Universidad Internacional de la Rioja (UNIR)

Escuela Superior de Ingeniería y Tecnología

Máster en en Ingeniería Matemática y Computación

Optimización de modelos BI-LSTM para la detección de violencia en video

Trabajo Fin de Estudios

presentado por: Cesar Antonio Madera Garcés

Dirigido por: Pablo Negre Rodriguez

Ciudad:Lima, Perú

Fecha: 23 de Marzo de 2025

Índice de Contenidos

Índice de Ilustraciones

Índice de Tablas

Resumen

Nota: En este apartado se introducirá un breve resumen en español del trabajo realizado (extensión máxima: 150 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

Palabras Clave: Se deben incluir de 3 a 5 palabras claves en español

Abstract

Nota: En este apartado se introducirá un breve resumen en español del trabajo realizado (extensión máxima: 150 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

Palabras Clave: Se deben incluir de 3 a 5 palabras claves en inglés

1. Introducción

Violence remains a critical global issue, with millions of incidents reported annually. According to the World Health Organization (WHO), interpersonal violence has remained as one of the top 10 causes of deaths per year in the Americas regions, sin with countless more cases of physical aggression and violent crimes going unreported (Organization, 2024). Latin America, in particular, has some of the highest violence rates, with countries like Peru, Chile, Brazil, Colombia, and Mexico experiencing significant challenges in crime prevention and public security (Bisca y cols., 2024). In 2024, the National Institute of Statistics and Geography (INEGI) reported 21.9 million legal-age victims only on Mexico, and 31.3 million crimes (INEGI, 2024). The rising availability of video surveillance and digital media presents an opportunity to develop automated systems capable of detecting and mitigating violent incidents in real time.

Artificial intelligence (AI) has emerged as a powerful tool in the field of video analysis, offering promising solutions for automated violence detection. Deep learning models, particularly convolutional neural networks (CNNs) and long-short term models (LSTM's), have demonstrated exceptional capabilities in processing spatiotemporal features from video data (Orozco, Buemi, y Berlles, 2021). By leveraging these technologies, AI-based systems can analyze video streams, recognize violent actions, and trigger alerts with high accuracy. However, challenges such as class imbalance, data scarcity, and false positives remain critical hurdles in real-world applications(Kulkarni, Batarseh, y Chong, 2021). This research aims to enhance the robustness and interpretability of AI-driven violence detection systems, contributing to safer environments in Mexico and beyond.

1.1. Justification

The escalating rates of violence in Latin America, particularly in Mexico as exposed in the prior section, underscore the urgent need for advanced surveillance systems capable of real-time incident detection. Traditional monitoring approaches, which depend on human oversight, are often inefficient due to cognitive fatigue and limitations in scalability(Marois, Hodgetts, Chamberland, Williot, y Tremblay, 2021). Artificial intelligence (AI), particularly deep learning, has demonstrated significant potential in automating violence detection through the integration of convolutional neural networks (CNNs) and long short-term me-

mory (LSTM) networks (Negre, Alonso, Prieto, Dang, y Corchado, 2024; Negre, Alonso, Prieto, Garcia, y Corchado, 2024; Abdali y Al-Tuma, 2019; Sharma, Sudharsan, Naraharisetti, Trehan, y Jayavel, 2021). While CNNs extract spatial features from video frames, LSTMs capture temporal dependencies, making them a powerful combination for analyzing dynamic scenes. However, the optimal design of these models remains an open challenge, as variations in CNN architectures and LSTM configurations directly affect detection accuracy, computational efficiency, and real-world applicability.

This study seeks to systematically investigate the trade-off between different CNN feature extractors and the number of LSTM cells to determine the most effective pipeline for violence detection. The choice of CNN influences feature extraction quality, while the number of LSTM cells impacts the model's ability to capture temporal patterns without incurring excessive computational costs. By optimizing this balance, the research aims to improve both the performance and efficiency of AI-driven violence detection systems. The outcomes will contribute not only to the academic advancement of spatiotemporal video analysis but also to the practical deployment of robust and scalable surveillance solutions, ultimately enhancing public safety in Mexico and beyond.

1.2. Work Approach

Having established the justification for this research, it is evident that the selection of feature extraction techniques and the number of LSTM cells play a crucial role in optimizing violence detection models. Existing approaches often overlook the trade-off between these two factors, potentially limiting performance in real-world applications. Therefore, this study aims to evaluate and refine the balance between CNN feature extractors and LSTM cell configurations to develop a more efficient and accurate pipeline for automated violence detection.

1.3. Document Structure

to be defined

2. Contexto y Estado del Arte

3. Identificación de Requisitos

4. Objetivos

4.1. Goals

In this chapter, the objectives and contributions of this thesis will be presented. A clear understanding of these aspects is essential to contextualize the scope and significance of this research. The following sections will provide a detailed discussion of the key goals pursued in this work, as well as the contributions it aims to make to the field.

- Main goal:
  - optimize the trade-off between modifying the CNN feature extractor and adjusting the number of LSTM cells to construct the most effective pipeline for violence detection
- Objetivos secundarios:
  - Assess the impact of various CNN architectures on the quality of extracted spatiotemporal features for violence detection.
  - Investigate how varying the number of LSTM cells affects temporal modeling and classification performance.
  - Identify the optimal balance between CNN feature extraction complexity and LSTM capacity to achieve the best performance with minimal computational cost.
  - Automate the labeling process by creating a real-time application for the pipeline.

4.2. Contributions

The main contribution of this thesis is the development of an optimized pipeline for violence detection that balances the complexity of CNN-based feature extraction and the number of LSTM cells to achieve superior accuracy and efficiency. By systematically analyzing the trade-offs between these two components, this work provides a structured approach to designing deep learning architectures for spatiotemporal violence recognition, improving both detection performance and computational feasibility. This contribution aims to advance AI-driven video analysis for real-time surveillance applications.

5. Metodología

6. Desarrollo del trabajo

7. Conclusiones y Trabajo Futuro

En la Ecuación (7.1)

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{7.1}$$

En la siguiente Tabla 7.1

1	2
22	11

Tabla 7.1: Tabla 1

En la siguiente Figura 7.1

Figura 7.1: Logo Unir

(Pimentel y Teixeira, 2016) (?, ?)

Referencias

Abdali, A. M. R., y Al-Tuma, R. F. (2019, 3). Robust real-time violence detection in video using cnn and lstm. *SCCS 2019 - 2019 2nd Scientific Conference of Computer Sciences*, 104-108. doi: 10.1109/SCCS.2019.8852616

Bisca, P. M., Chau, V., Dudine, P., Espinoza, R. A., Fournier, J.-M., Guérin, P., ... Salas, J. (2024, 11). Violent crime and insecurity in latin america and the caribbean – a macroeconomic perspective. *Departmental Papers, 2024*. Descargado de `https://www.elibrary.imf.org/view/journals/087/2024/009/article-A001-en.xml` doi: 10.5089/9798400288470.087.A001

INEGI. (2024). *Encuesta nacional de victimizaciÓn y percepciÓn sobre seguridad pÚblica (envipe) 2024* (Inf. Téc.). Autor. Descargado de `https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENVIPE/ENVIPE_24.pdf`

Kulkarni, A., Batarseh, F. A., y Chong, D. (2021). Chapter 5: Foundations of data imbalance and solutions for a data democracy. *ArXiv*.

Marois, A., Hodgetts, H. M., Chamberland, C., Williot, A., y Tremblay, S. (2021, 7). Who can best find waldo? exploring individual differences that bolster performance in a security surveillance microworld. *Applied Cognitive Psychology*, *35*, 1044-1057. doi: 10.1002/ACP.3837

Negre, P., Alonso, R. S., Prieto, J., Dang, C. N., y Corchado, J. M. (2024, 3). Systematic mapping study on violence detection in video by means of trustworthy artificial intelligence. *SSRN Electronic Journal*. Descargado de `https://papers.ssrn.com/abstract=4757631` doi: 10.2139/SSRN.4757631

Negre, P., Alonso, R. S., Prieto, J., Garcia, O., y Corchado, J. M. (2024, 5). Violence detection in video models implementation using pre-trained vgg19 combined with manual logic, lstm layers and bi-lstm layers. *SSRN Electronic Journal*. Descargado de `https://papers.ssrn.com/abstract=4832475` doi: 10.2139/SSRN.4832475

Organization, W. H. (2024). Monitoring health for the sdgs, sustainable development goals. *World Health Organization Journal*.

Orozco, C. I., Buemi, M. E., y Berlles, J. J. (2021, 6). Cnn–lstm con mecanismo de atención suave para el reconocimiento de acciones humanas en videos. *Elektron*, *5*, 37-44. doi: 10.37537/REV.ELEKTRON.5.1.130.2021

Pimentel, E. A., y Teixeira, E. V. (2016). Sharp hessian integrability estimates for nonlinear elliptic equations: An asymptotic approach. *Journal de Mathématiques Pures et Appliquées*, *106*(4), 744-767. Descargado de `https://www.sciencedirect.com/science/article/pii/S0021782416300101` doi: https://doi.org/10.1016/j.matpur.2016.03.010

Sharma, S., Sudharsan, B., Naraharisetti, S., Trehan, V., y Jayavel, K. (2021, 8). A fully integrated violence detection system using cnn and lstm. *International Journal of Electrical and Computer Engineering*, *11*, 3374-3380. doi: 10.11591/IJECE.V11I4.PP3374-3380

A. Apendices