

Chapter 65

Explainable Video Summarization for Advancing Media Content Production

Evlampios Apostolidis

 <https://orcid.org/0000-0001-5376-7158>

Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Georgios Balaouras

Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Ioannis Patras

School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

Vasileios Mezaris

 <https://orcid.org/0000-0002-0121-4364>

Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

ABSTRACT

This chapter focuses on explainable video summarization, a technology that could significantly advance the content production workflow of Media organizations. It starts by presenting the current state of the art in the fields of deep-learning-based video summarization and explainable video analysis and understanding. Following, it focuses on video summarization methods that rely on the use of attention mechanisms and reports on previous works that investigated the use of attention for explaining the outcomes of deep neural networks. Subsequently, it briefly describes a state-of-the-art attention-based architecture for unsupervised video summarization and discusses a recent work that examines the use of various attention-based signals for explaining the outcomes of video summarization. Finally, it provides recommendations about future research directions.

INTRODUCTION

The current practice in the Media industry for producing a video summary requires a professional video editor to watch the entire content and decide about the parts of it that should be included in the summary. This is a laborious task and can be really intensive and time-consuming in the case of long videos. Moreover, the constantly increasing engagement of users with video sharing platforms (e.g.,

DOI: 10.4018/978-1-6684-7366-5.ch065

YouTube, Vimeo, TikTok) and social networks (e.g., Facebook, Twitter, Instagram), that are used for posting online a variety of video content, such as educational, “how-to”/instructional, training, gaming, travelling, cooking and music playing videos, as well as commercials, movie trailers and sports highlights, led to the inclusion of these data distribution channels among the main communication means of Media organizations. However, these different communication means are usually associated with different specifications about the optimal or maximum video duration (Collyda et al., 2020). For example, videos posted on Facebook’s feed and YouTube are expected / recommended to be up to 2 min. long, videos posted on Instagram’s feed and Twitter are most commonly up to 30 sec., while videos posted on TikTok, Facebook and Instagram as stories are even shorter (i.e., 15 to 20 sec. long). This means that different summaries should be produced for a given video, which significantly increases the workload of the video editor.

Technologies for automated video summarization, aim to generate a short synopsis that summarizes the video content by selecting its most informative and important parts. The use of such technologies by Media organizations can drastically reduce the needed resources for media content production in terms of both time and human effort, and facilitate indexing, browsing, retrieval and promotion of their media assets. Despite the recent advances in the field of video summarization, which are tightly associated with the emergence of modern deep-learning network architectures (Apostolidis et al., 2021b), the outcome of a video summarization technology still needs to be curated by the video editor, in order to ensure that all the needed parts of the video were included in the video summary. This content production step could be further facilitated, if the video editor is provided with explanations about the suggestions made by the used video summarization technology. The provision of such explanations would allow a level of understanding about the functionality of this technology, thus increasing the editor’s trust in it and facilitating content curation.

Given the above, this chapter focuses on explainable video summarization, a technology that could significantly advance the content production workflow of Media organizations. It starts by presenting the current state of the art in the fields of deep-learning-based video summarization and explainable video analysis and understanding. Following, it focuses on video summarization methods that rely on the use of self-attention mechanisms for modelling frames’ dependence and estimating their importance. As a note, self-attention is a type of attention used in the Transformer Network (Vaswani et al., 2017) for modelling the relation between different elements of an input sequence in order to compute a representation of this sequence. In layman’s terms, the self-attention mechanism allows the elements of the input sequence to interact with each other, and takes their relationship into consideration to determine which of them requires greater attention and dynamically adjust their impact on the output. The chapter continues by reporting on previous works that investigated the use of attention for explaining the outcomes of deep neural networks. Most of them relate to the natural language processing (NLP) domain, but recently, attention was used to interpret the output of networks trained for image recognition and classification, and multimodal trajectory prediction. Subsequently, the chapter briefly describes a state-of-the-art attention-based architecture for unsupervised video summarization, and discusses a recent work that examines the use of various attention-based signals for explaining the outcomes of video summarization. Finally, it provides recommendations about future research directions on explainable video summarization, and concludes this report.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/explainable-video-summarization-for-advancing-media-content-production/331069

Related Content

Information Systems Education in Thailand: A Comparison Between the Views of Professionals and Academics

Ang Yang Angand Supawadee Jiwahasuchin (1998). *Journal of Global Information Management* (pp. 34-42).

www.irma-international.org/article/information-systems-education-thailand/51320

Research on the Impact of Information Security Certification and Concealment on Financial Performance: Impact of ISO 27001 and Concealment on Performance

Wenqing Wu, Kun Shi, Chia-Huei Wuand Jiayue Liu (2022). *Journal of Global Information Management* (pp. 1-16).

www.irma-international.org/article/research-on-the-impact-of-information-security-certification-and-concealment-on-financial-performance/281712

Gender and Employability Patterns amongst UK ICT Graduates: Investigating the Leaky Pipeline

Ruth Woodfield (2012). *Globalization, Technology Diffusion and Gender Disparity: Social Impacts of ICTs* (pp. 184-199).

www.irma-international.org/chapter/gender-employability-patterns-amongst-ict/62886

Modern Technology and Mass Education: A Case Study of a Global Virtual Learning System

Ahmed Ali (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1806-1821).

www.irma-international.org/chapter/modern-technology-mass-education/19075

Conceptualization of the Convergence Phenomenon to Develop an Applicable and Integrated Framework for the Emergence of Software-as-a-Service

Elizabeth A. Teracinoand DongBack Seo (2013). *Journal of Global Information Management* (pp. 1-16).

www.irma-international.org/article/conceptualization-of-the-convergence-phenomenon-to-develop-an-applicable-and-integrated-framework-for-the-emergence-of-software-as-a-service/99662