DETEKSI AKSI KEKERASAN PADA VIDEO MENGGUNAKAN ARSITEKTUR *DUAL-STREAM* CONVLSTM DENGAN OPTIMASI *FRAME GROUPING*

SKRIPSI

Diajukan Untuk Memenuhi Sebagian dari Syarat Memperoleh Gelar Sarjana Komputer Program Studi Ilmu Komputer

Oleh
Muhammad Fikri Kafilli
2107264

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2025

DETEKSI AKSI KEKERASAN PADA VIDEO MENGGUNAKAN ARSITEKTUR *DUAL-STREAM* CONVLSTM DENGAN OPTIMASI *FRAME GROUPING*

Oleh
Muhammad Fikri Kafilli
2107264

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memeroleh gelar Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

MUHAMMAD FIKRI KAFILLI

DETEKSI AKSI KEKERASAN PADA VIDEO MENGGUNAKAN ARSITEKTUR *DUAL-STREAM* CONVLSTM DENGAN OPTIMASI *FRAME GROUPING*

Disetujui dan disahkan oleh pembimbing:

Pembimbing I

Prof. Dr. Lala Septem Riza, M.T.

NIP. 197809262008121001

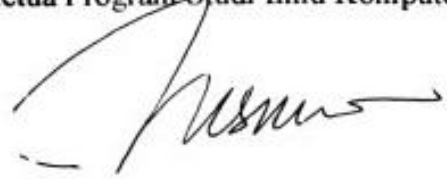
Pembimbing II

Yaya Wihardi, S. Kom., M. Kom.

NIP. 198903252015041001

Mengetahui,

Ketua Program Studi Ilmu Komputer



Dr. Muhamad Nursalman, M.T.

NIP. 197909292006041002

DETEKSI AKSI KEKERASAN PADA VIDEO MENGGUNAKAN ARSITEKTUR *DUAL-STREAM* CONVLSTM DENGAN OPTIMASI *FRAME GROUPING*

Oleh

Muhammad Fikri Kafilli

2107264

ABSTRAK

Sistem pengawasan konvensional sering kali bersifat reaktif dan tidak efisien dalam menangani insiden kekerasan, sehingga dibutuhkan solusi proaktif yang dapat memberikan peringatan secara dini. Penelitian ini bertujuan untuk merancang, membangun, dan menguji sebuah sistem deteksi kekerasan *real-time* yang tangguh dan efisien secara komputasi. Metode yang diusulkan menggunakan arsitektur *dual-stream* ConvLSTM yang secara paralel memproses fitur deteksi perubahan dan kerangka manusia. Efisiensi komputasi ditingkatkan melalui teknik *frame grouping* dan sebuah modul filter yang hanya menganalisis video saat terdeteksi adanya gerakan signifikan. Hasil eksperimen menunjukkan performa model dengan f1-*score* sebesar 78.62% dan penghematan waktu eksekusi GPU hingga 90.8% berkat penggunaan filter pada video dengan kamera statis dan aktivitas rendah. Sistem *end-to-end* yang diuji terbukti layak untuk penerapan *real-time*, dengan latensi pemrosesan maksimum 0.9891 detik. Penelitian ini berhasil mengembangkan sebuah kerangka kerja sistem deteksi kekerasan yang menyeimbangkan antara akurasi, ketangguhan, dan efisiensi komputasi, sehingga praktis untuk penerapan di dunia nyata.

Kata kunci: ConvLSTM, Deteksi Kekerasan, Deteksi Perubahan, *Frame Grouping*, Kerangka Manusia

VIOLENCE DETECTION IN VIDEO USING DUAL-STREAM CONVLSTM ARCHITECTURE WITH FRAME GROUPING OPTIMIZATION

Arranged by

Muhammad Fikri Kafilli

2107264

ABSTRACT

Conventional surveillance systems are often reactive and inefficient in handling violent incidents, necessitating proactive solutions capable of providing early warnings. This research aims to design, build, and test a robust and computationally efficient real-time violence detection system. The proposed method adopts a dual-stream ConvLSTM architecture that processes change detection and human skeleton features in parallel. Computational efficiency is enhanced through the frame grouping technique and a filter module that only analyzes video when significant motion is detected. Experimental results show a model performance with an f1-score of 78.62% and GPU execution time savings of up to 90.8% on videos with static cameras and low activity. The tested end-to-end system is proven viable for real-time application, with a maximum processing latency of 0.9891 seconds. This research successfully develops a framework for a violence detection system that balances accuracy, robustness, and computational efficiency, making it practical for real-world implementation.

Keywords: *Change Detection, ConvLSTM, Frame Grouping, Human Skeleton, Violence Detection*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR PUSTAKA

Akti, S., Tataroglu, G. A., & Ekenel, H. K. (2019). Vision-based fight detection from surveillance cameras. *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 1–6. https://doi.org/10.1109/ipta.2019.8936070

Al-Faris, M., Chiverton, J., Ndzi, D., & Ahmed, A. I. (2020). A review on computer vision-based methods for human action recognition. *Journal of Imaging*, 6(6), 46. https://doi.org/10.3390/jimaging6060046

Al-Madani, A. M., Mahale, V., & Gaikwad, A. T. (2023). Real-time detection of crime and violence in video surveillance using deep learning. *In Advances in intelligent systems research/Advances in Intelligent Systems Research* (pp. 431–441). https://doi.org/10.2991/978-94-6463-196-8_33

Algethami, N., & Redfern, S. (2020). A robust tracking-by-detection algorithm using adaptive accumulated frame differencing and corner features. *Journal of Imaging*, 6(4), 25. https://doi.org/10.3390/jimaging6040025

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. a. S., & Asari, V. K. (2019). A state-of-the-art survey on deep *learning* theory and architectures. *Electronics*, 8(3), 292. https://doi.org/10.3390/electronics8030292

Anagnostopoulos, C., & Krinidis, S. (2024). Sensors and advanced sensing techniques for *computer vision* applications. *Sensors*, 25(1), 35. https://doi.org/10.3390/s25010035

Aravinda, C. V., Al-Shehari, T., Alsadhan, N. A., Shetty, S., Padmajadevi, G., & Reddy, K. R. U. K. (2025). A novel hybrid architecture for video frame prediction: combining convolutional LSTM and 3D CNN. *Journal of Real-Time Image Processing*, 22(1). https://doi.org/10.1007/s11554-025-01626-w

Ariyandi, H. Z., Muhtadi, M., & Andreanto, D. D. (2025). Metode frame difference untuk deteksi gerakan tidur bayi berbasis computer vision. *edumatic jurnal pendidikan informatika*, 9(1), 21–30. https://doi.org/10.29408/edumatic.v9i1.29004

Benzyane, M., Azrour, M., Zeroual, I., & Agoujil, S. (2023). Investigating the influence of convolutional operations on LSTM networks in video classification. *Data & Metadata*, 2, 152. https://doi.org/10.56294/dm2023152

Biswas, M., Jibon, A. H., Kabir, M., Mohima, K., Sinthy, R., Islam, M. S., & Siddique, M. (2022). State-of-the-art violence detection techniques: a review. *Asian Journal of Research in Computer Science*, 29–42. https://doi.org/10.9734/ajrcos/2022/v13i130305

Bounoua, I., Saidi, Y., Yaagoubi, R., & Bouziani, M. (2024). Deep learning approaches for water stress forecasting in arboriculture using time series of remote sensing images: comparative study between ConvLSTM and CNN-LSTM models. *Technologies*, 12(6), 77. https://doi.org/10.3390/technologies12060077

Cadet, N. E., Osundare, N. O. S., Ekpobimi, N. H. O., Samira, N. Z., & Weldegeorgise, N. Y. W. (2024). AI-powered threat detection in surveillance systems: a *real-time* data processing *frame*work. *Open Access Research Journal of Engineering and Technology*, 7(2), 031–045. https://doi.org/10.53022/oarjet.2024.7.2.0057

Carreira, J., & Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics *dataset*. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* https://doi.org/10.1109/cvpr.2017.502

Çetinkaya, A., Baykan, Ö. K., & Kırgız, H. (2023). Analysis of machine *learning* classification approaches for predicting students' programming aptitude. *Sustainability*, 15(17), 12917. https://doi.org/10.3390/su151712917

Cheng, M., Cai, K., & Li, M. (2021). RWF-2000: An open large scale video database for violence detection. *2022 26th International Conference on Pattern Recognition (ICPR)*, 4183–4190. https://doi.org/10.1109/icpr48806.2021.9412502

De Gregorio, M., & Giordano, M. (2017). WISARDRP for change detection in video sequences. *The European Symposium on Artificial Neural Networks*. https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-133.pdf

Degardin, B., & Proenca, H. (2020). Human activity analysis: iterative weak/self-supervised learning frameworks for detecting abnormal events. *2020 IEEE International Joint Conference on Biometrics (IJCB).* https://doi.org/10.1109/ijcb48548.2020.9304905

Desai, M. M., & Mewada, H. K. (2021). Review on human pose estimation and human body joints localization. *International Journal of Computing and Digital Systems*, 10(1), 883–898. https://doi.org/10.12785/ijcds/100181

Dubey, S., & Dixit, M. (2022). A comprehensive survey on human pose estimation approaches. *Multimedia Systems*, 29(1), 167–195. https://doi.org/10.1007/s00530-022-00980-0

Durairaj, A., Madhan, E., Rajkumar, M., & Shameem, S. (2024). Optimizing anomaly detection in 3D MRI scans: the role of ConvLSTM in medical image analysis. *Applied Soft Computing*, 164, 111919. https://doi.org/10.1016/j.asoc.2024.111919

Durrani, A. U. R., Minallah, N., Aziz, N., Frnda, J., Khan, W., & Nedoma, J. (2023). Effect of hyper-parameters on the performance of ConvLSTM based deep neural network in crop classification. *PLoS ONE*, 18(2), e0275653. https://doi.org/10.1371/journal.pone.0275653

Gandapur, M. Q., & Verdú, E. (2023). CONVGRU-CNN: spatiotemporal deep learning for real-world anomaly detection in video surveillance system. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(4), 88. https://doi.org/10.9781/ijimai.2023.05.006

Gao, M., Zou, G., Li, Y., & Guo, X. (2024). Recent advances in computer vision: technologies and applications. *Electronics*, 13(14), 2734. https://doi.org/10.3390/electronics13142734

Garcia-Cobo, G., & SanMiguel, J. C. (2023). Human skeletons and change detection for efficient violence detection in surveillance videos. *Computer Vision and Image Understanding*, 233, 103739. https://doi.org/10.1016/j.cviu.2023.103739

Ghosh, D. K., & Chakrabarty, A. (2022). Two-stream *multi*-dimensional convolutional network for *real-time* violence detection. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2211.04255

Göncz, L., & Majdik, A. (2022). Object-based change detection algorithm with a spatial ai stereo camera. *Sensors*, 22(17), 6342. https://doi.org/10.3390/s22176342

Gong, W., Zhang, X., Gonzàlez, J., Sobral, A., Bouwmans, T., Tu, C., & Zahzah, E. (2016). Human pose estimation from monocular images: a comprehensive survey. *Sensors*, 16(12), 1966. https://doi.org/10.3390/s16121966

Goyette, N., Jodoin, P., Porikli, F. M., Konrad, J., & Ishwar, P. (2014). A novel video *dataset* for change detection benchmarking. *IEEE Transactions on Image Processing*, 23(11), 4663–4679. https://doi.org/10.1109/tip.2014.2346013

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: efficient convolutional neural networks for mobile vision applications. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1704.04861

Hu, L., Liu, S., & Feng, W. (2023). Skeleton-based action recognition with local dynamic spatial–temporal aggregation. *Expert Systems With Applications*, 232, 120683. https://doi.org/10.1016/j.eswa.2023.120683

Husein, A. M., Calvin, N., Halim, D., Leo, R., & William, N. (2019). Motion detect application with *frame* difference method on a surveillance camera. *Journal of Physics Conference Series*, 1230(1), 012017. https://doi.org/10.1088/1742-6596/1230/1/012017

Hwang, I., & Kang, H. (2023). Anomaly detection based on a 3D convolutional neural network combining convolutional block attention module using merged frames. *Sensors*, 23(23), 9616. https://doi.org/10.3390/s23239616

Javaid, M., Haleem, A., Singh, R. P., & Ahmed, M. (2024). Computer vision to enhance healthcare domain: an overview of features, implementation, and opportunities. *Intelligent Pharmacy*, 2(6), 792–803. https://doi.org/10.1016/j.ipha.2024.05.007

Jiang, T. (2023). rtmlib [Software]. *GitHub*. https://github.com/Tau-J/rtmlib

Jiang, T., Lu, P., Zhang, L., Ma, N., Han, R., Lyu, C., Li, Y., & Chen, K. (2023). RTMPOSE: Real-time *multi*-person pose estimation based on mmpose. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2303.07399

Kang, M., Park, R., & Park, H. (2021). Efficient spatio-temporal modeling methods for *real-time* violence recognition. *IEEE Access*, 9, 76270–76285. https://doi.org/10.1109/access.2021.3083273

Kim, J., Choo, H., & Jeong, J. (2024). Self-Attention (SA)-CONVLSTM Encoder–Decoder Structure-Based video Prediction for dynamic *motion* estimation. Applied Sciences, 14(23), 11315. https://doi.org/10.3390/app142311315

Khan, H., Yuan, X., Qingge, L., & Roy, K. (2025). violence detection from industrial surveillance videos using deep learning. *IEEE Access*, 1. https://doi.org/10.1109/access.2025.3531213

Kozłowski, M., Racewicz, S., & Wierzbicki, S. (2024). Image analysis in autonomous vehicles: a review of the latest ai solutions and their comparison. *Applied Sciences*, 14(18), 8150. https://doi.org/10.3390/app14188150

Lindroth, H., Nalaie, K., Raghu, R., Ayala, I. N., Busch, C., Bhattacharyya, A., Franco, P. M., Diedrich, D. A., Pickering, B. W., & Herasevich, V. (2024). Applied artificial intelligence in healthcare: a review of computer vision technology application in hospital settings. *Journal of Imaging*, 10(4), 81. https://doi.org/10.3390/jimaging10040081

Lu, P., Jiang, T., Li, Y., Li, X., Chen, K., & Yang, W. (2023). RTMO: Towards high-performance one-stage *real-time multi*-person pose estimation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2312.07526

Ludwig, K., Kienzle, D., & Lienhart, R. (2022). Recognition of freely selected keypoints on human limbs. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3530–3538. https://doi.org/10.1109/cvprw56347.2022.00397

Ma, G., & Zhang, Q. (2025). Research on person pose estimation based on parameter inverted pyramid and high-dimensional feature enhancement. *Symmetry*, 17(6), 941. https://doi.org/10.3390/sym17060941

Malakhov, A. (2016). Composable *multi-threading* for python libraries. *Proceedings of the Python in Science Conferences*, 15–19. https://doi.org/10.25080/majora-629e541a-002

Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies*, 12(2), 15. https://doi.org/10.3390/technologies12020015

Matsuzaka, Y., & Yashiro, R. (2023). AI-based *computer vision* techniques and expert systems. *AI*, 4(1), 289–302. https://doi.org/10.3390/ai4010013

Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: architectures, recent advances, and applications. *Information*, 15(12), 755. https://doi.org/10.3390/info15120755

Moreira, D., Barandas, M., Rocha, T., Alves, P., Santos, R., Leonardo, R., Vieira, P., & Gamboa, H. (2021). Human activity recognition for indoor localization using smartphone inertial sensors. *Sensors*, 21(18), 6316. https://doi.org/10.3390/s21186316

Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–36. https://doi.org/10.4097/kja.21209

Negre, P., Alonso, R. S., González-Briones, A., Prieto, J., & Rodríguez-González, S. (2024). Literature review of deep-learning-based detection of violence in video. *Sensors, 24*(12), 4016. https://doi.org/10.3390/s24124016

Ojha, R. R., Chawdary, H., & Saraswat, S. (2025). Enhancing public safety: *real-time* violence detection and notification system. *Procedia Computer Science*, 258, 2988–2995. https://doi.org/10.1016/j.procs.2025.04.558

Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security

systems: a systematic review. *PeerJ Computer Science*, 8, e920. https://doi.org/10.7717/peerj-cs.920

Ordóñez, F., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for *multi*modal wearable activity recognition. Sensors, 16(1), 115. https://doi.org/10.3390/s16010115

Padilla, R., Netto, S. L., & Da Silva, E. a. B. (2020). A survey on performance metrics for object-detection algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 237–242. https://doi.org/10.1109/iwssip48289.2020.9145130

Park, J., Mahmoud, M., & Kang, H. (2024). CONV3D-based video violence detection network using optical flow and rgb data. *Sensors*, 24(2), 317. https://doi.org/10.3390/s24020317

Porikli, F., & Yilmaz, A. (2012). Object detection and tracking. *In Studies in computational intelligence* (pp. 3–41). https://doi.org/10.1007/978-3-642-28598-1_1

Qiao, H., Wang, T., Wang, P., Qiao, S., & Zhang, L. (2018). A time-distributed spatiotemporal feature learning method for machine health monitoring with *multi*-sensor time series. *Sensors*, 18(9), 2932. https://doi.org/10.3390/s18092932

Qin, Y., Cao, J., & Ji, X. (2021). Fire detection method based on depthwise separable convolution and YOLOV3. *International Journal of Automation and Computing*, 18(2), 300–310. https://doi.org/10.1007/s11633-020-1269-5

Rai, N. M. (2025). Vision-based vehicle safety systems: statistical evaluation of efficacy and impact. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 11(2), 1787–1806. https://doi.org/10.32628/cseit25112544

Ramakrishnan, G., & Balamurugan, S. P. (2022). A comprehensive analysis of *motion* and *motion*less object detection in real time video surveillance using frame difference and background subtraction. *Annamalai University-IQAC*.

Rani, M., & Kumar, M. (2024). MobileNet for human activity recognition in smart surveillance using transfer learning. *Neural Computing and Applications*, 37(5), 3907–3924. https://doi.org/10.1007/s00521-024-10882-z

Reddy, R., G.Su*deep*thi, T.Vaishnavi, & Swapna.C. (2024). Smart surveillance for violence detection. *International Journal for Multidisciplinary Research*, 6(6). https://doi.org/10.36948/ijfmr.2024.v06i06.32682

Rehman, A., Saba, T., Khan, M. Z., Damaševičius, R., & Bahaj, S. A. (2022). Internet-of-things-based suspicious activity recognition using *multi*modalities

of computer vision for smart city security. *Security and Communication Networks*, 2022, 1–12. https://doi.org/10.1155/2022/8383461

Ren, B., Liu, M., Ding, R., & Liu, H. (2024). a survey on 3d skeleton-based action recognition using learning method. *Cyborg and Bionic Systems*. https://doi.org/10.34133/cbsystems.0100

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal*, 50(3), 419–430. https://doi.org/10.1002/bimj.200710415

Samkari, E., Arif, M., Alghamdi, M., & Ghamdi, M. a. A. (2023). Human pose estimation using deep learning: a systematic literature review. *Machine Learning and Knowledge Extraction*, 5(4), 1612–1659. https://doi.org/10.3390/make5040081

Sanchez-Caballero, A., Fuentes-Jiménez, D., & Losada-Gutiérrez, C. (2020). Exploiting the ConvLSTM: human action recognition using raw depth video-based recurrent neural networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2006.07744

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6). https://doi.org/10.1007/s42979-021-00815-1

Shao, Z., Cai, J., & Wang, Z. (2017). Smart monitoring cameras driven intelligent processing to big surveillance video data. *IEEE Transactions on Big Data*, 4(1), 105–116. https://doi.org/10.1109/tbdata.2017.2715815

Shavetov, S. V., Merkulova, I. I., Ekimenko, A. A., Borisov, O. I., & Gromov, V. S. (2019). Computer vision in control and robotics for educational purposes. IFAC-PapersOnLine, 52(9), 127–132. https://doi.org/10.1016/j.ifacol.2019.08.136

Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., & Woo, W. (2015). Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1506.04214

SIMFONI-PPA. (n.d.). https://kekerasan.kemenpppa.go.id/ringkasan

Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., & Singh, S. (2023). Facial expression recognition in videos using hybrid CNN & ConvLSTM. *International Journal of Information Technology*, 15(4), 1819–1830. https://doi.org/10.1007/s41870-023-01183-0

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM -- a tutorial into long short-term memory recurrent neural networks. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.1909.09586

Su, J., Her, P., Clemens, E., Yaz, E., Schneider, S., & Medeiros, H. (2022). Violence detection using 3d convolutional neural networks. *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–8. https://doi.org/10.1109/avss56176.2022.9959393

Sultonov, F., Park, J., Yun, S., Lim, D., & Kang, J. (2022). Mixer U-Net: an improved automatic road extraction from uav imagery. *Applied Sciences*, 12(4), 1953. https://doi.org/10.3390/app12041953

Talha, K. R., Bandapadya, K., & Khan, M. M. (2022). Violence detection *using* computer vision approaches. *2022 IEEE World AI IoT Congress (AIIoT)*, 544–550. https://doi.org/10.1109/aiiot54504.2022.9817374

Temkar, N. R. (2024). Automated violence detection in surveillance networks with deep learning. *Deleted Journal*, *28*(1s), 83–93. https://doi.org/10.52783/anvi.v28.2201

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3D convolutional networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, 4489–4497. https://doi.org/10.1109/iccv.2015.510

Traoré, A., & Akhloufi, M. A. (2020). 2D bidirectional gated recurrent unit convolutional neural networks for end-to-end violence detection in videos. *In Lecture notes in computer science* (pp. 152–160). https://doi.org/10.1007/978-3-030-50347-5_14

Ullah, F. U. M., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3d convolutional neural network. *Sensors*, 19(11), 2472. https://doi.org/10.3390/s19112472

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. https://doi.org/10.1007/s10462-020-09838-1

Vosta, S., & Yow, K. (2022). A cnn-rnn combined structure for real-world violence detection in surveillance cameras. *Applied Sciences*, 12(3), 1021. https://doi.org/10.3390/app12031021

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: a brief review. *Computational Intelligence and Neuroscience*, 2018, 1–13. https://doi.org/10.1155/2018/7068349

Wang, C., & Yan, J. (2023). A comprehensive survey of rgb-based and skeleton-based human action recognition. *IEEE Access*, 11, 53880–53898. https://doi.org/10.1109/access.2023.3282311

Wang, J., Zhao, D., Li, H., & Wang, D. (2024). Lightweight violence detection model based on 2d cnn with bi-directional *motion* attention. *Applied Sciences*, *14*(11), 4895. https://doi.org/10.3390/app14114895

Yadav, P., Gupta, N., & Sharma, P. K. (2022). A comprehensive study towards high-level approaches for weapon detection using classical machine *learning* and *deep learning* methods. *Expert Systems With Applications*, 212, 118698. https://doi.org/10.1016/j.eswa.2022.118698

Yao, S., He, Y., Zhang, L., Yang, W., Chen, Y., Sun, Q., Zhao, Z., & Cao, S. (2023). A CONVLSTM neural network model for spatiotemporal prediction of mining area surface deformation based on SBAS-INSAR monitoring data. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–22. https://doi.org/10.1109/tgrs.2023.3236510

Yuan, G., Gong, J., Deng, M., Zhou, H., & Xu, D. (2014). A moving objects detection algorithm based on three-frame difference and sparse optical flow. *Information Technology Journal*, 13(11), 1863–1867. https://doi.org/10.3923/itj.2014.1863.1867

Zhang, L., Lu, L., Wang, X., Zhu, R. M., Bagheri, M., Summers, R. M., & Yao, J. (2019). Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. *IEEE Transactions on Medical Imaging*, 39(4), 1114–1126. https://doi.org/10.1109/tmi.2019.2943841

Zhang, Y., Li, Y., & Guo, S. (2022). Lightweight mobile network for *real-time* violence recognition. *PLoS ONE*, 17(10), e0276939. https://doi.org/10.1371/journal.pone.0276939

Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., & Shah, M. (2023). Deep learning-based human pose estimation: a survey. *ACM Computing Surveys*, 56(1), 1–37. https://doi.org/10.1145/3603618

Zhou, P., Ding, Q., Luo, H., & Hou, X. (2018). Violence detection in surveillance video using low-level features. *PLoS ONE*, 13(10), e0203668. https://doi.org/10.1371/journal.pone.0203668