JAAFR INTERNATIONAL RESEARCH JOURNAL
JOURNAL OF ADVANCE AND FUTURE RESEARCH
JAAFR.ORG | ISSN : 2984-889X
An International Open Access, Peer-reviewed, Refereed Journal

Real-Time Anomaly Detection and Crowd Safety Management System

V.Bharath
III BCA C Student
Department of Computer Applications
VISTAS, Chennai

Dr. K.Dharmarajan Msc,M phil,Phd
Head of Department
Department of Computer Applications
VISTAS, Chennai

Abstract: Ensuring public safety in crowded environments such as stadiums, transportation hubs, and public events is a critical challenge. This project presents a Real-Time Anomaly Detection and Crowd Safety Management System that leverages artificial intelligence, computer vision, and data analytics to monitor and manage crowd behavior effectively. The system utilizes advanced deep learning models to analyze live video feeds from surveillance cameras and detect unusual or potentially dangerous activities such as overcrowding, sudden movements, violence, or unauthorized access. By employing techniques like object detection, motion analysis, and pattern recognition, the system can identify anomalies in real time with high accuracy. Additionally, the platform integrates alert mechanisms that notify authorities instantly when suspicious activities are detected, enabling rapid response and prevention of potential hazards. It also includes crowd density estimation and predictive analytics to anticipate risky situations before they escalate. To ensure reliability and ethical deployment, the system incorporates data privacy, secure data handling, and controlled access mechanisms. The solution is scalable and adaptable for use in smart cities, public gatherings, and emergency management scenarios. Overall, this system demonstrates the potential of AI-driven technologies in enhancing crowd monitoring, improving situational awareness, and ensuring public safety in real-time environments.

Keywords — Real-Time Anomaly Detection, Crowd Safety Management, Deep Learning, Computer Vision, Object Detection, Crowd Density Estimation, Surveillance Systems, YOLOv8, Optical Flow, Alert Mechanism, Smart Cities, Predictive Analytics, Public Safety.

I. INTRODUCTION

Public gatherings and high-density environments represent some of the most complex operational challenges in modern urban security management. Whether at a sporting stadium hosting tens of thousands of spectators, a busy transportation hub processing continuous flows of commuters, or a large-scale public festival, the sheer density and dynamism of crowd behaviour creates conditions where dangerous situations can arise, escalate, and cause harm within seconds. Traditional crowd management relies on human security personnel monitoring camera feeds, responding to radio communications, and making judgment calls under high cognitive load — an approach that scales poorly and is vulnerable to attention lapses, communication delays, and the inherent limitations of human reaction time.

The convergence of deep learning, high-resolution surveillance infrastructure, and edge computing has created a technological foundation capable of transforming this operational paradigm. Modern computer vision models can process video streams in real time, classify objects and actions with human-level accuracy, and deliver structured analytical outputs — crowd density estimates, anomaly scores, trajectory predictions — to decision-support systems that amplify the effectiveness of human security teams. The potential of these technologies to save lives, reduce injuries, and enable more efficient deployment of security resources is substantial.

This paper presents the design and implementation of a Real-Time Anomaly Detection and Crowd Safety Management System. The system analyzes live video from fixed surveillance cameras using a cascade of deep learning components: a YOLOv8-based person detection module, an optical flow motion analysis module, a crowd density estimation

network, and a graph neural network trained to identify anomalous interaction patterns among detected individuals. Outputs from these components are fused by a multi-modal anomaly scoring engine that generates per-zone safety scores at configurable temporal resolutions. When scores exceed adaptive thresholds, an alert generation module delivers notifications to authorities through multiple channels including mobile push notifications, dashboard alerts, and integration with building management systems.

The system is designed around three core principles: real-time responsiveness (processing latency under 500 milliseconds from camera capture to alert delivery), scalability (supporting simultaneous monitoring of up to 128 camera streams on a single GPU server), and ethical deployment (incorporating privacy-preserving processing, role-based access control, and immutable audit logging). Experimental evaluation on the UCF-Crime benchmark dataset and a live deployment at a university campus event demonstrates detection accuracy exceeding 94% with a false-positive rate below 5%.

II. RELATED WORK

Research in automated crowd analysis has evolved significantly over three decades, progressing from handcrafted feature-based methods to the deep learning approaches that now define the state of the art. Early work by Helbing and Molnar (1995) established the social force model for pedestrian dynamics, providing a physics-inspired mathematical framework for simulating and analysing crowd movement. While not directly applicable to real-time video analysis, the social force model introduced the concept of comfort distance, directional urgency, and interaction forces that continue to inform feature engineering in modern anomaly detection systems.

Crowd density estimation has been extensively studied as a prerequisite to crowd safety assessment. Zhang et al. (2016) introduced the Multi-Column Convolutional Neural Network (MCNN), which processes images at multiple scales to produce spatial density maps robust to perspective distortion and occlusion. Subsequent work by Li et al. (2018) with the CSRNet architecture demonstrated that dilated convolutions enable density estimation at significantly higher spatial resolution, improving the ability to distinguish crowd sub-regions with critically different density levels. The proposed system builds on this line of research by integrating CSRNet-derived density maps as a primary input to the anomaly scoring engine.

Anomaly detection in surveillance video has been approached through both reconstruction-based and classification-based methods. Hasan et al. (2016) demonstrated that autoencoders trained exclusively on normal event sequences develop a high reconstruction error for anomalous events not represented in the training distribution, enabling unsupervised anomaly detection without labelled anomaly samples. Liu et al. (2018) introduced the Future Frame Prediction model, which trains a network to predict the next frame in a normal video sequence and treats high prediction error as an anomaly indicator. These approaches inform the reconstruction-based component of the proposed system's multi-modal anomaly scoring engine.

Object detection frameworks have undergone rapid evolution. The YOLO (You Only Look Once) architecture family, introduced by Redmon et al. (2016) and refined through subsequent versions, achieves real-time object detection at frame rates compatible with surveillance applications. YOLOv8, the version employed in this system, provides state-of-the-art precision-recall performance on the COCO benchmark while supporting deployment on edge GPU hardware with inference latency below 15 milliseconds per frame on an NVIDIA Jetson Orin. This inference speed enables the proposed system to process multiple simultaneous camera streams without sacrificing detection fidelity.

Graph neural network approaches to crowd behaviour analysis have emerged recently as a powerful complement to image-level detection methods. Huang et al. (2020) modelled crowd participants as nodes in a dynamic graph, with edges encoding spatial proximity and interaction intensity. Graph convolutional operations propagate interaction features across the graph, enabling the detection of anomalous group dynamics such as the formation of a fight cluster or the stampede-precursor pattern of a rapidly expanding density gradient. The proposed system integrates a simplified graph interaction analysis module that processes detected person bounding boxes as a dynamic graph updated at 5-frame intervals.

III. EXISTING SYSTEM

Current crowd monitoring and anomaly detection solutions employed in operational public safety contexts exhibit several fundamental limitations that reduce their effectiveness in preventing incidents before they cause harm.

The dominant operational model remains human-monitored CCTV. Security personnel watch camera feeds displayed on multi-screen monitoring stations and are expected to identify unusual behaviour across dozens of simultaneous feeds. Research on video surveillance operator performance consistently demonstrates that sustained attention degrades significantly within 20 to 30 minutes of continuous monitoring, and that detection rates for abnormal events fall sharply as the number of simultaneously monitored screens increases beyond four. High-density crowd environments, where

the visual complexity of each frame is maximal, exacerbate this attentional bottleneck. Furthermore, human operators are unable to simultaneously compute quantitative metrics such as crowd density or approach speed that would provide objective early indicators of developing danger.

Rule-based video analytics systems, marketed as intelligent video surveillance (IVS) solutions by vendors such as Milestone Systems and Genetec, apply predefined heuristic rules to video streams. Typical rules include virtual line crossing, loitering detection based on dwell time thresholds, and object removal alerts. While these rules reduce operator workload for specific well-defined scenarios, they lack the flexibility to detect the diverse and context-dependent signatures of crowd anomalies. A rule specifying an alert when more than N persons occupy a defined region does not capture the difference between an orderly queue and a chaotic crush with the same headcount. The limitations of existing systems include:

- Human CCTV monitoring suffers from attention fatigue, limiting sustained detection performance to under 30 minutes
- Rule-based IVS systems cannot capture context-dependent crowd anomalies that require semantic understanding
- Most commercial systems lack real-time crowd density estimation at sub-zone granularity
- Existing platforms rarely integrate predictive analytics to anticipate dangerous crowd dynamics before they manifest
- Privacy and data governance mechanisms are inconsistent, creating compliance risks in public deployment
- Few solutions provide unified multi-camera correlation to detect coordinated or geographically distributed incidents

IV. SYSTEM ARCHITECTURE

The proposed system adopts a four-tier pipeline architecture that separates video ingestion, computer vision processing, anomaly reasoning, and alert delivery into distinct layers with well-defined interfaces. This layered design enables horizontal scaling of computationally intensive processing tiers independently from lightweight management and delivery tiers.

Tier 1: Video Ingestion Layer

The ingestion layer receives RTSP streams from IP surveillance cameras, decodes video frames using hardware-accelerated H.264/H.265 decoding via NVIDIA NVDEC, and distributes frames to per-camera processing queues. Each camera stream is processed at a configurable frame sampling rate (default 10 frames per second) to balance detection responsiveness against GPU resource consumption. A frame buffer retains the most recent 30 seconds of raw frames per camera in a ring buffer, enabling retrospective clip extraction for incident documentation without requiring continuous recording storage.

Tier 2: Computer Vision Processing Layer

This layer hosts four parallel deep learning modules applied to each sampled frame. The YOLOv8-x person detection module produces bounding boxes and confidence scores for all detected persons. The CSRNet crowd density estimation module produces a per-pixel density map from which zone-level headcount estimates are derived. The optical flow motion analysis module computes dense flow fields between consecutive frames and extracts flow magnitude histograms and directional entropy as motion feature vectors. The graph interaction module constructs a proximity graph from detected person bounding box centroids and applies two-layer graph convolution to produce a per-person interaction embedding. All four modules execute on a shared GPU with workload interleaved using CUDA streams.

Tier 3: Anomaly Scoring and Fusion Layer

The anomaly scoring layer receives outputs from all four Tier 2 modules and applies a weighted fusion function to produce a composite anomaly score in the range [0, 1] for each monitored zone at each time step. The fusion weights were determined empirically on the UCF-Crime validation set: density excess weight 0.30, motion entropy weight 0.25, interaction graph anomaly weight 0.30, and reconstruction error weight 0.15. Adaptive thresholds are computed per zone using a rolling 30-minute baseline of anomaly scores, enabling the system to calibrate thresholds to the baseline activity level of each specific zone rather than applying universal static thresholds.

Tier 4: Alert and Management Layer

When a zone's composite anomaly score exceeds its adaptive threshold, the alert layer generates a structured alert record containing zone identifier, timestamp, anomaly score, primary triggering component, and an automatically extracted incident clip. Alerts are delivered through mobile push notifications (Firebase Cloud Messaging), a real-time dashboard (React + WebSocket), email (SMTP), and optionally through integration with building management systems via a REST webhook. The management layer also provides a configuration interface for zone definition, threshold tuning, camera management, user access control, and reporting.

Figure 1: System Architecture — Video Ingestion → CV Processing (YOLOv8 + CSRNet + Optical Flow + Graph NN) → Anomaly Fusion → Alert Delivery

V. PROPOSED METHODOLOGY

The methodology of the proposed system is grounded in a multi-modal detection philosophy that combines four complementary evidence streams — person count, motion characteristics, social interaction patterns, and frame reconstruction fidelity — into a unified anomaly signal. Each evidence stream is independently informative for a different subset of anomaly types; their combination achieves coverage across the full spectrum of crowd safety events of interest.

Person detection is performed by YOLOv8-x fine-tuned on the CrowdHuman dataset, which provides diverse crowded-scene training examples with accurate annotations under high occlusion. Fine-tuning improves detection recall from 87.3% (ImageNet-pretrained weights) to 94.6% on a held-out crowded scene validation set. Detected bounding boxes are filtered by minimum confidence (0.45) and non-maximum suppression (IoU threshold 0.5) before being forwarded to downstream modules.

Crowd density estimation uses CSRNet applied to full-resolution frames at 2-second intervals. The density map is integrated over predefined zone polygons drawn by security administrators on the camera registration interface to produce per-zone headcount estimates. Zone-level density alerts are triggered when the estimated headcount exceeds a configurable capacity threshold, providing a direct metric for overcrowding detection that complements the more nuanced anomaly detection of the fusion layer.

Motion analysis extracts dense optical flow fields using the RAFT algorithm, which provides state-of-the-art flow accuracy at 20 frames-per-second throughput on GPU. From the flow field, the system computes three scalar features per zone: mean flow magnitude (indicating overall motion intensity), flow directional entropy (measuring the dispersion of movement directions, which is low in orderly movement and high in chaotic stampede precursors), and the 95th percentile flow magnitude (capturing the fastest-moving individuals who may indicate sudden flight responses). These features are normalised against zone-specific baselines and input to the fusion layer.

The graph interaction module models crowd participants as nodes in a k-nearest-neighbour graph (k=5) with edges weighted by inverse Euclidean distance between bounding box centroids. Two-layer graph convolution aggregates neighbour features and produces a per-node embedding. A pre-trained anomaly classifier applied to the mean-pooled graph embedding produces an interaction anomaly score that is elevated in the presence of clustering patterns characteristic of fights, falls, or coordinated disturbances. The model was trained on the RWF-2000 fight detection dataset, achieving 91.4% classification accuracy on the test split.

Figure 2: Methodology Flow — Frame Sampling → YOLOv8 Detection → CSRNet Density → RAFT Flow → Graph NN → Score Fusion → Adaptive Threshold → Alert

VI. MODULES DESCRIPTION

The system is organized into seven functional modules, each encapsulating a distinct operational responsibility and exposing well-defined interfaces to adjacent modules in the processing pipeline.

6.1 Camera Management Module

The camera management module handles the registration, configuration, and health monitoring of all connected IP cameras. Administrators register cameras by supplying RTSP stream URL, physical location, coverage zone polygon coordinates, and capacity thresholds through a web interface. The module continuously monitors stream liveness using heartbeat probes and automatically reconnects disconnected streams with exponential backoff. Camera metadata is stored in a PostgreSQL database and versioned to support configuration audit trails.

6.2 Person Detection Module

This module wraps the YOLOv8-x inference engine in a batched inference pipeline that processes frames from multiple cameras simultaneously using dynamic batching. Frames are grouped into batches of up to 8 by a scheduler that prioritises cameras with recent anomaly activity, ensuring that high-alert cameras receive more frequent inference cycles during incidents. Detected person bounding boxes are associated across consecutive frames using a ByteTrack multi-object tracker, maintaining consistent person identifiers across a tracking window for trajectory analysis.

6.3 Crowd Density Estimation Module

The density estimation module applies CSRNet to downsampled frames at 2-second intervals, producing 60×80 pixel density maps (corresponding to the 1920×1080 source frame with 32× spatial downsampling). Zone-level headcount estimates are computed by integrating density values over zone polygon masks. The module maintains a rolling 10-minute history of per-zone headcounts for trend analysis and delivers capacity utilisation percentages to the management dashboard in real time.

6.4 Motion Analysis Module

The RAFT optical flow computation is applied to consecutive frame pairs sampled at 5-frame intervals to balance flow accuracy against computation cost. Dense flow fields are zone-masked before feature extraction, ensuring that motion from adjacent zones does not contaminate zone-specific motion metrics. The module maintains per-zone rolling baselines of the three flow features with exponential smoothing (decay factor 0.95) to support adaptive normalisation. Sudden large deviations from baseline — such as the rapid increase in flow magnitude and directional entropy characteristic of a crowd surge — are flagged immediately to the fusion layer without waiting for the next scheduled fusion cycle.

6.5 Graph Interaction Analysis Module

The graph interaction module receives per-frame person bounding box centroids from the detection module and constructs the k-NN proximity graph using a spatial index for efficient nearest-neighbour lookup. Graph convolution is applied using pre-computed normalised adjacency matrices cached between frames, updating only the edges that change due to person movement. The resulting interaction embeddings and anomaly scores are logged per frame for visualisation on the management dashboard as a person-level heat map overlay on the live camera feed.

6.6 Anomaly Scoring and Alert Module

The fusion engine aggregates the four evidence streams at configurable intervals (default 2 seconds) using the weighted sum described in Section V. Adaptive thresholds are updated every 60 seconds using the Isolation Forest algorithm applied to the rolling 30-minute score history, automatically adjusting to diurnal patterns and event-specific baseline variations. When a zone exceeds its threshold, the alert module extracts a 30-second retrospective clip from the frame ring buffer, annotates it with bounding boxes and density overlays, and packages it with the alert metadata for delivery.

6.7 Privacy and Access Control Module

The privacy module applies automatic face blurring to all video clips and frames accessed through the management dashboard, using a dedicated SCRFD face detection model for localisation and Gaussian blurring for anonymisation. Raw unblurred footage is accessible only through a privileged investigator role that requires dual-administrator approval and generates an immutable access log entry. All data in motion is encrypted using TLS 1.3 and all stored data uses AES-256 encryption. Role-based access control assigns permissions across four levels: viewer, operator, supervisor, and administrator.

VII. IMPLEMENTATION

The system is implemented in Python 3.11 for all AI inference components and TypeScript/React for the management dashboard frontend. The core inference pipeline runs on Ubuntu 22.04 LTS with CUDA 12.1 and cuDNN 8.9. YOLOv8-x inference uses the Ultralytics Python library with TensorRT engine export for optimal GPU throughput. CSRNet and the graph interaction model are implemented in PyTorch 2.0 with torch.compile applied for compiled graph optimisation. RAFT optical flow uses the official PyTorch implementation.

The alert delivery backend is implemented in FastAPI with async request handling and Redis as a message broker for decoupled alert routing. The management dashboard is React 18 with Recharts for time-series visualisation, Leaflet for floor-plan zone mapping, and Socket.IO for real-time alert streaming. PostgreSQL 15 with PostGIS extensions stores camera configurations, alert records, and zone geometries. All components are containerised with Docker and orchestrated by docker-compose, with GPU passthrough configured for the inference containers via NVIDIA Container Toolkit.

The complete system is deployed on a server equipped with two NVIDIA RTX 4090 GPUs, 128 GB RAM, and a 10 GbE network interface. This hardware configuration supports simultaneous processing of 64 camera streams at 10 FPS each with total GPU utilisation of approximately 78% under full load. The system can be scaled horizontally by adding inference nodes and load-balancing the camera stream assignments using a round-robin distribution policy managed by the camera management module.

VIII. TESTING AND RESULTS

The system was evaluated through three complementary testing activities: benchmark evaluation on the UCF-Crime dataset, controlled scenario testing in a simulated crowd environment, and a live pilot deployment at a university campus event.

8.1 UCF-Crime Benchmark Evaluation

The UCF-Crime dataset contains 1,900 surveillance videos spanning 13 anomaly categories including fighting, robbery, assault, and traffic accidents. Evaluation used the standard area-under-curve (AUC) metric on the test split. The proposed system achieved an AUC of 87.4% on the full 13-class evaluation, a frame-level anomaly detection accuracy of 94.2%, and a false-positive rate of 4.8% on normal video clips. Crowd-specific categories (fighting, assault) achieved

the highest recall at 96.1%, reflecting the benefit of the graph interaction module for detecting group behaviour anomalies.

8.2 Controlled Scenario Testing

A testbed was constructed using eight cameras covering a 400 m² indoor area with 50 volunteer participants enacting predefined crowd scenarios: orderly queue, sudden dispersal, simulated fight (choreographed), zone overpopulation, and unauthorised access to a restricted sub-zone. All 20 scenario executions across five repetitions were detected correctly. Mean alert generation latency from scenario onset to alert delivery was 1.2 seconds, within the design target of 2 seconds. The false-positive rate during the 4-hour baseline recording with normal activity was 2.1%.

Table 1 compares the proposed system against two baseline configurations on the controlled scenario test.

Table 1: Detection Performance Comparison — Controlled Scenario Test

System Configuration	Detection Accuracy	False-Positive Rate	Alert Latency
Human CCTV Monitoring (4 screens)	74.0%	11.3%	38.4 seconds
Rule-Based IVS (density threshold only)	81.5%	8.7%	3.1 seconds
Proposed Multi-Modal System	100.0%	2.1%	1.2 seconds

8.3 Live Campus Event Pilot

The system was deployed at a university open day event attended by approximately 1,800 visitors over 6 hours, monitored by 12 cameras. The system generated 23 alerts during the event: 21 confirmed true positives (including 14 overcrowding alerts, 5 sudden-dispersal alerts, and 2 fight-precursor alerts that enabled security to intervene before physical contact), 1 false positive (a performer's theatrical movement flagged as anomalous), and 1 missed detection (a minor altercation in a camera blind spot). Security supervisors rated the alert quality as 4.4 out of 5 and indicated that the system meaningfully improved their situational awareness.

Table 2: System Performance Summary

Metric	Result
UCF-Crime AUC	87.4%
Frame-Level Detection Accuracy	94.2%
False-Positive Rate (UCF-Crime)	4.8%
Controlled Scenario Detection Rate	100%
Mean Alert Latency	1.2 seconds
Live Pilot True Positive Rate	91.3%
GPU Throughput (64 streams, 2× RTX 4090)	78% utilisation

IX. CONCLUSION

The Real-Time Anomaly Detection and Crowd Safety Management System presented in this paper demonstrates that multi-modal deep learning — combining person detection, crowd density estimation, optical flow motion analysis, and graph neural network interaction modelling — can achieve detection accuracy and response latency that substantially exceed both human monitoring and conventional rule-based video analytics systems.

The system achieved 94.2% frame-level detection accuracy on the UCF-Crime benchmark with a 4.8% false-positive rate, 100% detection across all controlled scenario executions with 1.2-second mean alert latency, and a 91.3% true-positive rate during live deployment at a real public event. The adaptive threshold mechanism proved effective in reducing false positives caused by expected fluctuations in crowd density and motion patterns, adapting automatically to diurnal variations and event-specific baseline characteristics.

The privacy and access control architecture ensures that the system can be deployed in public environments in compliance with data protection regulations, with automatic face anonymisation preventing the accumulation of personal biometric data outside formally authorised investigative contexts.

Future work will extend the system in three directions: integration of audio anomaly detection using microphone arrays to complement visual analysis; development of a crowd evacuation guidance module that triggers dynamic signage and public address announcements based on anomaly location and severity; and deployment of lightweight inference models

on edge compute hardware co-located with cameras to reduce bandwidth requirements and enable operation in environments with limited network connectivity.

REFERENCES

[1] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," Phys. Rev. E, vol. 51, no. 5, pp. 4282–4286, May 1995.

[2] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network," in Proc. IEEE CVPR, Las Vegas, NV, USA, 2016, pp. 589–597.

[3] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes," in Proc. IEEE CVPR, Salt Lake City, UT, USA, 2018, pp. 1091–1100.

[4] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning Temporal Regularity in Video Sequences," in Proc. IEEE CVPR, Las Vegas, NV, USA, 2016, pp. 733–742.

[5] W. Liu, W. Luo, D. Lian, and S. Gao, "Future Frame Prediction for Anomaly Detection — A New Baseline," in Proc. IEEE CVPR, Salt Lake City, UT, USA, 2018, pp. 6536–6545.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE CVPR, Las Vegas, NV, USA, 2016, pp. 779–788.

[7] C. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in Proc. IEEE CVPR, Vancouver, Canada, 2023.

[8] Z. Teed and J. Deng, "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow," in Proc. ECCV, Glasgow, UK, 2020, pp. 402–419.

[9] P. Huang, Y. Chang, and C. Fang, "Crowd Anomaly Detection Based on Social Force Model and Acceleration Feature," Sensors, vol. 20, no. 4, p. 1072, Feb. 2020.

[10] W. Zhang, P. Wang, and Y. Li, "RWF-2000: An Open Large Scale Video Database for Violence Detection," in Proc. IEEE ICIP, Abu Dhabi, UAE, 2020, pp. 4183–4187.