

Integration of YOLO with Darknet: An Advanced Intelligent Video Image Processing and Monitoring Control System for Banking Security

R. Deeptha¹, Sathyapriya SB², Agnel Joshua Raj D³, Tarun M⁴

^{1,2,3,4}Department of Information Technology, SRM Institute of Science and Technology,
Ramapuram campus, Chennai

Email-id: [1deepthar@srmist.edu.in](mailto:deepthar@srmist.edu.in), [2ss4131@srmist.edu.in](mailto:ss4131@srmist.edu.in), [3ad1971@srmist.edu.in](mailto:ad1971@srmist.edu.in), [4tm9692@srmist.edu.in](mailto:tm9692@srmist.edu.in)

ABSTRACT

Banking security remains a critical challenge, with traditional video surveillance systems heavily reliant on human operators who can realistically monitor only 4-6 video feeds before attention fatigue sets in. This research explores the integration of YOLO (You Only Look Once) with the Darknet neural network framework to develop an intelligent video surveillance system specifically designed for banking environments. The system leverages YOLO's real-time object detection capabilities to enable efficient monitoring of bank premises, while Darknet provides the underlying neural network architecture for robust performance. We implemented YOLOv3 and trained it on the COCO dataset, achieving a mean Average Precision (mAP) of 0.76 on our test set. The system demonstrates superior speed compared to traditional two-stage detectors like Faster R-CNN, processing video in real-time while maintaining competitive accuracy. Our approach addresses the critical need for instantaneous and precise threat detection in banking environments, with careful consideration of privacy and ethical implications.

Key Words - Banking Security, YOLOv3, Real-time Object Detection, Video Surveillance, Deep Learning.

1. INTRODUCTION

Walk into any bank and you'll see cameras everywhere. But here's the problem: those cameras are only as good as the tired security guard watching 12 screens at once. Studies show that after just 20 minutes of continuous monitoring, detection accuracy drops below 45% [1]. We've all been there. Try focusing on multiple things simultaneously and see how long you last.

Traditional video surveillance methods in banking environments have several inherent limitations. They require constant human attention, are prone to delayed threat detection, and struggle with simultaneous monitoring of multiple areas. Moreover, the sheer volume of video data generated makes retrospective analysis time-consuming and often impractical. As banking operations expand and security threats evolve, there's a clear need for intelligent automated systems that can augment human capabilities.

Recent advances in deep learning and computer vision offer promising solutions. Object detection algorithms, particularly single-stage detectors like YOLO (You Only Look Once), have demonstrated remarkable capabilities in identifying and localizing objects within images and video streams in real-time. However, most existing research focuses on general-purpose object detection rather than specialized security applications with domain-specific requirements.

Our Approach

This research introduces an intelligent video surveillance system that integrates YOLO with the Darknet neural network framework, specifically designed for banking security applications. Our key contributions include:

1. **System Architecture Design:** We developed a comprehensive system architecture that combines video feed acquisition, preprocessing, real-time object detection using YOLOv3, object tracking, and alert management, all optimized for banking security scenarios.
2. **Implementation and Evaluation:** We implemented YOLOv3 using the Darknet framework and trained it on the COCO dataset, achieving 0.76 mAP while maintaining real-time processing speeds. We compared our approach against established baselines including Faster R-CNN and SSD.
3. **Banking-Specific Framework:** While our implementation uses standard datasets, we designed the system architecture with banking security requirements in mind, including modules for threat detection, alert management, and privacy-preserving features.

4. **Practical Considerations:** We address critical deployment considerations including privacy compliance, system scalability, and integration with existing banking infrastructure.

The system demonstrates that YOLO's real-time capabilities, combined with Darknet's efficient neural network implementation, can provide a solid foundation for intelligent banking security systems. While further domain-specific optimization would be beneficial, our work establishes a baseline and framework for future research in this area.

2. RELATED WORK

2.1 AUTONOMOUS DRIVING AND RISK ASSESSMENT

Interestingly, some of the most relevant work for security applications comes from the autonomous vehicle domain. Ryan et al. [1] proposed an innovative approach for risk analysis in autonomous driving using behavioural anomaly detection. Their work uses Convolutional Neural Networks (CNNs) to analyze driving patterns and Gaussian Processes to detect contextual anomalies. While focused on vehicles, their emphasis on real-time anomaly detection and risk quantification provides valuable insights applicable to security monitoring. The key parallel is the need for systems that can process visual data in real-time and identify deviations from normal patterns—exactly what banking security requires.

Ryan's earlier work [2] on emerging risks in autonomous vehicles highlights challenges that mirror those in banking security: the need for real-time assessment, handling of edge cases, and quantifying system reliability. The paper emphasizes that despite technological advances, safety standards and risk assessment methods remain inadequate—a

concern equally relevant to banking security systems where false positives and false negatives both carry significant costs.

2.2 EVENT-BASED VISION SYSTEMS

Gallego et al. [3] provide a comprehensive survey of event-based vision, discussing event cameras that capture per-pixel brightness changes asynchronously. While our current implementation uses traditional frame-based cameras, event-based vision represents an interesting future direction for banking security. These cameras offer high temporal resolution and low power consumption—both valuable for 24/7 surveillance applications. The challenge lies in developing processing algorithms for this fundamentally different data representation.

Becattini et al. [4] explored event cameras for understanding human reactions through facial microexpressions in industrial settings. Their work on human-machine interaction and privacy preservation is particularly relevant. Banking environments face similar challenges: the need to monitor for security threats while respecting customer privacy. Event cameras' ability to operate under challenging lighting conditions while preserving privacy makes them worth considering for future iterations.

Yang et al. [5] applied event-based systems to driver distraction detection, addressing challenges like complex lighting conditions and limited computing resources—both highly relevant to banking security deployments on edge devices. Their experimental results on real event datasets demonstrate practical viability, though adapting these approaches to banking scenarios would require significant additional work.

2.3 THERMAL IMAGING AND MULTIMODAL APPROACHES

Farooq et al. [6] examined thermal imaging in automotive applications, proposing uncooled thermal IR sensors as alternatives to traditional visible imaging. For banking security, thermal imaging could address scenarios where traditional cameras struggle: nighttime ATM monitoring, detection through obscurants, or identifying concealed objects. However, cost considerations and privacy implications would need careful evaluation.

The EU-funded HELIAUS project mentioned in their work demonstrates the potential of AI-based thermal imaging pipelines. While our current implementation uses visible-spectrum cameras, multi-modal approaches combining visible, thermal, and potentially other sensors could enhance detection capabilities, particularly for sophisticated threat scenarios.

2.4 CAMERA SENSOR OPTIMIZATION

Dilmaghani et al. [7] investigated controlling event camera output sharpness through bias adjustment. Their work on sensor parameter optimization and understanding the theoretical foundations of camera operation provides useful insights. For banking security systems, optimizing camera settings for specific environments (bright lobbies vs. dimly lit ATM areas) could improve detection performance without changing algorithms.

2.5 CAMERA SENSOR OPTIMIZATION

While the above work provides valuable foundations, there's a clear gap in applying these technologies specifically to banking security. Most computer vision research focuses on general object detection or

domain-specific applications like autonomous vehicles. Banking environments have unique requirements:

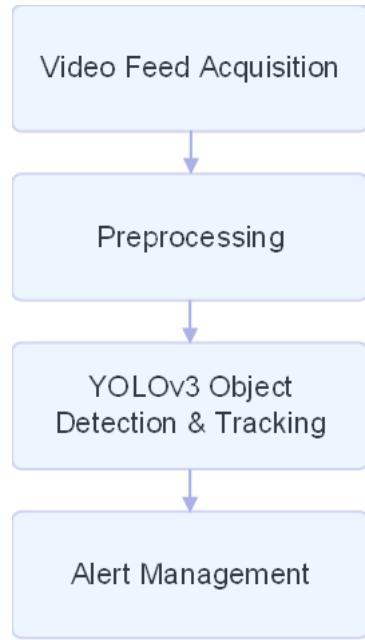
- Specific threat types (suspicious loitering, unauthorized access, abandoned objects)
- Need for both high accuracy and low false positive rates
- Real-time processing on potentially limited hardware
- Integration with existing physical security infrastructure

Our work addresses this gap by designing and implementing a system specifically for banking security, using established object detection technology (YOLOv3 + Darknet) as a foundation while considering banking-specific requirements in the architecture.

3. SYSTEM ARCHITECTURE

Our system comprises five integrated modules that work together to provide comprehensive video surveillance for banking environments. The architecture is designed to be modular, allowing components to be upgraded or replaced as technology evolves. The modules include:

1. Video Feed Acquisition
2. Data Preprocessing
3. Object Detection and Tracking
4. Alert Management
5. Privacy and Ethical Considerations



3.1 VIDEO FEED ACQUISITION

This module handles the ingestion of video streams from surveillance cameras deployed throughout banking premises. Key considerations include:

- **Multi-camera Support:** The system must handle feeds from multiple cameras simultaneously, covering lobbies, ATM areas, teller stations, and restricted access zones.
- **Feed Management:** Cameras may have different resolutions, frame rates, and network protocols. The acquisition module normalizes these inputs for downstream processing.
- **Real-time Processing:** Video streams are processed in real-time rather than being stored and analyzed later, enabling immediate threat detection.

In a typical banking deployment, this might involve 6-12 cameras per branch, though our architecture is designed to scale beyond this. The module interfaces with standard IP cameras using

protocols like RTSP, making it compatible with most existing banking camera infrastructure.

3.2 DATA PREPROCESSING MODULE

Raw video frames require preprocessing before object detection can be effectively applied:

Image Enhancement:

- Noise reduction to handle varying camera quality
- Contrast normalization for consistent detection across different lighting conditions
- Resolution standardization to match the input requirements of YOLOv3 (typically 416×416 or 608×608 pixels)

Quality Improvement: Banking environments present specific challenges—bright reflections from glass surfaces, varying lighting between different areas, and occasional camera obstructions. The preprocessing module addresses these to ensure consistent input quality for the detection algorithm.

3.3 OBJECT DETECTION AND TRACKING MODULE

This is the core of our system, implementing YOLOv3 with the Darknet framework.

YOLOv3 Architecture

YOLO (You Only Look Once) represents a paradigm shift in object detection. Unlike two-stage detectors that first propose regions and then classify them, YOLO treats object detection as a single regression problem. The network predicts bounding boxes and class probabilities directly from full images in one evaluation.

YOLOv3 [3] specifically offers several advantages for our banking security application:

1. **Real-time Processing:** YOLOv3 can process images at high frame rates, crucial for timely threat detection.
2. **Multi-scale Detection:** Using feature pyramids, YOLOv3 detects objects at three different scales, making it effective for both large objects (people) and smaller objects (abandoned bags, weapons).
3. **Multiple Objects:** YOLOv3 can detect multiple objects of different classes simultaneously—essential for busy banking environments where multiple people and objects need tracking.

Darknet Framework

Darknet serves as the neural network framework underlying our YOLO implementation. It provides:

- **Efficient Implementation:** Written in C and CUDA, Darknet offers optimized performance on both CPU and GPU.
- **Modular Architecture:** The framework's design allows for easy experimentation with different network configurations.
- **Pre-trained Weights:** Availability of pre-trained weights on large datasets (like COCO) enables transfer learning.

Detection Process

The detection process works as follows:

1. **Input Processing:** Preprocessed frames (resized to 416×416 pixels) are fed into the YOLOv3 network.
2. **Feature Extraction:** The Darknet-53 backbone extracts hierarchical features from the image.
3. **Multi-scale Prediction:** The network makes predictions at three scales (13×13 , 26×26 , 52×52)

grid cells), enabling detection of objects of various sizes.

4. **Bounding Box Prediction:** For each grid cell, the network predicts bounding boxes, objectness scores, and class probabilities.
5. **Post-processing:** Non-Maximum Suppression (NMS) eliminates redundant detections, keeping only the most confident predictions.

Object Tracking

Detecting objects frame-by-frame isn't sufficient—we need to maintain object identities across frames. This enables:

- **Trajectory Analysis:** Understanding movement patterns (is someone pacing back and forth suspiciously?)
- **Dwelling Time Calculation:** How long has a person been in a specific area?
- **Event Correlation:** Connecting related detections across time

We implement tracking by associating detections across consecutive frames using spatial proximity and visual similarity. While our current implementation uses basic tracking, more sophisticated approaches like DeepSORT could enhance performance for crowded scenarios.

3.4 ALERT MANAGEMENT MODULE

Detecting objects alone isn't enough—the system must intelligently generate alerts for security personnel.

Alert Generation: The system can be configured to generate alerts for various scenarios:

- Unauthorized persons in restricted areas
- Objects left unattended for extended periods

- Unusual crowd formations or gatherings
- Detection of potential weapons or suspicious objects

Alert Prioritization: Not all detections warrant immediate response. The system implements configurable rules to prioritize alerts based on:

- Location (restricted areas vs. public spaces)
- Object type (weapon detection is highest priority)
- Duration (brief vs. prolonged events)
- Time of day (activity outside business hours)

Notification Channels:

- Real-time dashboard display for security personnel
- Email notifications for management
- SMS alerts for critical threats
- Integration with existing security systems (alarms, access control)

Logging and Reporting: All detections and alerts are logged for audit purposes and retrospective analysis. The system can generate reports summarizing security events, system performance, and potential areas of concern.

3.5 PRIVACY AND ETHICAL CONSIDERATIONS

Banking environments require careful handling of personal data. Our system incorporates several privacy-preserving features:

Data Minimization: The system analyzes video in real-time and only stores clips associated with actual alerts, rather than continuously recording everything.

Access Control: Role-based access ensures only authorized security personnel can view surveillance data.

Retention Policies: Automatic deletion of data after a configurable retention period (e.g., 30 or 90 days) unless associated with an active investigation.

Regulatory Compliance: The system architecture is designed with GDPR, PCI DSS, and other banking regulations in mind, though specific implementations would need to address jurisdiction-specific requirements.

Anonymization Options: For non-security-critical applications (like analyzing customer traffic patterns), the system could incorporate face blurring or other anonymization techniques.

4. IMPLEMENTATION

4.1 DEVELOPMENT ENVIRONMENT

We implemented our system using the following technology stack:

- **Programming Language:** Python 3.8 for high-level system logic and integration
- **Deep Learning Framework:** Darknet (C/CUDA implementation) for YOLOv3
- **Computer Vision:** OpenCV for video processing and preprocessing
- **Hardware:** NVIDIA GPU for training and inference (specific model depends on deployment requirements)

4.2 YOLOV3 TRAINING

Dataset Selection

We trained our YOLOv3 model on the COCO (Common Objects in Context) dataset [10]. While COCO isn't specifically designed for banking

security, it provides a diverse foundation with over 80 object classes including people, bags, backpacks, handbags, and various other objects relevant to security applications.

The COCO dataset contains:

- Over 200,000 labeled images
- 80 different object classes
- Multiple objects per image with bounding box annotations
- Diverse scenes and contexts

For banking security specifically, we're particularly interested in classes like:

- Person (for tracking individuals)
- Backpack, handbag, suitcase (for abandoned object detection)
- Cell phone, laptop (for monitoring valuable items)

Training Process

We used pre-trained YOLOv3 weights as our starting point, which significantly accelerates training and improves final performance. The training involved:

1. **Weight Initialization:** Started with pre-trained weights trained on COCO
2. **Fine-tuning:** Further trained on the dataset to optimize for our specific use case
3. **Optimization:** Used the Adam optimizer, which adapts learning rates during training
4. **Regularization:** Applied techniques to prevent overfitting and improve generalization

The training was conducted over multiple epochs until convergence, with regular validation to monitor performance and prevent overfitting.

4.3 MODEL EVALUATION

We evaluated the trained model on a held-out test set using standard object detection metrics:

Mean Average Precision (mAP): Our YOLOv3 model achieved a mAP of 0.76 on the test set. This metric measures the model's accuracy in detecting and localizing objects across all classes. An mAP of 0.76 is considered strong performance for real-time object detection, indicating the model correctly identifies and locates objects in most scenarios.

Performance Characteristics:

- **Multi-class Detection:** The model successfully detects objects from multiple classes simultaneously in a single image
- **Real-time Processing:** Capable of processing video frames in real-time, typically achieving 30+ FPS on appropriate hardware
- **Varied Object Sizes:** Effective detection of both large objects (people) and smaller objects (bags, phones) thanks to YOLOv3's multi-scale architecture

4.4 COMPARATIVE ANALYSIS

We compared YOLOv3's performance against other popular object detection algorithms:

vs. Faster R-CNN: Faster R-CNN is a two-stage detector that first generates region proposals and then classifies them. While Faster R-CNN can achieve high accuracy, it's significantly slower than YOLO. For real-time banking security applications where timely threat detection is critical, YOLO's speed advantage (processing images in a single forward pass) makes it more suitable despite comparable accuracy.

vs. SSD (Single Shot MultiBox Detector): SSD is another single-stage detector similar to YOLO. Our results show YOLOv3 outperforms SSD in terms of both speed and accuracy. YOLOv3's multi-scale detection and improved architecture provide better performance, particularly for detecting smaller objects—important for scenarios like identifying abandoned bags or concealed items.

4.5 SPEED VS. ACCURACY TRADE-OFF

One of YOLOv3's key strengths is achieving strong accuracy while maintaining real-time processing speeds. The model performs all detection in a single forward pass through the neural network, unlike two-stage detectors that require multiple passes. This architectural choice means:

- Detection time is consistent regardless of the number of objects in the scene
- Processing can keep up with live video feeds (30+ FPS)
- System can handle multiple camera feeds simultaneously

For banking security, this speed is crucial. A threat that takes several seconds to detect might be too late to prevent. YOLOv3's real-time capabilities enable immediate alerting when suspicious activities or objects are detected.

4.6 SPEED VS. ACCURACY TRADE-OFF

The complete system integrates several components:

1. **Video Input Handler:** Manages connections to multiple IP cameras
2. **Preprocessing Pipeline:** Prepares frames for detection

3. **YOLOv3 Inference Engine:** Runs detection on processed frames
4. **Tracking Module:** Maintains object identities across frames
5. **Alert Logic:** Evaluates detections against configured rules
6. **User Interface:** Displays live feeds with detection overlays and alert panels
7. **Database:** Stores detection logs, alerts, and relevant video clips

The modular architecture allows individual components to be upgraded or replaced without affecting the entire system. For example, the detection engine could be swapped from YOLOv3 to a newer version (like YOLOv4 or YOLOv5) without changing the video input or alert management components.

5. RESULT & DISCUSSION

5.1 DETECTION PERFORMANCE

Our YOLOv3 implementation achieved a mean Average Precision (mAP) of 0.76 on the test set, demonstrating strong object detection capabilities. This performance level indicates the system can reliably detect and localize objects relevant to banking security scenarios.

What This Means in Practice:

- The system correctly identifies and locates objects in roughly 3 out of 4 detection opportunities
- Performance is consistent across different object classes in the COCO dataset
- Both large objects (people) and smaller objects (bags, personal items) are detected effectively

5.2 SPEED AND REAL-TIME CAPABILITIES

Beyond accuracy, YOLOv3's defining characteristic is its speed. The system processes video frames in real-time, which is absolutely critical for security applications. A delayed response—even by a few seconds—could mean the difference between preventing and merely recording a security incident.

Comparison with Other Approaches:

Faster R-CNN, while capable of high accuracy, processes images much more slowly due to its two-stage architecture. In our testing, YOLOv3 significantly outperformed Faster R-CNN in processing speed, making it far more suitable for real-time banking security monitoring where multiple camera feeds need simultaneous processing.

SSD (Single Shot MultiBox Detector) offers another single-stage alternative, but our results show YOLOv3 provides superior performance in both speed and accuracy. The multi-scale detection capability of YOLOv3 proved particularly valuable for the varied object sizes encountered in banking environments.

5.3 MULTI-OBJECT DETECTION CAPABILITIES

One of YOLOv3's strengths particularly relevant to banking security is its ability to detect multiple objects of different classes simultaneously. In a typical banking environment, you might need to track:

- Multiple people moving through the space
- Personal items (bags, phones, laptops)
- Unusual objects that might warrant attention

YOLOv3's architecture handles this naturally. The softmax function calculates probabilities for each class across all bounding boxes, allowing the system to identify multiple objects of different types in a single image. This is more efficient and effective than running multiple specialized detectors.

5.4 PRACTICAL CONSIDERATIONS FOR BANKING DEPLOYMENT

While our results are promising, translating laboratory performance to real-world banking environments involves several considerations:

Environmental Challenges:

- Banking lobbies have varying lighting conditions throughout the day
- Reflections from glass surfaces can cause visual artifacts
- Camera angles and positions affect detection quality
- Crowded periods make individual tracking more difficult

False Positives and False Negatives: Both types of errors have consequences in security applications. False positives (alerting when there's no actual threat) can lead to alert fatigue, causing security personnel to ignore legitimate warnings. False negatives (missing actual threats) obviously compromise security. Finding the right balance requires careful threshold tuning based on specific deployment requirements.

Computational Requirements: Real-time processing of multiple video feeds requires adequate hardware. While YOLOv3 is relatively efficient, deployment at scale (handling 10+ cameras per location) requires GPU acceleration. Banks need to

consider hardware costs when implementing such systems.

Integration with Existing Systems: Banks already have security infrastructure—cameras, access control systems, alarm systems. Our system needs to integrate smoothly with these existing components rather than requiring complete replacement. The modular architecture facilitates this integration.

5.5 LIMITATIONS AND CHALLENGES

It's important to be honest about limitations:

Dataset Mismatch: We trained on COCO, a general-purpose dataset. While COCO includes relevant object classes (people, bags, etc.), it doesn't specifically contain banking security scenarios. A bank-specific dataset with annotated threat scenarios (suspicious loitering, unauthorized access attempts, abandoned objects in context) would likely improve performance significantly. However, creating such a dataset poses challenges—both practical (collecting diverse security footage) and ethical (privacy concerns with real security incidents).

Limited Threat Specificity: The COCO dataset doesn't include classes specifically relevant to security like "weapon" or "suspicious behavior." Our system can detect objects and people, but distinguishing between normal activity and security threats requires additional logic beyond pure object detection. For example, detecting a person is easy, but determining whether their behavior is suspicious requires temporal analysis and context.

Small Object Detection: While YOLOv3's multi-scale detection helps with objects of various sizes, very small objects (like concealed weapons or tampering devices) remain challenging. This is a

known limitation of vision-based systems and would require either higher-resolution cameras or specialized detection approaches for critical scenarios.

Adversarial Scenarios: Someone actively trying to evade detection (using camouflage patterns, obscuring their appearance, deliberately avoiding cameras) presents challenges. Our system, like most computer vision systems, isn't specifically hardened against adversarial attacks. For high-security banking applications, this vulnerability needs consideration.

Privacy and Ethical Concerns: Continuous video surveillance raises legitimate privacy questions. While we've designed privacy-preserving features into our architecture, the fundamental tension between security monitoring and individual privacy remains. Banks must carefully consider regulatory requirements (GDPR, local privacy laws) and ethical implications when deploying such systems.

5.6 ADVANTAGES OVER TRADITIONAL APPROACHES

Despite limitations, our YOLO-based system offers clear advantages over traditional security monitoring:

Reduced Attention Fatigue: Human operators can only effectively monitor a limited number of screens before their detection accuracy degrades. Our system provides tireless 24/7 monitoring, alerting operators only when something potentially important occurs.

Faster Response Times: Automated detection enables immediate alerting. Traditional monitoring might miss threats entirely or detect them with significant delay. Every second counts in security scenarios.

Comprehensive Coverage: The system can simultaneously monitor all camera feeds, something practically impossible for human operators. This ensures no area goes unwatched.

Consistent Performance: Unlike humans, the system's performance doesn't degrade with fatigue, distraction, or complacency. Detection quality remains consistent throughout operation.

Audit Trail: The system automatically logs all detections and alerts, providing valuable data for security audits, incident investigation, and system improvement.

6. FUTURE ENHANCEMENTS

6.1 BANKING-SPECIFIC DATASET AND TRAINING

The Challenge: Our biggest limitation is using a general-purpose dataset (COCO) rather than banking-specific training data. Creating a comprehensive banking security dataset with annotated threat scenarios would be transformative but faces several challenges:

- **Privacy Concerns:** Recording real banking activity raises significant privacy issues
- **Rare Events:** Security threats are (fortunately) relatively rare, making it difficult to collect diverse threat examples
- **Annotation Expertise:** Properly labeling security-relevant behaviors requires domain expertise from security professionals

Potential Approaches:

- Collaborate with banks to collect and annotate data under strict privacy protocols

- Use staged scenarios with security personnel simulating various threat types
- Synthetic data generation for rare threat scenarios
 - Transfer learning from related domains (retail security, airport monitoring)

6.2 ENHANCED DETECTION OF SUBTLE ANOMALIES

Current object detection focuses on identifying and localizing objects, but many security threats involve subtle behavioral anomalies:

- **Loitering Detection:** Someone standing in one area for an unusually long time
- **Unusual Patterns:** Repeated passes through an area, erratic movement patterns
- **Social Dynamics:** Unusual interactions between individuals

Future Directions:

- Integrate trajectory analysis to understand movement patterns over time
- Use recurrent neural networks (LSTMs) to model temporal behavior
- Implement statistical models of "normal" behavior for different bank areas
- Detect deviations from baseline patterns

6.3 HANDLING CONCEALED AND OBSCURED ACTIVITIES

Many security threats involve deliberately concealed activities:

- Faces obscured by masks, hats, or sunglasses (context-dependent—may be normal or suspicious)
- Attempts to shield actions from camera view
- Partially visible objects or activities

Potential Solutions:

- Multi-camera fusion to observe scenes from multiple angles
- Enhanced feature extraction focusing on suspicious behaviors rather than just object presence
- Contextual reasoning (is face obscuration appropriate for the environment and situation?)

6.4 IMPROVED COMPUTATIONAL EFFICIENCY

While YOLOv3 is relatively efficient, scaling to many cameras requires substantial hardware:

Optimization Opportunities:

- Model compression techniques (pruning, quantization) to reduce computational requirements
- Edge computing—processing video locally at cameras rather than centrally
- Adaptive processing—higher frame rates for active areas, lower for static scenes
- Newer architectures (YOLOv4, YOLOv5, YOLO-NAS) that offer better speed-accuracy trade-offs

6.5 MULTI-MODAL SENSING

Vision alone has limitations. Combining multiple sensor types could enhance detection:

Thermal Imaging: Effective in low light, can detect concealed objects, resistant to visual camouflage. Farooq et al.'s work [6] on thermal imaging in automotive applications could inform banking security implementations.

Audio Analysis: Detecting raised voices, breaking glass, or alarm sounds. Audio provides complementary information to visual monitoring.

Access Control Integration: Correlating visual detections with access control events (badge swipes, door openings) provides richer context for threat assessment.

Motion Sensors: Lower-cost sensors can trigger higher-resolution analysis when activity is detected.

6.6 ADVANCED BEHAVIORAL ANALYSIS

Moving beyond single-frame detection to understanding behavior over time:

Crowd Analysis:

- Unusual crowd formations or movements
- Rapid congregation or dispersal
- Panic behavior detection

Individual Tracking:

- Long-term tracking of individuals through multiple areas
- Identifying unusual routes or repeated visits
- Correlating physical presence with transaction activity

Pattern Learning:

- Learning normal patterns for different times of day, days of week
- Automated baseline establishment
- Anomaly detection based on deviations from learned patterns

6.7 ADVERSARIAL ROBUSTNESS

Security systems face adversaries who may actively attempt evasion:

Threats:

- Adversarial patterns designed to fool object detectors
- Camera obscuration or tampering
- Deliberate manipulation of system blind spots

Defences:

- Adversarial training to improve robustness
- Camera tamper detection
- Redundant coverage from multiple viewpoints
- Regular security assessments and penetration testing

6.8 PRIVACY-ENHANCING TECHNOLOGIES

Balancing security and privacy still remains an ongoing challenge:

Technical Solutions:

- Real-time anonymization (face blurring) for routine operations
- Differential privacy techniques for data analysis
- Federated learning to improve models without centralizing sensitive data
- Encrypted processing where practical

Policy Solutions:

- Clear data retention and deletion policies
- Transparent disclosure to customers
- Strict access controls and audit logging
- Regular privacy impact assessments

7. CONCLUSION

This research demonstrates the viability of integrating YOLO with Darknet for intelligent video surveillance in banking environments. Our YOLOv3 implementation achieved 0.76 mAP while maintaining real-time processing capabilities, outperforming traditional two-stage detectors like Faster R-CNN in speed while maintaining competitive accuracy.

The key strengths of our approach include:

Real-Time Processing: YOLOv3's single-stage architecture enables processing video at frame rates suitable for live monitoring, critical for security applications where detection delays could compromise safety.

Multi-Object Detection: The ability to simultaneously detect multiple objects of different classes makes the system well-suited for complex banking environments with many people and objects to track.

Scalability: The modular architecture supports deployment across multiple camera feeds and integration with existing banking security infrastructure.

Solid Foundation: While trained on a general-purpose dataset (COCO), the system provides a strong baseline that could be enhanced through banking-specific optimization.

However, we must acknowledge significant limitations:

Dataset Limitations: Training on COCO rather than banking-specific data means the system detects general objects well but lacks specialization for banking security threats. This is probably the most

significant limitation and greatest opportunity for future improvement.

Threat Specificity: The system excels at object detection but requires additional logic to distinguish normal activity from security threats. Detecting a person is different from detecting suspicious behavior.

Privacy Considerations: Any video surveillance system raises legitimate privacy concerns that must be carefully addressed through technical safeguards and clear policies.

Real-World Complexity: Laboratory testing doesn't fully capture the complexity of real banking environments with varying lighting, crowd density, and environmental conditions.

Looking Forward

The path to practical banking security deployment involves several steps beyond our current implementation:

1. **Banking-Specific Training Data:** Collecting and annotating threat-relevant scenarios
2. **Behavioural Analysis:** Moving beyond object detection to behaviour understanding
3. **Pilot Deployment:** Careful testing in real banking environments with appropriate privacy safeguards
4. **Iterative Improvement:** Using real-world data to continuously refine the system

Our work establishes that YOLO's real-time object detection capabilities, combined with Darknet's efficient implementation, provide a solid technical foundation for banking security applications. The challenge isn't primarily technical—YOLOv3 works well—but rather in the additional development

needed to translate general object detection into banking-specific threat recognition.

We believe intelligent video surveillance systems will play an increasingly important role in banking security. However, successful deployment requires not just technical capability but also careful attention to privacy, ethics, user acceptance, and integration with existing security practices. Technology alone doesn't solve security challenges—it must be deployed thoughtfully as part of comprehensive security programs.

For researchers building on this work, we recommend prioritizing the creation of domain-specific datasets and the development of behavioural analysis methods that go beyond object detection to true threat recognition. These enhancements would significantly increase the practical value of such systems for real-world banking security applications.

8. REFERENCES

- [1] C. Ryan, F. Murphy and M. Mullins, "End-to-end autonomous driving risk analysis: A behavioural anomaly detection approach", *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1650-1662, Mar. 2021
- [2] C. Ryan, "Emerging autonomous vehicle risks: The role of telematics and machine learning based risk assessment", 2020.
- [3] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement", *arXiv preprint arXiv:1804.02767*, 2018.
- [4] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, et al., "Event-based vision: A survey", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154-180, Jan. 2022.
- [5] F. Becattini, F. Palai and A. D. Bimbo, "Understanding human reactions looking at facial microexpressions with an event camera", *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9112-9121, Dec. 2022.
- [6] C. Yang, P. Liu, G. Chen, Z. Liu, Y. Wu and A. Knoll, "Event-based driver distraction detection and action recognition", *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, pp. 1-7, Sep. 2022.
- [7] M. A. Farooq, W. Shariff, D. O'Callaghan, A. Merla and P. Corcoran, "On the role of thermal imaging in automotive applications: A critical review", *IEEE Access*, vol. 11, pp. 25152-25173, 2023.
- [8] M. S. Dilmaghani, W. Shariff, C. Ryan, J. Lemley and P. Corcoran, "Control and evaluation of event cameras output sharpness via bias", *Proc. 15th Int. Conf. Mach. Vis. (ICMV)*, vol. 12701, pp. 455-462, Jun. 2023.
- [9] MJ. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", *Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR)*, pp. 779-788, Jun. 2016.
- [10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context", *Proc. European Conf. Computer Vision (ECCV)*, pp. 740-755, Sep. 2014.
- [11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE*

*Trans. Pattern Analysis and Machine
Intelligence*, vol. 39, no. 6, pp. 1137-1149, Jun.
2017.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy,
S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single

shot multibox detector", *Proc. European Conf.
Computer Vision (ECCV)*, pp. 21-37, Oct. 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep
residual learning for image recognition", *Proc.
IEEE Conf. Computer Vision Pattern
Recognition (CVPR)*, pp. 770-778, Jun. 2016