

CENet: Content-aware and Edge-aware Network for Salient Object Detection

Zhikuan Gao

School of Mathematics and Statistics, Nanjing University of Information Science & Technology,
Nanjing, 210044, China

(Received October 15, 2021, accepted January 13, 2022)

Abstract: Benefitting from Fully Convolutional Networks (FCNs), salient object detection methods have achieved prominent performance. However, there are still some challenges in this task: 1) lack of effective feature representation and integration make the result salient maps lose some regions of object, or bring some non-saliency regions. 2) suffering from the continuous pooling or stride operations, the predicted maps will lose some important spatial detail information, especially the boundary of object. To address these two problems, we propose the Content-aware and Edge-aware network (CENet) which contains three sub-modules: 1) we design a content-aware feature extraction module which uses a transformer block and channel-wise attention mechanism to capture the distinct content features and suppresses the non-saliency regions. 2) an edge-aware feature extraction module is introduced to learn the boundary features and predict the intact edge of the salient object. 3) a feature fusion module is proposed to integrate features from the first two module in a learning way. We also design a hybrid loss function which has better performance than the widely-used binary cross entropy loss. Results show that, our method can detect the intact salient object without losing regions of object or bring some non-saliency regions, and can also obtain the precise boundary. Experimented on several datasets, our method can achieve the state-of-art performance.

Keywords: Saliency object detection, Content-aware and edge-aware feature, Axial-attention transformer, Learning-based feature fusion one.

1. Introduction

Salient object detection aims at highlighting the most visually distinctive object in an image. Different from fixation prediction, which predicts the fixation points or locations that attract human attention mostly at firstly glance [1], salient object detection focuses on predict integrated object or regions and is used as a preprocessing step of many computer vision applications, such as semantic segmentation, image retrieval, image editing, image retargeting, person re-identification, video summarization , video salient object detection [2].

In the past few decades, lots of methods have been proposed for salient object detection. Conventional methods, which are motivated by the research on the human visual attention mechanism, usually use heuristic priors and hand-crafted feature such as color, contrast and texture. Although these methods performance well on some simple scenario, low-level features are not robust enough to distinguish the salient object and background regions in complex cases since the hand-crafted features have limit capability to capture the high-level semantic and structural information. Recently, the merge of Fully Convolutional Neural Networks (FCNs) [3] helps salience object detection make great strides and the FCNs becomes the widely used structure in salient object detection since its prominent capability to capture robust high-level features and stronger semantic information. Different from previous Convolutional Neural Networks (CNNs), FCN is an end-to-end structure which predicts pixel-wise outputs from arbitrary-size inputs at both learning and inference stage.

Motivated by this, more and more FCNs based methods were proposed and performed well on salient object detection task, however, there are still several problems[4-5]. To address above problems, we propose a content-aware and edge-aware network, referred to as CENet, for salient object detection. Our mainly contributions are summarized as follows:

- We present a novel network CENet for salient object detection, aiming to detect and segmentation the intact salient object with precise boundary. To obtain the intact object, we design a content-aware feature extraction module which learns more robust semantic information and more discriminative feature representation. To obtain the distinct edge, we design an edge-aware feature extraction module which transforms the HED edge maps into edge features and refine the boundary of predicted object.
- We propose a residual learning-based content-edge feature fusion module which integrate content features and edge features gradually. A batch concatenate operation is used during the fusion process to combine the different features wisely.
- We present a novel hybrid loss, which fuses binary cross entropy, focal loss, SSIM loss and IOU loss and has better performance than the widely used binary cross entropy loss, aiming to leads network to pay more attention on content and edge information.

2. Related Work

2.1. Traditional Methods

Early methods usually detect salient object using heuristic cues or hand-craft features, such as center prior, boundary and background, color contrast and texture. Besides, many graph-based methods are proposed for saliency detection. In [6], Shan et al. provide seeds for manifold using background weight map. In [7], Yang et al. construct a close-loop graph in which each node is a super pixel and propose a two-stage scheme. This way, the salient object detection problem is transformed to manifold ranking. Although these traditional approaches make great strides in salient object detection and performance well on simple scenario, they still hardly give a satisfying result in most complex scenes due to lack of the capability to capture high-level features.

2.2. FCN-based Methods

Recently, Convolutional Neural Network (CNN) has represented powerful capability of extracting high-level features and achieved prominent results in image classification. Motivated by this, some methods classify each pixel or super pixel based on CNN. They predict each pixel whether belongs to saliency regions or not. However, the saliency maps produced by these patch-wise based methods are usually coarse since the fully connected (fc) layers destroy the spatial structure of object.

To achieve pixel-wise prediction and obtain more precise results, in [3], Long et al. propose a Fully Convolutional Network (FCN) which is originally used to semantic segmentation. Several approaches begin to utilize the FCNs to generate end-to-end saliency map[10].

2.3. Edge-aware Guidance Methods

Existing methods can locate salient object precisely, however, how to segment the object from background with distinct boundary is still a challenge. To solve this problem, several methods detect salient object with edge-aware guidance[11-13].

2.4. Content-aware Guidance Methods

In addition to edge information, content information is also important for salient object detection. It