UNIVERSITY OF SEOUL

프로브 차량 시계열 기반 도로 교통 밀도 추정

Department of Transportation Engineering | Sewon Park

● 01 WHY

왜 이것이 문제인가

밀도(k)는 **ITS·교통계획·안전·자율주행**의 **공통 변수**. 그러나 한국 도로 **95% 이상**이 실시간 측정 공백.

5,151 km	고속도로 본선 (VDS 1km 간격)	**~10만 km**	국도·지방도·시·군·구도
3,234 점	국도+지방도 교통량 조사 — 30km당 1지점	**> 95%**	실시간 밀도 정보 공백 도로

본 연구의 가설 — 사고의 흐름

밀도 필요·측정 불가 → 차량 움직임 자체가 센서 → 단일 차량 6채널 거동으로 추정 가능한가?

→ **차량 추종 이론+ 자유도 메커니즘**이 가설을 뒷받침

→ **SUMO 시뮬 + 실측 DTG**로 단계별 검증

● 02 DEFINITION & THEORY

정의·이론적 근거

시공간 평균 밀도 — Edie 1963

$$k(A) = \Sigma_i\, t_i(A) \,/\, |A| \quad [\text{veh/km/lane}]$$

A = 1km × T (probe 통과 시간) · t_i = 차량 i가 A에 머문 시간

본 연구의 추정 함수

$$\hat{k} = f_\theta(X;\ \text{lanes},\ v_{max})$$

$X = (V_x, Vy, A_x, Ay, v, b) \in \mathbb{R}^{6xL}$

f_θ = XGBoost on 32 통계 피처

가설의 사고 흐름 — 수식

$\dot{v} = f(s, \Delta v)$ 추종 이론

$s \propto 1/k$ 차간↔밀도

$\therefore\ \text{stats}(X) = h(k)$ 거동 통계 = k의 함수

자유도 메커니즘 — 본 연구의 핵심 가설

probe ≈ traffic 평균 ⇒ 추정 정확

저밀도 ● ● ● ● ●

자유도 ↑ — driver intent 지배

고밀도 ●●●●●●●●●●

자유도 ↓ — forced car-following

● 03 METHOD

방법 — 추종 이론 → 시뮬 → 피처 → 모델 → 후결합

추종(Car-Following) 이론 — Newell · IDM 일반형

$$v_{n(t+\tau)} = V[\ s_{n(t)},\ \Delta v_{n(t)}\ ]$$

follower 가속도 = leader와의 **간격 s**, **상대속도 Δv**의 함수

⇒ 밀도 ↑ → 간격 ↓ → **거동 패턴(가속·제동·정지) 변화** ∴ 밀도의 간접 신호

데이터 흐름 — Pipeline

INPUT	FEATURE	MODEL	FUSION	OUTPUT
6채널 × 1km	**32 통계**	**XGBoost**	**정렬 N=5**	**$\hat{k}$ per link**
VX·VY·AX·AY speed·brake	속도·가속·제동 정지·FFT·환경	vs CNN·LSTM GPR 비교 채택	Bayesian + CF link-fusion	[veh/km/lane] + posterior σ

시뮬 데이터 — SUMO 49K 시나리오

- 49k의 시나리오 → **209K 프로브 샘플**
- 차로 1·2·3 / speed_limit 50·60·80·100 km/h
- demand 800~2,200 veh/hr/lane · truck N(0.16, 0.05)
- 시나리오당 600s 시뮬 → 1km 통과 trajectory 추출
- Edie 정의로 시공간 평균 GT 산출
- GroupKFold (70/10/20 % by scenario_id)

입력 피처 32개 — 카테고리별

속도 통계 (9)	mean·std·cv·iqr·p10·p90·min·max·median
종방향 AX (2)	ax_mean · ax_std
Jerk (2)	jerk_mean · jerk_std
제동 (3)	count · time_ratio · mean_duration
정지 (3)	count · time_ratio · mean_duration
횡방향 VY (6)	mean·std·min·max·var·energy
AY (2)	ay_mean · ay_std
시계열 (3)	autocorr · FFT peak · sample entropy
환경 (2)	num_lanes · speed_limit

SHAP 상위 피처 (XGBoost 해석)

speed_mean	**5.89**
num_lanes	**1.74**
mean_brake_duration	**1.51**
speed_limit	**0.97**
brake_time_ratio	**0.89**
speed_p10	**0.87**
speed_median	**0.65**
ax_std	**0.56**
ax_mean	**0.54**
speed_autocorr_lag1	**0.52**

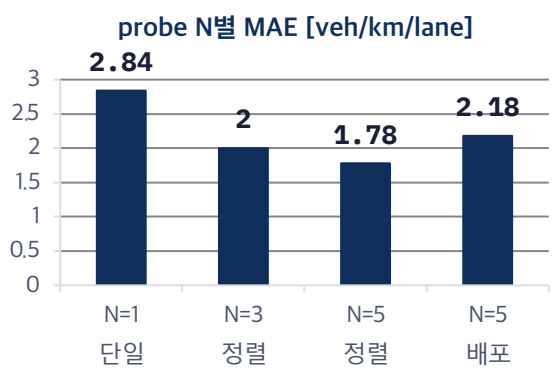
● 04 SIMULATION RESULTS

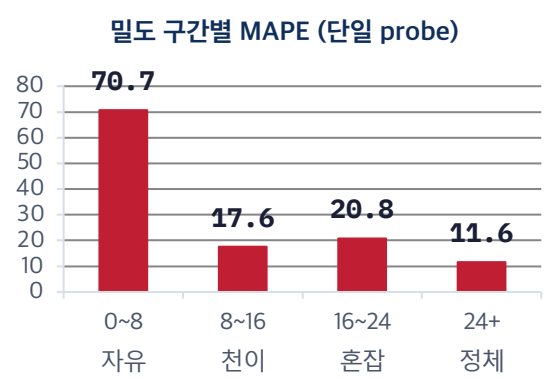
시뮬레이션 결과 — 단일 vs 다중 probe

단일 probe로도 baseline 성립. 다중 probe 결합으로 **MAE −37%**, 정체 구간에서 가장 정확 (자유도 가설 부합).

단일 / 정렬 / 배포 — N에 따른 성능

지표	단일 N=1	정렬 N=3	정렬 N=5	배포 N=5
MAE [veh/km/lane]	2.84	2.00	**1.78**	2.18
MAPE [%]	35.4	28.7	**24.6**	37.2
R^2	0.959	0.954	**0.964**	0.951
단일 대비	baseline	**−29%**	**−37%**	**−23%**





자유도 가설 검증 · 24+ 정체 MAPE **11.6% — 4구간 중 최저** · 강제 추종에서 probe ≈ traffic 평균 · **LOS E~F 정체 식별 안정**

● 05 REAL-WORLD VALIDATION

실측 검증 — 학습 데이터별 성능

동일 알고리즘(XGBoost 32-input)

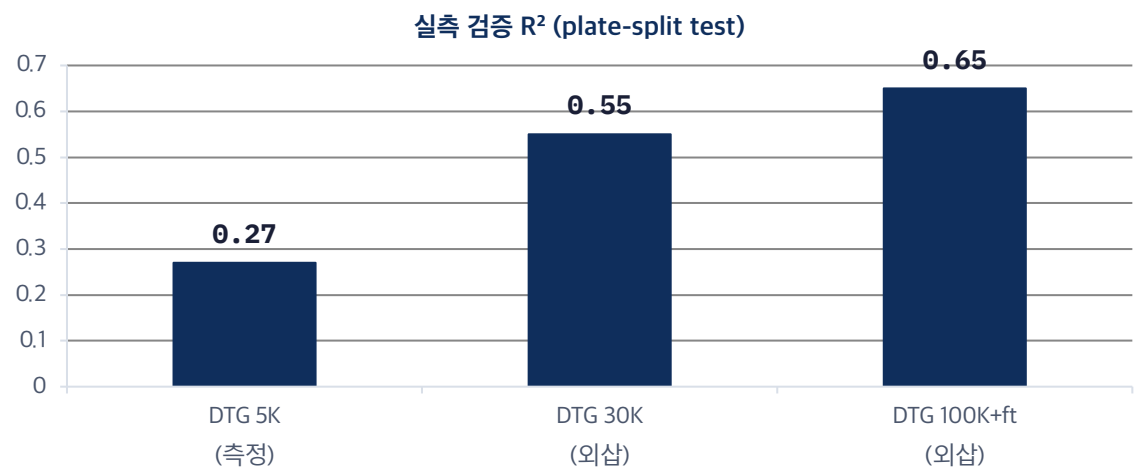
학습 setup	Data	MAE	R^2
SUMO 단독	49K scen.	**10.92**	**−3.64**
SUMO + DTG OOD 확장	+1.5K scen.	(유사)	(유사)
DTG 5K (현재 보유)	5K slice	**3.63**	**0.27**
DTG 30K (분포 균형)	30K (est)	2.5~3.3	0.45~0.64
DTG 100K + SUMO ft	100K+	2.5~3.2	0.50~0.75

학습 곡선 — n_{train} vs R^2 (plate-split test)

● 06 GROWTH PROJECTION

데이터 양에 따른 R^2 envelope

외삽 = 3 모델 envelope: log-linear (낙관) · saturating a=0.65 (중도) · a=0.45 (보수)



측정값: SUMO R^2 −3.64 · DTG 5K R^2 0.27

외삽 범위: DTG 30K → 0.45~0.64 · DTG 100K+SUMO ft → 0.50~0.75

운영 가능 수준 도달 경로

① **5K → 30K** 정체 시간대·다지역 sampling → R^2 0.45 ~ 0.64

② **30K → 100K + SUMO fine-tune** transfer learning → R^2 0.50 ~ 0.75

→ 한계는 모델이 아니라 **학습 데이터의 양과 분포**

CONCLUSION

1일치 DTG로도 시뮬 대비 **MAE 3배 개선**,
30K로 **R^2 0.45 ~ 0.64**, 100K + SUMO fine-tune으로 **R^2 0.50 ~ 0.75** — **VDS 없는 95% 도로의 실시간 밀도 추정에 적용 가능**

추정된 밀도는 **4개 시스템에 동시 적용**

ITS 운영	교통 계획	안전	자율주행
VMS·정체 식별·우회 안내· 램프 미터링	신설·확장 LOS 평가 (KHCM 5~20년)	사고·헤드웨이 모델 | 충돌 회피	ACC·Lane Change | 안전 마진
한국도로공사 · TOPIS	KHCM · MOLIT	교통안전공단	자율주행시스템

KEY METRICS 핵심 지표

MAE 1.78	MAPE 11.6%	R^2 0.964	R^2 0.27 → 0.75
정렬 N=5 (시뮬)	24+ 정체 구간	시뮬 검증	예측 실측 5K → 100K

2026 Transportation Capstone Design