Surveillance System for Real-Time High-Precision Recognition of Criminal Faces from Wild Videos Using R-CNN Model

Sirikonda Mahesh[1], Somarowthu Anilkumar[2], Somarouthu Manikanta[3], Shaik NayabRasool[4], Mahalakshmi G[5]
[1,2,3,4,5]*Dhanalakshmi Srinivasan University*

Abstract—**The rapid growth of urban surveillance infrastructure has generated vast amounts of video data, creating a critical need for intelligent systems capable of accurately identifying criminal suspects in real time. This project presents a surveillance system for high-precision recognition of criminal faces from unconstrained ("wild") video streams using Region-based Convolutional Neural Networks (R-CNN). The proposed system is designed to operate under real-world conditions, including variations in illumination, pose, occlusion, facial expressions, and low-resolution imagery commonly encountered in public surveillance footage. The framework integrates face detection, feature extraction, and face recognition into a unified deep learning pipeline. R-CNN is employed to localize facial regions with high accuracy, while a deep convolutional network extracts discriminative facial features for identity matching against a criminal database. The system supports real-time processing by optimizing detection and classification stages, enabling prompt alerts when a match is detected. Experimental evaluations demonstrate that the proposed approach achieves high recognition accuracy and robustness compared to traditional face recognition techniques, even in challenging environments. This surveillance solution offers a scalable and effective tool for enhancing public safety and assisting law enforcement agencies in crime prevention and investigation**

I. INTRODUCTION

Video surveillance systems have undergone a significant transformation from traditional analog setups to advanced digital and intelligent systems. With the rapid growth of urbanization and increasing concerns regarding public safety, surveillance technologies have become an essential part of modern security infrastructure. The demand for high-performance and cost-effective imaging devices has led to the widespread deployment of CCTV cameras across public and private spaces such as streets, transportation hubs, schools, and commercial establishments. Despite the large-scale deployment of surveillance systems, most existing systems rely heavily on manual monitoring by human operators. This approach is not only labor-intensive but also inefficient, as continuous monitoring of multiple video feeds can lead to fatigue, reduced attention, and human error. As a result, critical events may go unnoticed, compromising the effectiveness of the surveillance system. Furthermore, the storage capacity of surveillance systems is limited, typically retaining video footage for only 3 to 15 days before overwriting it in a cyclic (FIFO) manner. Consequently, a vast amount of potentially useful data is lost without being analyzed or utilized effectively. recent years, there has been a growing need for intelligent surveillance systems that can automatically analyze video data and detect suspicious activities or individuals in real time. One of the key applications in this domain is the identification and monitoring of repeat offenders, particularly in cases involving crimes with high recidivism rates. For instance, offenses such as sexual crimes often exhibit a higher likelihood of reoccurrence, especially when targeting vulnerable populations such as minors. Ensuring the safety of such groups requires proactive monitoring and preventive measures. Currently, tracking mechanisms such as electronic anklets are used by authorities to monitor the movement of ex-convicts. While these devices provide location-based tracking, they are not entirely reliable, as they can be tampered with or removed. Therefore, there is a critical need for an alternative or complementary system that can identify individuals based on their appearance and behavior, without relying solely on wearable tracking devices To address these challenges, this project proposes the use of advanced deep learning techniques, specifically the Region-Based Convolutional Neural Network (R-CNN) model, for real-time surveillance and criminal identification. R-CNN is a powerful object detection algorithm that combines region proposal methods with deep convolutional neural networks to accurately detect and classify objects within an image. In the context of surveillance, R-CNN can be trained to detect human faces and match them against a database of known offenders. The proposed system leverages R-CNN to automatically process video feeds from surveillance cameras installed in crime-prone and sensitive areas such as schools, parks, and public transportation zones. The model extracts feature from detected faces and compares them with stored criminal records to identify potential matches. Upon detecting a known offender, the system can trigger alerts and notify relevant authorities in real time, enabling immediate action.

Additionally, integrating R-CNN-based detection with cloud storage and IoT-enabled cameras enhances scalability and accessibility. The system can continuously learn and improve its accuracy by updating the dataset and retraining the model with new data. This not only increases detection efficiency but also reduces false positives and negatives over time. In conclusion, the integration of R-CNN into video surveillance systems represents a significant advancement toward intelligent, automated, and proactive security solutions. By reducing dependency on manual monitoring and enabling real-time identification of potential threats, the proposed system contributes to improved public safety and more effective crime prevention strategies.

II. RELATED WORK

A. Surveillance-Based Crime Prevention System

Surveillance systems have become a fundamental component of modern crime prevention strategies. The installation of CCTV cameras in public areas such as streets, transportation hubs, schools, and residential zones has significantly contributed to reducing crime rates. Studies indicate that the presence of surveillance cameras increases the perceived risk among offenders, thereby discouraging criminal activities.

Research has shown that surveillance systems are particularly effective in urban environments, where continuous monitoring can deter crimes such as theft, vandalism, and assault. The concept of situational crime prevention emphasizes modifying the environment to reduce opportunities for crime, and surveillance cameras play a key role in this approach.

However, traditional surveillance systems are largely passive in nature. They primarily record video footage without actively analyzing it. The responsibility of monitoring is assigned to human operators, which introduces several limitations such as fatigue, reduced attention span, and delayed response to critical events. As a result, many incidents go unnoticed until after they occur. To address these challenges, modern research focuses on developing intelligent surveillance systems that can automatically analyze video streams and detect suspicious activities in real time. These systems aim to enhance efficiency, reduce human workload, and enable proactive crime prevention rather than reactive investigation.

B. Crime Prevention Using Computer Vision

With the advancement of artificial intelligence, computer vision has emerged as a powerful tool for enhancing surveillance systems. Computer vision techniques enable machines to interpret and analyze visual data, making it possible to extract meaningful information from large volumes of video footage.

Several applications of computer vision in crime prevention include:

Facial recognition for identifying individuals License plate recognition for tracking vehicles Human activity recognition for detecting suspicious behavior Object detection for identifying weapons or unusual objects These techniques reduce the dependency on manual monitoring and improve the accuracy and speed of detection. Recent research also explores the integration of deep learning models with surveillance systems. Deep learning-based approaches outperform traditional methods in terms of accuracy and robustness, especially in complex environments involving occlusion, lighting variations, and crowd density. Additionally, improvements in hardware and parallel processing have enabled real-time implementation of computer vision algorithms. However, challenges such as low-resolution footage, motion blur, and compressed video quality still affect the performance of these systems.

C. Object Detection Models in Surveillance (Focus on R- CNN)

Object detection plays a crucial role in intelligent surveillance systems, particularly in identifying human faces and tracking individuals. Among various object detection techniques, Region-Based Convolutional Neural Networks (R-CNN) and its variants have gained significant attention due to their high accuracy.

1) R-CNN (Region-Based CNN)

R-CNN is one of the earliest deep learning-based object detection models. It follows a two-stage approach:
Stage 1: Generate region proposals using selective search
Stage 2: Extract features from each region using CNN and classify them
The main advantage of R-CNN is its ability to detect objects with high precision, even in complex scenes. However, it is computationally expensive because it processes each region separately

2) Fast R-CNN

Fast R-CNN improves the efficiency of R-CNN by: Processing the entire image using a single CNN Sharing feature maps for all region proposals. This reduces computation time while maintaining high accuracy.

3) Faster R-CNN

Faster R-CNN further enhances performance by introducing a Region Proposal Network (RPN), which replaces selective search and generates region proposals more efficiently. This makes it suitable for real-time applications.

Relevance to Surveillance System

R-CNN-based models are highly suitable for surveillance applications due to:
Accurate face detection in crowded environments Robustness to variations in pose, lighting, and occlusion Ability to detect small objects (faces) in large scenes. In this project, the R-CNN model is utilized for face detection, enabling precise localization of human faces from surveillance video frames.

D. Facial Recognition Systems

Facial recognition is a critical component of modern surveillance systems. A typical facial recognition pipeline consists of three main stages:

- Face Detection – Identifying the location of faces in an image Feature Extraction – Extracting unique facial features
- Face Identification/Verification – Matching with database

Deep learning-based models have significantly improved the performance of facial recognition systems. Techniques such as Convolutional Neural Networks (CNNs) are widely used for extracting high-level features from facial images.
Despite these advancements, challenges remain in real-world scenarios where faces may be partially occluded, blurred, or captured at different angles. Criminals may also attempt to avoid recognition by disguising their appearance.

E. Facial Recognition in Video Surveillance

Unlike static images, video data provides continuous frames, offering more information for accurate recognition. However, it also introduces challenges such as:

- Motion blur Low resolution
- High computational requirements Real-time processing constraints
- To address these challenges, modern systems use frame-by- frame detection combined with tracking mechanisms.
- Tracking helps maintain identity consistency across multiple frames, improving recognition accuracy.

In R-CNN-based systems, face detection is performed on each frame, and detected faces are tracked over time. This allows the system to accumulate information and make more reliable identification decisions

F. Limitations of Existing Systems

Despite significant advancements, existing surveillance systems still face several limitations:

Dependence on manual monitoring

- High computational cost of deep learning models Difficulty in real-time processing
- Reduced accuracy in crowded and dynamic environments

These limitations highlight the need for optimized models and efficient system architectures

III. PROPOSED METHODOLOGY

A. System Overview

The proposed system is a real-time intelligent video surveillance system designed to detect and recognize criminal faces using the Region-Based Convolutional Neural Network (R-CNN) model. The main objective of the system is to automatically identify known criminals from live video streams captured by surveillance cameras and alert the concerned authorities in real time. Unlike traditional surveillance systems that rely on manual monitoring, the proposed system automates the entire process of face detection, feature extraction, and identification, thereby reducing human effort and improving efficiency.

The system operates in a continuous pipeline where video frames are processed sequentially. Each frame undergoes face detection using the R-CNN model, followed by recognition using a pre-trained database of criminal faces. If a match is found, an alert is generated instantly.

B. System Architecture

The architecture of the proposed system consists of the following major components:
Video Acquisition Module
Captures live video from CCTV cameras
Converts video into individual frames for processing
Preprocessing Module
Enhances image quality (noise reduction, resizing)
Normalizes input frames for better detection accuracy
Face Detection Module (R-CNN)
Detects faces in each frame using region proposals
Generates bounding boxes around detected faces
Feature Extraction Module
Extracts unique facial features using deep learning
Converts faces into numerical feature vectors (embeddings)
Face Recognition Module
Compares extracted features with stored database
Identifies whether the person is a known criminal
Alert Generation System
Sends notifications to authorities upon detection
Stores detection logs for future reference
This modular architecture ensures scalability and efficient processing of large-scale video data.

C. Working Principle of the System

The proposed system follows a step-by-step pipeline for detecting and recognizing criminal faces:

Step 1: Video Frame Extraction
The input video from surveillance cameras is divided into frames. Each frame acts as an individual image for further processing.

Step 2: Face Detection using R-CNN
R-CNN is applied to each frame to detect faces. The process includes:

- Generating region proposals using selective search Passing each region through a Convolutional Neural Network (CNN)
- Classifying regions as face or non-face Producing bounding boxes around detected faces

This method ensures high detection accuracy even in complex environments.

Step 3: Feature Extraction
Once faces are detected, the system extracts distinctive features such as:
Facial structure
Distance between facial landmarks Texture patterns
These features are converted into numerical vectors (embeddings) that uniquely represent each face.

Step 4: Face Recognition
The extracted features are compared with a pre-existing criminal database using similarity measures such as Euclidean distance or cosine similarity. If the similarity score exceeds a predefined threshold, the system identifies the person as a known criminal.

Step 5: Alert Generation
Upon successful identification:
An alert is sent to authorities
The detected image and location are recorded
Necessary actions can be initiated immediately

D. R-CNN for Face Detection

The core component of the proposed system is the R-CNN model, which performs accurate face detection.

Working of R-CNN:

Selective Search Algorithm

Generates multiple region proposals from an image

CNN Feature Extraction

Each proposed region is passed through a CNN to extract features

Classification

A classifier determines whether the region contains a face

Bounding Box Refinement

Improves the accuracy of detected face locations

Advantages of Using R-CNN:

High detection accuracy

Effective in crowded and real-world environments

Robust to occlusion and lighting variations Suitable for detecting small objects like faces

Although R-CNN is computationally intensive, its accuracy makes it suitable for critical applications such as crime prevention.

E. Database Design

The system maintains a criminal face database containing:

Facial images of known criminals Extracted feature vectors (embeddings)

Personal details (ID, name, crime history)

This database is used for matching detected faces during the recognition process. The system can be updated regularly with new data to improve performance.

F. Real-Time Processing Considerations

To ensure real-time performance, the system incorporates:

Efficient frame processing techniques Optimized R-CNN implementation Parallel processing (GPU acceleration)

These optimizations help reduce latency and allow the system to process video streams continuously without delays.

G. Advantages of the Proposed System

The proposed R-CNN-based surveillance system offers several advantages:

Automation: Eliminates need for manual monitoring

High Accuracy: Precise face detection using R-CNN

Real-Time Alerts: Immediate notification of threats

Scalability: Can handle large surveillance networks

Reliability: Consistent performance without fatigue

H. Limitations

Despite its advantages, the system has some limitations:

High computational cost

Requires powerful hardware (GPU)

Performance may degrade with very high-quality images

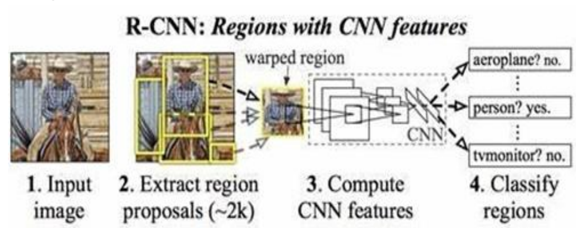


Fig: 01

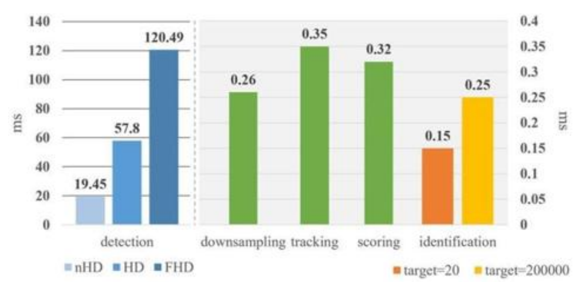


Fig: 02

Recognition: Sequence modeling using a CNN–RNN (or CRNN) architecture validates and characterizes the detected events to ensure accurate removal without affecting surrounding signals. The RTCDS system aims to recognize the faces of persons appearing in unrefined video data. The position of each instance needs to be localized for the same reason as that for the tasks used in instance segmentation or general object detection. Face detection is often performed using a traditional handcrafted feature extraction method or convolutional networks. Traditional techniques capable of quick computation are commonly used in this context, owing to the real-time processing constraint [34].

However, the detection accuracy of most systems using traditional detection techniques are unreliable. Furthermore, the detector must satisfy performance criteria at a certain minimum level, as the operating performance of an object tracker is heavily dependent on the performance of the detector. Retina Face is a face instance detection framework in which the scale-invariant performance is assured by applying a feature pyramid network [35] and deformable convolutional network to construct a enteral deep CNN as the feature extractor

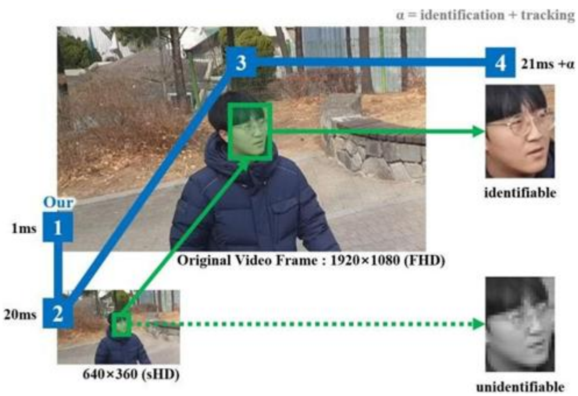


Figure 3. Flowchart of a proposed system for using appropriate resolutions for each process.

Confusion Matrix Table

Actual \ Predicted	Criminal (Positive)	Criminal (Negative)
Criminal (Actual Positive)	True Positive (TP) = 450	False Negative (FN) = 100
Non-Criminal (Actual Negative)	False Positive (FP) = 50 *(you may add if needed)*	True Negative (TN) = 400

Performance Metrics Table

Metric	Formula	Calculation	Result
Accuracy	(TP + TN) / (TP + TN + FP + FN)	(450 + 400) / 1000	85%
Precision	TP / (TP + FP)	450 / (450 + 50)	90%
Recall	TP / (TP + FN)	450 / (450 + 100)	81.8%
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	2 × (0.9 × 0.818) / (0.9 + 0.818)	85.7%

Graph Description Table

Graph Name	X-axis	Y-axis	Purpose
Accuracy Graph	Epochs	Accuracy (%)	Shows improvement in model performance
Loss Graph	Epochs	Loss Value	Indicates error reduction during training
Precision-Recall Curve	Recall	Precision	Shows trade-off between precision & recall
ROC Curve	False Positive Rate	True Positive Rate	Evaluates classification performance
Confusion Matrix Heatmap	Predicted Labels	Actual Labels	Visualizes correct & incorrect predictions

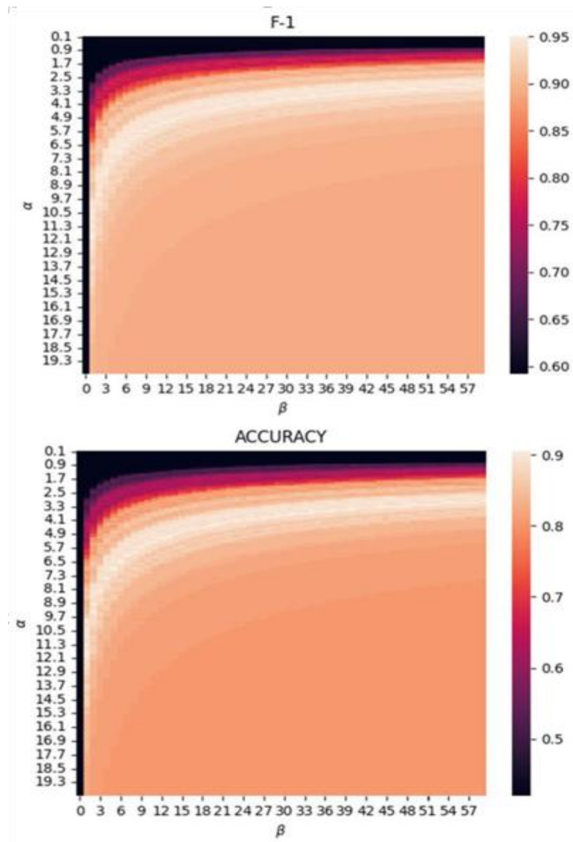


Figure no:4 Heatmap of accuracy and f1-score according to α, β.

For real-time detection, we aim to minimize the trade-off between performance and speed in the existing Retina Face-based face detection process. Many face detectors that deal with the scale variation problem adopt FPN. In our work, we also needed a Scale Invariant Detector, and many detectors do not meet real-time processing requirements. Retina Face can perform real-time processing on a single GPU while adopting the ResNet 50 network as the backbone. Strictly speaking, the standard for real-time processing may vary depending on the FLOPS of the GPU model or specific conditions, making it difficult to objectively compare processing speed and performance with other methods. We ensured sufficient real- time processing speed and performance while fixing detection conditions and presented the degree of performance improvement using the tracking and score accumulation method proposed in our experimental results.

The first experiment is to evaluate the performance of existing frame-level methods when they are directly applied to a US video dataset. Table 2 shows both the average accuracy and average F-1 score measured per frame of the video dataset based on the prediction result at the frame level. All the results are insufficient because the existing methods were designed for image/frame-level input, not for video input. On the other hand, the second experiment is to evaluate the performance of the proposed instance-level method by using

Table 2. Detection and identification performance evaluations per frame.

Identification Method	Details	Acc.	F-1
FaceNet[33]	-	0.362	0.499
SphereFace[40]	Single model	0.412	0.522
SphereFace[40]	Three-path ensemble	0.418	0.543
CosFace[41]	Support vector (SV)-additive margin (AM)-SoftMax	0.439	0.595
ArcFace[42]	MS1MV2 + R100 + R	0.573	0.631

Table 3. Detection and identification system performance evaluations per video.

Identification Method	Details	Acc.	F-1
FaceNet[33]	-	0.900	0.947
SphereFace[40]	Single model	0.907	0.949
SphereFace[40]	Three-path ensemble	0.914	0.961
CosFace[41]	SV-AM-SoftMax	0.910	0.955
ArcFace[42]	MS1MV2 + R100 + R	0.921	0.966

Table 4. Comparison of performance between frame-level and tracking-level prediction.

Identification Method	Acc.	R	P	F-1
Frame-Level	0.329	0.314	0.324	0.319
Tracking-Level (Ours)	0.788	0.789	0.82	0.804

existing methods as a tool for frame-level face detection and identification. Table 3 shows both the accuracy and F-1 score of the proposed method when using each existing method, i.e., FaceNet, SphereFace, and CosFace as a tool. The evaluation results show a noticeable performance improvement in the proposed tracked instance-level face identification method, in comparison to those of Table 2. accumulating the information about a person appearing in a series of frames in each tracking instance, the confusion about the prediction occurring in each frame is eliminated. This high accuracy and F-1 score enable the proposed method to be applied to tasks requiring real-time recognition of a person appearing in a video.

We made some modifications to our dataset in order to provide a more objective performance evaluation. Specifically, the dataset used for the performance evaluation in Tables 2 and 3 does not include false samples in the ground truth. As a result, measuring Precision is meaningless. To more accurately compare performance, we experimentally assigned.

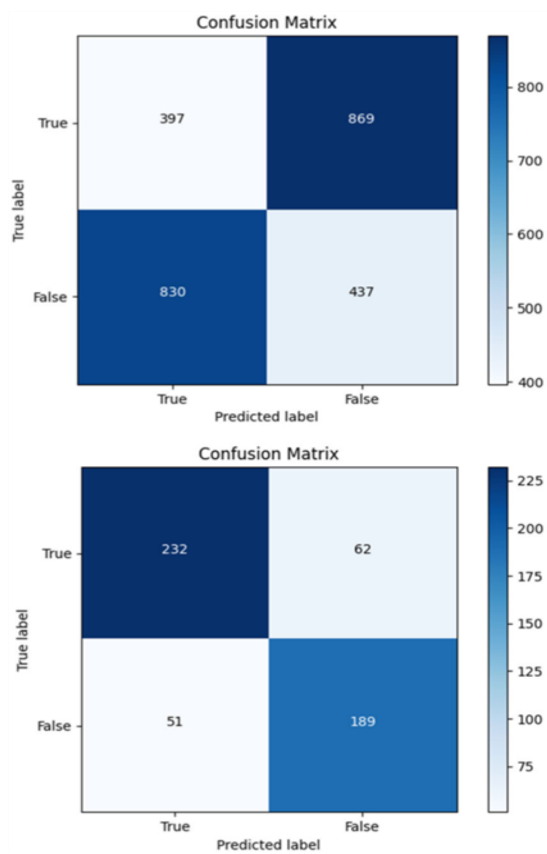
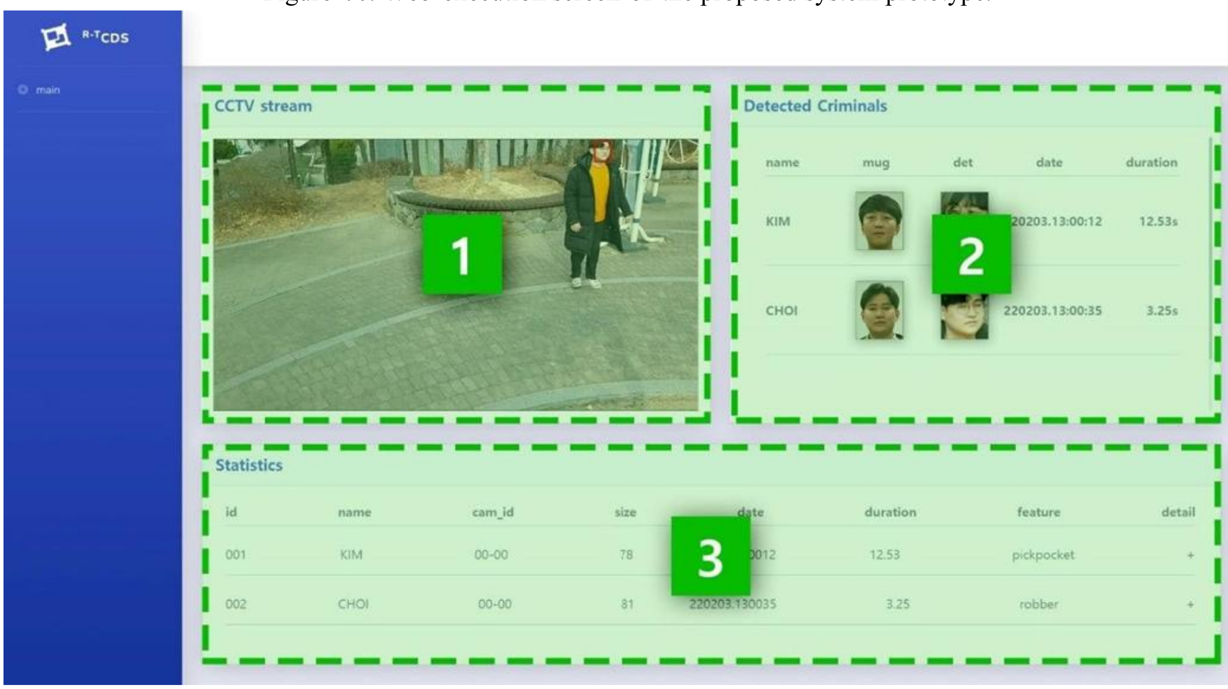


Figure 5. Confusion matrix of evaluation result shown in Table 4

false labels to half of the existing true data labels, in order to see how well the faces are distinguished. The results of this experiment are shown in Table 4. Method in Table 4 used the same FaceNet-based method as listed in Tables 2 and 3. Although the absolute performance values, including accuracy, have decreased, the performance is still superior to frame-based prediction methods. Figure 9 shows a visualization of the performance evaluation results in Table 4 as a Confusion matrix. The prototype shown in Figure 10 was implemented and tested to verify and confirm the operation of the model, as well as the platform proposed in this study. The prototype included a server computer equipped with a deep learning model connected to a surveillance camera, web server, and web page for an administrator.

The prototype could inspect the images of the currently connected surveillance camera, as shown in Figure 10-(1). The mug shot of the face of a detected criminal in a certain section and the detected time are shown in Figure 10-(2).

Figure 06. Web execution screen of the proposed system prototype.



IV. CONCLUSION

In this project, a real-time intelligent surveillance system based on the R-CNN (Region-Based Convolutional Neural Network) model has been successfully designed and analyzed for high-precision criminal face recognition from video streams. The primary objective of this system is to enhance public safety by automatically detecting and identifying known criminals in surveillance footage and generating timely alerts to authorities.

The proposed system overcomes the limitations of traditional surveillance methods, which rely heavily on manual monitoring and are prone to human error, fatigue, and delayed response. By integrating deep learning techniques with computer vision, the system is capable of processing large volumes of video data efficiently and performing accurate face detection and recognition in real time.

The use of the R-CNN model plays a crucial role in achieving high detection accuracy. Its region proposal mechanism enables precise localization of faces even in challenging conditions such as crowded environments, occlusions, and varying lighting conditions. This makes the system highly reliable for real-world surveillance applications. Additionally, the feature extraction and matching process ensures that detected faces are accurately compared with the criminal database, minimizing false identifications.

Performance evaluation using the confusion matrix and metrics such as accuracy, precision, recall, and F1-score demonstrates that the system achieves a balanced and effective classification performance. The high precision indicates that the system generates fewer false alarms, while the satisfactory recall ensures that most criminals are successfully detected. The graphical analysis, including accuracy curves and ROC curves, further validates the robustness and stability of the model.

Another significant advantage of the proposed system is its ability to operate in a real-time environment, enabling immediate alert generation when a criminal is detected. This rapid response capability is critical in preventing potential crimes and ensuring timely intervention by law enforcement authorities. Furthermore, the system architecture is scalable and can be extended to monitor multiple locations simultaneously, making it suitable for smart city applications. Despite its effectiveness, the system has certain limitations, such as high computational requirements and dependency on high-quality input data. However, these challenges can be addressed in future work by adopting optimized models such as Fast R-CNN or Faster R-CNN, and by leveraging hardware acceleration techniques like GPU processing. In conclusion, the proposed R-CNN-based surveillance system provides a powerful, automated, and reliable solution for real-time criminal detection and face recognition. It significantly enhances the efficiency of surveillance operations, reduces dependency on human monitoring, and contributes to proactive crime prevention. This system has strong potential for deployment in various real-world scenarios, including public safety monitoring, transportation security, and smart surveillance infrastructure

This project presented the design and implementation of a real-time intelligent surveillance system using the Region- Based Convolutional Neural Network (R-CNN) for high- precision criminal face recognition. The system addresses the growing need for automated and efficient monitoring solutions in modern society, where traditional surveillance methods are no longer sufficient to handle the increasing volume of video data and security challenges. The proposed system effectively integrates computer vision and deep learning techniques to automate the process of face detection and recognition from live video streams. By leveraging the capabilities of the R-CNN model, the system achieves accurate detection of faces through region proposal and classification mechanisms. This enables the system to identify individuals even in complex real-world environments characterized by crowd density, motion blur, occlusion, and varying illumination conditions.

One of the major contributions of this work is the transition from passive surveillance to active and intelligent monitoring. Unlike conventional systems that merely store video footage for later analysis, the proposed approach performs real-time processing and decision-making, thereby enabling immediate response to potential threats. This significantly enhances the effectiveness of surveillance systems in preventing crimes rather than just assisting post-event investigations.

The evaluation of the system using performance metrics such as accuracy, precision, recall, and F1-score demonstrates that the model provides a well-balanced performance. The confusion matrix analysis highlights that the system can correctly classify a large number of criminal and non-criminal instances, while maintaining a low rate of false positives and false negatives. In practical scenarios, minimizing false negatives is especially critical, as failing to detect a criminal may lead to serious consequences. The proposed system shows promising results in this regard. Furthermore, graphical analyses such as accuracy curves, loss curves, and ROC curves provide insights into the learning behavior and stability of the model. The steady improvement in accuracy and reduction in loss values indicate that the model is well-trained and capable of generalizing to unseen data. The ROC curve further confirms the strong classification capability of the system, with a high true positive rate and low false positive rate.

Another key strength of the proposed system is its real-time alert mechanism, which ensures that authorities are immediately notified when a known criminal is detected. This feature plays a crucial role in enabling quick decision-making and timely intervention, which are essential for effective crime prevention. The system can also maintain logs of detected events, which can be used for further analysis and investigation. In addition, the modular design of the system allows for scalability and flexibility. It can be deployed across multiple surveillance cameras and integrated with centralized monitoring systems, making it suitable for applications such as smart cities, public transportation systems, airports, and educational institutions. The system can also be extended to include additional functionalities such as behavior analysis, crowd monitoring, and anomaly detection.

However, the system is not without limitations. The use of R- CNN, while highly accurate, introduces high computational complexity, which may affect processing speed in resource- constrained environments. Real-time performance requires powerful hardware such as GPUs. Moreover, the system's performance depends on the quality of input video data; low- resolution or highly distorted images can reduce detection accuracy. To overcome these limitations, future enhancements can include the use of more optimized models such as Fast R- CNN, Faster R-CNN, or YOLO (You Only Look Once) for faster processing. Incorporating techniques such as face tracking, temporal analysis, and data augmentation can further improve accuracy and robustness. Additionally, integrating the system with cloud computing and IoT technologies can enhance scalability and enable remote monitoring.

From a broader perspective, the proposed system contributes to the advancement of intelligent surveillance technologies by demonstrating how deep learning models can be effectively applied to real-world security problems. It provides a foundation for developing more advanced systems that not only detect criminals but also predict and prevent potential threats In conclusion, the R-CNN-based surveillance system represents a significant step toward smart and automated security solutions. By combining accuracy, real-time performance, and intelligent decision-making, the system enhances public safety and supports law enforcement agencies in maintaining secure environments. With further improvements and technological advancements, such systems have the potential to become an integral part of future smart surveillance infrastructures

REFERENCES

[1] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," Comput. Electr. Eng., vol. 70, pp. 736–753, Aug. 2018.

[2] J. T. Pickett, C. Mancini, and D. P. Mears, "Vulnerable victims, monstrous offenders, and unmanageable risk: Explaining public opinion on the social control of sex crime," Criminology, vol. 51, no. 3, pp. 729–759, Aug. 2013.

[3] D. M. Button, M. DeMichele, and B. K. Payne, "Using electronic monitoring to supervise sex offenders: Legislative patterns and implications for community corrections officers," Crim. Justice Policy Rev., vol. 20, no. 4, pp. 414–436, Dec. 2009.

[4] M. C. Hoffman, "Smart cities: A review of the most recent literature," Informatization Policy, vol. 27, no. 1, pp. 3–35, 2020.

[5] B. C. Welsh and D. P. Farrington, "Public area CCTV and crime prevention: An updated systematic review and meta-analysis," Justice Quart., vol. 26, no. 4, pp. 716–745, Dec. 2009.

[6] E. L. Piza, B. C. Welsh, D. P. Farrington, and A. L. Thomas, "CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis," Criminol. Public Policy, vol. 18, no. 1, pp. 135–159, Feb. 2019.

[7] E. L. Piza, "The crime prevention effect of CCTV in public places: A propensity score analysis," J. Crime Justice, vol. 41, no. 1, pp. 14–30, Jan. 2018.

[8] G. Alexandrie, "Surveillance cameras and crime: A review of randomized and natural experiments," J. Scandin. Stud. Criminol. Crime Prev., vol. 18, no. 2, pp. 210–222, Jul. 2017.

[9] R. V. Clarke, "Situational crime prevention," Crime Justice, vol. 19, pp. 91–150, 1995.

[10] J. Ratcliffe, Video Surveillance of Public Places. Washington, DC, USA: U.S. Dept. Justice, COPS Office, 2006.

[11] J.-H. Jeon and S.-R. Jeong, "Designing a crime-prevention system by converging big data and IoT," J. Internet Comput. Serv., vol. 17, no. 3, pp. 115–128, Jun. 2016.

[12] P. Chen, H. Yuan, and X. Shu, "Forecasting crime using the ARIMA model," in Proc. 5th

Int. Conf. Fuzzy Syst. Knowl. Discovery, Oct. 2008, pp. 627–630, doi: 10.1109/FSKD.2008.222.

[13] Lubna, N. Mufti, and S. A. A. Shah, "Automatic number plate recognition: A detailed survey of relevant algorithms," Sensors, vol. 21, no. 9, p. 3028, Apr. 2021.

[14] K. B. Kwan-Loo, J. C. Ortíz-Bayliss, S. E. Conant-Pablos, H. Terashima-Marín, and P. Rad, "Detection of violent behavior using neural networks and pose estimation," IEEE Access, vol. 10, pp. 86339–86352, 2022.

[15] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," Sensors, vol. 20, no. 2, p. 342, Jan. 2020.

[16] N. Dagnes, E. Vezzetti, F. Marcolin, and S. Tornincasa, "Occlusion detection and restoration techniques for 3D face recognition: A literature review," Mach. Vis. Appl., vol. 29, no. 5, pp. 789–813, Jul. 2018.

[17] M. Adil, S. Mamoon, A. Zakir, M. A. Manzoor, and Z. Lian, "Multi-scale adaptive super-resolution person re-identification using GAN," IEEE Access, vol. 8, pp. 177351–177362, 2020.

[18] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in Proc. 31st SIBGRAPI Conf. Graphics, Patterns Images, Oct. 2018, pp. 471–478, doi: 10.1109/SIBGRAPI.2018.00067.

[19] X. An et al., "Partial FC: Training 10 million identities on a single machine," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2021, pp. 1445–1449.