Abstract

Saliency plays a key role in various computer vision tasks. Extracting salient regions from images and videos has been a well-established problem in computer vision. Determining salient regions in an image or video has a lot of immediate applications, such as Anomaly detection in videos, efficient data compression and other derived applications, such as robot vision, salient object determination, context-aware image and video retouching, tracking, person re-identification.

With the advent of deep convolutional neural networks, many of the existing problems have witnessed a significant boost in performance. Segmentation tasks in particular, which require a general understanding of the scene, achieve a very high-performance gain in terms of IoU (Intersection over Union). The generalizable capability of the filters learned by these neural networks is suitable for tasks for which the network was not even trained. In this thesis, we explore three applications, image segmentation, video segmentation and automatic anomaly detection from videos.

For the salient object segmentation in images, we explore two novel recurrent attention based methods. The soft attention method uses a recurrent gating mechanism to extract the salient object segmentation from an image. This technique weighs certain parts of the image more than the others to refine the segmentation outputs gradually. The refinement procedure does not require increased parameters since the recurrent gates have shared weights. The hard attention method tackles the problem by iteratively attending to image patches in a recurrent fashion and subsequently enhancing the predicted segmentation mask. Saliency features are estimated independently for every image patch which is further combined using an aggregation strategy based on a Convolutional GRU based network. The proposed approach works in an end-to-end manner, removing background noise and false positives incrementally. Through extensive evaluation on various benchmark datasets, we show superior performance to the existing approaches without any post-processing.

The task of video object segmentation suffers from a number of challenges such as 3D parallax, camera shake, motion blur, cluttered background etc. To handle the challenges associated with video segmentation we developed a novel unsupervised end-to-end trainable, fully convolutional deep neural network for object segmentation. Our model, though does not use temporal information, is robust and scalable across scenes, as it is tested in an unsupervised manner and can easily infer which objects constitute the foreground of the image. We perform better than all methods using handcrafted features, and close to deep methods using temporal information.

The final task is the automatic extraction of anomalous events from a given video. For this, we design a novel method to extract outliers from motion alone. We employ a stacked LSTM encoder-decoder structure to model the regular motion patterns of the given video sequence. The discrepancy between the motion predicted using the model and the actual observed motion in the scene is measured to detect anomalous activities. We show on-par performance with the existing state-of-the-art methods on the benchmark datasets.