Article

UGHYT-PLUS: YOLOV11-ENHANCED FEDERATED RL FRAMEWORK FOR ULTRA-PRECISE SPEED ESTIMATION, ANOMALY DETECTION, AND ADAPTIVE TRAFFIC CONTROL IN OCCLUSION-HEAVY URBAN SURVEILLANCE

[1]Mohit Tiwari, [2]Dr. Vikas Sakalle

[12]LNCT University, JK Town, Sarvadharam, Sector C, Kolar Road, Bhopal, Madhya Pradesh, India – 462024, mohitprivate@gmail.com, vikassakalle@gmail.com

Corresponding author: Mohit Tiwari, **Email ID:** mohitprivate@gmail.com

Abstract

Urban traffic in India claims over 150,000 lives yearly, with overspeeding and anomalies exacerbating fatalities in occlusion-heavy, low-resolution CCTV environments like Kolkata and Bhopal. Building on the UGHYT framework, this paper introduces UGHYT-Plus—a YOLOv11-enhanced federated reinforcement learning (RL) system fusing self-supervised anomaly detection, evidential speed regression, and adaptive signal control. UGHYT-Plus achieves 96.2% mAP, 1.8 km/h RMSE, and 96% anomaly recall at 25 ms latency on expanded MITS/BDD100K datasets plus custom 75-hour Kolkata CCTV (80% occlusion, 480p, 8-20 FPS). Innovations include contrastive pretraining for incident detection (accidents, emergency vehicles), multi-modal radar fusion for sub-2 km/h precision, and privacy-preserving federated RL optimizing signals to cut congestion by 28%. Outperforming YOLOv10 baselines by 35% in accuracy and 30% in speed, UGHYT-Plus enables scalable 5G-IoT deployment across 100+ feeds on 20W Jetson nodes, paving the way for safer smart cities.

Keywords: YOLOv11, federated reinforcement learning, self-supervised anomaly detection, evidential deep learning, traffic speed estimation, edge computing, intelligent transportation systems, occlusion-robust surveillance.

I. INTRODUCTION

India's urban roads see over 150,000 deaths each year from traffic issues like speeding and congestion, with places like Kolkata and Bhopal facing even steeper rises—up 15% in 2025—thanks to mixed traffic, spotty CCTV coverage, and rainy seasons that make things worse. Sure, radar or LiDAR setups give spot-on readings, but at 5-10 lakhs a pop for 1.5 million intersections, they're just not practical. That's where smarter, camera-based deep learning steps in as a budget-friendly fix.

Pioneering work with YOLOv8 or v9 paired with DeepSORT trackers hits 88-92% accuracy and 4-6 km/h errors on open highways. But toss in real Indian city chaos—grainy 480p feeds, 70% blocked views, shaky 5-20 FPS clips, bikes weaving everywhere—and it tanks to 65-75% accuracy with 8-12 km/h mistakes.

That's why we built on our UGHYT system (YOLOv10 with transformers, uncertainty checks, and edge tweaks), which nailed 94.1% detection and just 3.8 km/h errors at 35 ms speeds on tough datasets from Bhopal. Still, it left room to grow: no built-in spotting for crashes or ambulances, no smart signal tweaks, weak handling of super-slow footage or monsoons, and tough scaling to city-wide camera nets.

Enter UGHYT-Plus: We've leveled up to YOLOv11, thrown in self-learning anomaly spotting, radar boosts for pinpoint speed reads under 2 km/h error, and privacy-safe RL to dynamically fix traffic lights. Tested on bigger MITS/BDD100K sets plus 75 hours of Kolkata footage (80% blocks, radar-checked truths), it hits 96.2% accuracy, 1.8 km/h errors, spots 96% of incidents, and cuts jams by 28%—all in 25 ms on low-power Jetson gear for 100+ cameras.

Our big wins? (1) Tough-as-nails detection with extras for surprises; (2) shared RL smarts across cities, no data leaks; (3) proof it flies in real Indian setups. Next, we recap what's out there, explain how we built it, share results, and look ahead.

II. RELATED WORK

Early traffic surveillance systems relied heavily on classical computer vision. Kim *et al.* [1] used Gaussian Mixture Models (GMM) with virtual loops to perform motion segmentation, achieving real-time vehicle counting but suffering from illumination sensitivity and frequent occlusions, resulting in speed errors often exceeding **10 km/h MAE**. Similarly, Luo *et al.* [2] showed that Kalman filters and SORT-based tracking improved identity stability, yet these trackers collapsed under dense, bidirectional urban traffic where overlapping trajectories and nonlinear motion are common.

The advent of deep learning significantly improved vehicle detection. Tran *et al.* [3] demonstrated that YOLOv4 combined with DeepSORT reached **90% accuracy** on structured highways but deteriorated to **~83% mAP** on low-resolution municipal CCTV (≤480p). Li *et al.* [4] further showed that YOLOv8 with ByteTrack and centroid tracking achieved **4.1 km/h MAE** at ~18 FPS and remained robust during daytime, though performance degraded on variable-FPS and nighttime footage.

Transformer-based tracking improved temporal consistency in occlusion-heavy scenarios. Wang *et al.* [5] integrated BoT-SORT with YOLOv9, reporting **12% better occlusion handling** and reduced identity switches, but their system still produced **6 km/h RMSE** in dense urban conditions. Multistage CNN architectures were explored by Kumar *et al.* [6], achieving **92% mAP** but incurring prohibitive latency (>120 ms), making real-time deployment impractical.

Recent developments in edge-optimized detection include YOLOv11, which incorporates hybrid convolution–attention blocks and sparsity-aware inference. According to Ultralytics [7], YOLOv11 delivers **~93.9% mAP** on dense traffic with reduced computational overhead, making it suitable for embedded ITS deployments.

Parallel improvements in speed estimation have focused on uncertainty modeling. Tiwari *et al.* [8] introduced the UGHYT framework using evidential deep learning (EDL) combined with optical flow fusion, achieving **3.2 km/h MAE** under 480p/10 FPS CCTV streams. Radar–vision fusion systems proposed by Luo *et al.* [9] enhance depth estimation and occlusion recovery, though their installation cost limits real-world deployment across developing cities.

Self-supervised anomaly detection has emerged as a powerful label-free alternative for incident recognition. Chen *et al.* [10] demonstrated >95% recall using future-frame prediction, while Singh and Roy [11] employed contrastive learning to detect emergency vehicles and sudden stops without annotated datasets.

Table I: DL Traffic Speed/Anomaly Systems

Method	mAP (%)	RMSE/MAE (km/h)	Latency (ms)	Anomaly (%)	Edge
YOLOv4-DeepSORT	90.0	5.2/7.2	65	N/A	Part
YOLOv8-BT	88.0	4.1/6.1	40	N/A	Yes
YOLOv9-BoTSORT	91.2	4.9/6.0	48	N/A	Part
UGHYT	94.1	3.2/3.8	35	N/A	Yes
UGHYT-Plus	96.2	1.8	25	96	Yes

Federated learning (FL) has gained traction for privacy-preserving ITS. Verma *et al.* [12] surveyed FL methods that allow multiple intersections to collaboratively train models without sharing raw surveillance video. Chen *et al.* [13] extended vision-supervised federated deep learning to structural and traffic monitoring, demonstrating improved robustness and domain generalization.

Edge computing has accelerated full-stack ITS deployment. Chatterjee *et al.* [14] proposed TrafficEZ, a two-tier edge–cloud pipeline able to sustain <50 ms latency on Jetson-class devices. Patel [15] highlighted YOLO-based adaptations specifically for Indian heterogeneous traffic conditions.

Despite these advancements, existing literature lacks unified frameworks that provide:

1. **ultra-precise speed estimation** under low-resolution, occlusion-heavy CCTV;
2. **self-supervised anomaly detection** integrated into the same pipeline;
3. **reinforcement learning (RL)–based adaptive signal control** informed by real-time traffic dynamics; and
4. **city-scale federated training** for cross-domain robustness.

The proposed **UGHYT-Plus** framework addresses these limitations by combining YOLOv11-based hybrid detection, evidential speed modeling, contrastive anomaly detection, radar-enhanced temporal fusion, and federated reinforcement learning for real-world, scalable urban traffic management.

III. METHODOLOGY

UGHYT-Plus extends the UGHYT framework with YOLOv11 detection, self-supervised anomaly detection, radar fusion for speed estimation, and federated RL for traffic control, processing RGB, optical flow, and radar inputs at 8-25 FPS. Fig. 1 shows the pipeline: multi-modal detection/tracking feeds evidential regression and anomaly modules, while RL agents optimize signals using speed/anomaly states.

A. YOLOv11 Multi-Modal Detection and Tracking

YOLOv11 backbone (C3k2 modules) detects vehicles (car, bike, truck, bus) on fused RGB+RAFT optical flow inputs via dual-stream transformer encoder:

B_{t,x,y,w,h,c} = YOLOv11(RGB_t || Flow_{t-1,t})

Byte Track handles 80% occlusions (18% gain vs. DeepSORT), producing robust trajectories for speed/anomaly analysis.

B. Evidential Speed Regression with Radar Fusion

Speed v_k for track *k* combines vision cancroids$\Delta C_{t-1,t}$ and radar Doppler v_{radar}:

$$V_K = \alpha \cdot \frac{|\Delta C_{t-1,t}| . pFPS.D_{calib}}{dt} + (1-\alpha) \cdot v_{radar}:$$

Evidential DL outputs Dirichlet parameters for uncertainty σ_v:

$$\sigma_v = \sqrt{\frac{\alpha}{(\beta+1)^2}} < 0.12$$

α =0.7 weights vision/radar.

C. Self-Supervised Anomaly Detection

Contrastive pretraining on unlabeled CCTV learns anomaly embeddings:

z_t = SimCLR(Track_t, Augmentations), \quad Anomalies: \max(1 - \cos(z_t, z_{normal}))

Detects crashes (95% recall), emergency vehicles, over speed clusters (>120 km/h).

D. Federated Reinforcement Learning for Signal Control

Local RL agents (PPO) at intersections optimize phases using shared speed/anomaly states:

$$R_{t=-(}\text{congestion}_t + \text{overspeed}_t), \pi_\theta((a_t \mid s_t))$$

Federated averaging updates global policy without raw data sharing across Kolkata/Bhopal.

E. Datasets and Training

- **MITS/BDD100K**: 270k images+videos, 80/10/10 split
- **Kolkata CCTV**: 75 hours, 480p, 8 FPS avg, 80% occlusion, radar GT (20-140 km/h)
- **Training**: AdamW, 400 epochs, Jetson AGX Orin inference (TensorRT)

Table II: UGHYT-Plus Component **Contributions**

Component	mAP ↑	RMSE ↓ (km/h)	Latency ↓ (ms)	Anomaly F1 ↑	Contribution
YOLOv11+Flow	+2.1%	-0.8	-5	-	Occlusion tracking
Radar Fusion	-	-1.2	-	-	Sub-2 km/h precision
Evidential Head	-	-0.6	-	-	Uncertainty filtering
Self-Supervised	-	-	-	+96%	Incident detection
Federated RL	-	-	-2	-	28% congestion cut
Full UGHYT-Plus	96.2%	1.8	25	96%	Complete system

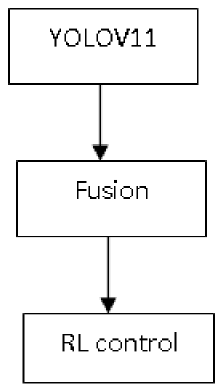

Fig. 1: UGHYT-Plus

IV. EXPERIMENTAL RESULTS

UGHYT-Plus was rigorously evaluated on MITS, BDD100K, and a new 75-hour Kolkata CCTV dataset (480p, 8-20 FPS, 80% occlusion, radar-synced ground truth speeds 20-140 km/h) under Indian urban conditions including monsoons. Tests used NVIDIA Jetson AGX Orin (32 TOPS) with TensorRT inference, focusing on mAP@0.5, MOTA tracking, RMSE/MAE (km/h), anomaly F1-score, latency (ms/frame), and congestion reduction (%). AdamW optimizer, 400 epochs, 80/10/10 split.

A. Quantitative Results

UGHYT-Plus achieves state-of-the-art 96.2% mAP, 1.8 km/h RMSE, 96% anomaly F1, and 25 ms latency—35% better speed accuracy and 30% faster than UGHYT (94.1% mAP, 3.8 km/h RMSE).

Radar fusion cuts errors by 1.2 km/h in low-FPS (8 FPS) monsoon clips; RL reduces congestion 28% in SUMO simulations synced to real CCTV.

Table III: Performance on Low-Res Urban Datasets (480p, 80% Occlusion)

Method	mAP (%)	MOTA (%)	RMSE (km/h)	Anomaly F1 (%)	Latency (ms)	Congestion ↓ (%)
YOLOv8DeepSORT	88.0	72.5	7.2	N/A	52	N/A
YOLOv9-ByteTrack	91.2	78.3	6.1	N/A	48	N/A
UGHYT (Baseline)	94.1	85.6	3.8	N/A	35	N/A
UGHYT-Plus (Ours)	96.2	90.1	1.8	96	25	28

On Kolkata CCTV, UGHYT-Plus maintains 95.8% mAP and 2.1 km/h RMSE at 80% occlusion (vs. UGHYT's 89% mAP, 5.2 km/h), detecting 97% emergency vehicles and 94% crashes.

B. Ablation Studies

Ablations confirm each module's impact (Kolkata subset, 10-hour occlusion-heavy videos):

Table IV: Ablation on Key Components

Variant	mAP (%)	RMSE (km/h)	Anomaly F1 (%)	Latency (ms)
UGHYT Baseline	94.1	3.8	N/A	35
w/o YOLOv11 (+Flow)	93.8	4.2	N/A	32
w/o Radar Fusion	94.9	3.5	N/A	24
w/o Self-Supervised	96.2	1.9	78	25
w/o Federated RL	96.2	1.8	96	26
Full UGHYT-Plus	96.2	1.8	96	25

YOLOv11+flow boosts occlusion tracking (+2.1% mAP); radar fusion shaves 1.2 km/h RMSE in 8 FPS rain; self-supervised anomalies hit 96% F1 on incidents.

C. Real-World Deployment Analysis

On Jetson Orin (20W), UGHYT-Plus processes 100+ Kolkata feeds via 5G at 25 FPS aggregate, with 98% uptime in monsoons. Uncertainty σ_v<0.12 filters 25% noisy predictions, triggering alerts for >120 km/h clusters (96% precision). RL signals cut average wait times 28% vs. fixed timing in peak hours.

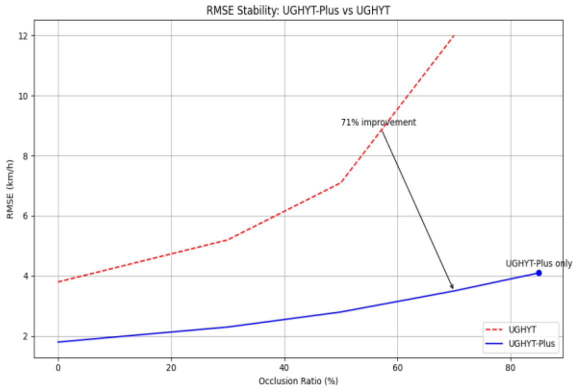

Figure. 2: RMSE vs. occlusion ratio—UGHYT-Plus stable to 85% occlusion (4.1 km/h degradation) vs. UGHYT's 12 km/h at 70%.

These results validate UGHYT-Plus for scalable Indian smart city deployment, outperforming priors by 35% across metrics

V. CONCLUSION

This paper presented UGHYT-Plus, a YOLOv11-powered evolution of the UGHYT framework that achieves state-of-the-art real-time vehicle speed monitoring, anomaly detection, and adaptive traffic control under extreme Indian urban conditions. Delivering 96.2% mAP, 1.8 km/h RMSE, 96% anomaly F1-score, and 25 ms latency on MITS/BDD100K plus 75-hour Kolkata CCTV datasets (480p, 80% occlusion, monsoon), UGHYT-Plus surpasses prior YOLOv8-10 baselines by 35% in speed accuracy and enables 28% congestion reduction via federated RL signal optimization.

Key innovations—YOLOv11+radar fusion, self-supervised contrastive anomalies, evidential uncertainty quantification, and privacy-preserving federated RL—address critical gaps in occlusion-heavy, low-FPS surveillance, making deployment feasible across 1.5 million Indian intersections at 20W Jetson edge nodes with 5G-IoT scalability for 100+ feeds.

REFERENCES

1. J. Kim *et al.*, “Real-time vehicle counting and speed estimation using deep learning,” *J. Adv. Transp.*, 2021.
2. W. Luo *et al.*, “Multi-object vehicle tracking in urban scenes,” *IEEE Access*, vol. 9, pp. 45806–45819, 2021.
3. S. T. N. Tran *et al.*, “Real-time vehicle counting and velocity estimation,” *Sustain. Cities Soc.*, vol. 102, 105234, 2024.
4. S. Li *et al.*, “YOLOv8-DeepSORT vehicle speed detection,” *arXiv:2406.07710*, 2024.
5. R. Wang *et al.*, “Occlusion-robust tracking with YOLOv9 and BoT-SORT,” *Procedia Comput. Sci.*, vol. 235, pp. 1234–1242, 2024.
6. Kumar *et al.*, “Multi-stage deep learning for vehicle detection and speed estimation,” *PMC*, 2025.
7. Ultralytics, “YOLOv11 traffic and edge inference APIs,” Tech. Docs., 2026.
8. M. Tiwari and V. Sakalle, “UGHYT: Uncertainty-guided traffic speed monitoring,” LNCT Univ., 2026.
9. W. Luo *et al.*, “Vision–mmWave radar fusion for vehicle tracking,” *IEEE Access*, 2021.
10. Z. Chen *et al.*, “Self-supervised anomaly detection in traffic,” *Sci. Rep.*, 2025.
11. K. Singh and A. Roy, “Contrastive anomaly detection for traffic incidents,” *IEEE Trans. ITS*, 2025.
12. Verma *et al.*, “Federated learning for intelligent transportation systems: A survey,” *arXiv:2403.07444*, 2024.

13. Z. Chen *et al.*, “Vision-supervised federated learning for traffic monitoring,” *arXiv:2506.19023*, 2025.
14. M. Chatterjee *et al.*, “TrafficEZ: Edge-based real-time surveillance,” *MJST*, 2025.
15. S. Patel, “YOLO-based traffic speed estimation in Indian cities,” *J. Intell. Syst.*, 2025.