Chapter 2: Literature Review

Review of existing literature on deep learning in anomaly detection and video surveillance

**Road Accident**

Chen et al. [1] propose an RHD solution based on CVIS for intelligent transportation systems. Road hazards (RH) pose significant threats to safety and economies. Using meta-learning, they generalize critical features from RH data and design a lightweight RHD model for efficient inference on onboard computing devices (OCD). Experimental results show 90.2% accuracy in detecting RH, with an inference time of 14.7ms.

Thakare et al. [2] introduce a process where, following accident localization, post-processing is conducted to ascertain the context and severity of the incident. To streamline the formation of temporal relations, real-time object detection and their positions are emphasized. Their study presents a method for object interaction-based localization and description of road accident events using deep learning.

Kukade Jyoti, Swapnil Soner, and Sagar Pandya et al. [3] implemented various CNN architectures by combining VGG16 with ConvoLSTM and LSTM for violence detection in real-time video feeds, achieving an impressive 99% accuracy. The preprocessing steps involve transforming the video into a time sequence of images displayed in YUV format, subsampling the images, and grouping them into batches. Additionally, data augmentation is applied to enhance the size and quality of the photos.

Al-Dahash et al. [4] provide an in-depth analysis of recent advancements in vehicle collision identification and alert systems. They discuss various techniques, including sensor-based and vision-based methods, highlighting their benefits in reducing accident rates and enhancing driving safety. Despite notable progress, challenges such as accuracy, reliability, and cost-effectiveness remain unresolved.

Chen et al. [5] propose a real-time automobile collision detection system leveraging deep learning. Their approach involves training a deep neural network using data from multiple sensors, such as accelerometers, gyroscopes, and webcams. Through extensive testing with real collision incidents, they demonstrate the system's high accuracy and rapid response times, suggesting promising prospects for enhancing collision detection systems.

The European Union's (EU) Road Accident Policy Concept 2021-2030 (EU, 2020) provides an in-depth description of the EU's strategy for enhancing roadway safety during the ensuing 10 years. The strategy includes several difficult objectives, such as a 50% decrease in deaths and serious harm on EU highways by 2030. The Framework underlines the importance of using technology in achieving these objectives, such as collision detection and warning systems. The EU's approach is based on the Safe System Approach, which acknowledges that accidents are inevitable but aims to minimize their effects by improving road safety for vehicles, pedestrians, and cyclists [6]. The causes and effects of fatal collisions in the United States are thoroughly examined in the National Highway Traffic Safety Administration's (NHTSA) report on fatal motor vehicle crashes (NHTSA, 2019). The paper emphasizes the significance of lowering fatal collision rates through enhanced traffic safety measures, including the application of technology. According to the analysis, human error is a major factor in many fatal incidents, but technology may help by offering early warning systems and automated emergency braking [7]. An intelligent road vision warning system based on binocular vision was created by the authors of [21] and may be used with driverless cars to assist reduce traffic accidents.

Wang et al. [8] introduce an intelligent accident detection and warning system for urban traffic safety. Their method integrates sensors with machine learning algorithms for real-time accident detection and classification. Designed to integrate seamlessly with existing traffic management systems, the system is adaptable to various environments. Evaluation with real-world data from a major Chinese city confirms the system's high accuracy and quick response capabilities.

Zhang et al. [9] propose an intelligent vehicle collision detection system, integrating sensors with fuzzy logic for collision detection and classification. Their approach, designed for various vehicles, prioritizes affordability and ease of deployment. Evaluation with simulated collision scenarios shows high accuracy and minimal false alarms.

Loke et al. [10] utilize video processing to detect accidents from car-mounted cameras. Analyzing video frames, they identify accident-related motion patterns and trigger alerts. Their system achieves a notable 95% accuracy in accident detection, applicable to both road and vehicle cameras.

Tawari, A., Choudhury, P., et al. [11] devised a smartphone-based accident detection system utilizing data from the phone's accelerometer. This system could detect abrupt changes in motion, signaling a potential accident. Additionally, it included GPS-based position monitoring to aid emergency responders in locating the accident scene. Their work resulted in the creation of an accident detection system leveraging smartphone accelerometer data, which analyzed velocity and directional changes to detect accidents, achieving an impressive 94% accuracy rate.

Alavi, S.A.A., Ahmadi, H., Farhadi, H., et al. [12] employed machine learning algorithms to devise a crash detection system. This system, which integrated data from CCTV cameras and sensors, achieved an impressive accuracy rate of 97%.

Khan et al. [13] analyzed dashcam data using image processing to identify accidents based on motion patterns. Their system achieved 91% accuracy by integrating optical motion and edge identification algorithms, primarily analyzing dashcam images.

Li et al. [14] employed sensors like motion detectors and gyroscopes for real-time collision detection. Their system, utilizing machine learning to analyze sensor data, achieved 90% accuracy by identifying movement patterns indicative of accidents and triggering an alert mechanism.

Gupta et al. [15] proposed a real-time autonomous accident detection system using computational intelligence techniques to analyze CCTV footage. Road accidents persist as a significant and deadly issue, exacerbated by factors like speeding and distracted driving. Early accident detection offers the potential to save lives and improve overall road safety.

Thakare Kamalakar Vijay [16] presented a system where they grouped videos based on similarity. They proposed a new dataset MP-RAD, designed on a gaming platform. The 2 branch DCNN model is used for feature extraction, these extracted spatiotemporal features are then fused with rank rank-based pooling strategy which is classified later with fully interconnected layers. They annotated each frame with spatial or temporal tags. The AUC of 77.25% was achieved.

Djenouri et al. [17] employ image processing to remove noise by detecting vehicles and outliers. They propose a hybrid RESNET and regional convolutional neural network framework for accident identification, achieving superior classification accuracy. The authors optimize learning hyperparameters using evolutionary computing techniques, including genetic algorithm-based methods like crossover mutation, enhancing system efficiency.

Fang et al. [18] propose a method for traffic accident detection using self-supervised consistency learning in driving scenarios. This approach integrates appearance, motion, and context consistency learning into a self-supervised architecture. Dashboard cameras assist in establishing spatial frame relationships, while Convolutional LSTM and gated RNN classify item position and spatial-temporal consistency, achieving 67.8% accuracy for the A3D and DADA datasets.

YOLO v3, MIF model, and UFIR filter are used to create accident culpability reports in addition to identifying the speed and collision angle of wrecked cars and trajectories [19]. [20] combines a multilayer neural network and spatiotemporal feature encoding to identify accidents in a VANET environment with the aid of car-mounted cameras.

**References:**

[1] Chen, Chen, Guorun Yao, Lei Liu, Qingqi Pei, Houbing Song, and Schahram Dustdar. "A cooperative vehicle-infrastructure system for road hazards detection with edge intelligence." IEEE Transactions on Intelligent Transportation Systems (2023).

[2] Thakare, Kamalakar Vijay, Debi Prosad Dogra, Heeseung Choi, Haksub Kim, and Ig-Jae Kim. "Object Interaction-Based Localization and Description of Road Accident Events Using Deep Learning." IEEE Transactions on Intelligent Transportation Systems 23, no. 11 (2022): 20601-20613.

[3] Kukade Jyoti, Swapnil Soner, and Sagar Pandya. "Autonomous Anomaly Detection System for Crime Monitoring and Alert Generation." Journal of Automation, Mobile Robotics and Intelligent Systems 16, no. 1 (2022): 62-71.

[4] Al-Dahash, H., Al-Dabbagh, M., & Al-Rizzo, H. (2019). Vehicle collision detection and warning system: a review of the state of the art. Journal of Traffic and Transportation Engineering, 6(1), 1-18.

[5] Chen, X., Zhou, X., Qiu, Y., Zhang, X., & Gao, Y. (2019). Real-time vehicle collision detection based on deep learning. Sensors, 19(11), 2590.

[6] European Union. (2020). European Union Road Safety Policy Framework 2021-2030: Next steps towards "Vision Zero". European Commission.

[7] NHTSA. (2019). Fatal Motor Vehicle Crashes: Overview. National Highway Traffic Safety Administration.

[8] Wang, W., Wu, H., Tang, L., & Wang, L. (2020). An intelligent accident detection and warning system for urban traffic safety. Journal of Advanced Transportation, 2020.

[9] Zhang, H., Ma, S., & Hu, Z. (2019). Design and implementation of an intelligent vehicle collision detection system. IEEE Access, 7, 117574-117583.

[10] Loke, S.W., Kamarudin, L.M., & Yee, K.W. (2016). Development of an intelligent accident detection and notification system. Procedia Computer Science, 76, 414-419.

[11] Tawari, A., Choudhury, P., Mishra, D., & Mishra, S. (2019). Smartphone-based accident detection system using accelerometer sensor data. Journal of Big Data, 6(1), 34.

[12] Alavi, S.A.A., Ahmadi, H., Farhadi, H., & Mansouri, A. (2021). Intelligent vehicle-based accident detection using a combination of video and sensor data. Journal of Intelligent Transportation Systems, 25(1), 1-11.

[13] Khan, A.U., Hamid, E., Kausar, R., & Aslam, N. (2018). Automated vehicle accident detection and notification system using image processing. 2018 3rd International Conference on Communication and Information Processing (ICCIP), 1-5.

[14] Li, C., Yu, Z., & Wu, J. (2020). A novel approach to accident detection in urban transportation based on multi-sensor data fusion. Sensors, 20(19), 5496.

[15] Gupta, Sajal, Manish Rawat, and A. S. Rao. "Accident Detection and Prediction with Notification Alert System." Advances in Mechanical Engineering and Technology: Proceedings of 6th International Conference on Advanced Production and Industrial Engineering (ICAPIE)-2021. Singapore: Springer Singapore, 2022.

[16] Vijay, T.K., Dogra, D.P., Choi, H., Nam, G. and Kim, I.J., 2022. Detection of Road Accidents Using Synthetically Generated Multi-Perspective Accident Videos. IEEE Transactions on Intelligent Transportation Systems.

[17] Djenouri, Youcef, Gautam Srivastava, Djamel Djenouri, Asma Belhadi, and Jerry Chun-Wei Lin. "Hybrid RESNET and Regional Convolution Neural Network Framework for Accident Estimation in Smart Roads." IEEE Transactions on Intelligent Transportation Systems 23, no. 12 (2022): 25335-25344.

[18] Fang, Jianwu, Jiahuan Qiao, Jie Bai, Hongkai Yu, and Jianru Xue. "Traffic accident detection via self-supervised consistency learning in driving scenarios." IEEE Transactions on Intelligent Transportation Systems 23, no. 7 (2022): 9601-9614.

[19] Sui, Yifan, Shaodong Zhou, Zhiyang Ju, and Hui Zhang. "A Vision-Based System Design and Implementation for Accident Detection and Analysis via Traffic Surveillance Video." *IEEE Canadian Journal of Electrical and Computer Engineering* 45, no. 2 (2022): 171-181.

[20] Zhou, Zhili, Xiaohua Dong, Zhetao Li, Keping Yu, Chun Ding, and Yimin Yang. "Spatio-temporal feature encoding for traffic accident detection in VANET environment." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 10 (2022): 19772-19781.

[21] Han, Zidong, Junyu Liang, and Jianbang Li. "Design of intelligent road recognition and warning systems for vehicles based on binocular vision." *IEEE Access* 6 (2018): 62880-62889.

[22] <https://morth.nic.in/sites/default/files/RA_2021_Compressed.pdf>

**Fight Detection**

Kaur and Singh et al. [23] address the urgent need for intelligent surveillance systems to detect violence in public environments. They review common frameworks and deep learning techniques, particularly CNNs and LSTMs, for violence detection. Additionally, they discuss popular datasets and challenges in this domain, providing valuable insights for future research.

Yao and Hu et al. [24] present a comprehensive survey of violent behavior detection (VioBD) in intelligent video surveillance systems. They outline the fundamental principles and challenges of VioBD and categorize existing approaches into traditional, end-to-end deep learning, and hybrid deep learning frameworks. Additionally, they discuss public datasets used for evaluating VioBD methods and highlight open research problems and future trends in this field.

Mohtavipour et al. [25] propose a deep violence detection framework for video surveillance. Their approach utilizes handcrafted features fed into a convolutional neural network (CNN) across spatial, temporal, and spatiotemporal streams. Trained on diverse datasets, their CNN outperforms existing methods in accuracy and processing time.

Ullah et al. [26] offer a comprehensive review on vision-based violence detection (VD) in surveillance videos, addressing the need for automatic VD systems in crowded environments. They cover machine learning strategies, NN-based pattern analysis, dataset challenges, evaluation metrics, and future research directions in the VD literature.

Mumtaz et al. [27] review violence detection (VD) techniques in smart city surveillance, tracing the shift from manual feature engineering to deep learning-based models. They discuss localization strategies for detected violence, compare image processing and machine learning approaches with complex models, and analyze datasets to outline strengths, weaknesses, and future directions in VD.

Pang et al. [28] tackle violence detection in videos by merging visual and audio data. Their neural network includes an attention module for weighted feature generation, a fusion module for integration, and a mutual learning module. Results on the XD-Violence dataset showcase their method's superiority over existing approaches.

Omarov et al. [29] underscore the necessity of assessing violence detection model performance using multiple metrics like precision, recall, F1-score, AUC-ROC, false alarm, and missing alarm. They elaborate on each metric's significance in evaluating violence detection model effectiveness.

Irfanullah et al. [30] propose real-time violence detection in surveillance videos using CNNs. Overcoming challenges such as defining violent objects and scarcity of labeled datasets, their MobileNet model surpasses AlexNet, VGG-16, and GoogleNet in accuracy (96.66%) and loss (0.1329%). It also exhibits faster computation, proving effective in violence detection on the hockey fight dataset.

Bianculli et al. [31] address the need for automated violence detection in videos to alleviate the workload of security personnel. They introduce a dataset comprising 350 high-resolution MP4 video clips, labeled as non-violent or violent. The non-violent category intentionally includes actions prone to causing false positives. Clips feature performances by non-professional actors, varying from 2 to 4 actors per clip.

Peixoto et al. [32] propose a framework for violence detection in videos, evaluating it on the Hockey Fight and Movies datasets. They compare results with various baselines and conduct a user study to gauge human annotators' perception of violence levels. Different dCNNs are trained to detect distinct violence aspects, later combined using a tailored network for general violence classification.

Harvey et al. [33] explore five methods for video classification on UCF101, including frame-based ConvNet classification, a time-distributed ConvNet with RNN, 3D convolutional networks, and ConvNet feature extraction with separate RNN or MLP. Constraints include eliminating optical flow, subsampling to 40 frames, minimal preprocessing, and fitting models within 12 GiB GPU memory.

Bermejo et al. [34] employ STIP and MoSIFT video descriptors along with the bag-of-words framework to classify video sequences as fight or non-fight. They introduce a new dataset of hockey video clips and achieve nearly 90% accuracy on fight detection. The paper highlights MoSIFT's superior performance over STIP on action movie clips.

Peixoto et al. [35] introduce a violence detection method by dissecting it into sub-concepts like fights, explosions, blood, and gunshots. Utilizing custom CNNs, they analyze each sub-concept's traits and guide their representation. Through a decision neural network combining visual and auditory feature detectors, they craft a versatile violence detector adaptable to diverse cultural contexts and user preferences, bolstering its robustness and usability.

Vijeikis et al. [36] propose an efficient violence detection architecture for surveillance cameras, comprising a spatial feature extractor using a U-Net-like network, LSTM-based temporal feature extractor, and a dense layer classifier. Despite its computational lightness, the model achieves good results, with an average accuracy of 0.82 ± 2% and precision of 0.81 ± 3%.

Garcia-Cobo and SanMiguel et al. [37] introduce a violence detection method in surveillance videos leveraging human skeletons and change detection. They evaluate the model on three datasets, demonstrating competitive performance with accuracies of 90.25%, 88.00%, and 62.00%, respectively. The proposed model exhibits robustness and generalizability across various datasets, promising for surveillance applications.

Soliman et al. [38] introduce a deep neural network for video violence recognition, using pre-trained VGG-16 for spatial features and LSTM for temporal features. Their model achieves near state-of-the-art accuracy and introduces the Real-Life Violence Situations dataset. Fine-tuning on this dataset yields 88.2% accuracy.

**References:**

[23] Kaur, Gurmeet, and Sarbjeet Singh. "Violence Detection in Videos Using Deep Learning: A Survey." Advances in Information Communication Technology and Computing: Proceedings of AICTC 2021 (2022): 165-173.

[24] Yao, Huiling, and Xing Hu. "A survey of video violence detection." Cyber-Physical Systems 9, no. 1 (2023): 1-24.

[25] Mohtavipour, Seyed Mehdi, Mahmoud Saeidi, and Abouzar Arabsorkhi. "A multi-stream CNN for deep violence detection in video sequences using handcrafted features." The Visual Computer 38, no. 6 (2022): 2057-2072.

[26] Ullah, Fath U. Min, Mohammad S. Obaidat, Amin Ullah, Khan Muhammad, Mohammad Hijji, and Sung Wook Baik. "A comprehensive review on vision-based violence detection in surveillance videos." ACM Computing Surveys 55, no. 10 (2023): 1-44.

[27] Mumtaz, Nadia, Naveed Ejaz, Shabana Habib, Syed Muhammad Mohsin, Prayag Tiwari, Shahab S. Band, and Neeraj Kumar. "An overview of violence detection techniques: current challenges and future directions." Artificial intelligence review 56, no. 5 (2023): 4641-4666.

[28] Pang, Wen-Feng, Qian-Hua He, Yong-jian Hu, and Yan-Xiong Li. "Violence detection in videos based on fusing visual and audio information." In ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 2260-2264. IEEE, 2021.

[29] Omarov, Batyrkhan, Sergazi Narynov, Zhandos Zhumanov, Aidana Gumar, and Mariyam Khassanova. "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review." PeerJ Computer Science 8 (2022): e920.

[30] Irfanullah, Tariq Hussain, Arshad Iqbal, Bailin Yang, and Altaf Hussain. "Real time violence detection in surveillance videos using Convolutional Neural Networks." Multimedia Tools and Applications 81, no. 26 (2022): 38151-38173.

[31] Bianculli, Miriana, Nicola Falcionelli, Paolo Sernani, Selene Tomassini, Paolo Contardo, Mara Lombardi, and Aldo Franco Dragoni. "A dataset for automatic violence detection in videos." Data in brief 33 (2020): 106587.

[32] Peixoto, Bruno, Bahram Lavi, Paolo Bestagini, Zanoni Dias, and Anderson Rocha. "Multimodal violence detection in videos." In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2957-2961. IEEE, 2020.

[33] Harvey, Matt. "Five video classification methods implemented in Keras and TensorFlow." *Coastline Automation* (2017).

[34] Bermejo Nievas, Enrique, Oscar Deniz Suarez, Gloria Bueno García, and Rahul Sukthankar. "Violence detection in video using computer vision techniques." In *Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29-31, 2011, Proceedings, Part II 14*, pp. 332-339. Springer Berlin Heidelberg, 2011.

[35] Peixoto, Bruno M., Bahram Lavi, Zanoni Dias, and Anderson Rocha. "Harnessing high-level concepts, visual, and auditory features for violence detection in videos." *Journal of Visual Communication and Image Representation* 78 (2021): 103174.

[36] Vijeikis, Romas, Vidas Raudonis, and Gintaras Dervinis. "Efficient violence detection in surveillance." *Sensors* 22, no. 6 (2022): 2216.

[37] Garcia-Cobo, Guillermo, and Juan C. SanMiguel. "Human skeletons and change detection for efficient violence detection in surveillance videos." *Computer Vision and Image Understanding* 233 (2023): 103739.

[38] Soliman, Mohamed Mostafa, Mohamed Hussein Kamal, Mina Abd El-Massih Nashed, Youssef Mohamed Mostafa, Bassel Safwat Chawky, and Dina Khattab. "Violence recognition from videos using deep learning techniques." In 2019 *Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 80-85. IEEE, 2019.