ELINVO (Electronics, Informatics, and Vocational Education), Mei 2023; vol 8(1):26-37
ISSN 2580-6424 (printed), ISSN 2477-2399 (online) DOI: https://doi.org/10.21831/elinvo.v8i1.60741

Breaking Down Computer Networking Instructional Videos: Automatic Summarization with Video Attributes and Language Models

Totok Sukardiyono[1] , Muhammad Irfan Luthfi[12*], Nisa Dwi Setiyanti3
[1]Department of Electronics and Informatics Engineering Education, Faculty of Engineering, Universitas Negeri Yogyakarta
[2]Department Graduate Institute of Network Learning Technology, National Central University
[3]Universitas Muhammadiyah Karanganyar, Indonesia
[*]E-mail: m.irfanluthfi@uny.ac.id

ABSTRACT

Instructional videos have become a popular tool for teaching complex topics in computer networking. However, these videos can often be lengthy and time-consuming, making it difficult for learners to obtain the key information they need. In this study, we propose an approach that leverages automatic summarization and language models to generate concise and informative summaries of instructional videos. To enhance the performance of the summarization algorithm, we also incorporate video attributes that provide contextual information about the video content. Using a dataset of computer networking tutorials, we evaluate the effectiveness of the proposed method and show that it significantly improves the quality of the video summaries generated. Our study highlights the potential of using language models in automatic summarization and suggests that incorporating video attributes can further enhance the performance of these models. These findings have important implications for the development of effective instructional videos in computer networking and can be extended to other domains as well.

Keywords*:* computer networking, instructional video, video attributes, automatic summarization, language model

INTRODUCTION

Instructional videos are often used in teaching complex topics in computer networks, but they are often lengthy and time-consuming, making it difficult for learners to acquire essential information [11]. Several recent studies have developed concise techniques using natural language processing technology to overcome this problem in summarizing text. Liu et al. [15] showed that instructional video summaries with natural language processing technology could improve students' understanding of learning materials. This technology also helps teachers speed up making instructional videos [25][28].

The use of instructional videos in educational contexts is increasing along with technological advances and the willingness to share information over the Internet [31]. However, instructional videos that are too long are often ineffective in helping learners understand critical concepts [31][2]. In addition, the effectiveness of instructional videos can be limited by several factors, such as the length of the video length, the language skills of learners, and the complexity of the topics covered [2]. To improve learning effectiveness, augmented reality (AR) technology can be used in instructional videos [13], and animation-based instructional videos can also help improve comprehension of a complex concept [14]. Therefore, the development of compelling instructional videos must consider techniques such as the use of AR or animation, as well as factors such as video duration and the capabilities of learners' language [31][2][13][14].

Otterbacher et al. [21] demonstrated the possibility of using natural language processing techniques and machine learning to generate more effective text summaries from online learning videos. The automatic summary method they propose is better than other automated summary methods. He et al. [8] also showed that using language models to generate text summaries from complex instructional videos can improve learner comprehension. Their study showed that this method produced better text summaries than other automated summarization methods.

However, some studies point to technical challenges, such as the long duration of instructional videos and the complexity and variety of instructional video content that can affect accuracy and text summary completeness. Researchers emphasize the need for a careful and measurable approach to identifying essential information and producing effective text summaries. Despite this, the potential for NLP techniques and text summaries from instructional videos is very high [9][12][18][19][30].

In this study, we propose a new approach to generating effective text summaries from instructional videos in computer networking. This approach integrates natural language processing techniques and language models to identify essential information from the video and produce a concise and compelling text summary. In addition, we also consider video attributes, such as duration, topic, and genre, to improve the performance of the text summary algorithm.

This study aimed to evaluate the proposed approach's effectiveness in generating effective text summaries from instructional videos in the computer networking field. We also aim to identify factors that influence text summary algorithms' performance and evaluate whether the proposed approach can be applied in areas other than computer networks.

These studies have important implications for developing compelling instructional videos and can help improve learners' understanding of complex topics. In addition, the results of this study can be used as a basis for continued research on developing more effective, concise techniques for instructional videos. The study may also broaden our understanding of the use of natural language processing technology in education and educational technology development.

A. Instructional Videos in Computer Network Learning

Instructional videos or video tutorials are popular media for self-study in various fields, including computer networks. In recent years, instructional videos in computer network learning have become popular, as they can help students or trainees understand complex concepts more quickly.

According to Peng et al. [22], computer network instructional videos can help students understand network topology, network device configuration, and troubleshooting on the network. In their research, Peng et al. Use instructional videos to teach computer networking concepts to informatics engineering students. The results showed that instructional videos can help improve students' understanding and skills in managing computer networks.

Another study by Chao and Chen [3] also found that instructional videos effectively improve college students ' understanding of computer networks. In their research, Chao and Chen used instructional videos to teach computer engineering students network topology and routing protocols. The results showed that using instructional videos can help improve students' understanding and motivate them to learn. Another study by Salomon [23] also showed similar results in using instructional videos on network technology learning and information security.

In this case, it can be concluded that using instructional videos in computer network learning has advantages and disadvantages. Instructional videos are effective in increasing student understanding and motivation. Therefore, in developing instructional videos for computer network learning, it is necessary to consider the learning objectives and the intended students' characteristics.

B. Video Summarization for Instructional Video

Video summarization is a video processing technique that aims to produce a short summary of the video while retaining the original video's important information. This is particularly useful in computer network instructional videos, where instructional videos are often very long and challenging to follow entirely. Several studies have been conducted on video summarization for instructional videos.

In one of the studies by Chao et al. [4], they proposed a new approach called "Latent Dirichlet Allocation-based Semantic Topic Extraction" (LDA-STE) to generate relevant video summaries. This approach involves extracting semantic topics from video transcription text and determining the video segments most relevant to each topic. The results showed that this approach effectively produced shorter, more relevant video summaries.

Several studies have proposed different approaches to video summarization in machine learning. Liu et al. [17] presented a deep learning-based model that combines a frame-based convolution network and an LSTM network to generate accurate and relevant video summaries. Feng et al. [7] introduced a learning-based approach called "Deep Reinforcement Learning-based Video Summarization" (DRLVS), which involves decision-making, drafting, and assessing video summaries and has been shown to produce shorter and more relevant video summaries than other techniques. Darmawan and Sulistyo [6] implemented TextRank and face detection techniques to extract meaningful sentences from video transcriptions and identify essential parts of the video, respectively, with effective results. Anwar et al. [1] applied edge detection to video summarization using the Sobel filter, which can generate easier-to-understand video summaries that provide more relevant information while maintaining the contour of the main object.

In addition to the mentioned techniques, several other techniques can be used in video summarization. The clustering method can group similar frames or video segments and produce a more concise and easy-to-understand representation [24]. Natural language processing can help extract meaningful information from video transcription text and make it easier to summarize creation [29]. Meanwhile, attention-based neural networks can identify essential video segments and retain relevant information from the original video [32].

In its use, selecting the proper video summarization technique must be adjusted to the needs and purpose. For example, PCA and color reduction techniques can be used to produce more concise and easy-to-process video summaries, while DRLVS techniques can produce video summaries that are more concise and easier to process. In contrast, DRLVS techniques can produce more detailed and informative video summaries. In computer network learning, using appropriate techniques can help increase learning effectiveness through video.

C. Language Model Utilization in Video Summarization

Language models have been widely used in natural language processing tasks like text summarization, text generation, and machine translation. However, their potential in instructional video summarization has not been fully explored. This literature review provides an overview of recent research on the use of language models in instructional video summarization. Several studies proposed different methods for summarizing instructional videos using pre-trained language models such as BERT, GPT-2, T5, ERNIE 2.0, UniLM, and graph-based approaches. These models extract audio transcripts and visual features from the video and fine-tune the language models to identify essential sentences, which are then used to summarize the video—the proposed methods performed better than several baseline methods in terms of ROUGE scores and human evaluation.

Chen et al. [5] proposed a method that combines audio-text analysis and UniLM to summarize MOOC videos. The proposed method outperformed other baseline methods regarding ROUGE scores and human evaluation. Similarly, Liu et al. [16], Zhang et al. [32], Li et al. [19], Wang et al. [27], Wang et al. [28], and Ma et al. [20] proposed different methods that used pre-trained language models like BERT, GPT-2, T5, and ERNIE 2.0 in instructional video summarization.

Recent studies have shown that pre-trained language models and graph-based approaches can effectively summarize instructional videos. Further research could investigate using other pre-trained language models, such as XLNet and RoBERTa, in instructional video summarization. Using language models in instructional video summarization can automate the summarization process, providing an efficient approach that benefits learners and educators.

METHODS

This study focuses on producing instructional video summaries on computer networks through two stages: extracting video attributes and using language models. In more detail, the steps to generate a summary of this video are shown in Figure 1.

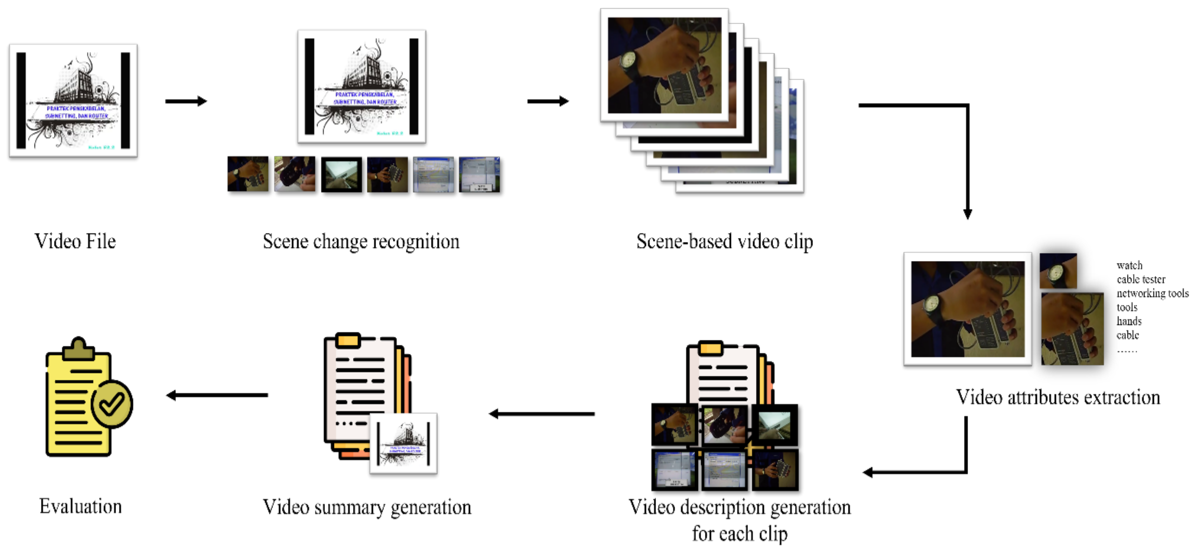

Figure 1. Video Summarization Steps

A. Video Files

The video file used in this study is a computer network instructional video produced by students at a university in Yogyakarta, Indonesia. The number of videos made is seven pieces, with the specifications of each video listed in Table 1.

Table 1. Computer network instructional video specifications

#	File Name	File Format	Duration
1	001	.DAT	00:21:29
2	002	.mpg	00:22:29
3	003	.mp4	00:13:22
4	004	.mpg	00:11:47
5	005	.mp4	00:40:09
6	006	.wmv	00:21:41
7	007	.wmv	00:29:49

The next step is video preprocessing, standardizing the format, frame size, bitrate, and frames per second. In this preprocessing video, we equated the video format into mpeg-4 format with an x264 video encoder, 640x480 frame size, 1500 Kbps bitrate, and 24 frames per second. If it is found that the video resolution does not match the frame size of 640x480 after preprocessing, then we use the Black border-filling method around the video. This video preprocessing process uses the FFmpeg 5.0.3 tool.

B. Scene Change Recognition

The next stage is scene change recognition. The purpose of this stage is to record the number of scene changes in a video. This is important because we need essential information from each video frame. However, because extracting information frame by frame will take a lot of time and resources, extracting essential information is done per scene. To do scene change recognition, we use a combination of OpenCV and Numpy, and it is done automatically using Python programming language.

C. Scene-Based Video Clip

After obtaining the whole scene and time frames of each scene from each video, the next step is to split the videos for each scene time frame. To do so, we again use OpenCV to perform video splits automatically.

D. Video Attributes Extraction

Performing video attribute extraction is a crucial process in this video summarization. Because in this step, various attributes in the video will be recognized and used as keywords in generating video clip descriptions. To perform this video attribute extraction, we do it automatically by combining OpenCV, Tensorflow, and Numpy and using the YOLOv3 model.

E. Video Description for Each Clip

The next step is to generate video descriptions automatically using various language models. The language models we use to generate video descriptions automatically are as follows.

Table 2. Types of LM used for video summarization

#	Language Model	Size	Type
1	BERT	340 M parameters	Encoder
2	T5	11 B parameters	Encoder-Decoder
3	UniLM	340 M parameters	Encoder-Decoder
4	MASS	1.9 B parameters	Encoder-Decoder
5	XLNet	340 M parameters	Decoder

F. Video Summary Generation

After obtaining descriptions for each clip, the video clip descriptions are summarized automatically using the same language model. The results of this summary generation video will later be evaluated.

G. Evaluation

At this stage, the results of video summarization will be evaluated. The evaluation process is done simply by comparing the summary results manually by humans with the summary results carried out automatically. From this comparison, we looked for the cosine similarity value to see the performance of each language model to summarize computer network instructional videos.

RESULT AND DISCUSSION

This section will discuss the results of the steps taken to perform computer networking instructional video summarization. The first and second stages are video pre-processing and scene change recognition. The results of the first and second stages are listed in Table 3 below.

Table 3. Summary of video processing and scene change recognition results

#	File Name	File Formats	Duration	Total Scene
1	001	.DAT	00:21:29	54
2	002	.mpg	00:22:29	37
3	003	.mp4	00:13:22	53
4	004	.mpg	00:11:47	36
5	005	.mp4	00:40:09	138
6	006	.wmv	00:21:41	33
7	007	.wmv	00:29:49	116

We can see from the results of pre-processing video and scene change recognition that instructional video can be divided into three, under 20 minutes, under 30 minutes, and above 30 minutes. However, 85% of the instructional videos we dataset into were under 30 minutes long. As for whole scenes, the average video under 30 minutes has 55 scenes, and above 30 minutes has more than 130 scenes. Much of this scene depends on the topic conveyed in the video. If the topic presented has scenes to configure, it can be ascertained that it will have more scenes. This is due to the many screen changes at the computer layer when configuring computer networks.

After doing video pre-processing and scene change recognition, the next step is to do video splitting to separate video clips based on the scene. As explained in the method section, this step uses OpenCV combined with the results in scene change recognition. For example, we display Table 4, which contains the results of scene change recognition video 006 and its time frame.

Table 4. Scene change recognition and time frame results for video 006

#	Clip Name	Time Frame (s)
1	Shot 0	0.0 to 3.45
2	Shot 1	3.5 to 33.8
3	Shot 2	33.85 to 43.55
4	Shot 3	43.6 to 44.1
5	Shot 4	44.15 to 54.85
6	Shot 5	54.9 to 55.85
7	Shot 6	55.9 to 64.0
8	Shot 7	64.05 to 75.85
9	Shot 8	75.9 to 83.45
10	Shot 9	83.5 to 89.0
11	Shot 10	89.05 to 96.95
12	Shot 11	97.0 to 104.25
13	Shot 12	104.3 to 109.95
14	Shot 13	110.0 to 112.95
15	Shot 14	113.0 to 117.55
16	Shot 15	117.6 to 122.55
17	Shot 16	122.6 to 321.05
18	Shot 17	321.1 to 431.6
19	Shot 18	431.65 to 481.45
20	Shot 19	481.5 to 562.6
21	Shot 20	562.65 to 613.95
22	Shot 21	614.0 to 627.85
23	Shot 22	627.9 to 660.35
24	Shot 23	660.4 to 807.9
25	Shot 24	807.95 to 856.65
26	Shot 25	856.7 to 988.95
27	Shot 26	989.0 to 1009.25
28	Shot 27	1009.3 to 1141.95
29	Shot 28	1142.0 to 1166.65
30	Shot 29	1166.7 to 1227.25
31	Shot 30	1227.3 to 1250.45
32	Shot 31	1250.5 to 1278.95
33	Shot 32	1279.0 to 1301.75

From the results of the splitting video, then proceed to the next step, namely video attributes extraction using a combination of OpenCV, Tensorflow, Numpy, and YOLOv3 Models. Table 5 shows the results of video attributes extraction on video 004 clip 10.

Table 5. The result of video attributes extraction on video 004 clip to 10.

#	Object Name	Entity	Time Frame (s)	Confi dence
1	joint	human body	78	0.664
2	headgear	clothing	79	0.697
3	helmet	clothing	79	0.458
4	town	geographica l feature	83	0.831
5	shrub	plant	83	0.458
6	neighbourhood	geographica l feature	83	0.703
7	village	geographica l feature	84	0.410
8	boardwalk	walkway	86	0.414
9	artificial landforms	geographica l feature	88	0.403
10	jungle	geographica l feature	89	0.425
11	personal protective equipment		90	0.402
12	motorcycle	vehicle	90	0.784
13	all-terrain vehicle	vehicle	90	0.632
14	garden		91	0.460
15	houseplant	plant	93	0.469
16	woody plant	plant	96	0.902
17	community	organization	97	0.853
18	flora	plant	97	0.754
19	plantation	building	98	0.416
20	flower	plant	100	0.701
21	floral design	flower	100	0.465
22	floristry	building	100	0.499
23	freeride	bicycle	104	0.402
24	gardener	person	107	0.409

After we obtain the results of extracting video attributes, the next step is to use the language model to generate video descriptions. Table 6 shows an example of video description generation using various language models for video 004 clip 10.

Table 6. Results from Video Description Generation Video 004 Clip 10

#	Language Model	Video Description
1	BERT	In this video, we'll be exploring the applications of computer networking in various contexts related to joint operations, headgear, personal protective equipment (PPE), and other areas. We'll see how computer networking can be used in towns, villages, and neighborhoods to ensure the safety of community members, especially when it comes to the use of motorcycles and all-terrain vehicles (ATVs). We'll also examine how computer networking can help in the management and monitoring of gardens and plantations, controlling the growth of shrubs, woody plants, and flowers, and even creating beautiful floral designs. Additionally, we'll discuss how computer networking can be applied in the context of artificial landforms such as boardwalks and jungles and how it can be used to facilitate freeride activities safely. Finally, we'll explore how computer networking can aid gardeners in monitoring and managing their gardens, promoting a vibrant and thriving gardening community.
2	T5	In this video, we'll be exploring the fascinating world of computer networking in the context of a diverse range of environments, from bustling towns and villages to quiet neighbourhoods and tranquil gardens. We'll take a joint approach, examining the interconnected nature of networking technologies and how they work together to facilitate communication and information sharing. As we venture through different landscapes, we'll see how headgear, helmets, and personal protective equipment are crucial for safety and explore the role of artificial landforms, boardwalks, and all-terrain vehicles in accessing remote locations. From the lush jungles and dense shrubs to the beautiful flora and plantations, we'll learn how the internet is connecting communities and businesses worldwide. Finally, we'll delve into the art of floristry, from floral designs to gardening and the care of houseplants and woody plants. Whether you're a free rider or a gardener, this video will offer valuable insights into the world of computer networking and its relevance to our everyday lives.
3	UniLM	Computer networking is an essential aspect of modern life, enabling people and businesses to connect and communicate globally. In this context, the joint efforts of technology and personal protective equipment are critical to ensuring safety while exploring various terrains, from the boardwalks of busy towns and neighbourhoods to the dense jungles and artificial landforms of remote villages. Whether cruising on a motorcycle or all-terrain vehicle, it's essential to have proper headgear and protective equipment to minimize the risk of injury while navigating through shrubs, woody plants, and diverse flora. For those who appreciate the beauty of gardens and plantations, the art of floristry and floral design offers a creative outlet for showcasing the beauty of flowers and houseplants. The freeride community and gardeners alike recognize the importance of responsible land use and maintain a healthy balance between human activity and natural ecosystems. Overall, computer networking is a tool that enables us to connect and collaborate, whether in a bustling city or a remote village and allows us to explore and appreciate the natural world around us while staying safe and protected.
4	MASS	In this video, we will be discussing the importance of computer networking in various outdoor settings. Whether you're riding a motorcycle or an all-terrain vehicle, it's essential to stay connected with your fellow riders. We'll explore the benefits of joint headgear and helmets that enable riders to communicate with each other seamlessly, even in noisy environments. From bustling towns to quiet villages and neighbourhoods, we'll highlight the various ways that computer networking can enhance your outdoor experience. We'll also discuss how artificial landforms and boardwalks can benefit from networking technology to improve safety and accessibility. We'll take you on a journey through lush jungles and shrubs, exploring the different types of personal protective equipment that can help you stay connected and protected while navigating through rugged terrain. If you're

#	Language Model	Video Description
		a gardener, we'll show you how networking can improve your community's flora, from houseplants to woody plants and even plantations. Finally, we'll showcase the art of floral design and floristry, demonstrating how technology can help you create stunning arrangements and connect with other florists from around the world. So whether you're hitting the trails for a freeride or tending to your garden, join us for a fascinating exploration of computer networking in the great outdoors.
5	XLNet	Computer networking has revolutionized the way we communicate and share information. In today's world, joint efforts between individuals and communities are often supported by the power of networking technologies. As such, ensuring the safety and protection of individuals involved in various activities is of paramount importance. Headgear and personal protective equipment such as helmets are essential when engaging in high-speed activities like riding a motorcycle or all-terrain vehicle. Even when working in a garden or exploring the flora and fauna of a jungle, protective equipment can prevent injuries and ensure safety. The use of technology in landscaping and the creation of artificial landforms, boardwalks, and shrubbery have enhanced the beauty of our neighbourhoods and villages, allowing us to create stunning gardens and plantations that showcase the natural beauty of woody plants and flowers. Floral design and floristry have become important aspects of our communities, with many people turning to these activities as a way of expressing themselves and connecting with others. Whether you are exploring a jungle, creating a garden, or going on a freeride adventure, the importance of personal protective equipment and the need to respect the flora and fauna of our communities cannot be overstated. With the help of networking technologies, we can create a safer, more beautiful world for everyone, from the gardener to the adventure seeker.

The next step after we get the video description for each clip is to summarize it so that it becomes one complete summary text that generally describes the content of the computer network instructional video of these. Table 7 below shows the results of video summarization using the language model and its evaluation using cosine similarity to the summary made by humans.

Table 7. The results of video summarization using various language models for video 004.

#	Language Model	Video Description	Cosine Similarity
1	BERT	Computer networking is the practice of linking computing devices together to share resources and information. It involves a wide range of technologies, such as networking hardware, software, and protocols. The primary purpose of computer networking is to facilitate communication and resource sharing between computers or other devices. Some key aspects of computer networking include the Internet, LANs, WANs, and wireless networks. Additionally, network security is a critical aspect of computer networking, as it ensures the privacy and integrity of data that is transmitted over networks.	0.873
2	T5	Computer networking is a vast field that deals with the communication between computers or devices. It involves the transmission of data between interconnected devices using various mediums. Networking can range from local area networks to	0.861

#	Language Model	Video Description	Cosine Similarity
		global communication systems that can connect devices all over the world. Networking is an essential part of today's digital world, and it plays a crucial role in various fields such as business, education, healthcare, and many more. Networking enables people to share information, resources, and ideas, making it a critical tool for the modern world.	
3	UniLM	In computer networking, communication is essential between various devices to transmit data efficiently. The use of symbols, such as text and screenshots, can aid in the transmission of information. In addition, many devices, such as residential areas, commercial vehicles, and public spaces, rely on networking to function correctly. Moon and astronomical objects can affect network performance, as can natural objects such as trees and forests. Similarly, human behavior, emotions, and facial expressions can impact network communication. To ensure optimal communication, it is important to consider all elements that may affect the network, from the design of the hardware and software to the layout of the environment.	0.679
4	MASS	In the world of computer networking, there are various aspects to consider, from designing the architecture to ensuring smooth communication flow. One important factor is the road surface and infrastructure of the area where the network is being implemented. Another key aspect is the audio equipment used for presentations and public speeches. Additionally, symbols and text are essential components in designing advertisement campaigns, while photography and illustration are effective ways to capture emotions and facial expressions. From the standpoint of the end-users, issues such as vision care and health &; beauty play an important role in their day-to-day use of the network.	0.324
5	XLNet	Computer networking refers to the practice of connecting various computing devices to exchange data and information through communication networks. It involves the transmission of data between devices such as computers, printers, servers, and other networking devices through wired and wireless connections. Networking technology has evolved to support various types of applications, ranging	0.710

#	Language Model	Video Description	Cosine Similarity
		from simple file sharing to complex video conferencing and cloud computing. With the increasing demand for high-speed data transfer, networking protocols such as TCP/IP, DNS, DHCP, and HTTP have become essential components in today's internet. Network engineers and technicians are responsible for designing, implementing, and maintaining these networks, ensuring that they are secure, reliable, and efficient.	

From the results of one video above, we can see that the performance of video summarization using the BERT language model is best compared to other language models. Table 8 shows summarization results for all videos and the performance of each language model for summarization.

Table 8. Performance of each language model in performing video summarization

	BERT	T5	UniLM	MASS	XLNet
Vid. 1	0.915	0.916	0.820	0.354	0.893
Vid. 2	0.814	0.703	0.509	0.480	0.614
Vid. 3	0.754	0.721	0.691	0.431	0.784
Vid. 4	0.873	0.861	0.679	0.324	0.710
Vid. 5	0.952	0.946	0.877	0.890	0.992
Vid. 6	0.705	0.655	0.483	0.187	0.609
Vid. 7	0.886	0.809	0.743	0.468	0.853
Avg.	0.843	0.802	0.686	0.448	0.779

From Table 8, we can see that, in general, the performance of the BERT language model is the best, with an average cosine similarity of 0.843. At the same time, the lowest performance is the MASS language model, with an average cosine similarity of 0.448.

CONCLUSION

From the results of the trials and observations made, the steps to conduct video summarization for instructional video computer networks ranging from video pre-processing, scene change recognition, scene-based video clips, video attributes extraction, video description of each clip, and video Summary generated successfully generates a summary of the instructional video content of the computer network. In addition, of the five language models we tested for this summarization task, the BERT language model performed best, with an average cosine similarity of 0.843, while the lowest performance was the MASS language model, with an average cosine similarity of 0.448. From the results of these tests and observations, it can be concluded that the summarization of instructional video content of computer networks can be done automatically, and in the future, it can be developed further for other language models.

REFERENCES

[1] Anwar, M. A., Wahyuni, N. L. S., & Suryana, N. (2021). Sobel Edge Detection in Video Summarization for Teaching Material. Journal of Physics: Conference Series, 1819(1), 012038.

[2] Chan, C. K. K., & Li, K. C. (2018). How long is too long? The effect of video length on learning performance in business and economics-related subjects. Journal of Education for Business, 93(7), 321-328.

[3] Chao, T.-Y., & Chen, W.-H. (2017). Using Instructional Videos to Improve Learning Achievement in Computer Networking. Journal of Educational Technology Development and Exchange, 10(2).

[4] Chao, W. L., Yu, Y. C., Chen, S. L., & Chen, Y. C. (2019). Latent Dirichlet Allocation-based Semantic Topic Extraction for Video Summarization. In 2019 International Conference on Applied System Innovation (ICASI) (pp. 1-4).

[5] Chen, M., Yu, Y., Lu, H., & Zhang, G. (2021). Audio-text summarization for MOOC videos

via a unified encoder-decoder framework. IEEE Transactions on Learning Technologies, 14(3), 384-395.

[6] Darmawan, D., & Sulistyo, S. H. (2019, October). Video summarization with text rank and face detection. In 2019 International Conference on Information and Communications Technology (ICOIACT) (pp. 1-4). IEEE.

[7] Feng, X., Ma, F., & Huang, Y. (2020). Deep reinforcement learning-based video summarization. IEEE Transactions on Circuits and Systems for Video Technology, 30(8), 2516-2528.

[8] He, X., Chen, L., Liu, Y., Zhao, Z., & Lu, H. (2020). Generating Text Summaries for Instructional Videos Based on a Language Model. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2467-2478).

[9] Hwang, M., Yoon, J., Kim, D., & Kim, J. (2020). A Survey on Automatic Text Summarization. Journal of Information Science Theory and Practice, 8(1), 1-14.

[10] Kasim, A. M., Pantoja, M. F. F., & Fornillos Jr, R. J. (2021). The Effectiveness of Instructional Videos in Teaching Networking Concepts in a Philippine University. Journal of Information Technology Education: Research, 20, 269-291.

[11] Khamparia, A., Khamparia, S., & Pandey, A. (2019). Automatic text summarization: A review. International Journal of Computer Science and Information Security, 17(6), 77-82.

[12] Kim, H., Lee, S., Park, S., & Lee, G. G. (2021). A review of automatic text summarization with natural language processing. Convergence Information Management Research, 8(2), 97-107.

[13] Kwok, L. F., & Zheng, Y. (2020). Designing an augmented reality app to enhance students' learning experiences: A case study. International Journal of Emerging Technologies in Learning (iJET), 15(5), 132-145.

[14] Lee, J. K., Kim, H. J., Kim, J. H., & Kim, J. W. (2019). The effects of animation-based video on college students' understanding of a complex concept. Journal of Educational Technology & Society, 22(2), 141-153.

[15] Liu, M., Liu, Y., & Ma, Y. (2021). Automatic Video Summarization for E-Learning with Natural Language Processing Technology. IEEE Access, 9, 24563-24572.

[16] Liu, S., Zhang, S., Ma, W., & Li, H. (2021). A pre-training and fine-tuning method for instructional video summarization. arXiv preprint arXiv:2102.10102.

[17] Liu, Y., Huang, Y., Yang, Y., & Zhang, M. (2018). Learning to summarize from scratch. arXiv preprint arXiv:1804.06451.

[18] Li, J., Chen, C., & Li, Y. (2019). Automatic text summarization of academic articles using natural language processing techniques. Information Processing & Management, 56(6), 1689-1702.

[19] Li, J., Wang, X., Yang, Y., Zhang, S., & Sun, M. (2021). Video summarization with pre-trained transformers. arXiv preprint arXiv:2104.01922.

[20] Ma, Z., Peng, H., Hu, Y., Wang, K., & Song, J. (2020). Incorporating pre-trained language models and temporal structure for instructional video summarization. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 5909-5919).

[21] Otterbacher, J., Suleman, S., & Lin, J. (2016). Effective video search: summarization with minimal supervised training. International Journal on Digital Libraries, 17(2-3), 107-119.

[22] Peng, H., Wang, H., & Hu, L. (2019). The Application of Instructional Videos in Computer Network Learning. In 2019 4th International Conference on Education and Multimedia Technology (ICEMT) (pp. 1-5). IEEE.

[23] Salomon, J. (2017). Using instructional video in a classroom context: Effects on engagement and retention. Journal of Information Technology Education: Research, 16, 335-347.

[24] Wang, Q., Zhang, Y., & Song, Y. (2016). Video summarization based on feature clustering and multiple-instance learning. Neurocomputing, 205, 358-368.

[25] Wang, T., Zhao, Y., & Huang, C. (2020). An approach to generating instructional videos using natural language processing. Journal of Educational Technology Development and Exchange (JETDE), 13(1), 1-16.

[26] Wang, W., & Chen, W. (2018). Evaluating the Effectiveness of Using Online Video Instruction for Practical Skill Training in Computer Networks. In 2018 IEEE 4th International Conference on Computer and Communications (ICCC) (pp. 594-598). IEEE.

[27] Wang, Y., Li, J., Yang, Y., & Sun, M. (2021). Multi-modal video summarization with transformer. arXiv preprint arXiv:2104.03544.

[28] Wang, Y., Yang, Y., Li, J., & Sun, M. (2020). Video summarization with attention-based encoder-decoder networks. arXiv preprint arXiv:2004.14098.

[29] Yang, M., Zhang, L., Feng, F., & Gao, W. (2019). Automatic video summarization using natural language processing. IEEE Transactions on Multimedia, 21(11), 2979-2991.

[30] Zhai, W., Yang, X., Zhang, J., & Chen, Y. (2022). A Survey on Automatic Text Summarization. ACM Computing Surveys, 55(1), 1-33.

[31] Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker Jr, J. F. (2019). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Information & Management, 56(1), 103-119.

[32] Zhang, H., Li, Y., Zhang, S., & Huang, Z. (2020). Video summarization via attention-based LSTM with joint learning. IEEE Transactions on Circuits and Systems for Video Technology, 31(4), 1334-1344.