KEC Conference 2026
ISSN: 3021-940X (Print) DOI: 10.3126/injet.v3i2.95661

Single-Camera AI-Based Traffic Analytics for Heterogeneous Intersections: A Deep Learning Framework for Microscopic Data Extraction from Smartphone Video

Bibek K.C.[1*], Ayushma Pokhrel[1], Deepika Pandey[1], Dikshya Karna[1], Dipika Dahal[1], Ramesh Marikhu[2], Ramesh Bala[1]

[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, bibekkc2027@gmail.com*
[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, ayushma.pokhrel1@gmail.com*
[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, deepikapandey566@gmail.com*
[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, dikshyakarnaofficial75@gmail.com*
[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, dahaldipika0@gmail.com*
[2]*Almonds A.I. Company Pvt. Ltd., Bhaktapur, Nepal Email:ramesh@almondsai.com*
[1]*Department of Civil Engineering, Khwopa College of Engineering, Bhaktapur, Nepal, bala.ramesh@khwopa.edu.np*

Abstract

In countries like Nepal, where traffic is heterogeneous, mixed, and non-lane-disciplined, the data collection methods used in transportation planning remain outdated and incapable of producing the microscopic parameters needed to understand local driving behaviors. This paper demonstrates a computer vision framework that converts footage from a single elevated smartphone camera (4K, 24 fps) into microsimulation-ready traffic data at the Koteshwor intersection, a busy three-legged junction in Kathmandu. The framework deploys YOLOv8x for vehicle detection, ByteTrack for multi-object tracking supported by a three-tier identity recovery mechanism, and per-approach planar homography, along with a post-processing pipeline for axial-length reclassification, speed correction, Wiedemann 74 safety-distance extraction, and complementary acceleration/deceleration metrics. From two independent peak-hour datasets yielding 31,882 total vehicle trajectories, 29 stratified validation clips covering 1,805 manually counted vehicles resulted in a 100% GEH pass rate and a class-wise Mean Absolute Percentage Error (MAPE) of 2.48%. The framework outputs volumetric counts across seven vehicle classes, Origin-Destination (OD) matrices, speed distributions, and trajectory-level microscopic parameters, demonstrating that a single smartphone setup can produce richer and more reliable traffic data than traditional manual surveys.

Keywords: Traffic analytics, YOLOv8, ByteTrack, heterogeneous traffic, Origin-Destination, computer vision, microsimulation, Wiedemann 74, homography, Koteshwor, Nepal

1. Introduction

The rapid urbanization of the Kathmandu Valley has led to a transportation crisis, with vehicle registrations increasing from 150,000 to over 570,000 in a single decade [1]. At major bottlenecks like Koteshwor intersection, infrastructure-heavy interventions including road widening have been implemented, but congestion persists. Addressing this pressure requires data-driven mobility strategies built on detailed microscopic analysis [2]. Extracting accurate naturalistic trajectory data serves as the foundation for this shift, aiding driver modeling and the development of robust traffic simulations [3].

In developing countries like Nepal, motorcycles dominate, accounting for approximately 70–80% of total traffic [1]. Their rapid acceleration, diamond-shaped maneuvering, and seepage behavior reduce average headways and increase saturation flow, significantly influencing intersection dynamics [4]. At dense intersections like Koteshwor, these conditions are further aggravated by frequent occlusion from larger vehicles and the absence of lane discipline. Under these conditions, manual counting becomes unreliable: documented evidence shows that classification-specific errors exceed 10% in high-volume, multi-class traffic [5], a finding corroborated by the validation performed in this study (Section 5.2).

* *Corresponding author*

The use of AI-driven computer vision to extract traffic data from video addresses these limitations directly. The YOLO family of Deep Neural Networks (DNNs) has emerged as the leading detection approach due to its high accuracy and computational efficiency [6], yet available pre-trained models lack local training data for regional vehicle classes, and motorcycle density causes severe identity fragmentation during tracking. Recent region-specific frameworks such as TRAMON [7] have improved detection accuracy for mixed traffic to 98–99% by training on localized datasets, yet their scope is limited to volumetric counting and does not extend to Origin-Destination (OD) extraction, speed profiling, or microscopic driving-behavior parameter estimation. To the authors' knowledge, no existing end-to-end framework handles the complete pipeline from raw video to transport-engineering-ready outputs for heterogeneous traffic.

This study develops and validates an integrated computer vision framework that converts raw smartphone video into trajectory-level microscopic data using a single camera, generating volumetric counts across seven localized classes, OD matrices, and empirical Wiedemann 74 microscopic driving-behavior parameters for microsimulation calibration. By combining trajectory data extracted via YOLOv8 and ByteTrack with homography-based spatial calibration, this framework provides a scalable, data-driven methodology for intersection evaluation, validated at the operationally complex Koteshwor junction in Kathmandu, Nepal.

2. Literature Review

YOLO-based detectors have been tested extensively for real-world traffic counting. Rouf et al. modified YOLOv7 to reduce missed detections of small, distant vehicles on urban highways [8]. In the Nepalese context, Kunwar et al. demonstrated the value of microsimulation-based evaluation under heterogeneous traffic conditions, highlighting the calibration challenges that arise from mixed vehicle classes and the absence of lane discipline [9]. Getting from frame-to-frame detections to stable trajectories is a complex tracking challenge. Bewley et al. showed the viability of Kalman filters paired with Hungarian matching in the SORT algorithm [10], but severe occlusion in motorcycle-heavy traffic causes simple association logic to fail. Zhang et al. tackled this with ByteTrack, which brings low-confidence detections back into the matching pool so partially hidden vehicles are retained rather than discarded [11].

Converting pixel-level trajectories into transport parameters requires mapping image coordinates to real-world dimensions. Al-Farisi et al. used single-camera setups paired with homography to project pixel tracks into a bird’s-eye coordinate plane [12]. Hietbrink demonstrated a similar technique to convert pixel displacements into physical distances, enabling localized OD extraction from a single camera [13]. Detailed trajectory data is critical for mixed-traffic modeling [20, 21], yet extracting empirical Wiedemann 74 safety-distance parameters from single-camera video remains unaddressed. This paper fills that gap by extending single-camera trajectory extraction to include empirical microscopic driving-behavior parameter estimation alongside volumetric counts, OD matrices, and speed profiling within a unified pipeline.

3. Study Area and Data Collection

The study area is the Koteshwor Intersection (27.678578°N, 85.349438°E), a major arterial node connecting central Kathmandu to Bhaktapur District via the Arniko Highway, meeting the Ring Road at this junction [14]. It functions as a three-legged, predominantly police-controlled intersection with heavy heterogeneous traffic and frequent weaving. Three approach arms were established: Tinkune (North/West), carrying Ring Road traffic; Bhaktapur (East), carrying Arniko Highway traffic; and Lalitpur (South), carrying commuter traffic from Patan municipality.

Video was recorded in 4K at 24 fps using an iPhone 16 mounted on the second-floor balcony of an adjacent building at approximately 7.5 m height, selected to maximize the field of view across all three approaches while maintaining sufficient pixel density for motorcycle detection; the roof level was inaccessible due to the building’s truss structure. Recording was performed during morning peak hours on two consecutive days: Day 1 (09:30–10:30 AM, August 13, 2025) and Day 2 (09:45–10:45 AM, August 14, 2025). The iPhone 16 was chosen for its optical image stabilization and sustained 4K capability, representing consumer hardware readily available to local transport agencies.

4. Methodology

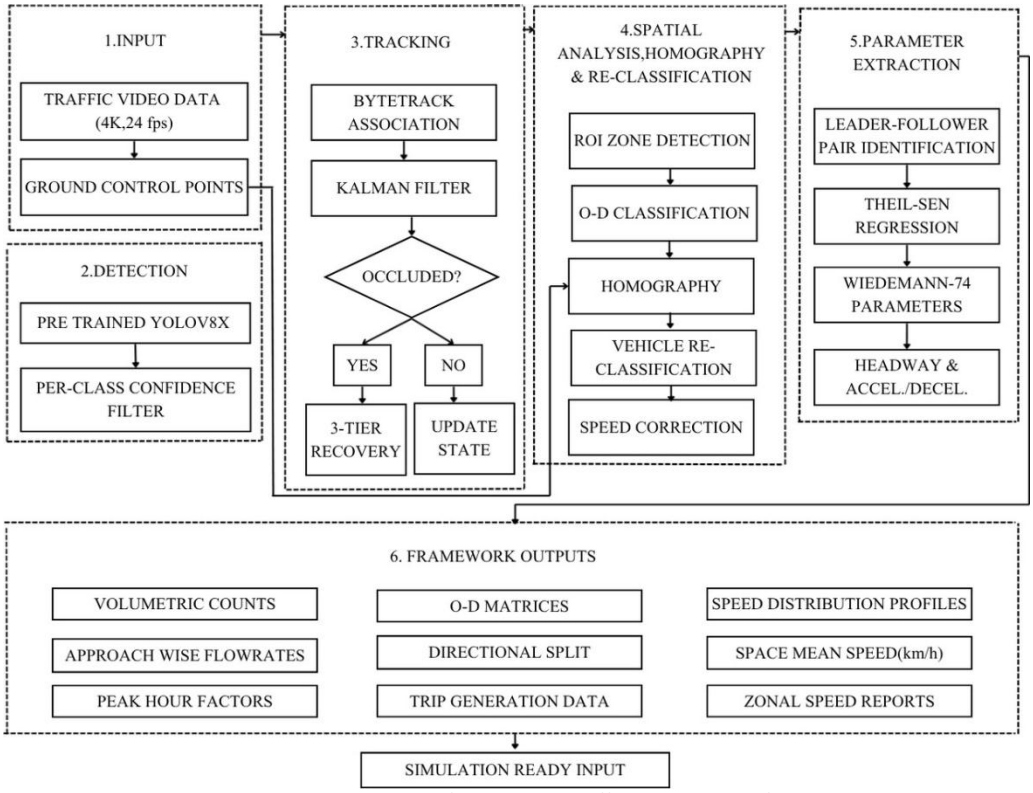

Figure 1. System architecture of the proposed traffic data extraction framework

The proposed framework operates through a multi-stage pipeline, as detailed in Figure 1.

4.1 Detection, Tracking, and Identity Recovery

YOLOv8x handles primary vehicle detection with class-specific confidence thresholds: motorcycles require lower thresholds (0.14–0.15) to capture occluded instances, whereas larger vehicles like buses and trucks are set higher (0.38–0.40). The NMS IoU is set at the default 0.70 to draw distinct bounding boxes for tightly grouped vehicles.

On the Lalitpur–Bhaktapur approach, two-wheelers tracked near the exit zone appeared distant and small, and lateral occlusion from adjacent vehicles obscured their edges, causing the standard pass to miss them. A Fast Super-Resolution Convolutional Neural Network (FSRCNN) [15] is applied to upscale those distant regions before a secondary scan, recovering over 60% of otherwise undetected motorcycles. The NMS IoU for this secondary pass is set at 0.30 to prevent duplicate boxes.

ByteTrack [11] manages multi-object tracking by associating high-confidence detections (>0.6) first, followed by a second pass for low-confidence detections (0.1–0.6). However, the severe occlusion caused by heavy heterogeneous traffic still fragments trajectories. To reconnect broken tracks, a custom three-tier recovery mechanism was implemented:

- **Lightweight ReID:** A 128-dimensional color-histogram feature vector is extracted and compared using cosine similarity to match lost detections back to their original trajectories [16].
- **DBSCAN [17]:** Cluster centroids maintained throughout a vehicle's track allow completely occluded motorcycles to be re-associated once they reappear, using a neighborhood radius (epsilon) of 0.45 and minimum cluster size of two points.
- **Occlusion-aware Kalman Filter:** A position-velocity model [x, y, vx, vy] dynamically widens its search radius (Q) by 1.3× during blind spots. Per-approach tuning prioritizes forward movement (Qy > Qx) on Lalitpur for accelerating traffic, while Bhaktapur uses higher measurement smoothing (R=50.0) to cancel 4K video jitter. In practice, the Kalman filter contributes most to recovery by maintaining predictions through extended blind spots, followed by ReID for moderate occlusion

gaps. DBSCAN provides supplementary recovery when both prior tiers fail, particularly for motorcycles that reappear with altered spatial profiles after full occlusion by larger vehicles.

4.2 Spatial Analysis, Homography, Reclassification, and Speed

Trajectory extraction was restricted to stable visibility zones where turning movements are most identifiable. OD paths were classified using ROI polygons for entry and exit boundaries, cross-checked with vector heading analysis and a secondary angle check relative to the x-axis. Two planar homography zones were created for perspective correction: a Central Zone covering the primary intersection area, and a separate T–B Zone for the distant Tinkune–Bhaktapur direction, each calibrated using a 3×3 homography matrix.

Vehicle reclassification relies on computed physical axial length, supported by bounding box width and height-to-width ratios. The raw YOLO output often misclassifies untrained regional vehicles based on the visible angle (e.g., frontal views of a Hiace yield "car," rear views of an LCV yield "truck"). By projecting bounding boxes into world-space and computing physical dimensions, the four generic YOLO classes were mapped into seven transport-relevant categories: Motorcycle (MC), Car, Bus, Heavy Commercial Vehicle (HCV), Hiace, LCV, and Tempo. Classification thresholds were adjusted to account for reprojection error, which increases marginally toward the periphery of each homography zone.

While the Hiace and LCV share a similar axial length (5-7 m), separating them using true physical width is restricted by the oblique camera angle; ground-plane ($z = 0$) homography introduces perspective errors along the transverse axis [18]. However, consistent 2D bounding box distortion patterns under this viewing angle allow the reclassification logic to separate them in the image plane, with validation MAPE of 2.7% for Hiace and 3.1% for LCV. Finally, the mapped world coordinates were used to extract continuous speed distributions across all classes and directions.

4.3 Microscopic Driving-Behavior Parameter Extraction

Leader-follower pairs were identified by checking velocity-vector alignment ($\cos > 0.7$) within 4 m lateral bands, producing 1,964 pairs and 138,460 frame-level observations spanning five directions, grouped into four extraction datasets by vehicle-orientation similarity for convergence testing.

Comparing results across directions differing in vehicle orientation and calibration geometry verifies that velocity-domain parameters are not artifacts of a specific camera angle. For standstill distance (ax), where YOLO and NMS lack the resolution to isolate bumper-to-bumper gaps below 1 m, a detection-bias correction was applied using constraints validated by regional literature.

5. Validation

5.1 Clip-Level Count Validation

The AI counts were validated against manual counts using stratified clips from both recordings, adopted because long-duration manual counting from video is prone to undercounting and observer fatigue [5]. Day 1 contributed 17 clips and Day 2 contributed 12 (Table 1). The class-level breakdown (Table 2) demonstrates high accuracy: motorcycles achieved a MAPE of 0.2%, with Hiace and LCV at 2.7% and 3.1%, respectively.

Table 1. Clip-level validation summary.

Metric	Day 1	Day 2	Combined
Validation clips	17	12	29
Vehicles (manual)	971	834	1,805
Vehicles (AI)	979	824	1,803
Count diff.	+8	-10	-2
Mean MAPE (%)	2.46	2.51	2.48
GEH<5 pass	17/17	12/12	29/29

Table 2. Per-class counting accuracy (29 clips).

Class	Manual	AI	Diff	MAPE%	% Total
MC	1,385	1,388	+3	0.2	76.7
Car	271	267	-4	1.5	15.0
Bus	70	69	-1	1.4	3.9
Hiace	37	38	+1	2.7	2.0
LCV	32	31	-1	3.1	1.8
Tempo	10	10	0	0.0	0.6
Total	1,805	1,803	-2	0.1	100

5.2 Hourly OD Validation

One full hour of Day 1 footage was manually counted and compared against the AI output. The system reported 16,010 vehicles with an overall error of 3.3% relative to the manual baseline, and individual approach errors ranging from 0.1% to +6.7% (Table 3). All five movements passed the GEH < 5 criterion, with the highest individual GEH of 3.53 (Bhaktapur–Tinkune) and an aggregate GEH of 4.09, both within the standard GEH < 5 acceptance criterion. An equivalent hourly validation for Day 2 was not repeated because duplicating a full-hour manual count at intersection-level volumes above 15,000 vehicles was operationally infeasible within the final validation schedule.

Table 3. Hourly OD validation, Day 1 (same-footage manual count).

OD Movement	AI	Man.	Diff	Err%	GEH	Pass
Bhaktapur–Tinkune	7,781	7,473	+308	+4.1	3.53	PASS
Tinkune–Bhaktapur	3,023	3,026	-3	-0.1	0.05	PASS
Lalitpur–Tinkune	2,427	2,346	+81	+3.5	1.66	PASS
Lalitpur–Bhaktapur	1,870	1,752	+118	+6.7	2.77	PASS
Tinkune–Lalitpur	909	900	+9	+1.0	0.30	PASS
Total	16,010	15,497	+513	+3.3	4.09	5/5

5.3 Spatial Calibration and Speed Verification

Homography zones configured for Day 1 yielded RMS reprojection errors below 1.0 m (Central 9-point: 0.43 m, T–B: 0.005 m). Given a calibration scale of approximately 0.013 m/pixel (from a 31.2 m reference span measuring 2,339 pixels), an error of 1.0 m corresponds to roughly 75 pixels, less than 2% of the 4K frame width, within established tolerances for single-camera trajectory analysis [12]. The mapping calculation was verified using synthetic trajectories with predefined durations at 24 fps; the homography transformer returned speeds matching the deterministic ground truth. As independent spatial evidence, the homography-derived axial lengths of reclassified vehicles consistently fall within manufacturer-specified dimensions [19]: motorcycles map to 1.86–2.01 m, cars and tempos to 3.4–5.4 m, Hiace and LCV to 5.1–7.0 m, and buses to 7.7–12.0 m. Since speed is computed from the same coordinate transformation applied to consecutive trajectory points, dimensionally consistent length recovery across an order-of-magnitude size range (1.86–12.0 m) confirms the spatial reliability of the underlying homography.

5.4 Baseline Comparison and Trajectory Diagnostics

To isolate the contribution of the proposed framework's components, the same YOLOv8x + ByteTrack detector-tracker backbone was evaluated in two configurations on identical clip intervals and ROI gates. The vanilla baseline used COCO-trained YOLOv8x at default confidence and NMS-IoU thresholds with untuned ByteTrack association, i.e. no upstream detection tuning, no identity recovery, and no downstream refinement. The proposed framework adds per-class confidence thresholds and tuned NMS-IoU at the detection stage (Section 4.1), a three-tier identity-recovery cascade (IoU, Re-ID feature, and Kalman prediction gates) at the tracking stage (Section 4.1), and homography-based reclassification, export-quality filtering, and trajectory refinement

downstream (Section 4.2). Sixteen manually counted 30 s clips were used for four common-visibility movements: Bhaktapur-Tinkune, Tinkune-Bhaktapur, Tinkune-Lalitpur, and Lalitpur-Tinkune. The Lalitpur-Bhaktapur movement was excluded because severe adjacent occlusion in the single-camera view required the FSRCNN-assisted super-resolution detection pass described in Section 4.1 and produced short exit-zone tracks rather than comparable common-visibility trajectories; it is therefore reported only in Table 5 diagnostics. Because the vanilla detector resolves only the four COCO super-categories (motorcycle, car, bus, truck), manual counts and both outputs were collapsed into four common families for a fair comparison: two-wheelers, car/light vehicles, buses, and truck/heavy vehicles. Movement-level absolute count error was 35 for the proposed framework and 284 for the vanilla baseline, an 8.1x reduction (Table 4); class-wise absolute error followed the same pattern (25 vs. 362 across the four families). GEH < 5 was met on 16/16 clips for the proposed framework and 5/16 for the baseline. The largest relative baseline degradation within the common-visibility set occurred on Tinkune-Bhaktapur, where occlusion and small/distant two-wheelers reduced the vanilla count to 37 against 101 manually counted vehicles; the proposed framework retained 98 vehicles and passed GEH in all four clips.

Table 4. Limited baseline count comparison over four common-visibility movements (four 30 s clips per movement; proposed framework vs. vanilla YOLOv8x + ByteTrack).

Movement	Manual	Proposed	Vanilla	Abs. error (Prop. / Van.)	GEH < 5 (Prop. / Van.)
B-T	543	561	417	18 / 126	4/4 / 0/4
T-B	101	98	37	3 / 64	4/4 / 1/4
T-L	209	207	187	2 / 22	4/4 / 3/4
L-T	130	118	58	12 / 72	4/4 / 1/4
Overall	**983**	**984**	**699**	**35 / 284**	**16/16 / 5/16**

Frame-level identity ground truth was not annotated, so MOT benchmark metrics (IDF1, HOTA, MOTA) are not reported. The diagnostics reported here are trajectory length and continuity, which are the properties the microscopic-extraction step consumes. Across the full 09:30-10:30 export, 16,010 trajectories were retained, with 10th, 50th, and 90th percentile lengths of 62, 130, and 234 frames and a median calibrated travel distance of 17.9 m (Table 5). Median continuity over the audited short clips was 1.00. At 24 fps the lower-tail track exceeds 2.5 s of detected motion, sufficient for the microscopic driving-behavior extraction of Section 4.3. The L-B movement has the lowest median calibrated distance (5.18 m), reflecting its short FSRCNN-assisted exit-zone recovery window under severe adjacent occlusion. Consistent with the stable-visibility-zone scope defined in Section 4.2, Table 5 should be interpreted as trajectory-length and continuity diagnostics rather than as a corridor-coverage assessment.

Table 5. Trajectory diagnostics from the final combined Day 1 trajectory export (09:30-10:30 AM).

Movement	Tracks	Frames p10 / median / p90	Median distance (m)
B-T	7,781	71 / 142 / 248	25.05
T-B	3,023	48 / 82 / 173	15.44
T-L	909	71 / 125 / 196	24.55
L-T	2,427	68 / 110 / 205	15.24
L-B	1,870	66.9 / 201 / 300.1	5.18
Overall	**16,010**	**62 / 130 / 234**	**17.9**

6. Results and Discussion

6.1 Traffic Volume, OD Matrices, and Modal Composition

Day 1 (09:30–10:30 AM) documented 16,010 vehicles across five OD movements (the Bhaktapur–Lalitpur movement was unobservable due to camera constraints). Day 2 (09:45–10:45 AM) documented 15,872 vehicles, a cross-day difference of just 0.86%, demonstrating operational reproducibility. Slight positive bias in some OD movements likely reflects minor manual undercounting or occasional AI track fragmentation,

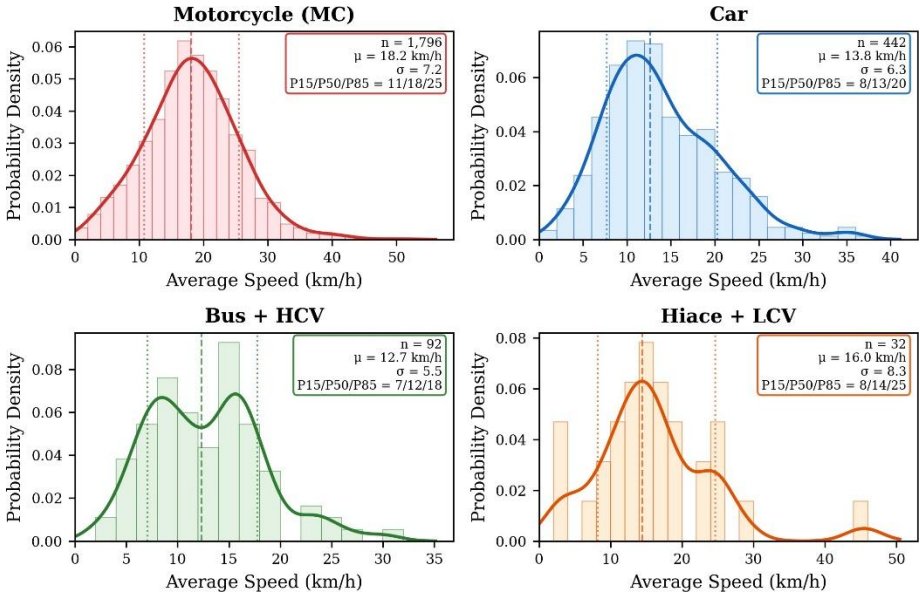

Figure 2. Speed distribution curves by vehicle class for the Lalitpur–Tinkune direction, Day 1

whereas undercounts result from complete physical occlusion. Table 6 details the Day 1 OD matrix; Table 7 the modal split, with motorcycles comprising 78.4%.

For context, the TRAMON framework [7], trained on a dedicated 282,000-image Vietnamese dataset, achieved 98–99% aggregate counting accuracy on heterogeneous traffic but was limited to volumetric totals. The present framework achieves comparable counting accuracy (MAPE 2.48%) without any region-specific training data, while additionally producing OD matrices, speed distributions, and microscopic driving-behavior parameters from the same video source.

Table 6. OD matrix, Day 1 (09:30–10:30 AM).

Origin ↓ \ Dest →	Bhaktapur	Tinkune	Lalitpur	Total
Bhaktapur	—	7,781	—	7,781
Tinkune	3,023	—	909	3,932
Lalitpur	1,870	2,427	—	4,297

Table 7. Class-wise vehicle distribution, Day 1

Vehicle Class	Count	Share (%)
MC	12,550	78.4
Car	2,460	15.4
Bus + HCV	649	4.1
Hiace + LCV	288	1.8
Tempo / 3-Wh	63	0.3
Total	**16,010**	**100.0**

6.2 Speed Profiles (Lalitpur–Tinkune)

Speed distributions were computed for every direction. The Lalitpur–Tinkune approach (n = 2,427) is detailed here (Fig. 2) due to clear class-level separation. Denoting 15th, 50th, and 85th percentiles as P15, P50, P85: motorcycles logged a mean of 18.2 km/h (P15/P50/P85: 11/18/25 km/h), cars 13.8 km/h (8/13/20 km/h), and buses/HCVs 12.7 km/h (7/12/18 km/h). The MC > Car > Bus hierarchy reflects physical dynamics where two-wheelers exploit lateral gaps that restrict larger vehicles. The 5.5 km/h operational gap between motorcycles and heavy vehicles underscores the necessity of class-specific microsimulation calibration. On the Bhaktapur–Tinkune approach (n = 7,781), saturated flow compressed all classes into a narrow 15–18 km/h band, eliminating the class-level speed differentiation visible under less congested conditions.

6.3 Microscopic Driving-Behavior Parameters

Table 8 lists W74 safety-distance parameters (ax, bx_add, bx_mult) separately from complementary acceleration/deceleration calibration metrics and cross-direction convergence statistics. The values align with established South Asian car-following research [3, 20], identifying aggressive driving dynamics intrinsic to Kathmandu intersections.

Table 8. Extracted microscopic driving-behavior parameters: W74 safety-distance parameters (ax, bx_add, bx_mult) and complementary acceleration/deceleration calibration metrics. * Constrained: confidence-weighted blend of measured and literature-validated estimates.

Parameter	Value	Default	Range (4 dirs.)	CV%	n
bx_add	1.63	2.00	1.00–1.90	25.2	25,660
bx_mult	2.09	3.00	1.08–2.33	29.9	25,660
Accepted decel (m/s^2)	1.23	—	1.21-1.26	1.6	4,497
Cooperative decel P75 (m/s^2)	1.14	—	1.06-1.18	4.5	8,827
Queue disch. P50 (m/s^2)	2.50	—	2.39–2.68	4.2	3,327
Queue disch. P85 (m/s^2)	3.55	—	3.47–3.68	2.5	3,327
ax (MC) (m)*	0.47	2.00	0.20–0.59	—	852
ax (Car) (m)*	1.02	2.00	0.80–1.21	—	813

The W74 safety-distance multipliers (bx_add = 1.63, bx_mult = 2.09) fall well below European defaults (2.00 and 3.00), depicting the tight following behavior characteristic of heterogeneous traffic. Their moderate cross-direction CV (25.2% and 29.9%) reflects individual following variation; pooled medians over n = 25,660 observations serve as the most reliable estimate. The complementary accepted-deceleration metric converges tightly (CV = 1.6%) across all directions, verifying that braking intensity is orientation-invariant.

Because YOLO and NMS struggle to isolate bounding boxes separated by less than 1 m, standstill distance (ax) metrics are constrained estimates. The motorcycle ax (0.47 m) is the midpoint of the bias-corrected lower bound (0.30 m) and raw pooled 15th percentile (0.59 m, n = 365); the midpoint was adopted because NMS suppression systematically inflates measured gaps, creating an asymmetric positive bias that makes the raw percentile an upper bound. The bounds of 0.20–0.59 m align with the 0.20–0.50 m motorcycle standstill gaps in comparable literature. At 0.47 m, motorcycle ax is less than half the car value (1.02 m), highlighting the necessity of per-class extraction for heterogeneous traffic models. The complementary acceleration/deceleration metrics converge tightly across directions (CV < 5%).

7. Conclusion

We presented a deep learning pipeline that extracts trajectory-level traffic data from a single elevated smartphone camera at low cost. Two independent peak-hour recordings at the Koteshwor intersection produced 31,882 vehicle trajectories in total. Counting accuracy was validated on 29 stratified clips against manual tallies, achieving a 100% GEH pass rate with a class-wise MAPE of 2.48%. The 138,460 frame-level observations allowed extraction of VISSIM-compatible Wiedemann 74 safety-distance parameters plus acceleration/deceleration calibration targets; these values can be used directly in VISSIM and as field calibration targets for SUMO or Paramics after model-specific mapping. Among the extracted parameters, motorcycle standstill distance was 0.47 m, less than half the corresponding car figure of 1.02 m, while accepted deceleration varied by only 1.6% across orientations. Both results reinforce that European defaults hard-coded in simulation packages do not hold for South Asian mixed traffic and must be replaced with field-calibrated values. For a country like Nepal, where motorcycle-dominated, non-lane-disciplined conditions differ fundamentally from the traffic regimes these defaults were derived from, a single smartphone recording session can produce validated traffic counts, OD matrices, speed distributions, and microscopic driving-behavior parameters that would otherwise require weeks of manual survey and specialized equipment. Source code and anonymized trajectory data will be made available by the corresponding author upon reasonable request.

8. Limitations and Future Work

The single elevated camera was optimized for the intersection core rather than full corridor coverage. As a result, one OD pair and some upstream/downstream approach segments fell outside continuous view, and trajectories should be interpreted as validated intersection-core tracks within stable visibility zones rather than complete entry-to-exit vehicle histories. Reclassification accuracy for Hiace and LCV is constrained by overlapping axial-length ranges, and YOLO's bounding boxes are less reliable for vehicle types absent from its training data. Standstill distance and close-range spatial parameters carry measurement uncertainty from perspective-dependent bounding-box errors and NMS suppression. Independent speed validation using radar or GPS probes was not performed; although the homography produced dimensionally consistent vehicle lengths across the full-size spectrum [19], direct speed validation would further strengthen the spatial outputs.

A practical next step is to curate a region-specific annotated image dataset for local vehicle classes, enabling future detector adaptation and reducing reliance on generic COCO categories. Future work will also extend the framework to synchronized multi-camera and drone/UAV viewpoints to cover all approaches continuously, reduce occlusion, and validate full entry-to-exit trajectories while preserving the low-cost single-camera workflow as a deployable baseline. Meanwhile, the trajectory and OD data can feed directly into VISSIM or SUMO calibration, and the high-resolution outputs support work on reinforcement learning-based signal control and naturalistic driver behavior studies.

Acknowledgements

The authors thank the Department of Civil Engineering at Khwopa College of Engineering for resources and guidance, and our supervisors for their mentorship. Thanks are also due to local authorities for their cooperation during video data collection at the Koteshwor intersection.

References

[1] JICA, "Data Collection Survey on Traffic in Kathmandu Valley," Japan International Cooperation Agency, 2017.

[2] B. K.C. et al., "A Multi-Faceted Approach to Urban Congestion in Developing Nations: Theory, Practice, and the Kathmandu Experience," JACEM, vol. 11, pp. 141–155, 2025.

[3] R. B. Amrutsamanvar, B. R. Muthurajan, and L. D. Vanajakshi, "Extraction and analysis of microscopic traffic data in disordered heterogeneous traffic conditions," Transp. Lett., vol. 13, no. 1, pp. 1–20, 2021.

[4] P. Neupane, "Estimation of Motorcycle Equivalent Unit Using Multiple Linear Regression and Impact of Motorcycle on Saturation Flow Rates," Master's thesis, Tribhuvan University, IOE, Pulchowk Campus, 2014.

[5] P. Zheng and M. McDonald, "An Investigation on the Manual Traffic Count Accuracy," Procedia – Soc. Behav. Sci., vol. 43, pp. 226–231, 2012, doi: 10.1016/j.sbspro.2012.04.095.

[6] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments," IEEE Access, vol. 8, pp. 1935–1944, 2020, doi: 10.1109/ACCESS.2019.2961944.

[7] T. H. Nguyen, T. Nguyen, and D. Vu, "TRAMON: An automated traffic monitoring system for high density, mixed and lane-free traffic," IATSS Res., vol. 47, no. 4, pp. 538–549, 2023, doi: 10.1016/j.iatssr.2023.10.001.

[8] M. A. Rouf, Y. Iwahori, Q. Wu, H. Wu, X. Yu, and A. Wang, "Real-time Vehicle Detection, Tracking and Counting System Based on YOLOv7," Embedded Self Organising Syst., vol. 10, no. 7, pp. 4–8, 2023.

[9] D. B. Kunwar, R. Pradhananga, S. Parajuli, and S. Shrestha, "Enhancing Traffic Operations and Safety Performance at Roundabout Intersections: A Simulation-Based Case Study of Dhalkebar Roundabout on the East-West Highway," J. Transp. Syst. Eng., vol. 1, no. 1, 2025.

[10] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in Proc. IEEE ICIP, 2016, pp. 3464–3468.

[11] Y. Zhang et al., "ByteTrack: Multi-object tracking by associating every detection box," in ECCV, 2022, pp. 1–21.

[12] M. Al-Farisi et al., "Vehicle Speed Estimation Using Consecutive Frame Approaches and Deep Image Homography," IEEE Access, vol. 12, pp. 97444–97456, 2024.

[13] N. Hietbrink, "Visitracker: Feature extraction by single-camera tracking for origin/destination information," Master's thesis, Univ. Amsterdam, 2002.

[14] JICA, "Preparatory survey for Koteshwor intersection improvement project in Nepal," Japan International Cooperation Agency, 2024.

[15] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in ECCV, 2016, pp. 391–407.

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in Proc. IEEE ICIP, 2017, pp. 3645–3649.

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. KDD, 1996, pp. 226–231.

[18] S. Lee, Y. Kim, and J. Kwon, "CNN-Based Vehicle Bottom Face Quadrilateral Detection Using Surveillance Cameras for Intelligent Transportation Systems," Sensors, vol. 23, no. 15, p. 6688, 2023.

[19] Tata Motors, TVS Motor Company, and Toyota Motor Corporation, "Vehicle specification catalogues and technical datasheets." Accessed: Jun. 4, 2026. [Online]. Available: https://www.tatamotors.com, https://www.tvsmotor.com, https://www.toyota.com.

[20] V. Kanagaraj, G. Asaithambi, T. Toledo, and T.-C. Lee, "Trajectory data and flow characteristics of mixed traffic," Transp. Res. Rec., vol. 2491, no. 1, pp. 137–147, 2015.

[21] N. Raju, S. S. Arkatkar, S. Easa, and G. Joshi, "Data-driven approach for modeling the nonlane-based mixed traffic conditions," J. Adv. Transp., vol. 2022, Art. no. 6482326, 2022.