Journal of Education for Pure Science- University of Thi-Qar
Vol .14, No.3 (2024)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: https://doi.org/10.32792/jeps.v14i3.546

A new approach to video summary generation

Zainab Shaker Matar Al-Husseini[1], Sahar R. Abdul Kadeem[2] and Ali H. Naser[3]

[1]Faculty of Archaeology, Department of Archeology and Islamic Civilization, University of Thi-Qar, Nasiriyah, Iraq.

[2] National University of Science and Technology, Dhi Qar, Iraq.
[3] Dhi-Qar Education Directorate.

1) zainab.shaker@utq.edu.iq , https://orcid.org/0000-0002-1927-8569
2) sahar@nust.edu.iq
3) prog.ali95@gmail.com

* Corresponding email: zainab.shaker@utq.edu.iq

Received 02/04 /2024, Accepted 14/06 /2024, Published 01 / 09 /2024

Abstract:

This paper describes a video summary approach for an actual video file that provides a rapid way to assess its information. This is a difficult task to identify relevant or informative sections. When the original video demands knowledge of its content, we build a video summary via extracting in-depth characteristics of each original video. Video summarization deals with the generation of a condensed version of the original video by including meaningful frames or segments while eliminating redundant information provides a concise and accurate representation of the original video clips by providing the most representative summary with increased interest through the application of a matching function between the current and prior frames.

The primary goal of this work is to give a strategy for recording video clips using security cameras and converting them into a short video to provide a clear analysis of the video by reducing excess and extracting the key video contents into the real video. In a lengthy video clip, viewers may not have enough time to see the full thing. The viewer may be interested in seeing a specific topic under the

label "Important". It may be composed of pieces that may be shared on social media and communicate the semantics of long videos, particularly video snippets that must be submitted to Internet. To browse, will be needed a network with adequate bandwidth. That is, non-affiliated video streaming, indexing, and retrieval of video content makes it valuable to viewers.

Keywords: Video summarization, image filters, compare function.

1. INTRODUCTION

Video summarizing is a technique that allows big video datasets to have their primary tales and material condensed into a brief original video [1]. Video summarization has a multitude of potential applications. For instance, if house occurrences are captured on security video, the purpose would be to condense the significant abnormal event footage into a few minutes so that it would be easier to comprehend. The input for video summary is a video that contains all of the original content. The goal is to select essential frames or tiny portions of content from the original input video in order to create a summary that expresses the main ideas without requiring viewers to watch the entire clip [2]. It possible that viewers won't have enough time to finish the lengthy video. A viewer may find it interesting to watch on the specific topic under "important," depending on what they are looking for [3]. The goal of extracting a video's representative visual features for social media sharing is to effectively portray the semantics of the original long video, and this has attracted a lot of attention recently [4]. In the modern digital age, a vast number of videos have been produced and distributed via various streaming media. In particular, these videos are uploaded to the cloud or the internet, therefore viewing them requires a high bandwidth network. More focus is being placed on video summarizing, which provides a succinct and accurate summary of the original video segments by displaying the most representative synopsis. This is an excellent way to conserve time, storage, network, and multimedia infrastructure, among other resources [5]. There are actually two categories of video summaries: Whether it's static video abstract, consisting of a series of key frames, or dynamic video skimming, comprising a set of dynamically produced audio-video sub clips, the goal is to gather the most noteworthy or captivating video segments that encapsulate the core of the source clips. In the actual world even if we have a ten of videos with a lot of content, not every frame may be equally valuable, or the content may be repetitive or unhelpful. However, collecting the relevant material to turn a brief but engaging video into a finished product without repeating or omitting any of the input material's contents is a challenging task while working on video summarizing. Numerous studies have been conducted on video summarization. Since this summary is a part of the subjectivity of knowing, the problem is under-constraint [6].

In this work, the article describes how a lengthy video clip is divided into shorter segments and narratives, thereby representing data and creating a duplicate of the video that closely resembles the original material. It also investigates the methods or approaches used to summarize the video while working with the application area by utilizing a novel way to build a video summary that summarizes possible information linked to distraction from the whole video content supplied as input. By employing a matching function to summarize the video and determine how much has changed from the previous frame to the present frame. The major goal is to develop a method that can record brief videos from security camera clips and edit them into a coherent analysis by cutting out unnecessary material and identifying the key points of the original video.

2. LITERATURE REVIEW

The way neural networks are designed has a big influence on how well models work. But creating architectures by hand takes time and is limited by human skill. By automating the neural architecture design process, NAS algorithms prospect to decrease the amount of operator intervention required. Examines the challenges early NAS approaches faced and offers a thorough evaluation of their advantages and disadvantages. This review, in contrast to earlier ones, groups NAS algorithms according on their search space, optimization techniques, and assessment standards [7]. The study makes major advances to the field by illuminating the intricate processes involved in effectively summarizing videos. By carefully examining key concepts and methods, the authors pave the way for advancements in the field of visual communication research. Their work can be very helpful to researchers and practitioners who wish to understand more about the mathematical foundations of video summarizing techniques [8]. Exhibit effective strategy is in producing thorough summaries in a variety of media. Through alignment and attention to pertinent data, the suggested model outperforms other models in summarization tasks. The model's capacity to identify semantic linkages and provide cogent summaries is improved by the incorporation of dual contrastive losses. Through the provision of prospective pathways for future inquiry and application in diverse sectors needing thorough information synthesis. This research advances the field of multimodal summarization [9]. The model learns to choose crucial frames by utilizing reinforcement learning, and the 3D spatio-temporal U-Net efficiently captures spatial information and temporal connections. The model can produce educational and concise video summaries thanks to the integration of different strategies that advance the field of video summarization [10]. Different effective methods for producing appropriate summaries and comments through video review. This work provides important information in providing a resource for researchers and scholars who want to understand and apply deep learning techniques [11]. The model's

precise ability to capture images and its well-defined features enable it to generate summaries for educational videos. Hierarchical modeling further improves video accuracy through updatable and scalable frameworks and is good at producing summaries from long-term video datasets. This development has Increase the accuracy and precision of clips[12].When DRL algorithms are combined with computer vision algorithms, This study demonstrates how this combination has improved video summaries in this field, which are an important resource for all researchers who want to employ DRL in their projects to solve CV algorithms problems[13].These strategies, which combine classical techniques and DL algorithms in the field of summarizing video clips, have proven their worth and effectiveness in improving models [14]. This work demonstrates the rapid development and updates that have occurred in this field over recent years in order to solve the challenges that have recently emerged and thus increase the accuracy of the clips, improve them, and make them interpretable [15]. They contribute to the creation of more enlightening summaries by providing insights into the underlying causal mechanisms driving video material. Causalainer contributes to the larger objective of explainable artificial intelligence by improving the transparency and interpretability of video summarizing algorithms by including causal explanations. This work provides new opportunities to enhance the comprehension and reliability of automatic video summarization systems [16].

The papers that are presented offer a range of viewpoints on techniques for video summarizing and deep learning. Collectively, these works enhance the field and offer valuable methods and perspectives to computer vision and deep learning researchers.

3. Methodology

The proposed system starts with video recording. Through surveillance cameras, the video is in the RGB system, so we will notice the sharp contrast between the recorded videos, causing a difference in the intensity of lighting, so we transfer from the RGB system to the gray scale system [17]. Grayscale picture by applying the equation (1).

$$\text{Gray scale image} = (0.3\ R + 0.59\ G + 0.11\ B) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

The image was then enhanced using a medium filter; this was the first step. We used this technique on the grayscale image, and then we applied the Gaussian filter process to the image that was produced by the image improvement step in order to reduce the noise and increase the contrast of the final image [18]. As we mentioned, the equation (2) can be used to enhance the image.

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{x^2+y^2}{\sigma^2}\right) \quad \text{..... (2)}$$

And following that, we'll swap out the photo's pixels that are higher and equal to the threshold value for values 1 and 0, respectively, representing white and black, respectively. This will give us the binary image, and we can see from it that, by applying the equation (3), the white color in the image represents the object or shape, and the black color represents the background [19].

$$q(x,y) = f(x) = \begin{cases} 1 & if\ p(\&x,y) \geq 0 \\ 0 & if\ p(\&x,y) < 0 \end{cases} \quad \text{..... (3)}$$

After the process of obtaining the video and performing the operations of the pre-processing processes and converting the image to the binary system, we perform the process of summarizing the video from the original video by calculating the difference between the two frames, the previous frame and the subsequent frame, using the comparison function that calculates the difference between the first frame and the second frame. If they are similar, it retrieves the value Zero and if there is no similarity between the two frames, it retrieves the value of one through the value of the specified threshold for the frame. If the number of one in the binary image is less than the percentage of change, then the frames are similar and there is no difference to be neglected. But if the number of one is greater than the rate of change, the frames are different. It stores the new frame in the array by using the equation (4) [20].

Compare function = length * width * Variation rate ……..… (4)

The illumination can be balanced against the difference in illumination between the frames by taking the three lines at the top of the video and subtracting them to compare the difference in illumination by taking the average rate, the absolute amount of the frames by using the equation(5)[19].

Absolute value = [mean (mean(x)) - mean (mean (y))] ……...… (5)

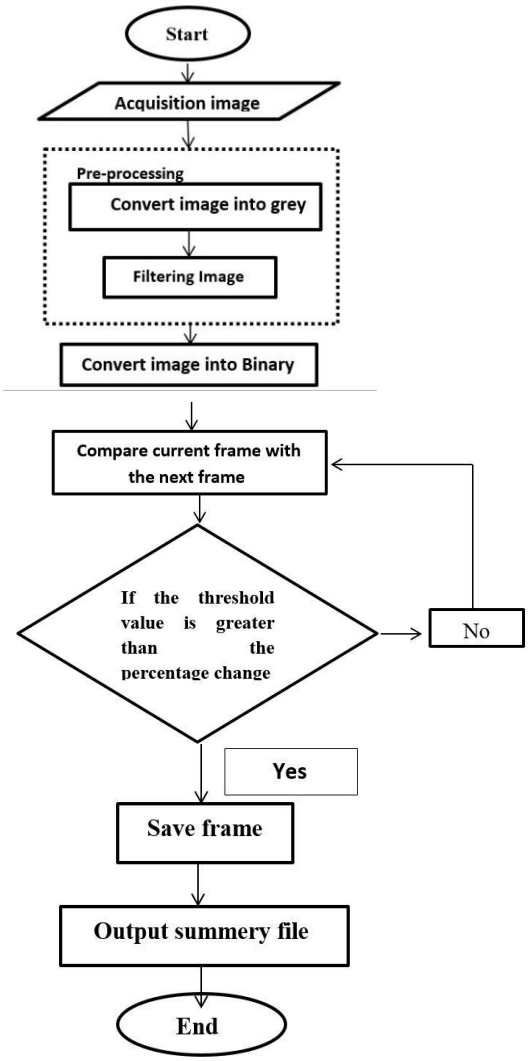

Figure (1): system flow chart system

4. Results

In this paper, the way to summarize the video can be affected by the times of night and day and different lighting conditions. When the video was captured of the same people and events repeatedly with the same rate of change, the result was different from the original video.

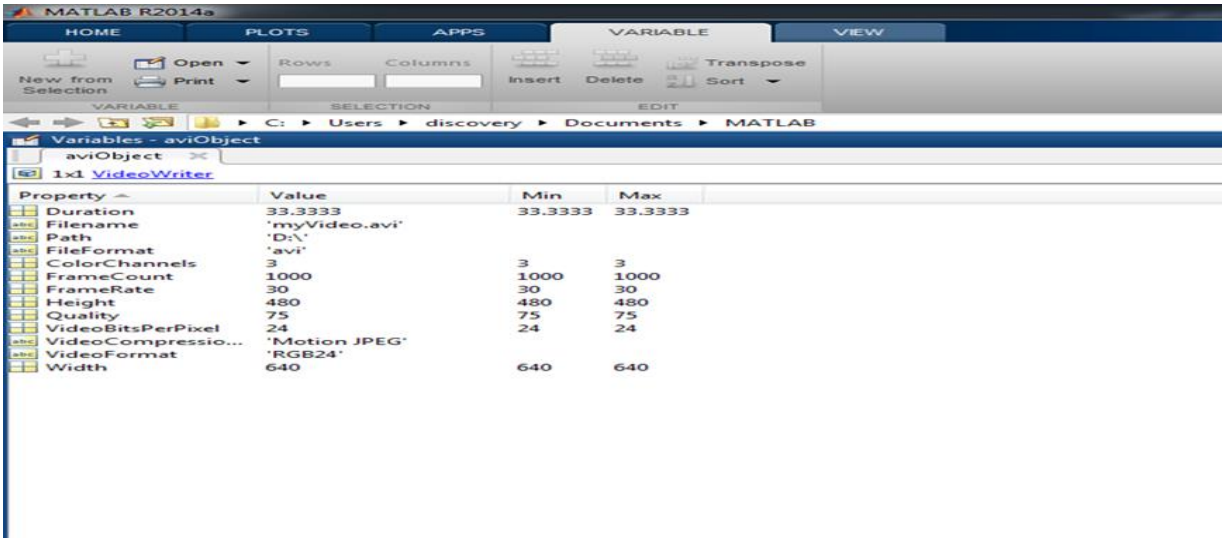

Figure (2): original video

Where the ratio is 0.02 and the original video 1000 frames while the result of the first video was in the daytime and the different lighting conditions for the same original video. Figure 2 is a summary of the same video where was the frame count 204. Shows in figure (3)

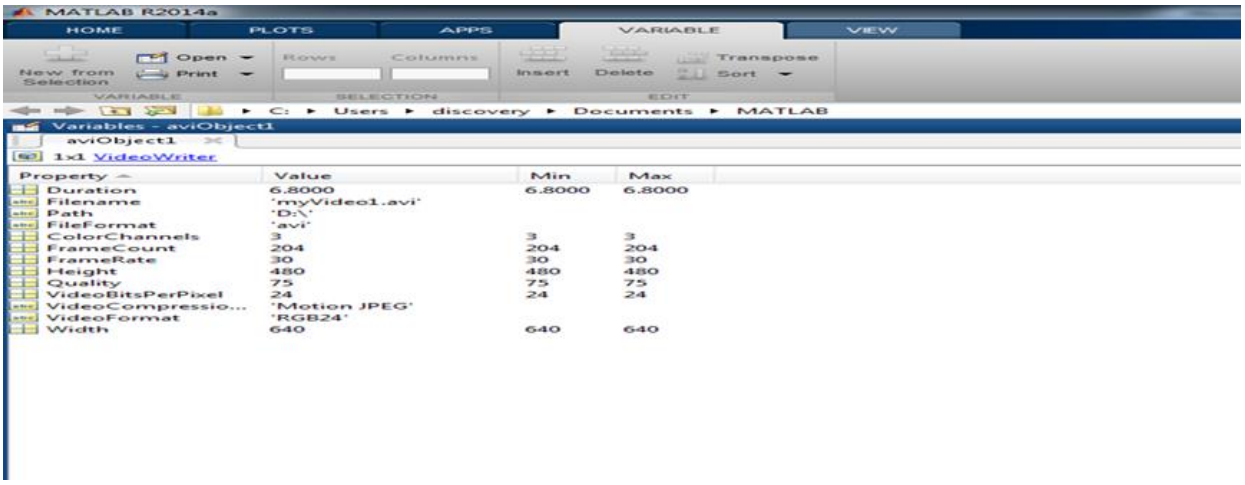

Figure (3): Summary video with 0.02 in daytime

While the result of the second video was at p count 152. Shows in figure (4)

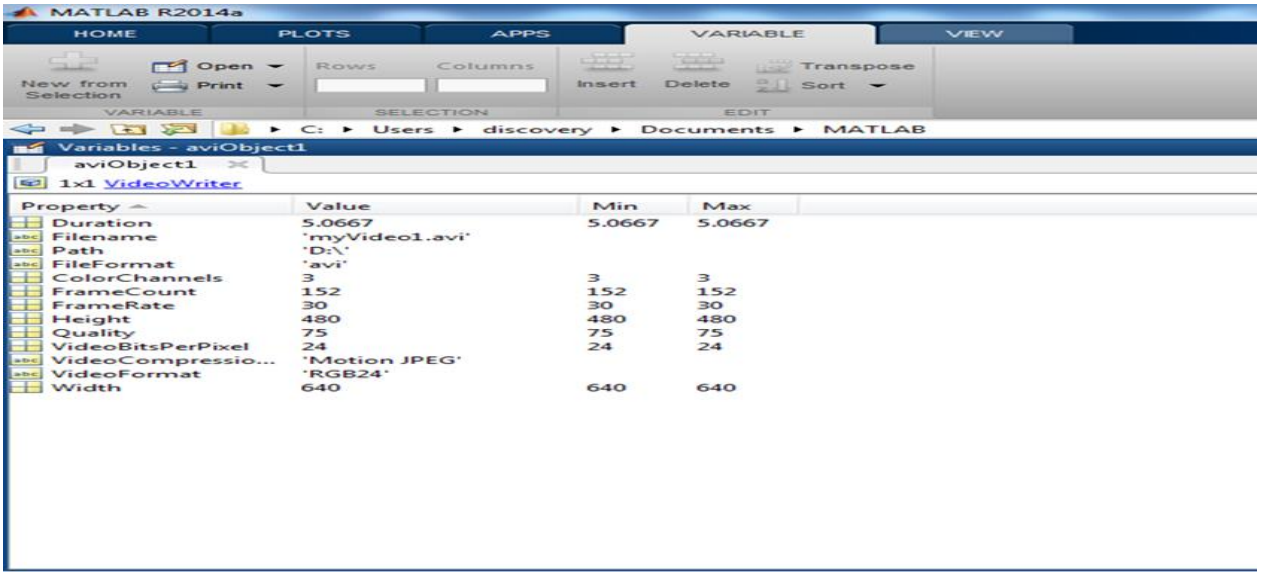

Figure (4): Summary video with 0.02 in night time

When changing the percentage of change 0.05 to the result was in the fourth figure, where the greater the percentage change, the fewer the number of identical frames where was the frame count is 17. Shows in figure (5)

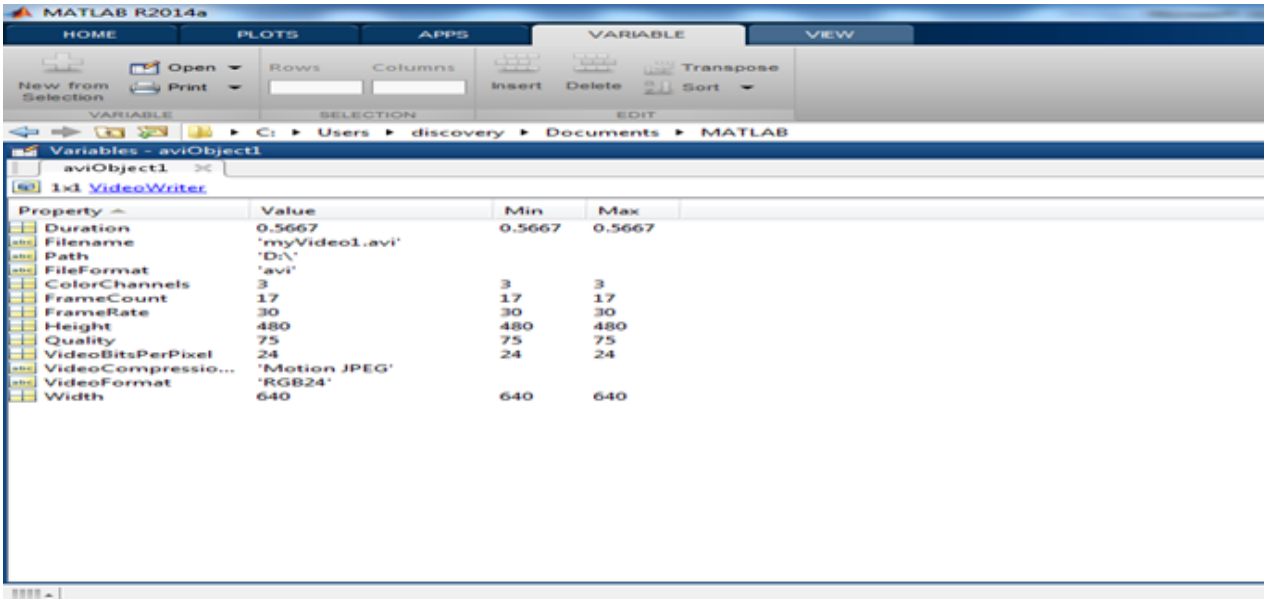

Figure (5): Summary video with 0.05

Previous studies used the Reconstructive Sequence Graph Network (RSGN) to encode frames and snapshots, where the encoding is done using long-term and short-term memory (LSTM). The results were that when using supervised equal 0.450 and unsupervised learning was 0.580, In another article, I also used deep learning and achieved major breakthroughs in many fields due to its powerful

capabilities. The result of using supervised learning was also equal to 53,1 and unsupervised learning was 53,6 in this article, the result was 6,80

5. Conclusion

A system capable of video processing and summarization can be developed through several techniques. This paper proposed a video summarization system with different conditions and a different percentage of change depending on the comparison function and the threshold value, our visual inspection of the obtained summarization results did not raise any apparent concerns. However, practitioners who wish to use our approach should be mindful of the sources of bias we have outlined above depending on the specific use case they are addressing, the model achieves competitive state-of-the-art performance on both general video summarization tasks and a medical video summarization task. The ultrasound video summarization method can be used for a variety of applications.

References

[1] Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I. (2021). Video summarization using deep neural networks: A survey. Proceedings of the IEEE, 109(11), 1838-1863.

[2] Narasimhan, M., Rohrbach, A., & Darrell, T. (2021). Clip-it! language-guided video summarization. Advances in neural information processing systems, 34, 13988-14000.

[3] Workie, A., Sharma, R., & Chung, Y. K. (2020). Digital video summarization techniques: A survey. Int. J. Eng. Technol, 9, 81-85.

[4] Jangra, A., Mukherjee, S., Jatowt, A., Saha, S., & Hasanuzzaman, M. (2023). A survey on multi-modal summarization. ACM Computing Surveys, 55(13s), 1-36.

[5] Zhao, B., Li, H., Lu, X., & Li, X. (2021). Reconstructive sequence-graph network for video summarization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(5), 2793-2801.

[6] Ji, Z., Zhao, Y., Pang, Y., Li, X., & Han, J. (2020). Deep attentive video summarization with distribution consistency learning. IEEE transactions on neural networks and learning systems, 32(4), 1765-1775.

[7] Ren, P., Xiao, Y., Chang, X., Huang, P. Y., Li, Z., Chen, X., & Wang, X. (2021). A comprehensive survey of neural architecture search: Challenges and solutions. ACM Computing Surveys (CSUR), 54(4), 1-34.

[8] Narwal, P., Duhan, N., & Bhatia, K. K. (2022). A comprehensive survey and mathematical insights towards video summarization. Journal of Visual Communication and Image Representation, 89, 103670.

[9] He, B., Wang, J., Qiu, J., Bui, T., Shrivastava, A., & Wang, Z. (2023). Align and attend: Multimodal summarization with dual contrastive losses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14867-14878).

[10] Liu, T., Meng, Q., Huang, J. J., Vlontzos, A., Rueckert, D., & Kainz, B. (2022). Video summarization through reinforcement learning with a 3D spatio-temporal u-net. IEEE transactions on image processing, 31, 1573-1586.

[11] Abdar, M., Kollati, M., Kuraparthi, S., Pourpanah, F., McDuff, D., Ghavamzadeh, M., ... & Porikli, F. (2023). A review of deep learning for video captioning. arXiv preprint arXiv:2304.11431.

[12] Zhu, W., Lu, J., Han, Y., & Zhou, J. (2022). Learning multiscale hierarchical attention for video summarization. Pattern Recognition, 122, 108312.

[13] Le, N., Rathour, V. S., Yamazaki, K., Luu, K., & Savvides, M. (2022). Deep reinforcement learning in computer vision: a comprehensive survey. Artificial Intelligence Review, 1-87.

[14] Issa, O., & Shanableh, T. (2023). Static video summarization using video coding features with frame-level temporal subsampling and deep learning. Applied Sciences, 13(10), 6065.

[15] Brauwers, G., & Frasincar, F. (2021). A general survey on attention mechanisms in deep learning. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3279-3298.

[16] Huang, J. H., Yang, C. H. H., Chen, P. Y., Chen, M. H., & Worring, M. (2023). Causalainer: Causal explainer for automatic video summarization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2629-2635).

[17] Yuan, Y., Zhang, N., Han, C., Yang, S., Xie, Z., & Wang, J. (2022). Digital image processing-based automatic detection algorithm of cross joint trace and its application in mining roadway excavation practice. International Journal of Mining Science and Technology, 32(6), 1219-1231.

[18] Aulia, J., Radila, Z., Azhary, Z. A., Nasution, A. M., Pratama, D. Y., Indriawati, K., ... & Tresna, W. P. (2023). The Bullet Launcher with A Pneumatic System to Detect Objects by Unique Markers. Journal of information and communication convergence engineering, 21(3), 252-260.

[19] Lin, H. Y., Liu, H. W., Jang, F. J., & Tseng, C. H. (2021, January). BiLuNet: a multi-path network for semantic segmentation on X-ray images. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 10034-10041). IEEE.

[20] Ren, Z., Fang, F., Yan, N., & Wu, Y. (2022). State of the art in defect detection based on machine vision. International Journal of Precision Engineering and Manufacturing-Green Technology, 9(2), 661-691.