Shaldenko, O., & Zdor, K. (2024). Neuro-mathematical fusion for shot change detection in video sequences. *Actual Issues of Modern Science. European Scientific e-Journal, 29*, 15-24. Ostrava: Tuculart Edition, European Institute for Innovation Development.

DOI: 10.47451/inn2024-03-02

The paper is published in Crossref, ICI Copernicus, BASE, Zenodo, OpenAIRE, LORY, Academic Resource Index ResearchBib, J-Gate, ISI International Scientific Indexing, ADL, JournalsPedia, Scilit, EBSCO, Mendeley, and WebArchive databases.

Oleksii Shaldenko, Candidate of Technical Sciences (Ph.D.), Associate Professor, Department of Digital Technologies in Energy, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Kyiv, Ukraine.
ORCID 0000-0001-6730-965X
Kostiantyn Zdor, Ph.D. Student, Assistant, Department of Digital Technologies in Energy, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Kyiv, Ukraine.
ORCID 0009-0008-7640-1499, Scopus 57890097900

Neuro-mathematical fusion for shot change detection in video sequences

Abstract: Shot change detection in visual media plays a pivotal role in various domains, including cinema, surveillance, and digital content organization. Traditional rule-based algorithms have shown limitations in handling the complexities of modern video content, prompting the exploration of computational intelligence approaches. This article presents a deep investigation of shot change detection, covering from traditional mathematical techniques to neural network methodologies. Through a series of experiments, we investigate the efficacy of a mathematical approach based on histograms and subsequently demonstrate the potential of integrating Long Short-Term Memory (LSTM) networks. Our findings reveal that combining mathematical precision with neural networks enhances shot change detection accuracy and efficiency, paving the way for practical real-time applications in domain of video processing and analysis. These improvements underscore the importance of adaptability and innovation in addressing the evolving challenges of visual media processing while emphasizing the importance of ethical considerations in algorithmic decision-making processes. Overall, this article invites researchers to explore the intersection of mathematical rigor and neural networks in the realm of shot change detection, offering insights into future directions and opportunities in visual perception.

Keywords: shot change detection, neural networks, Long Short-Term Memory (LSTM), video content analysis.

Introduction

In the vast landscape of visual media, the ability to determine subtle transitions between scenes is crucial for content-based analysis. Whether it is in the realms of cinema, surveillance, or digital content organization, the accurate detection of shot changes within video sequences holds significant importance. Shot change detection serves as the fundamental building block for various downstream tasks, including video summarization, content-based retrieval, and scene analysis.

Traditionally, shot change detection relied on rule-based algorithms, often incorporating simple mathematical techniques such as frame differencing or histogram analysis (*Lin et al, 2010*).

While effective in some scenarios, these methods often falter in the face of complex visual dynamics, such as rapid camera motion, lighting variations, or intricate scene compositions.

This article explores the intersection of mathematics and neural networks within the field of shot change detection. It undertakes research that follows the historical progression of shot change detection methodologies, from traditional mathematical models to the forefront of contemporary research, which is characterized by advanced neural network techniques (*Soucek & Lokoč, 2020*). Throughout this exploration, we dissect the complexities of mathematical algorithms and scrutinize the architectural sophistication of neural networks, enhancing their individual capabilities and constraints.

Moreover, we dive into the inherent limitations and challenges in the domain, navigating through the complexities of lighting fluctuations, camera motions, and scene complexities (*Abdulhussain et al, 2018*). Nevertheless, within these challenges, we discern promising trends and prospective routes.

Problem statement

Within the domain of visual media processing, the precise identification of shot changes within video sequences presents a notable challenge, carrying implications across numerous industries and applications. Conventional rule-based algorithms, grounded in elementary mathematical methodologies, frequently encounter difficulties in handling the complex characteristic of contemporary video content, including swift scene transitions, dynamic camera motions, and elaborate scene arrangements. While these traditional techniques may prove beneficial in specific contexts, their inherent limitations in adaptability and robustness often restrain their effectiveness in practical, real-world scenarios.

Moreover, with the escalating volume and diversity of video data across digital platforms, there is a growing demand for more advanced and scalable shot change detection solutions. In this context, development of neural network-based approaches signifies a new phase in computational intelligence, offering improved accuracy, flexibility, and generalization capabilities. However, the combination of mathematical and neural methodologies introduces its own array of challenges, encompassing algorithmic complexity, data accessibility, and computational resource demands.

Hence, the current challenge lies in the pursuit of ideal shot change detection methods that effectively combine mathematical precision with neural innovation. Solving this challenge requires a diverse strategy, including creation innovative algorithmic frameworks, investigation varied feature representations, and detailed examination of performance across diverse datasets and scenarios.

To summarize, the problem statement underscores the urgent necessity to push forward the frontier of shot change detection by harmonizing mathematical and neural methodologies. Addressing this challenge directly enables researchers to unlock novel routes in visual content analysis, enhancing applications spanning from video summarization and content retrieval to surveillance and beyond.

Proposed approach

To expedite and refine the detection of shot boundaries in videos, we propose employing mathematical algorithms to gather crucial information from each frame and utilize it for boundary detection (*Joyce & Liu, 2006*). This algorithm involves segmenting frames into blocks and generating visual representations in different color spaces, followed by histogram computation.

Let *L* denote the count of frames, B_i denote the *i*-th block in the frame, and *C* represent the number of blocks created during the splitting process, each block maintaining the same shape (*Figure 1*) (*Park et al., 2016*).

Subsequently, representations are generated for each block B_i in different color spaces. After experimentation, we determined that the optimal color spaces were grayscale and HSV (Hue, Saturation, Value). Notably, we found that utilizing only saturation and value from the HSV spectrum sufficed. These representations are denoted as B_i^{gray}, $B_i^{saturation}$ and B_i^{value} respectively (*Zedan et al, 2016*).

Histograms are then computed for each block in each data representation, denoted as H_i^{gray}, $H_i^{saturation}$, and H_i^{value}. Each histogram compresses data counts into the range $[0; C_h]$ to manage data compression (*Mas & Fernandez, 2006*).

Additionally, for each B_i^{gray}, we calculate edges using the Sobel-Feldman operator, resulting in B_i^{sobel}, followed by histogram computation denoted as H_i^{sobel} (*Figure 2*) (*Huan et al., 2008*).

The distance between histograms is calculated as:

$$d(a,b) = \sqrt{\sum_{j=1}^{C_h} (a_j - b_j)^2}$$

where *a* and *b* represent histograms.

Subsequently, distances between histograms are combined into a single list:

$$d_i = d(H_i^{gray}, H_{i+1}^{gray}) \cup d(H_i^{saturation}, H_{i+1}^{saturation}) \cup d(H_i^{value}, H_{i+1}^{value}) \\ \cap\, d(H_i^{sobel}, H_{i+1}^{sobel})$$

Thus, the difference between the same blocks in two frames can be calculated as:

$$D_i = \frac{1}{4C_h} \sum_{j=1}^{j=4C_h} (d_{ij})$$

yielding a single value denoting the distance (*Mohanta et al., 2012*).

Next, distances between frames are computed:

$$D_i^{frame} = \bigcup_{j=1}^{C} (D_{ij})$$

Distances between neighbouring frames can be calculated as:

$$D = \bigcup_{i=1}^{L} (D_i^{frame})$$

Anomaly detection techniques are applied to the histograms to identify deviations from the expected distributions. Let $\overline{D}$ represent the mean value of D and σ represent the standard deviation of D. This enables the identification of blocks between frames that deviate from the distribution:

$$D_{ij}^{map} = \begin{cases} 1: D_{ij} > \bar{D} + \sigma \\ 0: D_{ij} \leq \bar{D} + \sigma \end{cases}$$

Subsequently, deviations for all differences between frames based on deviated blocks are determined as:

$$A = \left\{ D_i \mid \sum_{j=1}^{C} (D_{i,j}) > \overline{D^{map}} + \sigma^{map} * k \right\}$$

where $\overline{D^{map}}$ and σ^{map} represent the mean value and standard deviation for distribution D^{map} respectively, and k is a coefficient determining anomaly detection threshold sensitivity.

This approach amalgamates block-based analysis, multi-view representation in different color spaces, histogram calculation, and anomaly detection to detect shot changes in video sequences.

To enhance the anomaly detection process, Long Short-Term Memory (LSTM) networks can be utilized (*Lindemann et al., 2021*). LSTM networks, a type of recurrent neural network (RNN), are adept at modeling sequential data and capturing long-term dependencies. In the context of shot change detection, LSTM networks can effectively analyze the temporal evolution of histogram features across consecutive blocks in the video sequence.

To train the neural network based on LSTM, we utilize D as input data where D_i represents timestamps and the number of features equals C. This allows us to replace the anomaly detection algorithm based on mean values and standard deviation with a neural network.

In conclusion, our proposed method for detecting shot boundaries in videos employs a comprehensive approach, leveraging mathematical algorithms and anomaly detection techniques. By segmenting frames into blocks and generating visual representations in grayscale, HSV, and Sobel-Feldman spaces, we enhance the accuracy of shot detection. Furthermore, our approach integrates histogram computation and anomaly detection to identify deviations between frames, thus effectively capturing shot changes. Moreover, by incorporating Long Short-Term Memory (LSTM) networks, we enhance the temporal analysis of sequential data, enabling more efficient shot change detection. This mix of techniques presents a robust framework for accurate and efficient shot boundary detection in video sequences, with potential applications in various domains.

Experiment 1

For our initial experiment, we utilized the SHOT dataset comprising 853 short videos, totaling 960,794 frames and containing 6,111 shots (*Zhu et al., 2023*). This dataset was selected due to its diverse range of videos and inclusion of challenging shot boundaries, including gradual transitions (*Figure 3*).

Algorithm 1. Compare results for the mathematical approach with TransNet V2, AutoShot@F1, AutoShot@Precision.

```
Input:
    SHOT dataset.
Output:
    Precision and F1 score.
```

Prosedure:

Step 1. Load dataset.

Step 2. For each video calculate prepare frames by splitting to blocks, converting to color spaces, calculating edges with Sibel-Feldman operator and calculating histograms.

Step 3. Calculate differences between neighboring frames.

Step 4. Determine deviating blocks and identify deviating transitions for each video.

Step 5. Calculate precision and F1 score.

During the experiment, we varied the number of blocks, color spaces, histogram sizes, and threshold coefficients, eventually settling on 64 blocks. These blocks were positioned as a grid to gather frame information without overlap.

We explored different combinations of color spaces, finding that the gray color space could effectively replace the RGB color space. Additionally, HSV proved valuable for detecting anomalies, although the Hue channel contained redundant information similar to the gray color space and was subsequently discarded.

Based on our experiments, we opted to reduce the histogram size to 64 bins, as the default size of 256 values yielded excessive information potentially lost during L2 distance calculation.

However, our findings indicated that the proposed mathematical approach using histograms for scene detection was ineffective and unreliable (*Table 1*). This experiment serves as an initial step towards developing more sophisticated shot change detection algorithms and applications in the realm of video processing and analysis.

Thus, our initial experiments on shot boundary detection utilizing the SHOT dataset and a mathematical approach yielded valuable insights into the complexities of scene transitions in videos. Despite our thorough exploration of various parameters such as block sizes, color spaces, and histogram sizes, our findings revealed limitations in the efficacy of the proposed method. While we successfully identified optimal configurations for certain elements like block size and color space selection, our approach utilizing histograms for scene detection proved inadequate compared to existing methods such as TransNet V2 and AutoShot@F1. These results underscore the need for further research and refinement in shot change detection algorithms, pointing towards the direction of leveraging more advanced techniques for improved accuracy and reliability in video processing and analysis.

Experiment 2

Building upon the initial experiment, we conducted a second experiment to assess the effectiveness of integrating Long Short-Term Memory (LSTM) networks for anomaly detection in scene change detection. This experiment aimed to showcase how LSTM-based anomaly detection enhances the accuracy and robustness of shot change detection compared to the solely mathematical histogram-based approach. The same algorithm was employed, but step 4 was replaced with training an LSTM-based neural network.

To train the neural network, we processed all videos using the algorithm from the first experiment, resulting in sequences with a size of (frame count, 64). Subsequently, we prepared short sequences with a length of 32 frames that ended with or without a scene change. To improve robustness against false positives, we included samples where the scene changes closely resembled model output (*Figure 4*)

We opted to minimize the size of the neural network to expedite the process and reduce overfitting. Consequently, our model consisted of one LSTM layer with 4 units, followed by a dense layer with 64 units and an output layer (*Figure 5*).

As a result, combination of mathematical approach with neural networks allowed us to achieve 88.9% precision and 88.8% F1 score accuracy (*Table 2*) which was 4.7% better then AutoShot@F1. We achieved comparable results with TransNetV2 and AutoShot architectures, but this approach also offers advantages in compact size and low computational requirements. The neural network we developed utilized around 1000 FLOPs per timestamp, making it suitable for real-time recognition.

Conclusions

The paper explores he dynamic landscape of shot change detection, traversing the realms of traditional mathematical approaches and cutting-edge neural network methodologies. Through a series of experiments and analyses, we have delved into the intricacies of scene change detection, highlighting the challenges, advancements, and potential future directions in this critical domain of visual media processing.

The initial experiment underscored the limitations of relying solely on mathematical techniques, particularly histogram-based approaches, for accurate and robust shot change detection. While foundational, these methods proved insufficient in handling the complexities inherent in modern video content, necessitating the exploration of more sophisticated solutions.

Building upon this groundwork, the second experiment showcased the transformative potential of integrating Long Short-Term Memory (LSTM) networks with mathematical algorithms. By leveraging the temporal dependencies in video sequences, LSTM-based anomaly detection achieved state-of-art accuracy and efficiency of shot change detection, exceeding AutoShot F1 score by 4.7% while offering advantages in model size and computational requirements.

With a streamlined architecture comprising one LSTM layer with four units, a dense layer with 64 units, and an output layer, our model demonstrated remarkable efficiency, requiring only around 1000 FLOPs per timestamp. This compact design not only facilitates real-time recognition but also reduces the computational burden, making it suitable for resource-constrained environments.

Our findings underscore the importance of combining mathematical rigor with the adaptability of neural networks, signaling a promising future for the field of shot change detection. As we continue to push the boundaries of research in this domain, the fusion of traditional methodologies with cutting-edge technologies promises to unlock new avenues for advancements in visual media processing and analysis.

References:

Abdulhussain, S. H., Ramli, A. R., Saripan, M. I., Mahmmod, B. M., Al-Haddad, S. A. R., & Jassim, W. A. (2018). Methods and Challenges in Shot Boundary Detection: A Review. *Entropy (Basel)*, *20*(4), 214. https://doi.org/10.3390/e20040214

Huan, Zh., Xiuhuan, L., & Lilei, Yu. (2008). Shot boundary detection based on mutual information and Canny Edge Detector. *2008 International Conference on Computer Science and Software Engineering*, 1124-1128. https://doi.org/10.1109/CSSE.2008.939

Joyce, R., & Liu, B. (2006). Temporal segmentation of video using frame and histogram space. *IEEE Transactions on Multimedia*, *8*(1), 130-140. https://doi.org/10.1109/TMM.2005.861285

Lindemann, B., Maschler, B., Sahlab, N., & Weyrich, M. (2021). A survey on anomaly detection for technical systems using LSTM networks. *Computers in Industry*, *131*, 103498. https://doi.org/10.1016/j.compind.2021.103498

Lin, W. & Sun, M.-T., Li, H. & Hu, H.-M. (2010). A new shot change detection method using information from motion estimation. *PCM 2010*, *Part II*, *LNCS 6298*, 264-275. https://doi.org/10.1007/978-3-642-15696-0_25

Mas, J., & Fernandez, G. (2006, October 04). Video shot boundary detection based on color histogram. *Digital Television Center*. Barcelona: La Salle School of Engineering, Ramon Llull University. https://www-nlpir.nist.gov/projects/tvpubs/tvpapers03/ramonlull.paper.pdf

Mohanta, P., Saha, S., & Chanda, Bh. (2012). A model-based shot boundary detection technique using frame transition parameters. *IEEE Transactions on Multimedia*, *14*, 223-233. https://doi.org/10.1109/TMM.2011.2170963

Park, S., Son, J., & Kim, S.-J. (2016). Study on the effect of frame size and color histogram bins on the shot boundary detection performance. *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 1-2. https://doi.org/10.1109/ICCE-Asia.2016.7804726

Soucek, T., & Lokoč, J. (2020). TransNet V2: An effective deep network architecture for fast shot transition detection. *ArXiv, abs/2008.04838*. https://paperswithcode.com/paper/transnet-v2-an-effective-deep-network

Zhu, W. et al. (2023). AutoShot: A short video dataset and state-of-the-art shot boundary detection. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2238-2247. https://doi.org/10.1109/CVPRW59228.2023.00218

Zedan, I., Elsayed, Kh., & Emary, E. (2016). Abrupt cut detection in news videos using dominant colors representation. *2016 International Conference on Advanced Intelligent Systems and Informatics*, 320-331. https://doi.org/10.1007/978-3-319-48308-5_31

Appendix

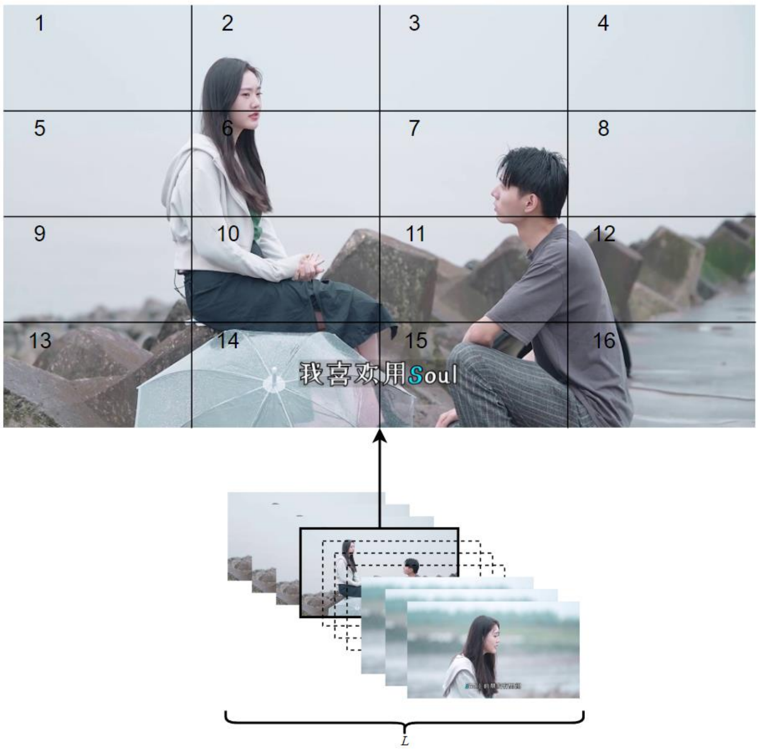

Figure 1. Example of splitting frame to blocks

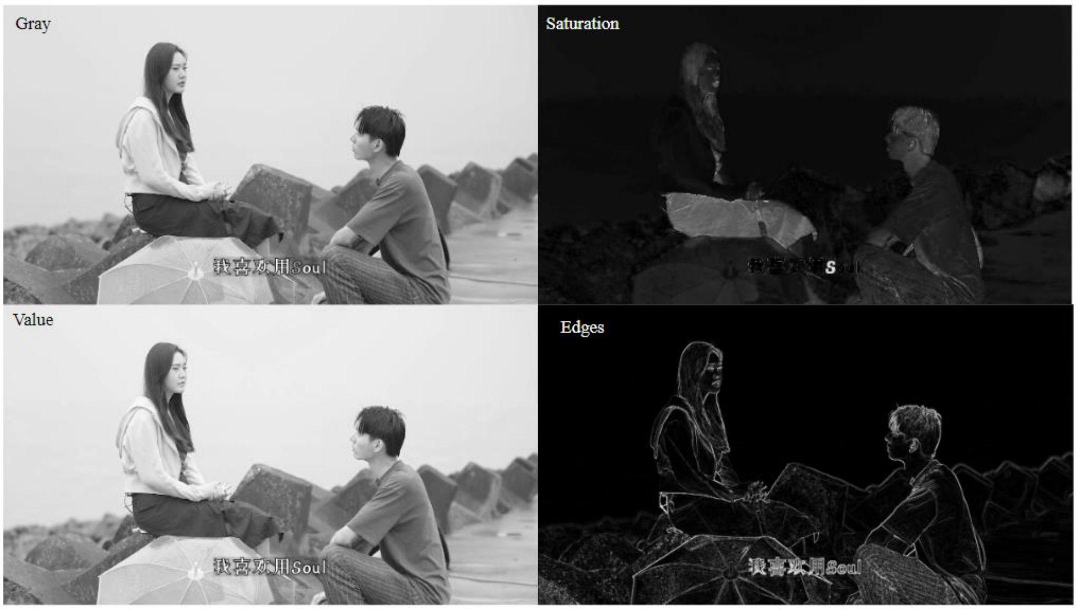

Figure 2. Visualization of different color spectrums and edges

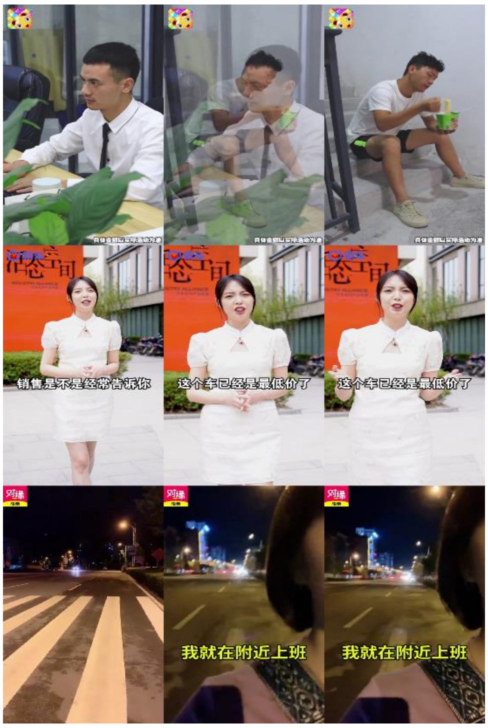

Figure 3. Example of SHOT dataset transitions

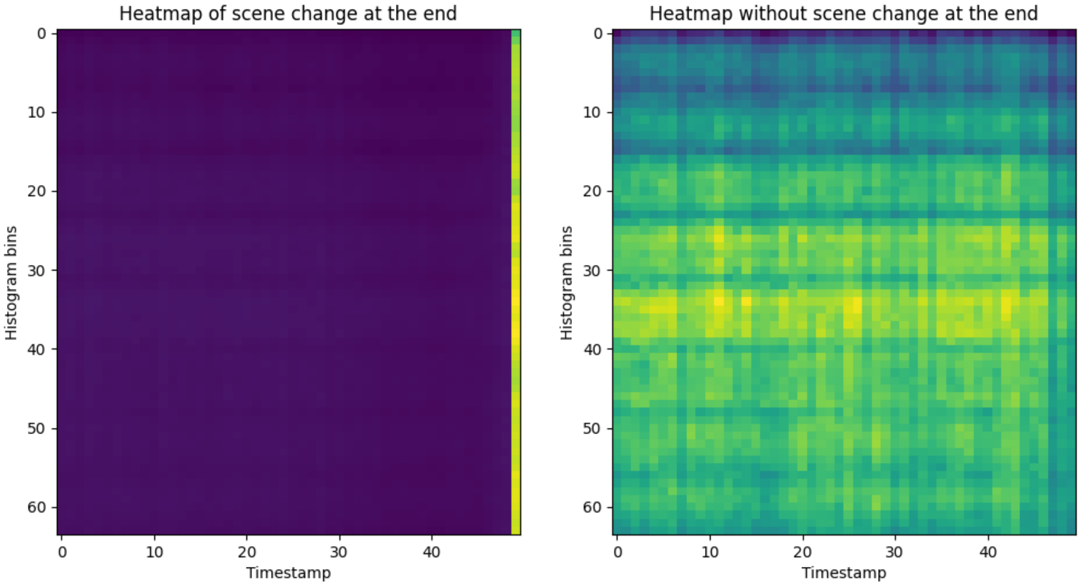

Figure 4. Heatmaps that accumulated train values for true and false labels respectively

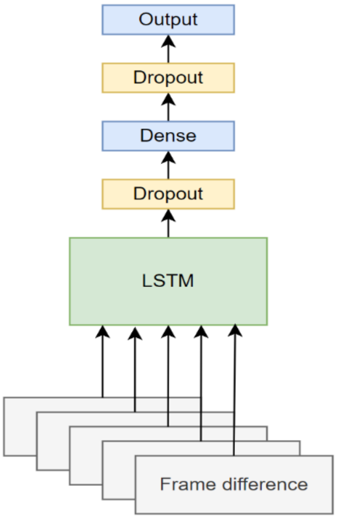

Figure 5. LSTM-base neural network architecture for shot boundary detection

Table 1. Comparison of mathematical approach with state-of-art models

Method	TransNetV2	AutoShot@F1	AutoShot@ Precision	Mathimatical approach
F1	0.799	**0.841**	0.826	0.473
Prec.	0.904	0.923	**0.939**	0.448

Table 2. Comparison between the mathematical approach and the mathematical approach enhanced with neural networks against state-of-the-art models

Method	TransNetV2	AutoShot@F1	AutoShot@ Precision	Mathimatical approach	Mathimatical with NN approach
F1	0.799	0.841	0.826	0.473	**0.888**
Prec.	0.904	0.923	**0.939**	0.448	0.889