УДК 004.8:[004.032.26+004.932
DOI: 10.31673/2412-9070.2024.062091

К. А. ЗДОР, аспірант;
ORCID: 0009-0008-7640-1499
О. В. ШАЛДЕНКО, канд. техн. наук, доцент,
ORCID: 0000-0001-6730-965X
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ

НЕЙРО-МАТЕМАТИЧНИЙ ПІДХІД ДЛЯ ВИЯВЛЕННЯ ЗМІН ПЛАНІВ У ВІДЕОПОСЛІДОВНОСТЯХ

Виявлення зміни плану (shot) у візуальних медіа відіграє важливу, фактично вирішальну роль у різних сферах, включаючи кіно, відеоспостереження та організацію цифрового контенту. Традиційні математичні алгоритми мають недостатню точність при аналізі сучасного відеоконтенту, що спонукає до дослідження підходів штучного інтелекту. У цій статті представлено дослідження алгоритмів виявлення зміни кадру, що охоплює традиційні математичні методи та застосування нейронних мереж. Була проведена серія експериментів та досліджено ефективність математичного підходу, заснованого на гістограмах у комбінації з рекурентними нейронними мережами. У результаті експериментів було визначено, що рекурентні нейроні мережі в комбінації із трансформацією даних за допомогою математичних підходів досягають високої точності навіть для відеоконтенту з складними переходами планів. Отримані результати свідчать про ефективність поєднання математичних підходів з нейронними мережами та їх актуальність до вирішення складних задач пов'язаних з обробкою відеоконтенту.

Ключові слова: нейронна мережа, рекурентна нейронна мережа, аналіз, відеоконтент, обробка інформації, штучний інтелект.

Вступ

Враховуючи постійно зростаючий обсяг відеоданих, здатність визначати переходи між планами є дуже важливою для аналізу відеоконтенту, таких як точне виявлення зміни планів для кіно, відеоспостереження та іншого цифрового контенту. Виявлення зміни плану є основою для різних задач, таких як узагальнення відео, пошук на основі контенту та аналіз планів та сцен.

Традиційно виявлення зміни плану ґрунтувалося на математичних алгоритмах, які часто включали прості математичні методи, такі як розрахунок різниці між кадрами або аналіз гістограм [1]. Хоча ці методи можуть бути ефективними у деяких сценаріях, вони часто виявляються недостатньо точними перед складними візуальними переходами, такими як швидкий рух камери, зміна освітлення або складні композиції планів.

Ця стаття досліджує поєднання математичних підходів та нейронних мереж у галузі виявлення зміни плану. У статті також розглядається історичний розвиток методології виявлення зміни кадру від традиційних математичних підходів до сучасніших досліджень із застосуванням нейронних мереж [2]. У дослідженні аналізуються складності математичних алгоритмів і розглядаються нейронні мережі з подальшою оцінкою їх можливостей та обмежень.

Також ця стаття заглиблюється у внутрішні обмеження та виклики в цій галузі, досліджуючи складності змін освітлення, рухів камери та складності сцен [5]. Незважаючи на ці виклики, було виявлено перспективні тенденції та можливі шляхи розвитку у сфері виявлення зміни плану.

Постановка задачі

У сфері обробки візуальних даних точне визначення змін плану у відеопослідовностях є важливою задачею, що може використовуватися у багатьох галузях та застосунках. Традиційні алгоритми, які базуються на математичних методах, часто демонструють погіршення якості в обробці сучасного відеоконтенту, який може містити в собі швидкі переходи між планами, динамічні рухи камери та складні композиції сцен. Хоча ці традиційні техніки можуть бути корисними при вирішенні деяких задач, через їх внутрішні обмеження в адаптивності та стійкості, вони часто показуюсь погану ефективність у реальних умовах.

Також зі зростанням обсягу та різноманітності відеоданих на цифрових платформах, зростає потреба в більш досконалих і адаптивних рішеннях для виявлення зміни плану. Із урахуванням цих потреб розробка підходів на основі нейронних мереж має на меті покращити точність, гнучкість і здатність до узагальнення при розпізнаванні планів. Однак поєднання математичних і нейронних підходів створює свої власні виклики, включаючи алгоритмічну складність, доступність даних та вимоги до обчислювальних ресурсів.

Задача полягає в пошуку ефективного методу поєднання математичних підходів з нейронною мережею з метою точного та швидкого виявлення зміни планів. Вирішення цієї задачі потребує нових підходів, які включають в себе розробку нових алгоритмів, дослідження можливих трансформацій кадрів з метою виділення особливостей та аналіз точності для різних типів переходів між планами. Розв'язання цієї задачі дозволяє застосовувати нові підходи до аналізу відеоконтенту у сферах, які включають аналіз відео, пошук контенту та аналіз відеоспостереження.

Запропонований підхід

Для розробки ефективного алгоритму виявлення переходів планів у відео, було запропоновано використовувати математичні алгоритми для збору важливої інформації з кожного кадру та використання цієї інформації для виявлення меж плану на відео [11]. Алгоритм передбачає розбиття кадрів на блоки та створення візуальних представлень у різних колірних просторах, після чого виконується обчислення гістограм.

Нехай L позначає кількість кадрів, B_i позначає і-й блок у кадрі, а C представляє кількість блоків, створених під час процесу поділу, причому кожен блок зберігає ту саму форму, як показано на рис. 1 [6].

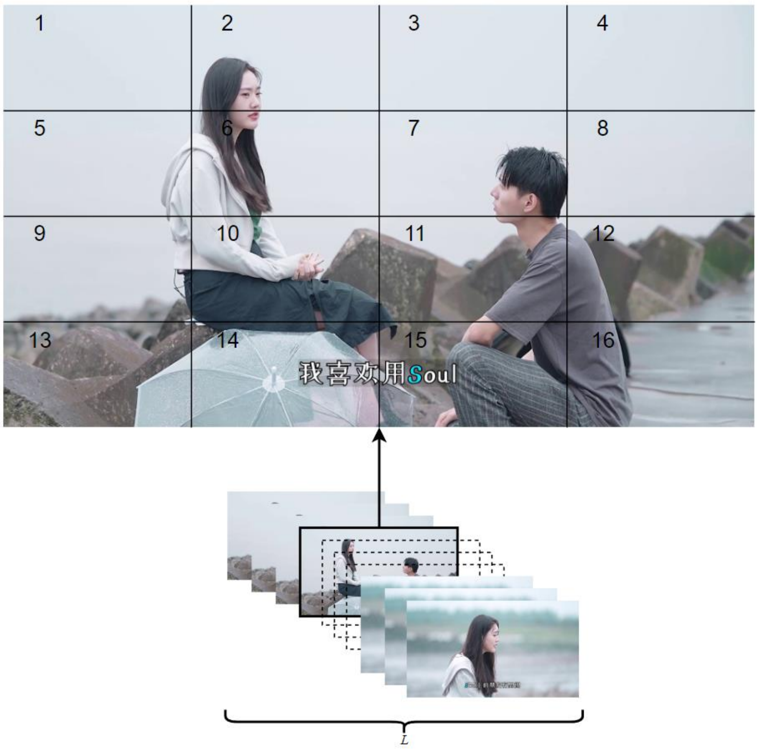

Рис. 1. Приклад розбиття кадру на блоки

Для кожного блоку B_i у різних колірних просторах створюються відповідні представлення. Експериментальним шляхом було визначено, що оптимальними колірними просторами є градації сірого та HSV (Hue, Saturation, Value). Також було виявлено, що достатньо використовувати лише насиченість і яскравість із спектру HSV. Ці представлення позначаються як B_i^{gray}, $B_i^{saturation}$ та B_i^{value} відповідно [7]. Потім обчислюються гістограми для кожного блоку в кожному представленні даних, позначені як H_i^{gray}, $H_i^{saturation}$ та H_i^{value}. Кожна гістограма стискає кількість даних у діапазон [0; C_h] [8]. Крім того, для кожного B_i^{gray} ми обчислюємо контури за допомогою оператора Собеля-Фельдмана, отри, муючи B_i^{sobel}, після чого виконується обчислення гістограми, яке позначається як H_i^{sobel}, як показано на рис. 2 [10].

Рис. 2. Представлення зображення у різних колірний

Відстань між гістограмами обчислюється за допомогою формули:

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

де *a* та *b* представляють гістограми.

Потім відстані між гістограмами об'єднуються в один список:

$$d_i = d\left(H_i^{gray}, H_{i+1}^{gray}\right) \cup d\left(H_i^{saturation}, H_{i+1}^{saturation}\right) \cup d\left(H_i^{value}, H_{i+1}^{value}\right) \cap d\left(H_i^{sobel}, H_{i+1}^{sobel}\right) \tag{2}$$

Отже, різницю між відповідними блоками у двох кадрах можна обчислити як:

$$D_i = \frac{1}{4C_h} \sum_{j=1}^{j=4C_h} \left(d_{ij}\right), \tag{3}$$

отримуючи єдине значення, що позначає відстань між цими блоками [9].

Далі обчислюються відстані між кадрами:

$$D_i^{frame} = \bigcup_{j=1}^{C} \left(D_{ij}\right) \tag{4}$$

Відстань між сусідніми кадрами можна обчислити за допомогою наступної формули:

$$D = \bigcup_{i=1}^{L} \left(D_i^{frame}\right) \tag{5}$$

Техніки виявлення аномалій застосовуються до гістограм для виявлення відхилень від очікуваних розподілів. Нехай $\underline{D}$ представляє середнє значення *D*, а σ — стандартне відхилення *D*. Це дозволяє ідентифікувати блоки між кадрами, які відхиляються від загального розподілу:

$$D_{ij}^{map} = \left\{\frac{1: D_{ij} > \underline{D} + \sigma}{0: D_{ij} \leq \underline{D} + \sigma}\right\} \tag{6}$$

Потім відхилення для всіх різниць між кадрами на основі відхилених блоків визначаються як:

$$A = \left\{ D_i \mid \sum_{j=1}^{C} \left(D_{i,j}\right) > \underline{D^{map}} + \sigma^{map} * k \right\}, \tag{7}$$

де $\underline{D^{map}}$ та σ^{map} це середнє значення та стандартне відхилення для розподілу D^{map} відповідно, а k є коефіцієнтом, який визначає чутливість порогу виявлення аномалій.

З метою покращення точності при виявлені переходів між планами було вирішено замінити частину виявлення аномалій на рекурентну нейрону мережу типу Long Short-Term Memory (LSTM) [3]. Для навчання нейронної мережі на основі LSTM було використано D як вхідні дані, де D_i представляє часові мітки, а кількість ознак дорівнює C. Це дозволяє замінити алгоритм виявлення аномалій, що ґрунтується на середніх значеннях і стандартних відхиленнях, на нейронну мережу. В результаті вдалось досягнути точності влучання 88.9% та точності F1 88.8% [12].

Далі з метою покращення точності було вирішено провести експерименти з різними типами рекурентних нейронних мереж, застосовуючи різну глибину моделей та кількість параметрів.

Експеримент

Рис. 3. Приклади переходів у наборі даних SHOT

Для експерименту було використано набір даних (SHOT), що складається з 853 коротких відео, загальною кількістю 960,794 кадрів і 6,111 планів [4]. Цей набір даних був обраний через різноманітність відео та наявність складних переходів планів, включаючи поступові пере-ходи, як показано на рис. 3.

Алгоритм навчання рекурентних моделей для порівняння результатів з TransNet V2, AutoShot@F1, AutoShot@Precision має наступний вигляд:

Крок 1. Завантажити набір даних SHOT.

Крок 2. Підготувати кадри для кожного відео: розділити на блоки, перетворити у колірні простори, обчислити контури за допомогою оператора Собеля-Фельдмана, а також гістограми (формула 1).

Крок 3. Розрахувати різниці між сусідніми трансформованими кадрами (формула 5).

Крок 4. Навчити нейронну мережу на основі визначених різниць між кадрами.

Крок 5. Обчислити точність влучності та F1-оцінку.

Для навчання нейронної мережі всі відео було оброблено та на їх основі було згенеровано послідовності для навчання, розмір яких дорівнює 50 кадрам, як показано на рис. 4. З метою зменшення хибних виявлень було згенеровано додатково зразки, де зміна сцен була близько до виходу моделі.

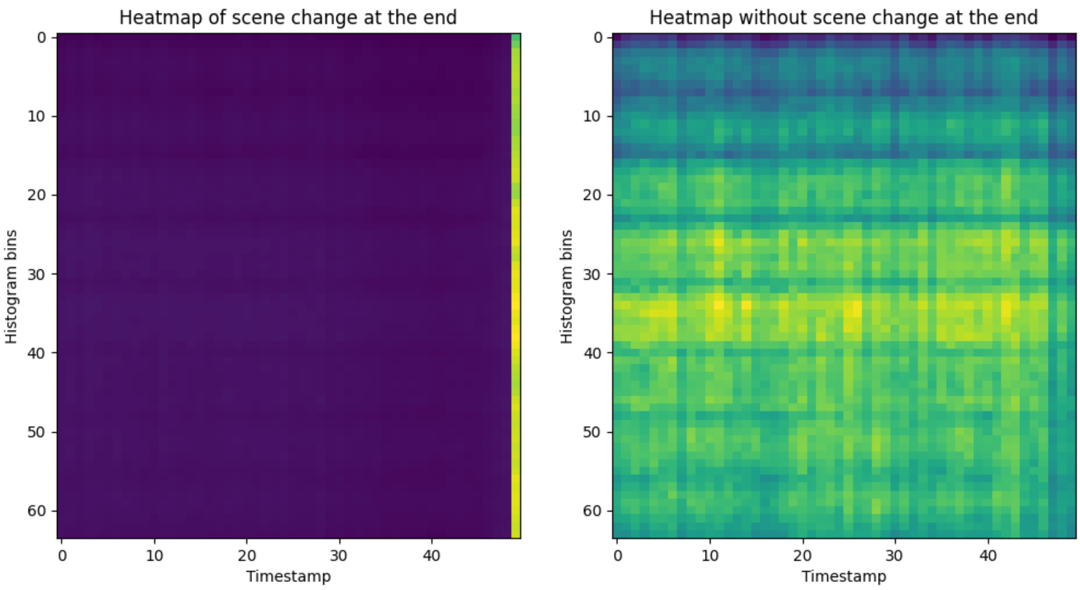

Рис. 4. Теплові карти тренувальних даних для істинних і хибних міток відповідно

Перед початком тренування нейронної мережі було вирішено знайти оптимальні параметри для математичної частини алгоритму. Для цього було використано алгоритм з застосуванням математичного знаходження аномалій (формули 6-7). В результаті експериментів було визначено, що оптимальною кількість блоків для розбиття зображення є 64 однакові блоки, що розташовані у вигляді сітки, яка рівномірно покриває зображення та не має накладань. Також було досліджено різні набори колірних просторів. Найкращі результати були досягнуті при використані RGB та HSV. Додаткові експерименти показати, що колірний простір RGB можна замінити на градацію сірого без втрати точності. Також при аналізі впливу кожного каналу з колірного простору HSV було встановлено, що канал Hue містив надлишкову інформацію, яка не впливала не результат. Тому, було вирішено використовувати градацію сірого та канали Saturation і Value з колірного простору HSV з метою покращення швидкодії. Також було протестовано різні розміри гістограм. В результаті було встановлено, що гістограма по всіх значеннях (256) містить багато надлишкової інформації, яка втрачалася під час обчислення евклідової відстані. Провівши експерименти для різних розмірів гістограми було встановлено, що найкраща точність досягається при зменшені розміру гістограми до 64 значень.

При першому поєднані нейронної мережі було вирішено використовувати просто нейрону модель, яка складалася з одного LSTM-шару, за яким слідував повнозв'язний шар із 64 одиницями та вихідний шар, як показано на рис. 5.

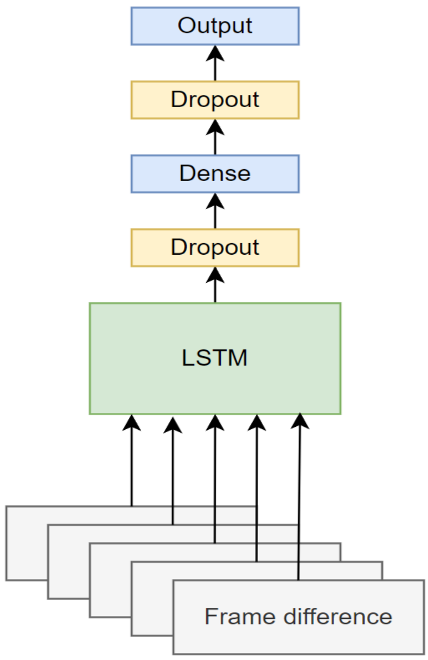

Рис. 5. Архітектура нейронної мережі на базі LSTM для виявлення переходу планів

У результаті першого поєднання математичного підходу з нейронними мережами, вдалося досягти точності влучання 88,9% та F1-оцінки 88,8% (таблиця), що на 4,7% краще, ніж AutoShot@F1 [12].

Далі було вирішено зосередитись на покращені точності влучання з метою збільшення релевантності обраних елементів. В ході експериментів було вирішено протестувати різні типи рекурентних нейронних мереж, різну глибину та кількість параметрів в моделі. Для рекурентних шарів було вирішено використовувати LSTM та Gated Recurrent Unit (GRU). При порівняння LSTM та GRU було встановлено, що при використані одного рекурентного шару, модель, яка має в основі GRU, має кращу точність сходження та F1-оцінку, проте при збільшені кількості рекурентних шарів точність LSTM починає зростати та перевершує точність GRU. Також було встановлено, що при збільшені кількості рекурентних шарів більше двох, точність почала значно погіршуватися. Проте кількість параметрів для рекурентних нейронних мереж показала залежність, при якій при використані малої кількості параметрів зростає точність влучання та повнота, але падає F1-оцінка. З цього можна зробити висновок, що залежно від задачі можна максимізувати точність влучання або повноту, або сконцентруватися на більш збалансованому підході. Також було протестовано різну кількість повнозв'язних шарів, які слідували після рекурентних шарів. Максимізація точності влучання (98.2%) за рахунок втрати F1-оцінки (83.3%) була досягнути при використані двох повнозв'язних шарів, які слідували за двома LSTM шарами. Проте при збільшені кількості параметрів для шарів LSTM та трьох повнозв'язних шарів вдалося досягнути більш збалансованого результату, при якому точність влучання та F1-оцінка дорівнювали 93.9% та 88.5% відповідно (таблиця).

Порівняння математичного підходу з використання нейронних мереж з найсучаснішими моделями

Method	F1	Precision
TransNetV2	0.799	0.904
AutoShot@F1	0.841	0.923
AutoShot@Precision	0.826	0.939
Mathematical with NN approach(2 LSTM(8) and 2 Layers)	0.833	0.982
Mathematical with NN approach(2 LSTM(128) and 3 Layers)	0.885	0.939

У ході експериментів було досягнуто показників які перевищують точність архітектур TransNetV2 та AutoShot, при цьому цей підхід також має переваги у можливості вибирати між підходами, які включають в себе максимізацію точності влучання, наповнення або F1-оцінки. Також цей підхід має перевагу у компактному розмірі моделі та низьких обчислювальних вимогах. Розроблені нейронні мережі використовують від 6 kFLOPs до 500 kFLOPS на один фрейм, що дозволяє використовувати цей підхід для розпізнавання у реальному часі.

Висновки

У статті було досліджено різні підходи до виявлення змін планів, включаючи як традиційні математичні підходи, так і застосування нейронних мереж. У процесі експериментів було проведено поглиблене дослідження складнощів виявлення змін планів, виклики, досягнення та потенційні напрямки розвитку у сфері обробки відеоданих.

Проведений експеримент продемонстрував потенціал інтеграції мереж Long Short-Term Memory (LSTM) з математичними алгоритмами. Використовуючи тимчасові залежності у відеопослідовностях та виявлення аномалій за допомогою нейронних мереж на основі LSTM, вдалося розробити два варіанти нейронних мереж. Перший варіант нейронної мережі максимізує точність влучання, що дозволяє перевищити точність влучання AutoShot@Precision на 4.3%. Другий варіант нейронної мережі максимізувати F1-оцінку та перевищує F1-оцінку AutoShot@F1 на 4.4% при цьому зберігаючи точність влучання на рівні AutoShot@Precision.

Також розроблений підхід має компактний розмір та низькі обчислювальні вимоги, використовуючи 6-500 kFLOPS на один фрейм, що дозволяє використовувати цей підхід для розв'язання задач у реальному часі.

Список літератури

1. Lin, Weiyao & Sun, Ming-Ting & Li, Hongxiang & Hu, Hai-Miao. (2010). A New Shot Change Detection Method Using Information from Motion Estimation. 264-275. 10.1007/978-3-642-15696-0_25.

2. Souček, Tomáš & Lokoč, Jakub. (2020). TransNet V2: An effective deep network architecture for fast shot transition detection.

3. Lindemann, Benjamin & Maschler, Benjamin & Sahlab, Nada & Weyrich, Michael. (2021). A survey on anomaly detection for technical systems using LSTM networks. Computers in Industry. 131. 103498. 10.1016/j.compind.2021.103498.

4. W. Zhu et al., "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), Vancouver, BC, Canada, 2023, pp. 2238-2247, doi: 10.1109/CVPRW59228.2023.00218.

5. H. Abdulhussain, Sadiq & Ramli, Abd Rahman & Saripan, M Iqbal & Mahmmod, Basheera & Al-Haddad, Syed Abdul Rahman & Jassim, Wissam. (2018). Methods and Challenges in Shot Boundary Detection: A Review. Entropy. 20. 10.3390/E20040214.

6. Park, Soyoung & Son, Jeongwoo & Kim, Sun-Joong. (2016). Study on the effect of frame size and color histogram bins on the shot boundary detection performance. 1-2. 10.1109/ICCE-Asia.2016.7804726.

7. Zedan, Ibrahim & Elsayed, Khaled & Emary, Eid. (2016). Abrupt Cut Detection in News Videos Using Dominant Colors Representation. 10.1007/978-3-319-48308-5_31.

8. Mas, Jordi & Fernandez, Gabriel. VIDEO SHOT BOUNDARY DETECTION BASED ON COLOR HISTOGRAM.

9. Mohanta, Partha & Saha, Sanjoy & Chanda, Bhabatosh. (2012). A Model-Based Shot Boundary Detection Technique Using Frame Transition Parameters. Multimedia, IEEE Transactions on. 14. 223-233. 10.1109/TMM.2011.2170963.

10. Huan, Zhao & Xiuhuan, Li & Lilei, Yu. (2008). Shot Boundary Detection Based on Mutual Information and Canny Edge Detector. 1124-1128. 10.1109/CSSE.2008.939.

11. Joyce, Robert & Liu, Bede. (2006). Temporal Segmentation of Video Using Frame and Histogram Space. Multimedia, IEEE Transactions on. 8. 130 - 140. 10.1109/TMM.2005.861285.

12. 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.

K. Zdor, O. Shaldenko

NEURO-MATHEMATICAL FUSION FOR SHOT CHANGE DETECTION IN VIDEO SEQUENCES

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. To test these approaches we decided to use the SHOT dataset which contains 853 short videos. This dataset provides a good variety of shot transitions that include difficult transitions like dissolve or zoom transitions that allow testing our approaches on modern-type videos. Through a series of experiments, we investigate the efficacy of a mathematical approach based on using various color spaces, histograms, and anomaly detection. Subsequently, we demonstrate the potential of integrating Long Short-Term Memory (LSTM) networks that replace the mathematical anomaly detection algorithm. 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 the 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, information processing, artificial intelligence.