**PROJECT REPORT**

AI-Driven Crowd Management Crowd and Anomaly Detection

**Introduction :**

Video anomaly detection is a critical task in computer vision, particularly in applications involving surveillance, public safety, and crowd management. With the increasing prevalence of CCTV systems, the ability to automatically detect unusual or suspicious behavior in real-time has become a pressing need. Anomaly detection systems can significantly enhance public safety by identifying events such as violence, theft, or accidents, which require immediate attention.

This project implements and attempts to enhance the methodology proposed in the research paper *"Video Anomaly Detection System Using Deep Convolutional and Recurrent Models"* by Maryam Qasim and Elena Verdu. The system leverages deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to extract both spatial and temporal features from video feeds. These features enable the accurate classification of normal and anomalous activities. Using the UCF-Crime dataset, this report documents the steps taken to develop and optimize the system, addresses challenges encountered during the implementation, and outlines future directions for achieving state-of-the-art performance.

**Key Objectives :**

* **Base Paper Implementation:** Reproducing the framework outlined in the paper to verify its functionality and accuracy, thus ensuring a strong foundation for further enhancements.
* **Performance Optimization:** Adapting and optimizing the existing model to enhance its accuracy and robustness, particularly when applied to large-scale datasets like UCF-Crime.
* **Innovative Model Design:** Exploring and integrating novel architectural improvements to improve performance metrics and generalizability.
* **Scalability and Efficiency:** Addressing computational challenges to enable real-time, scalable anomaly detection without compromising performance.

**Datasets Used :**

1. **UCF-Crime Dataset:**
   * **Description:** A comprehensive dataset featuring diverse classes of anomalies, including assault, abuse, explosion, arson, burglary, and more. It includes hundreds of videos, each annotated with specific labels indicating normal or anomalous behavior.
   * **Challenges:** The dataset’s large size and diversity present challenges in processing and model generalization. Additionally, its imbalanced class distribution requires careful handling to prevent biased predictions.

**Implementation Details :**

**Base Paper Framework:** The framework proposed in the base paper integrates CNN and SRU architectures:

* **Convolutional Neural Network (CNN):** ResNet-18/50 architectures are employed to extract high-level spatial features from individual video frames. These features capture critical visual information required for anomaly detection.
* **Simple Recurrent Unit (SRU):** Designed to efficiently capture temporal dependencies across video sequences while maintaining a lower computational overhead compared to traditional recurrent models like LSTMs or GRUs. SRUs allow the model to analyze sequences of frames to detect temporal patterns of anomalies.
* **Optimization Metrics:** Model performance is evaluated using accuracy, F1 score, and the area under the receiver operating characteristic curve (AUC), providing a comprehensive measure of classification capability.

**Steps Taken:**

1. **Dataset Preparation:**
   * Organized the UCF-Crime dataset into training, validation, and test splits to ensure balanced and unbiased evaluation.
   * Preprocessed video frames using data generators to optimize memory usage and facilitate batch processing during training.
2. **Model Development:**
   * Implemented CNN-SRU model variants, including ResNet-18 and ResNet-50, to compare their feature extraction and temporal modeling capabilities.
   * Fine-tuned pre-trained models by selectively unfreezing specific layers, allowing adaptation to the anomaly detection domain. This approach balanced the retention of learned features with domain-specific training.
   * Introduced advanced performance metrics, such as precision-recall curves, to complement AUC and F1 score evaluations.
3. **Optimization:**
   * Applied techniques such as reducing batch size, lowering learning rates, and incorporating dropout layers to mitigate overfitting.
   * Conducted extensive experiments to balance computational efficiency and model performance, ensuring scalability for large datasets.

**Challenges Encountered :**

* **Dataset Size:**
  + The UCF-Crime dataset’s size posed significant memory and computational challenges during preprocessing and training phases. This was mitigated using data generators and by video frame resizing.
* **Long Training Times:**
  + Training on large datasets required substantial computational resources. Techniques such as distributed training, optimized batch sizes, and leveraging high-performance GPUs helped address these issues.
* **Overfitting:**
  + Disparities between training and validation accuracies highlighted overfitting tendencies. Techniques such as data augmentation, reduced learning rates, early stopping, and the addition of dropout layers were employed to address this.

**Key Learnings :**

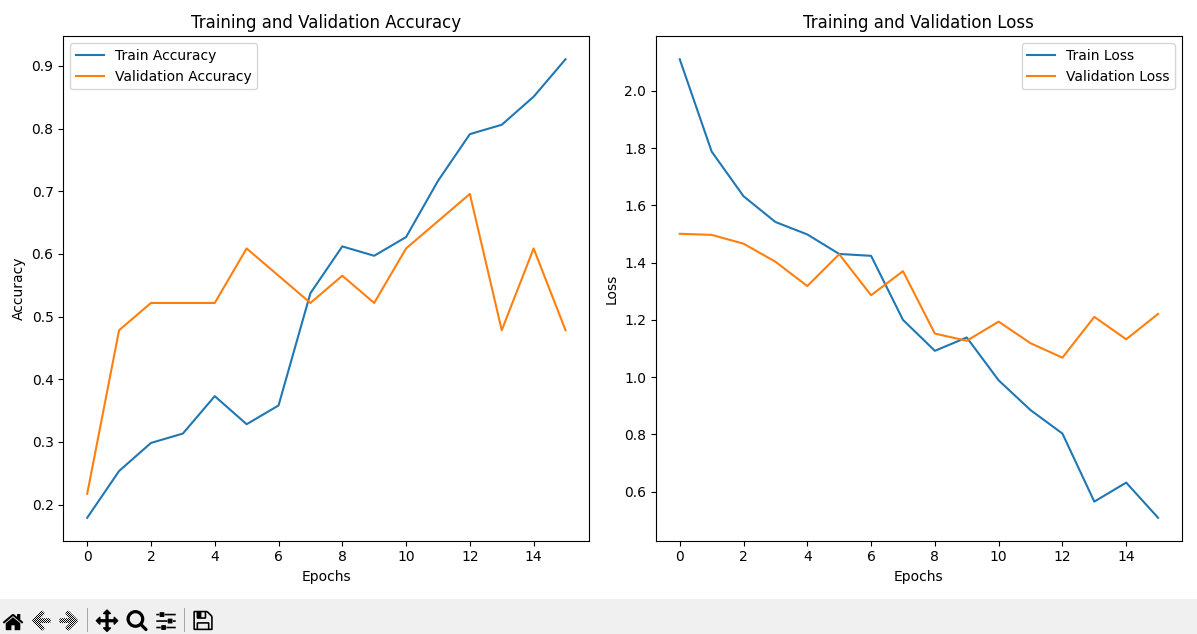
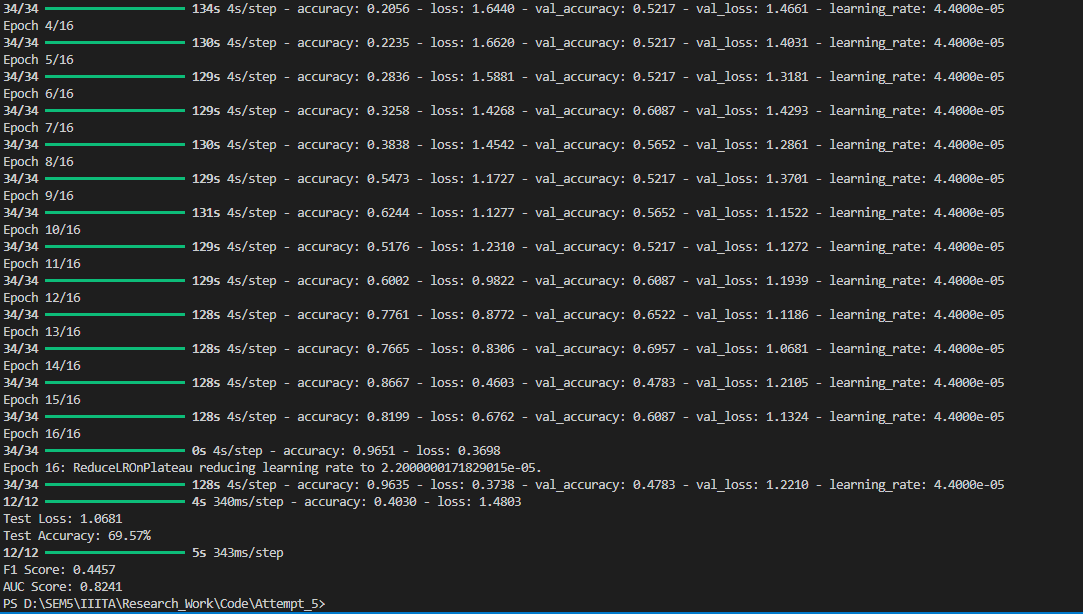
Base Model Insights (ResNet50 + SRU):

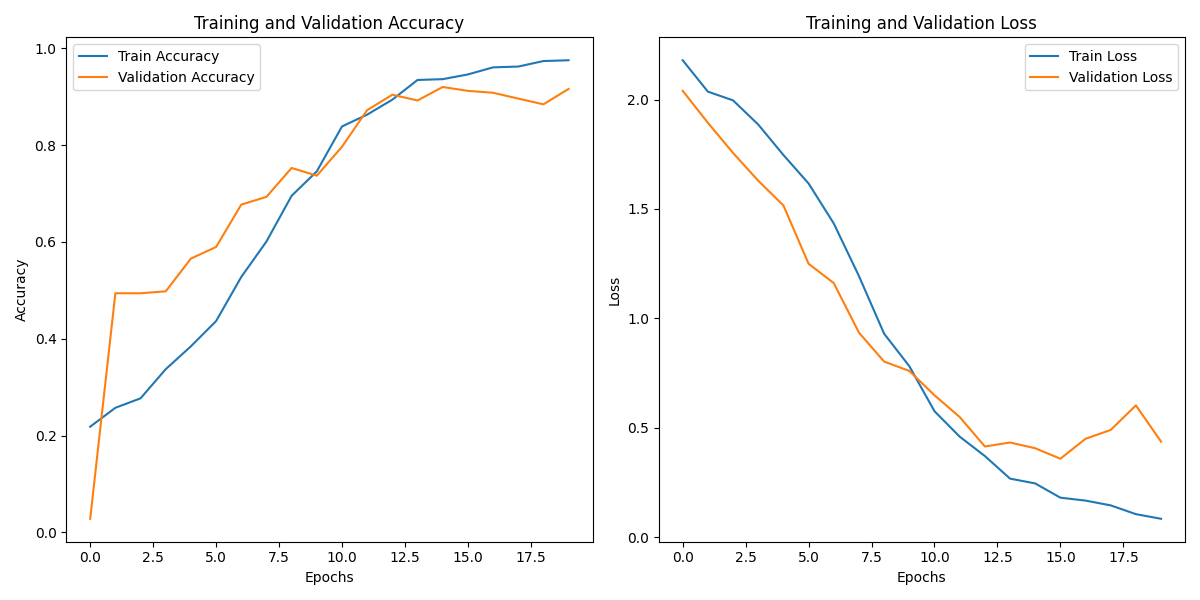
* The integration of ResNet50 for spatial feature extraction and a custom SRU layer for temporal dependency modeling demonstrated the effectiveness of combining pre-trained CNNs with lightweight RNNs.
* Unfreezing specific layers in the ResNet50 model for fine-tuning allowed the framework to adapt pre-learned features to anomaly detection, improving classification accuracy.

New Model Contributions (TimeDistributed CNN + SimpleRNN):

* Replacing SRU with SimpleRNN and introducing a TimeDistributed wrapper for convolutional layers facilitated frame-wise spatial feature extraction with explicit temporal sequence modeling. This modular design enhanced interpretability and reduced model complexity.
* The simpler architecture reduced training overhead, making it a suitable alternative for environments with limited computational resources while maintaining comparable accuracy.

**Results :**

* **Base Paper Reproduction:**
  + Successfully achieved a training accuracy of 94% and test accuracy of 69% on the UCF-Crime dataset. These results validate the base framework’s potential for anomaly detection tasks.
* **Model Variants:** 
  + TimeDistributedCNN + SimpleRNN demonstrated improved anomaly classification capabilities by efficiently combining spatial and temporal features with test accuracy of 91% and AUC score of 0.9568 on the UCF-Crime Dataset.
* **Base Model’s Graph:**
* **Base Model’s Accuracy, AUC & F1 Score:**
* **New Model’s Graph:**

****

* **New Model’s Accuracy, AUC & F1 Score:**

**Test Loss:** 0.3578

**Test Accuracy:** 91.24%

**F1 Score:** 0.7258

**AUC Score:** 0.9568

**Future Directions :**

* **Model Optimization:**
  + Implement advanced techniques such as pruning and quantization to reduce model size and improve inference speed.
* **Attention Mechanisms:**
  + Integrate spatiotemporal attention layers to better capture anomalies in complex video sequences. These layers can dynamically allocate computational focus to the most relevant regions in the video
* **New Architectures:**
  + Experiment with Transformer-based models, which excel in capturing long-range dependencies and offer state-of-the-art results in sequence modeling tasks.
* **Multi-Modal Integration:**
  + Combine audio and video data to develop a comprehensive anomaly detection framework, capable of detecting events with multimodal cues, such as screams or crashes accompanying visual anomalies.
* **Real-Time Application:**
  + Adapt the model for deployment in real-time surveillance systems by addressing latency and computational efficiency. This includes streamlining inference pipelines and leveraging edge-computing devices.

**Conclusion :**

The implementation and optimization of a video anomaly detection system using deep convolutional and recurrent models highlight the potential of hybrid architectures in addressing complex surveillance challenges. While the initial results are promising, achieving state-of-the-art performance requires further advancements in model optimization, attention mechanisms, and real-time scalability. These efforts will pave the way for robust, real-time video anomaly detection systems capable of significantly enhancing public safety and operational efficiency.

**References :**

1. Francisco Luque Sánchez, Isabelle Hupont, Siham Tabik, Francisco Herrera, “Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects”, Information Fusion, Volume 64, 2020, pp 318-335
2. Kumar, M., Patel, A.K. & Biswas, M. Real-time detection of abnormal human activity using deep learning and temporal attention mechanisms in video surveillance. *Multimed Tools Appl* **83**, 55981–55997 (2024).
3. Charan Kumar, Pc, et al. Using Existing CCTV Networks for Crowd Management and Crime Prevention using Deep Learning. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024.
4. Zhou, H., Yu, J., & Yang, W. (2023). Dual Memory Units with Uncertainty Regulation for Weakly Supervised Video Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(3), 3769-3777. <https://doi.org/10.1609/aaai.v37i3.25489>
5. Qasim, Maryam, and Elena Verdu. Video anomaly detection system using deep convolutional and recurrent models. *Results in Engineering* 18 (2023): 101026.