K-v2, which has 11 common categories, and proposed a Transformer-based method that can provide a global sensory field [3].

He et al. proposed a new module for edge prediction and an effective edge-aware convolutional network module to extract the global shape of glass objects and obtain the final prediction [4].

However, the above studies were conducted for many types of transparent objects, including plastic products and resin products. These objects have different properties from glass, so the above studies achieved excellent performance on most small transparent objects but not windows, doors, and other glass objects. The method in this study extracts the boundary features of glass, uses a Transformer-based network for training, and uses the learned features to detect glass areas.

3. Glass detection method

Compared to CNNs that perceive the field of view locally, we use the self-attention in Transformer to provide a good global perception

of the field of view. Therefore we use a Transformer combined with CNN, which can combine local information and global information better compared to pure CNN.

Fig 1 depicts the proposed method, where the boundary features and RGB features of the image are first extracted and then input to the Transformer network.

In the Transformer network, the image is first partitioned into small matrices by the embedding part; then, the position embedding records the position information after the partition. Subsequently, the position embedding records the position information of the segmented image and inputs it into the Transformer Encoder and Transformer Decoder to calculate the query, key, and value of the features and finally output the result.

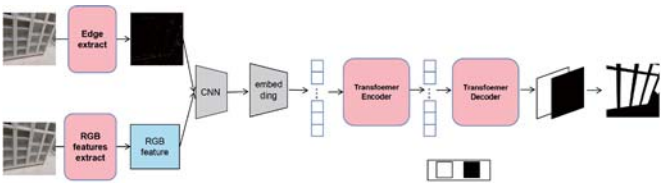

Figure 1. Overview of glass detection method

4. Experiments

The dataset we use is Trans10k-v2, which contains 11 classes of transparent objects and 5000 training data, 1000 validation data and 4428 test data. The red boxes in Fig 2 mark the boundary features of glass objects; the yellow boxes are not features of glass.

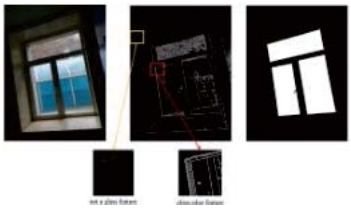

Figure 2. Glass boundary feature detection

Fig 3 shows part of our experimental results, which prove that the proposed method is effective for detecting glass.

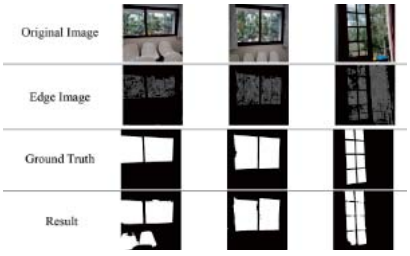

Figure 3. Results of our proposed method

The first line is the input image, and the second line is the boundary information extracted from the input image. The third line is the ground truth image, and the last line is the result.

5. Conclusion

Herein, we propose a method for detecting glass area base on Transformer networks. The results prove that the proposed method can detect glass area in images.

The current detection of the position of the glass object is accurate. However, the extraction of boundary information features is imperfect, which makes the boundary of its detection results very coarse.

In the future, we will study the Transformer network more extensively and design a more efficient glass object segmentation method for better application in scene modeling, glass reflection denoising, and other fields.

Acknowledgements

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [22ZH1200, The research of the fundamental media · contents technologies for hyper-realistic media space].

References

[1] Mei, Haiyang, et al. "Don't hit me! glass detection in real-world scenes." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[2] Cao, Yang, et al. "FakeMix augmentation improves transparent object detection." *arXiv preprint arXiv:2103.13279* (2021).

[3] Xie, Enze, et al. "Segmenting transparent object in the wild with transformer." *arXiv preprint arXiv:2101.08461* (2021).

[4] He, Hao, et al. "Enhanced boundary learning for glass-like object segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

[5] Zhang, Jiaming, et al. "Trans4Trans: Efficient transformer for transparent object segmentation to help visually impaired people navigate in the real world." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.