Multimed Tools Appl (2018) 77:20415–20453
https://doi.org/10.1007/s11042-017-5438-7

Video scene analysis: an overview and challenges on deep learning algorithms

Qaisar Abbas[1] · **Mostafa E. A. Ibrahim**[1,2] · **M. Arfan Jaffar**[1]

Received: 26 April 2017 / Revised: 29 September 2017 / Accepted: 20 November 2017 /
Published online: 9 December 2017

Abstract Video scene analysis is a recent research topic due to its vital importance in many applications such as real-time vehicle activity tracking, pedestrian detection, surveillance, and robotics. Despite its popularity, the video scene analysis is still an open challenging task and require more accurate algorithms. However, the advances in deep learning algorithms for video scene analysis have been emerged in last few years for solving the problem of real-time processing. In this paper, a review of the recent developments in deep learning and video scene analysis problems is presented. In addition, this paper also briefly describes the most recent used datasets along with their limitations. Moreover, this review provides a detailed overview of the particular challenges existed in real-time video scene analysis that has been contributed towards activity recognition, scene interpretation, and video description/captioning. Finally, the paper summarizes the future trends and challenges in video scene analysis tasks and our insights are provided to inspire further research efforts.

Keywords Deep learning · Computer vision · Video processing · Activity classification · Scene interpretation · Video description · Video captioning

1 Introduction

Video scene analysis is an automatic process to recognize humans and objects from live-video sequences. In last several decades, the computer vision and artificial intelligence areas have been considered as an active research domain for the development of automatic application. Particularly, the area of video analysis consists of human action recognition, activity classification, scene interpretation, and video description or captioning. Human activity recognition is

✉ Qaisar Abbas
qaisarabbasphd@gmail.com

[1] Department of Computer Science, Al Imam Muhamad Ibn Saud Islamic University, Riyadh, Kingdom of Saudi Arabia

[2] Benha Faculty of Engineering, Benha University, Qalubia, Benha, Egypt