5 Appendix

5.1 Traffic camera distribution and traffic detector setting

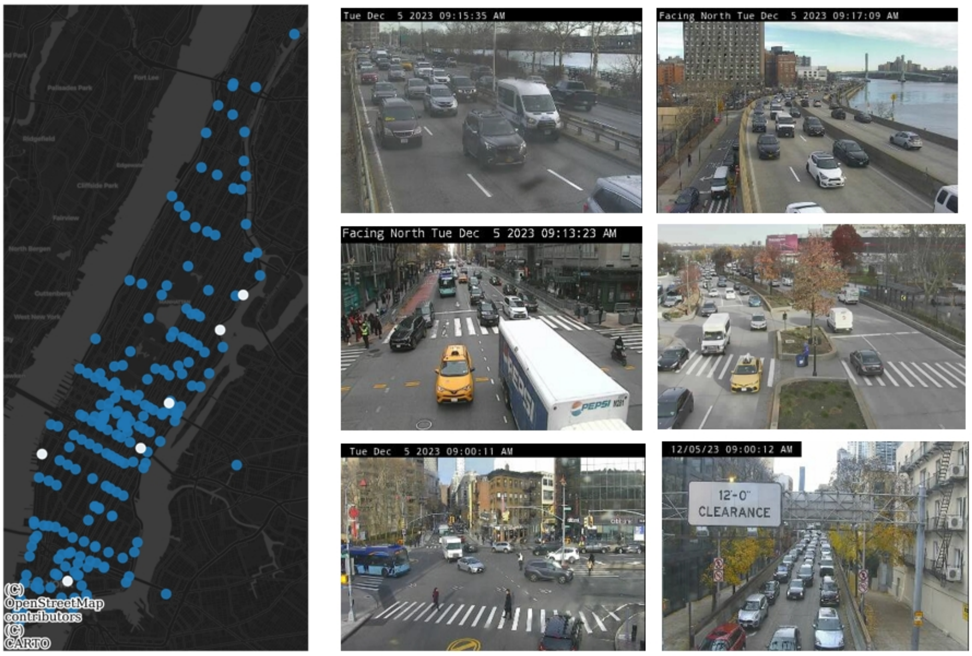
A total of 914 traffic cameras (https://webcams.nyctmc.org/api/cameras/) provide live feeds in NYC, with 331 located in Manhattan. After manually reviewing each camera to exclude those with poor views, 309 cameras are used for analysis, among which 267 update their views every 2 seconds, and the remaining update every 5 seconds. The resolutions of videos vary from 360 $\times$240 pixels to 1920 $\times$1080 pixels, with over 80% at the latter resolution. Extended Data Figure 1 (a) presents the spatial distribution of these cameras along with some snapshots taken on December 5, 2023, from 09:00-10:00 a.m. As shown, traffic cameras cover a variety of road types and intersections, containing rich on-road vehicles and traffic information with high temporal resolution and dense spatial coverage. For this study, we analyze traffic camera footage from four one-week periods in 2024 (the first week of January, April, August, and December) to determine fleet composition and signal timing. These patterns are then applied to other days, assuming that fleet composition and signal timing remain relatively stable over time. Beginning in January 2025, we expand data collection to a weekly basis to closely monitor the effects of congestion pricing.

One main challenge is linking cameras to the road network, as a single camera may cover multiple roads or directions, particularly at intersections. To address this, we first map the cameras on Google Maps based on their spatial coordinates. Next, we manually compare the Google Maps Street View with the camera footage to identify the road segment sharing the same view. After locating the road segment, separate sets of detectors are assigned to each road and matched to the corresponding segment (Extended Data Figure 1 (b)). To accommodate different camera configurations, four parallel detectors are placed per direction to optimize detection accuracy. We analyze the impact of detector location on recorded traffic volume and find a robust pattern when detectors are positioned near the stop line (Extended Data Figure 1 (c), C3 and C4). Conversely, placing detectors farther from the stop line (Extended Data Figure 1 (c), C1 and C2) results in underestimations of traffic volume, as vehicles appear smaller and are less likely to be detected by the computer vision algorithm.

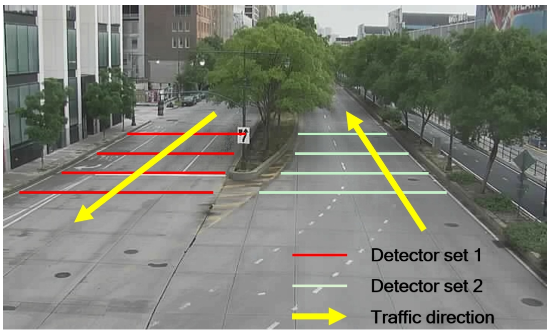
5.2 Vehicle type classification

We use the Bing Image API to collect images. For each retrieved image, we use YOLOv8 to detect objects classified as cars, motorcycles, buses, or trucks, crop them out, and save them as individual images. To ensure dataset quality, images smaller than 50 $\times$50 pixels are also removed. In total, 6,382 images are collected.

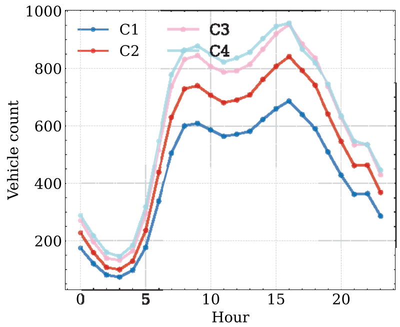
We test a total of 10 image classifiers from the *timm* library [71], including ViT (`vit_small_patch16_224`), Swin (`swin_base_patch4_window7_224`), ConvNeXt (`convnext_tiny`), RepVGG (`repvgg_a2`), Inception-v4 (`inception_v4`), ResNet-50 (`resnet50`), DenseNet-201 (`densenet201`), InceptionNeXt (`inception_next_tiny`), Xception (`xception71`), and EfficientNet-v2 (`efficientnetv2_rw_t`). Images are resized to 224×224 pixels by cropping and reflection padding before training. All models are pre-trained on the ImageNet-1k dataset [72] and fine-tuned for the new dataset



(a) Map and view of cameras



(b) Detector setting



(c) Hourly volume by detector

Extended Data Fig 1: | **Map and view of cameras and detectors.** (a) Six cameras (marked in white) are randomly selected to show their snapshots: Worth Street @ Bowery; FDR Dr @ 96 Street; 2 Ave @ 42 St; FDR Dr @ 111 ST; 11 Ave @ W 23rd St; 62 St @ QBB Upper-Level exit ramp. (b) Locations of traffic detectors. Four detectors are placed for each direction. (c) Traffic volume collected by different detectors.

using the *timm* framework [71]. The training parameters are as follows: batch size = 32, training/validation/testing ratio = 8:1:1, number of epochs = 15, and early stopping if validation loss does not decrease 10^{-3} for 5 epochs. The initial learning rate is chosen based on the loss-learning rate curve showing the sharpest downward slope.

Extended Data Figure 2 (a) shows learning curves on the training dataset. The training losses decrease rapidly and converge after several epochs, albeit at varying convergence speeds. Extended Data Figure 2 (b) presents the relationship between the number of parameters and testing accuracy for all models. As shown, Swin [73] achieves the highest accuracy (93.1%) but has the most parameters (the lowest efficiency). ConvNeXt [60] achieves a slightly lower accuracy (92.9%) but has significantly fewer parameters (much higher efficiency). Considering both accuracy and efficiency, ConvNeXt is selected as the final image classifier. The class-wise performance of ConvNext, as illustrated in Extended Data Figure 2 (c), demonstrates high accuracy values (81-100%) along the diagonal, with particularly high performance in identifying motorcycles (98.8%), refuse trucks (98.0%), and school buses (100%). Some vehicle types, such as intercity versus transit buses and passenger versus single-unit trucks, remain challenging to distinguish, likely due to their similar shapes and appearances.

5.3 Traffic volume validation

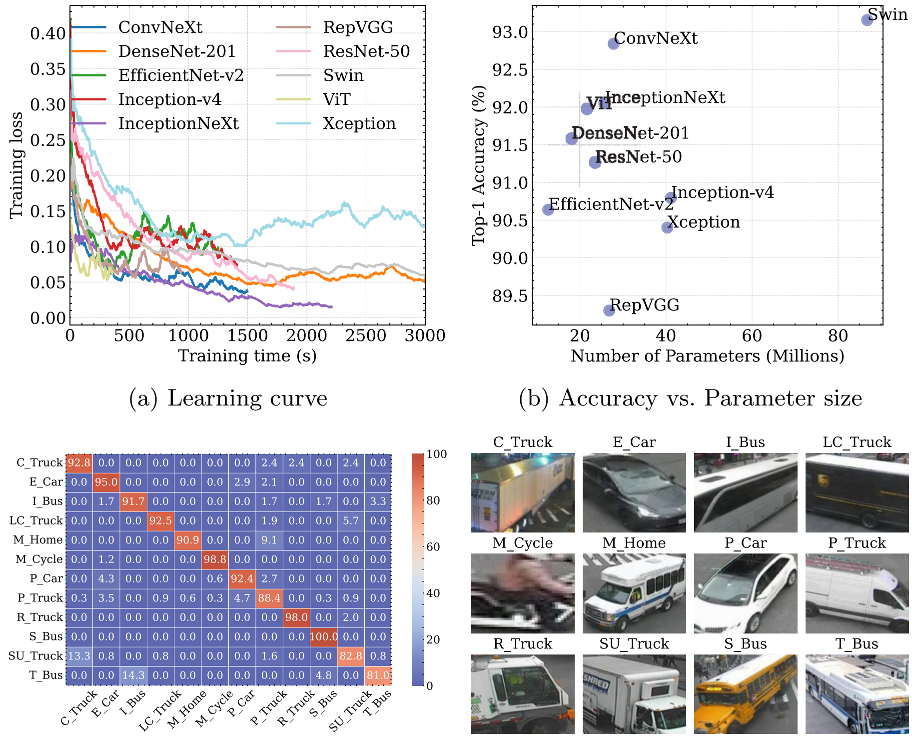
To account for potential underestimation due to low FPS, we weight the camera-based traffic volume by its frame update frequency (1/FPS). We then compare the camera-based volume with the ground truth for validation. The NYC Department of Transportation uses automated traffic recorders (ATR) to collect traffic volume. Since ATR counts are not year-round and the number of recorded days per location varies annually, we compare the hourly averages for the same day of the week within the same month between camera-based and ATR-based volumes for matching links. A total of 60 cameras can find a corresponding ATR, and the MAPE for the hourly average volume can achieve 16.31%, with a Pearson correlation of 0.93 (Figure 1 (b)).

The validation is also conducted on an hourly basis. The results show that camera-based volumes tend to underestimate the ground truth during nighttime, which is mainly due to poor lighting conditions that affect the performance of computer vision algorithms. During the daytime, camera-based volumes exhibit much higher accuracy, with an average MAPE of 13.07%. However, daytime estimates tend to slightly overestimate the ground truth, likely due to the complexity of traffic conditions and mixed directions of traffic flows within one camera's view.

Note that given the absence of key camera configuration details, such as height, view angle, and focal length, and the limitations of tracking algorithms for low-FPS, low-resolution videos, we opt not to use camera footage to compute traffic speed, as this process requires precise distance estimation.

5.4 Fleet composition

We show the spatial distribution of hourly volume by vehicle types in Extended Data Figure 3. For each vehicle type, data from three time periods are presented: midnight (03:00–04:00 a.m.), morning peak (08:00–09:00 a.m.), and afternoon peak (04:00–05:00 p.m.). Significant spatiotemporal differences in fleet composition are observed. For instance, buses are predominantly concentrated in Midtown, trucks are mainly concentrated in suburban areas, and passenger cars are mainly located along the outer



(a) Learning curve

(b) Accuracy vs. Parameter size

(c) Confusion matrix (ConvNeXt)

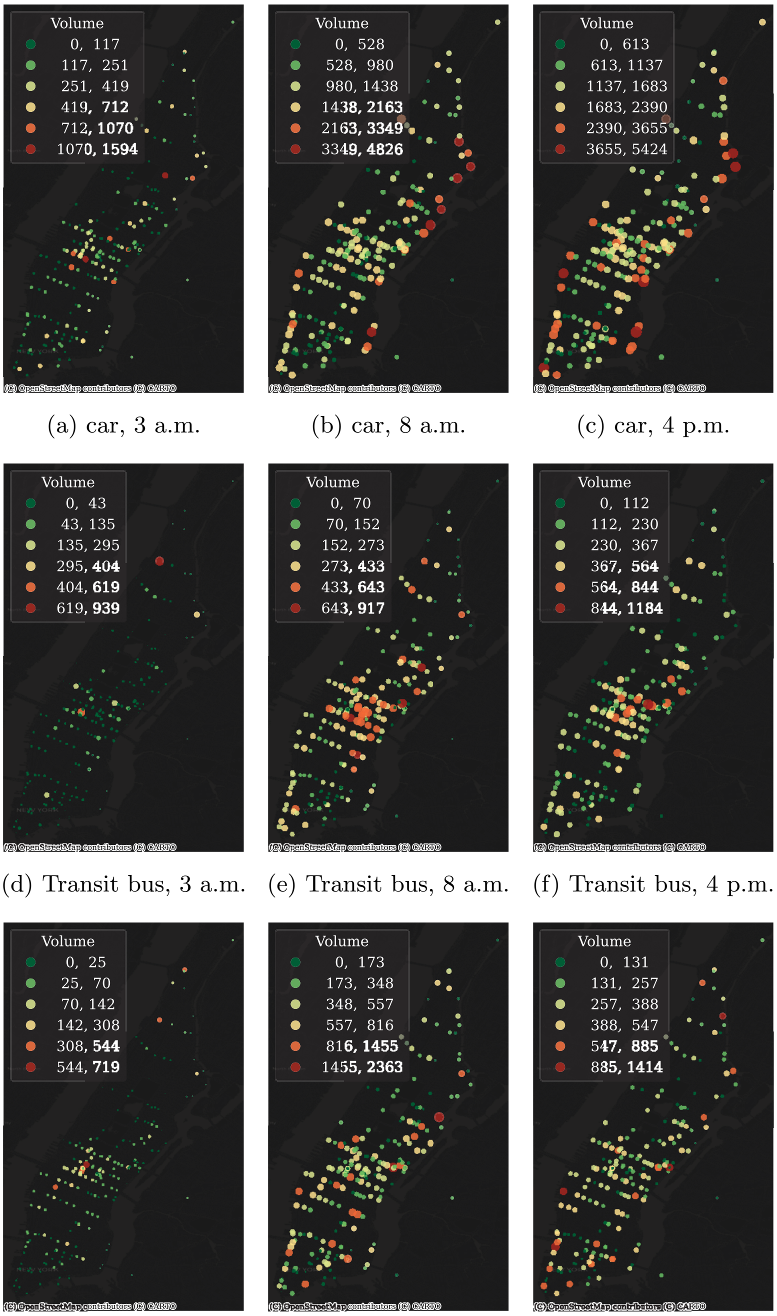
(d) Captured vehicles by type

Extended Data Fig 2: | **Vehicle type classifier performance and outcome.** (a) Learning curves for the ten image classifiers. (b) Scatter plot showing the number of parameters (M) versus Top-1 Accuracy (%). (c) Confusion matrix for ConvNeXt. Twelve vehicle types are considered: combination truck (C_Truck), electric car (E_Car), intercity bus (I_Bus), light commercial truck (LC_Truck), motorcycle (M_Cycle), motorhome (M_Home), passenger car (P_Car), passenger truck (P_Truck), refuse truck (R_Truck), single-unit truck (SU_Truck), school bus (S_Bus), and transit bus (T_Bus). (d) Examples of extracted images from camera footage for each vehicle type.

rings. These variations underscore the importance of accounting for hyperlocal fleet composition when estimating on-road traffic emissions.

5.5 Signal timing inference

For validation, we manually check 60 intersections (10 for each type) and report the errors in Extended Data Table 1. Note that the road network is sourced from OSM, which includes highways, trunks, primary roads, secondary roads, residential roads, tertiary roads, and unclassified roads. No signalized intersections are observed on highways or trunks. In addition, for simplification, we combine “unclassified” and “tertiary” roads as “residential”.



(a) car, 3 a.m. (b) car, 8 a.m. (c) car, 4 p.m.

(d) Transit bus, 3 a.m. (e) Transit bus, 8 a.m. (f) Transit bus, 4 p.m.

(g) Truck, 3 a.m. (h) Truck, 8 a.m. (i) Truck, 4 p.m.

Extended Data Fig 3: | **Spatial map of passenger cars, transit buses, and trucks.** (a-c) Hourly passenger cars; (d-f) Hourly transit buses. (g-i) Hourly trucks.

Extended Data Table 1: Summary of signal timing

Type	**Total count**	**Max. cycle**	**Min. cycle**	**Avg. cycle**	**Ground truth**	**MAE**
primary-residential	952	125	70	97	100	3
residential-secondary	910	115	60	76	80	4
residential-residential	414	90	45	60	65	5
secondary-secondary	205	135	70	88	90	2
primary-secondary	198	140	90	113	110	3
primary-primary	183	145	110	133	135	2

5.6 OD matrix

The OD matrix data is obtained through the Data for Good program by *Cuebiq* [61]. Aggregated mobility data are provided by Cuebiq, a location intelligence platform. Data is collected from anonymized users who have opted in to provide access to their location data anonymously, through a CCPA and GDPR-compliant framework. Through its Social Impact program, Cuebiq provides mobility insights for academic research and humanitarian initiatives. The Cuebiq responsible data sharing framework enables research partners to query anonymized and privacy-enhanced data, by providing access to an auditable, on-premise Data Cleanroom environment. All final outputs provided to partners are aggregated in order to preserve privacy.

To calculate population weighting factors, we compare the census block group (CBG)-level population size (from the 5-year American Community Survey 2020) with the number of devices assigned to each CBG during the year 2020. Home Census Block Group assignments are determined using three variables: the number of days spent at a location in the past month, the daily average number of hours spent there, and the time of day (nighttime/daytime) spent at the location [35, 61]. We find a strong positive correlation between the total population and the total number of devices ($\rho = 0.73$), with an average penetration rate of approximately 21.49% annually.

5.7 Dynamic traffic assignment (DTA)

We used DTALite to conduct the DTA. DTALite is an open-source mesoscopic traffic simulator that includes both static traffic assignment and dynamic traffic simulation to reflect the impact of road capacity constraints [36, 65]. Three traffic stream models, namely, point queue model (for secondary, tertiary, and residential roads), spatial queue model (for primary and secondary roads), and simplified kinematic wave models (for motorways), are embedded in the mesoscopic simulator to describe queuing behavior at bottlenecks with tight capacity constraints. Developed in C++, DTALite supports parallel computing on shared-memory multi-core systems, enabling large-scale network simulations with high efficiency. In this section, we briefly outline the inputs prepared for the DTALite simulation.

Simulation network: Road networks from OSM cannot be directly used for traffic assignment due to missing details such as the number of lanes, free flow speed, and link capacity, as well as their unstandardized network format, such as mixed use of directed and undirected links. To address these, we use the *osm2gmns* package [63] to convert the OSM network into an assignment-ready format. *osm2gmns* enriches OSM by adding essential link attributes such as lanes, free flow speed, and capacities, and standardizes the network by converting bi-directional roads into pairs of directed links. Finally, we include 418,602 links covering 31,767 miles in the simulation, which is a large-scale network that would typically require several hours to run on traditional traffic simulators.

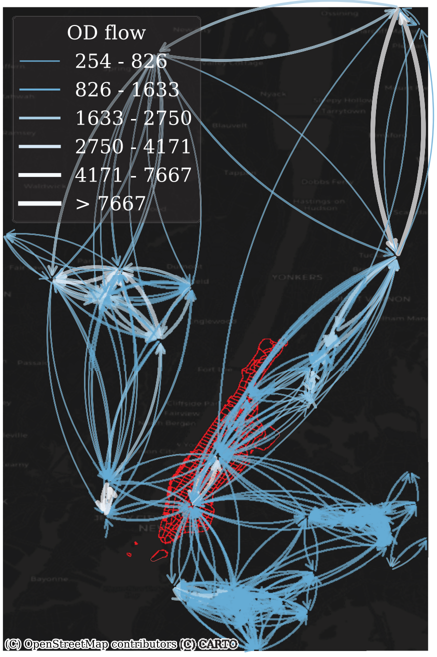
OD matrix: The simulation includes 2,744 TAZs within the modeled area. The connection between TAZ and the road network is via the centroid connector. To reduce the computational load of large-scale network simulation, we focus on Manhattan as the core area and aggregate OD flows outside Manhattan based on their importance. The importance of an OD pair is determined by the total traffic volume passing through Manhattan. After aggregation, the final network includes 401 zones within Manhattan and 556 zones (originally 2,343) outside Manhattan. As shown in Extended Data Figure 4 (a-b), traffic volume outside Manhattan is sparse and primarily concentrated on major roads, whereas more detailed traffic patterns are simulated within Manhattan.

OD matrix calibration: As shown in Extended Data Figure 4 (c-d), OD matrix calibration results in a steady decrease in MAPE but also causes a sharp increase in the user equilibrium (UE) gap after a certain number of iterations. This is due to the trade-off in the objective function of OD matrix calibration. The model aggressively minimizes the gap between observed and assigned traffic volumes while compromising the UE condition. To avoid this, we terminate the iteration at 38, where the UE gap remains below 0.1%, while MAPE also remains at low values of 20.18%.

Speed-density calibration: We compare road traffic speed before and after speed-density calibration in Extended Data Figure 5. After calibration, the R^2 improves significantly, ranging from 0.892 to 0.942 across different road types. In contrast, R^2 before calibration is low, demonstrating that default parameters without calibration fail to accurately capture road-level speed-density relationships.

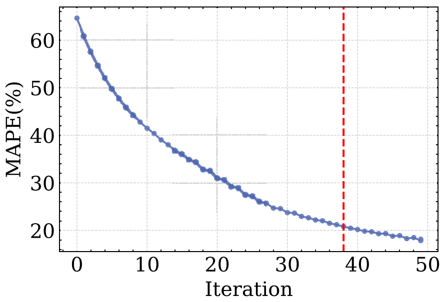
5.8 MOVES

MOVES mainly bases on operating modes (OpMode) to determine emission rates. Each operating mode is classified based on vehicle-specific power (VSP), speed, and acceleration. Mode 0 represents deceleration and braking. Mode 1 represents idle conditions at very low speeds. Modes 11 and 21 correspond to coasting conditions, differentiated by speed ranges. Modes 12–16 and 22–40 represent various cruise and acceleration states, further segmented by VSP intervals and speed bins. Extended Data Figure 6 provides an example of MOVES emission rates for different source types (all in the model year 2020) in each OpMode bin. High speeds, moderate accelerations at high speeds, and rapid accelerations at moderate or high speeds push on-road activity into higher VSP bins, leading to higher fuel consumption and emission rates in the emission calculation.

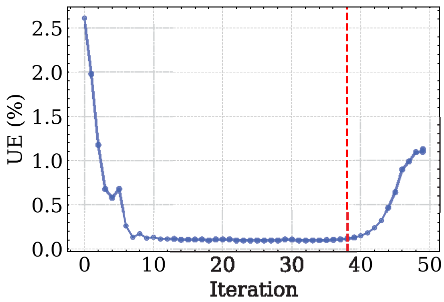


(a) Driving OD matrix

(b) Assignment volume

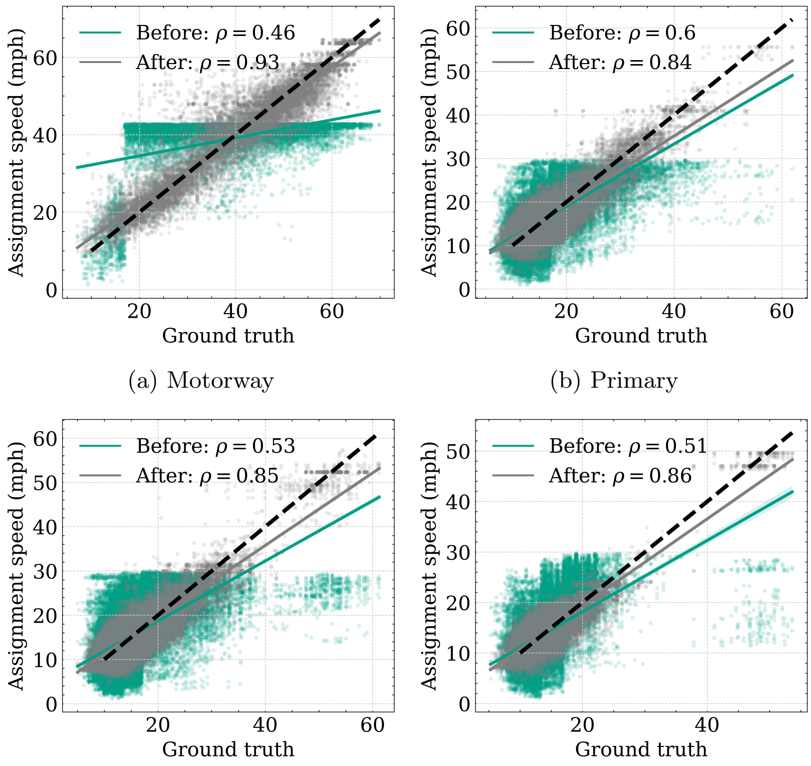


(c) Change of volume MAPE



(d) User equilibrium (UE) gap

Extended Data Fig 4: | **The DTA outcomes of the whole simulation network.** The volume is based on the average value during the morning peak 08:00-09:00 a.m. OD flow is aggregated at the census tract level, and only class labels greater than 1 are plotted for better visualization. In the real simulation, OD flow is aggregated at the TAZ level and all OD flows are inputted. Although only Manhattan City is the focus (red area), we considered nearby OD flows to include passing-through travels. In (c-d), the red line represents the iteration where we terminate the DTA and calibration.



(a) Motorway (b) Primary

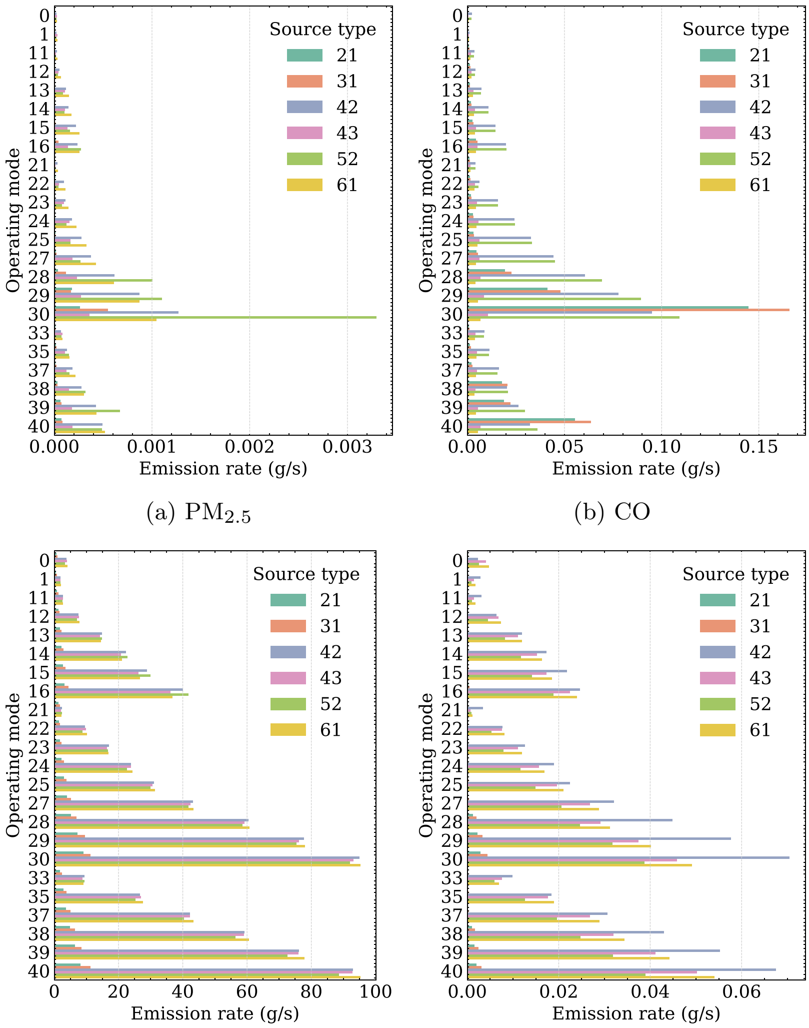
(c) Secondary (d) Residential

Extended Data Fig 5: | **Speed-density calibration.** Comparison of assigned vs. observed traffic speed by motorways (a), primary roads (b), secondary roads (c), and residential roads (d), before (teal) and after (gray) calibration. Each point represents the hourly speed on a single road. The black dashed line indicates the 1:1 reference line.

5.9 Scenario analysis

Extended Data Figure 7 shows the spatial distribution of four types of emissions during three periods: Morning peak (08:00-09:00 a.m.), Afternoon peak (04:00-05:00 p.m.), and Midnight (03:00-04:00 a.m.). As shown, the emission rates of most of the on-road traffic emissions decrease significantly during midnight, while the morning peak and afternoon peak show much more substantial concentration, with midtown and cross-city bridges showing the highest density. This is mainly due to the variation of traffic volume across the hours of the day and the spatial locations.

Extended Data Figure 8 shows the spatial distribution of CO_2 changes during four different events. A substantial decrease is observed across all road segments in Manhattan City during the snowstorm, Henri flooding, and COVID-19, while a substantial increase is observed during Black Friday, although the degree of change varies significantly across different road segments. These results highlight the time sensitivity of our method due to the use of MPLD to compute the OD matrix for traffic assignment.

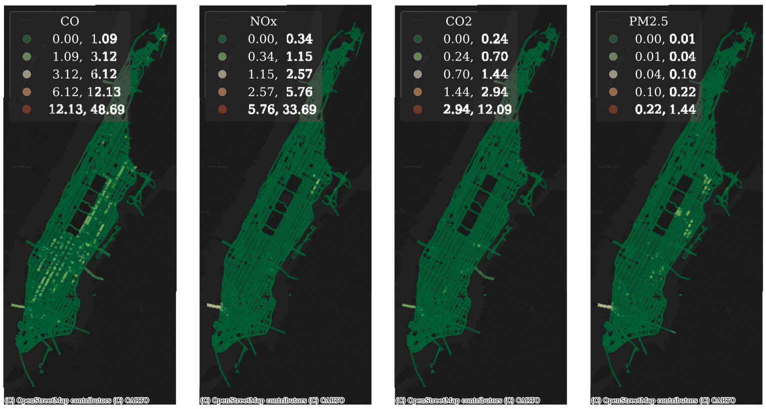


(a) $PM_{2.5}$ (b) CO

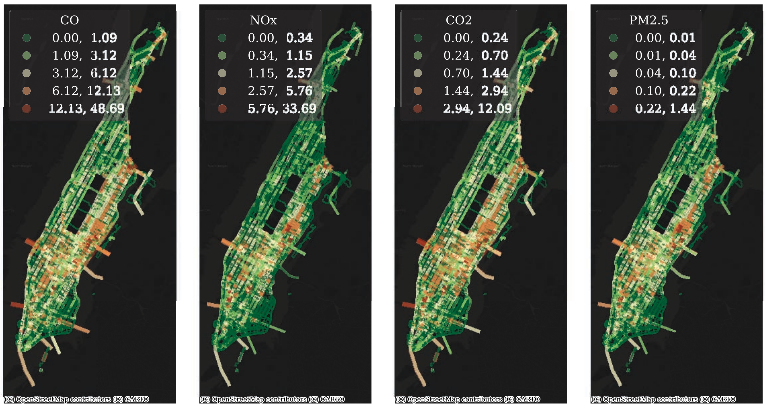
(c) CO_2 (d) NO_x

Extended Data Fig 6: | **Emission rate by operating mode.**

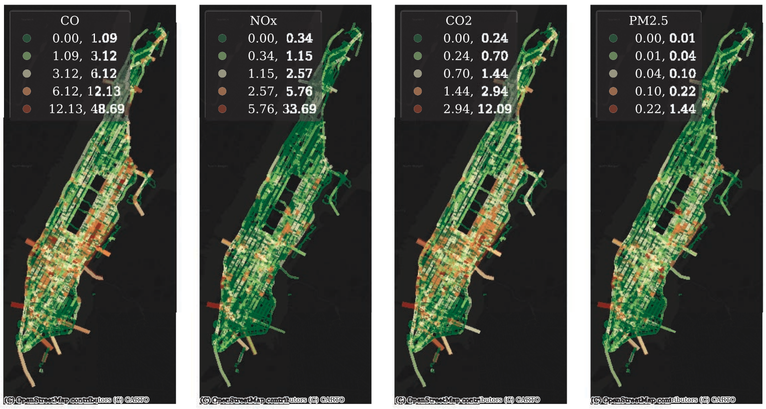
Extended Data Figure 9 shows the change in speed and volume after the implementation of congestion pricing in Manhattan. Panel (a) shows the percentage change in traffic speed across different road types over 2 to 8 weeks after the announcement of congestion pricing. Speed increased consistently over time across all road types, with the most significant increases observed on motorways and primary roads (15%). Residential and secondary roads saw smaller but still positive speed gains (8–12%). Panel (b) shows the percentage change in traffic volume by vehicle type over the same period. Truck volumes decreased the most, reaching nearly -14% by week 8, followed by car volumes, decreasing around -10%. Bus and other vehicle categories experienced



(a) 03:00 a.m. (b) 03:00 a.m. (c) 03:00 a.m. (d) 03:00 a.m.

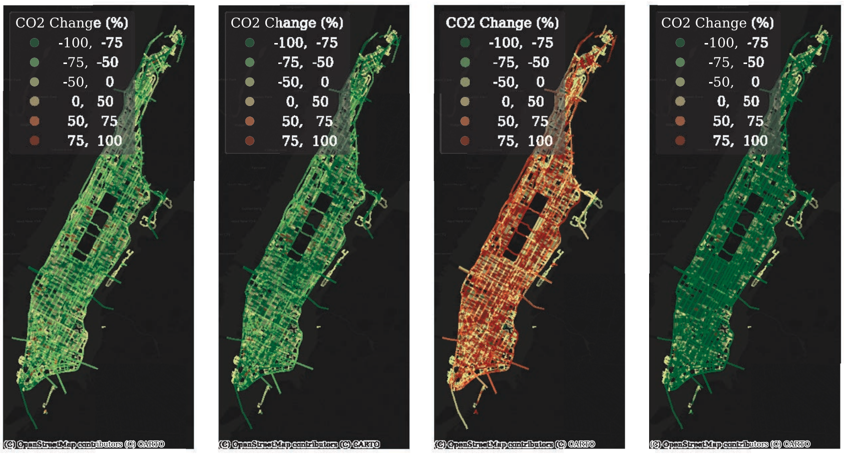


(e) 08:00 a.m. (f) 08:00 a.m. (g) 08:00 a.m. (h) 08:00 a.m.



(i) 04:00 p.m. (j) 04:00 p.m. (k) 04:00 p.m. (l) 04:00 p.m.

Extended Data Fig 7: | **Spatial distribution of emission density.** CO_2 is measured in ton·h^{-1}·$mile^{-1}$ and the others are measured in kg·h^{-1}·$mile^{-1}$.



(a) Snow storm (b) Henri (c) Black Friday (d) COVID-19

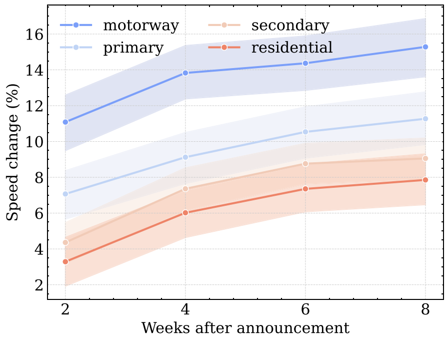
Extended Data Fig 8: | **Event-induced CO_2 emissions changes.** The daily average total emission on the same day of the week without events is used as the baseline.

relatively small changes. The greater reductions in truck traffic are largely due to its highest toll rates under the congestion pricing policy [12].

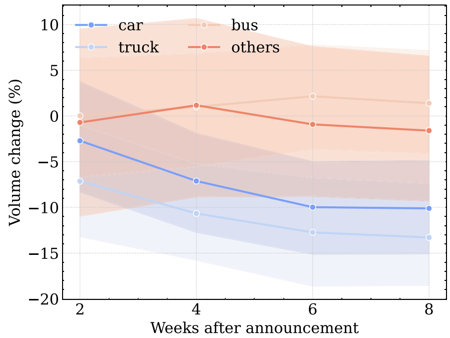
The observed changes are consistent with multiple independent data sources [43, 44]. For instance, data from the E-ZPass system show that the total number of vehicles crossing the Lincoln Tunnel declined by 8.18% in January 2025 compared to January 2024, with the reduction increasing to 11.71% in February 2025. Similarly, vehicle counts through the Holland Tunnel decreased by 4.99% in January and 9.72% in February 2025, relative to the same months in the previous year. Additionally, a study based on Google Maps traffic trends [44] reported significant increases in average traffic speeds in NYC central business district (CBD) after congestion pricing: highway speeds rose by 13%, arterial speeds by 10%, and local road speeds by 8%.

5.10 Computational setup

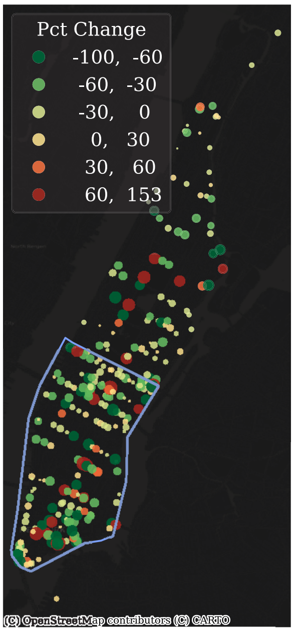
All data processing is conducted on an hourly basis with a one-day lag relative to real time due to the update frequency of the MPLD. Within the simulation, the temporal resolution is further refined to 5-minute intervals to capture finer traffic dynamics. Running on a local server with 4 `NVIDIA GEFORCE RTX 2080 Ti` GPUs, the average processing time for one hour of citywide data is 22 minutes, including 19 minutes for image processing, 2 minutes for dynamic traffic simulation, and 1 minute for emission estimation. This processing time, particularly for image processing, can be further reduced with more advanced GPUs or cloud-computing frameworks. These results indicate that, with fully live data streams, the framework is capable of generating near-real-time emission estimates within approximately 30 minutes.



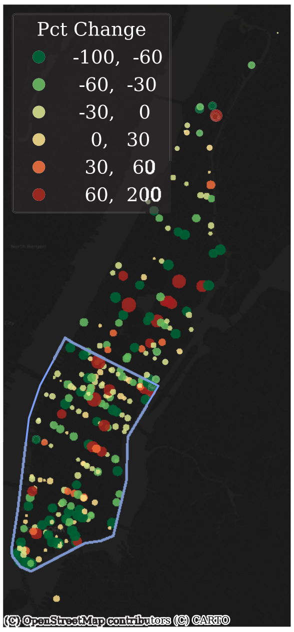
(a) Network speed change



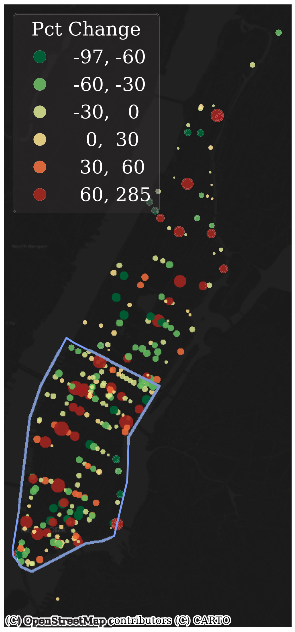
(b) Network volume change



(c) Passenger car change



(d) Truck change



(e) Pedestrian change

Extended Data Fig 9: | **Impact of congestion pricing on traffic speed and volume.** The daily average traffic volume and speed from the same day of the week and the corresponding week in 2024 are used as the baseline. (a-b) Temporal changes in network-wide average traffic speed (by road type) and volume (by car type). Shaded areas represent 95% confidence intervals. (c-e) Spatial distribution of traffic volume changes (in %) by week 8. Each point represents a camera location. Green indicates a decrease in volume, red indicates an increase. Marker size reflects the magnitude of absolute change. "Truck" represents the combined total of passenger trucks, single-unit trucks, and combination trucks. The blue boundary denotes the congestion relief zone.