Abstract

Noha Medhat Ghatwary

Learning spatiotemporal features for esophageal abnormality detection from endoscopic videos

Esophageal cancer is categorized as a type of disease with a high mortality rate. Early detection of esophageal abnormalities (i.e. precancerous and early cancerous) can improve the survival rate of the patients. Recent deep learning-based methods for Selected types of esophageal abnormality detection from endoscopic images have been proposed. However, no methods have been introduced in the literature to cover the detection from endoscopic videos, detection from challenging frames and detection of more than one esophageal abnormality type. In this paper, we present an efficient method to automatically detect different types of esophageal abnormalities from endoscopic videos. We propose a novel 3D Sequential DenseConvLstm network that extracts spatiotemporal features from the input video. Our network incorporates 3D Convolutional Neural Network (3DCNN) and Convolutional Lstm (ConvLstm) to efficiently learn short and long term spatiotemporal features. The generated feature map is utilized by a region proposal network and ROI pooling layer to produce a bounding box that detects abnormality regions in each frame throughout the video. Finally, we investigate a post-processing method named Frame Search Conditional Random Field (FS-CRF) that improves the overall performance of the model by recovering the missing regions in neighborhood frames within the same clip. We extensively validate our model on an endoscopic video dataset that includes a variety of esophageal abnormalities. Our model achieved high performance using different evaluation metrics showing 93.7% recall, 92.7% precision, and 93.2% F-measure. Moreover, as no results have been reported in the literature for the esophageal abnormality detection from endoscopic videos, to validate the robustness of our model, we have tested the model on a publicly available colonoscopy video dataset, achieving the polyp detection performance in a recall of 81.18%, precision of 96.45% and F-measure 88.16%, compared to the state-of-the-art results of 78.84% recall, 90.51% precision and 84.27% F-measure using the same dataset. This demonstrates that the proposed method can be adapted to different gastrointestinal endoscopic video applications with a promising performance.