

Bayesian Deep Learning Approach for Real-Time Lane-Based Arrival Curve Reconstruction at Intersection Using License Plate Recognition Data

Yang He¹, Chengchuan An¹, Jiawei Lu², Yao-Jan Wu³, Zhenbo Lu, and Jingxin Xia¹

Abstract—The acquisition of real-time and accurate traffic arrival information is of vital importance for proactive traffic control systems, especially in partially connected vehicle environments. License plate recognition (LPR) data that record both vehicle departures and identities are proven to be desirable in reconstructing lane-based arrival curves in previous works. Existing LPR data-based methods are predominantly designed for reconstructing historical arrival curves. For real-time reconstruction of multi-lane urban roads, it is pivotal to determine the lane choice of real-time link-based arrivals, which has not been exploited in previous studies. In this study, we propose a Bayesian deep learning approach for real-time lane-based arrival curve reconstruction, in which the lane choice patterns and uncertainties of link-based arrivals are both characterized. Specifically, the learning process is designed to effectively capture the relationship between partially observed link-based arrivals and lane-based arrivals, which can be physically interpreted as lane choice proportion. Moreover, the lane choice uncertainties are characterized using Bayesian parameter inference techniques, minimizing arrival curve reconstruction uncertainties, especially in low LPR data matching rate conditions. Real-world experiment results conducted in multiple matching rate scenarios demonstrate the superiority and necessity of lane choice modeling in reconstructing arrival curves.

Index Terms—Bayesian deep learning, real-time arrival curve reconstruction, real-time vehicle count estimation, license plate recognition data.

I. INTRODUCTION

ARRIVAL curves depict time-stamped cumulative vehicle arrivals across signal cycles, offering a more realistic and informative traffic arrival profile compared to the mean or cyclic arrival rates [1], [2], [3], [4]. Real-time arrival curves

are valuable in deriving timely performance measures (e.g., real-time queue length [5], [6], [7], [8]) and hence optimizing signal timing [9], [10] and connected and automated vehicle (CAV) trajectory [11], [12], [13], etc. However, estimating lane-specific real-time arrival curves on multi-lane urban roads is still a challenging task due to the complicated lane choice behaviors associated with different movements.

Previous studies extensively exploited data from fixed-location detectors and mobile sensors to acquire traffic arrival information. Fixed advance detectors, installed at a fixed distance upstream of the stop bar, are commonly used to detect arrivals. However, when the vehicular queue exceeds the detector location, known as the queue-over-detector (QOD) condition, newly arriving vehicles can not be detected, leading to under-estimated arrivals [14]. Despite further studies imposing uniform assumptions on arrivals during QOD, the estimated arrivals are approximated. Mobile sensors, such as floating cars and connected vehicles, offer arrival data with broader spatial coverage. Despite the expected increase, the penetration rate remains low for a relatively long term. As a result, only a few probe vehicle samples are available at intersections within a signal cycle, limiting their application primarily to characterizing the cyclic arrival patterns [15].

In recent years, License Plate Recognition (LPR) technology has seen widespread adoption in cities across China and globally, paving the way for innovative applications, such as queue length estimation [8], [16], [17], [18], dynamic vehicular demand estimation [19], [20], and commuting pattern analysis [21], etc. License Plate Recognition (LPR) data integrates the strengths of both fixed-location and mobile sensors while offering superior features. At the intersection level, LPR functions similarly to fixed detectors by identifying lane-based vehicle departures, but with the additional capability of recording each vehicle's unique license plate. At the network level, akin to mobile sensors, LPR enables vehicle tracking across multiple intersections by matching their license plate, yet it surpasses mobile sensors with substantially higher penetration rates. With the tracking ability and high penetration, LPR data have facilitated great advancements in effectively estimating cyclic arrival rates [2], [3], [4] and reconstructing the more challenging arrival curves [22], [23].

The acquisition of arrival curves is equivalent to determining the cumulative arrival indices of all vehicles. Ideally, for two

Received 21 September 2023; revised 28 July 2024 and 20 September 2024; accepted 31 October 2024. Date of publication 14 November 2024; date of current version 9 January 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 52272309 and Grant 52202398 and in part by the International Science and Technology Cooperation Project of Jiangsu Province under Grant BZ2023015. The Associate Editor for this article was C. K. Ahn. (Corresponding author: Jingxin Xia.)

Yang He, Chengchuan An, Zhenbo Lu, and Jingxin Xia are with the Intelligent Transportation System Research Center, Southeast University, Nanjing 211189, China (e-mail: yanghe@seu.edu.cn; ccan@seu.edu.cn; luzhenbo@seu.edu.cn; xiajingxin@seu.edu.cn).

Jiawei Lu is with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: lujiaweiwk@gmail.com).

Yao-Jan Wu is with the Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, Tucson, AZ 85721 USA (e-mail: yaojan@arizona.edu).

Digital Object Identifier 10.1109/TITS.2024.3491433

1558-0016 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

adjacent intersections monitored by LPR sensors, all these indices can be determined by matching license plates from the upstream to the downstream intersection. However, in practice, the arrival curve reconstruction problem is raised as partial unmatched vehicles exist owing to recognition failures, lack of upstream LPR sensors, mid-block traffic merging, etc. To tackle this, the existing approaches mainly formulated an interpolation-based method to infer the arrival indices of unmatched vehicles based on matched vehicles. For example, Zhan et al. [22] customized a cyclic arrival function-based Gaussian Process (GP) model to interpolate the arrival indices of unmatched vehicles while characterizing arrival uncertainties. Mo et al. [23] further improved the GP model by mending the erroneous license plates, enhancing reconstruction robustness in undesirable matching conditions.

However, applying the interpolation methods to real-time reconstruction presents challenges. These methods rely on previously matched vehicles to reconstruct arrival curves [22], [23], but the lane choices of real-time link arrivals are often undetermined. As a result, the intended direction of the vehicles traveling on the road remains unknown, leading to inaccurate real-time reconstruction. Therefore, modeling the lane choices of real-time arriving vehicles is pivotal for real-time reconstruction. Moreover, lane choice behaviors are often stochastic owing to different driving preferences. For example, through-going drivers may choose either the lane with the shortest queues or the one requiring the fewest manipulations. Consequently, characterizing the uncertainties in lane choice becomes essential, especially in data-limited scenarios where these behavioral patterns are difficult to observe.

This study proposes a Bayesian deep learning approach for real-time lane-based arrival curve reconstruction using LPR data, in which lane choice patterns and uncertainties are both characterized. Specifically, the learning process is designed to capture the relationships between partially observed link-based arrivals and lane-specific arrivals, which can be physically interpreted as lane choice proportion. Moreover, the lane choice uncertainties are modeled to minimize arrival curve reconstruction uncertainties, especially in low LPR data matching rate conditions.

The contributions of this study are summarized as follows

- 1) A Bayesian deep learning approach is proposed to reconstruct real-time lane-based arrival curves using LPR data, in which lane choice patterns are characterized. Specifically, the learning process is designed to capture the relationship between partially observed link-based arrivals and lane-specific arrivals.
- 2) Building on lane choice learning, Bayesian parameter inference techniques are used to robustly characterize lane choice uncertainties. This development effectively minimizes the arrival curve reconstruction uncertainties in low LPR data matching rate (e.g., 10 %) conditions.
- 3) The proposed approach is validated on a real-world dataset using both deterministic and probabilistic evaluation metrics. Experimental results indicate the superiority of the proposed lane-choice learning-based method and the necessity of modeling lane-choice uncertainties.

II. LITERATURE REVIEW

According to the sensing technologies acquiring traffic arrivals, the existing approaches can be broadly divided into two categories, fixed-location detector-based and mobile sensor-based. Fixed detector-based approaches capture the arrival data mainly by detecting the vehicle's presence using advanced detectors installed at a fixed distance upstream of the stop bar at the intersection [24]. However, when the queue tail exceeds the detector, known as the queue-over-detector (QOD) condition, new arrivals can not be detected due to continuous occupation of the detector [6], [7]. The detector occupation leads to underestimated traffic arrivals, especially in over-saturated conditions when queues can not dissipate in one signal cycle. To address this problem, researchers [14], [25] have assumed a uniform distribution of traffic arrival during the QOD condition, though this is only an approximation of actual arrivals.

With the popularization of ride-hailing services and the emergence of connected vehicle technologies, massive vehicle trajectory data from mobile sensors have gained great attention in traffic state monitoring and performance evaluation [26], [27]. Existing studies, assuming stationary arrival pattern during a historical period, have utilized historical data from multiple signal cycles to estimate cyclic arrival rates through various methods such as constant global index difference [28], EM-solved maximum likelihood estimation [5], [29], and kernel density estimation [15]. Prior studies have indicated that sufficient probe samples, typically at least one or two per signal cycle, are required to capture more refined arrival profiles [30], [31]. However, despite expected increases, the penetration rate of mobile sensors remains low in the current and near future. The unbalanced spatial distribution of mobile sensors further exacerbates these challenges, limiting their application to cyclic arrival pattern analysis.

As a special type of fixed-location detector, LPR systems not only record vehicle passing timestamps and lanes but also register informative vehicle license plate identities. In addition, LPR integrates the strength of mobile sensors by enabling vehicle tracking across the network through license plate matching. Further, it presents substantially higher penetration rates and spatial coverage than mobile sensors, making it desirable for both estimating cyclic arrival rates [2], [4] and reconstructing more challenging arrival curves [22], [23]. The existing arrival curve reconstruction methods mainly extracted lane-based arrival indices by matching the vehicle's license plate and then formulated interpolation-based models to infer the indices of those unmatched vehicles. Zhan et al. [22] customized an arrival function-based Gaussian process (GP) method to interpolate the cumulative arrival indices of unmatched vehicles while capturing the arrival uncertainties. The piece-wise linear arrival function is used by assuming a cyclic arrival pattern which may not always guaranteed in practice. Mo et al. [23] further improved the GP model by developing a data-mending approach to correct the erroneously recognized license plate, providing more matched vehicles and enhancing model robustness in limited data conditions. In summary, the existing studies have effectively leveraged

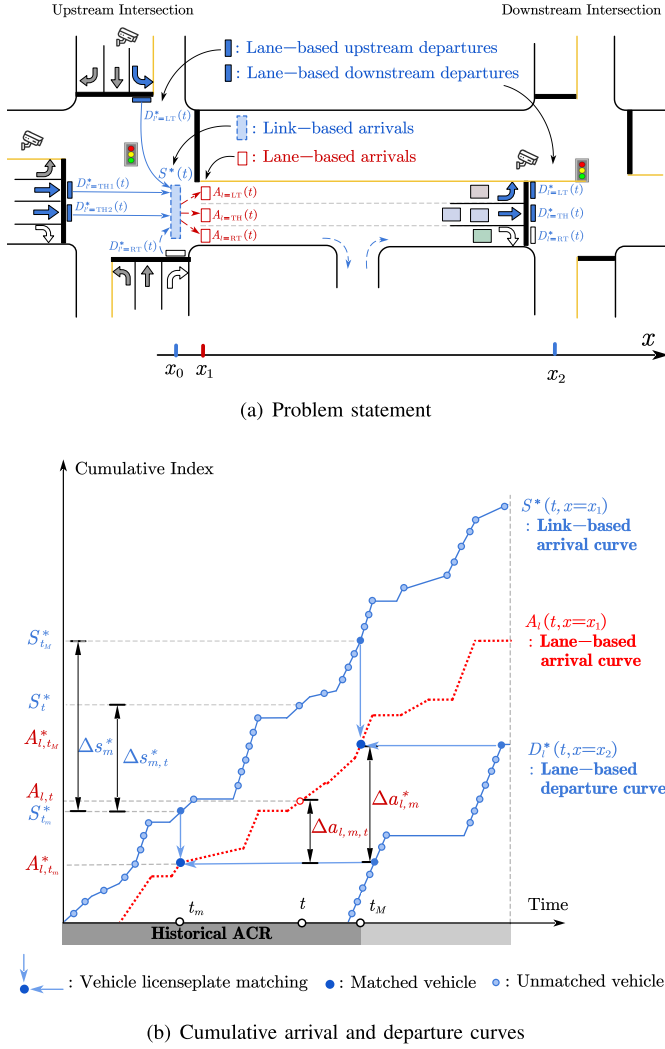


Fig. 1. Problem statement and the relationship between cumulative departure and arrival curves.

LPR data to reconstruct arrival curves, accounting for their stochastic natures and practical constraints. Despite the great success, these interpolation-based approaches still rely on historical matched vehicles regarding real-time reconstruction. The informative real-time link-based arrival data available from LPR systems are not exploited, which weakens their capability in real-time applications.

III. PROBLEM STATEMENT

A. Lane-Based Arrival Curve Reconstruction (ACR)

Fig. 1(a) describes a common scenario of the license plate recognition (LPR) data collection at two adjacent intersections. LPR cameras are usually installed to monitor the vehicle departures of left-turn and through traffic under the protective phase (i.e., hollow rectangles in Fig. 1(a)), while the right-turn traffic under the permitted phase (i.e., hollow rectangles in Fig. 1(a)) is not monitored. By indexing the recognized departing vehicles recorded in LPR data, the lane-based cumulative departure curves at upstream and downstream intersections can be obtained. By matching the license plates of departed

vehicles from the upstream to the downstream intersection, the lane-based arrival indices of partial vehicles can be determined, as shown in Fig. 1(b). The arrival curve reconstruction problem is raised as not all the vehicles can be matched owing to recognition failures, lack of upstream LPR sensors, mid-block traffic merging, etc. There are several assumptions used in this study

- **Assumption 1:** The vehicle travel time within the intersection is ignored. As a result, the traffic departures from the stop-line of the upstream approaching lane (i.e., $x = x_0$) are approximated to the link-based traffic arrivals at the upstream section of the target link (i.e., $x = x_1$),
- **Assumption 2:** The mid-block merging traffic arrivals are regarded as the link-based traffic arrivals at the upstream section of the link (i.e., $x = x_1$) since they will eventually depart from the downstream lanes.
- **Assumption 3:** The upstream arriving traffic arrivals that diverge to side streets in the mid-block section are not considered in this study, as they do not pass the downstream lanes.

To obtain the lane-based traffic arrival curve, there are four types of vehicle cumulative curves of concern:

- 1) Lane-based upstream departure curves at the section of upstream approaching lane x_0 , $D_{l'}^*(t, x = x_0)$, where l' is the upstream lane heading to the target link, e.g., $l' \in \{LT, TH1, TH2, RT\}$ in Fig. 1.
- 2) Link-based arrival curve at the upstream section of target link x_1 , $S^*(t, x = x_1)$,
- 3) Lane-based departure curves at downstream section x_3 , $D_l^*(t, x = x_3)$, where l is the downstream lane, e.g., $l \in \{LT, TH, RT\}$ in Fig. 1.
- 4) Lane-based arrival curve at the upstream section x_1 , $A_l(t, x = x_1)$, where l is the downstream lane, e.g., $l \in \{LT, TH, RT\}$ in Fig. 1.

Based on Assumption 1, the upstream link-based arrival curve is approximated to the sum of upstream lane-based departures that head to the target link, i.e.,

$$S^*(t, x = x_1) = \sum_{l'} D_{l'}^*(t, x = x_0), \quad (1)$$

where l' is the upstream lane heading to the target link, e.g., $l' \in \{LT, TH1, TH2, RT\}$ in Fig. 1. This study aims to estimate real-time lane-based traffic arrival curves at the upstream section of the target link $A_l(t, x = x_1)$ based on the upstream link-based arrival curve $S^*(t, x = x_1)$ and downstream lane-based departure curves $D_l^*(t, x = x_3)$ using LPR data from upstream and downstream intersections. Furthermore, based on the estimated upstream arrivals, the traffic arrivals at arbitrary sections between the upstream and downstream can be determined using a free-flow speed, i.e.,

$$A_l(t, x) = A_l\left(t - \frac{x - x_1}{v_f}, x_1\right), \quad x_1 \leq x \leq x_2, \quad (2)$$

where v_f is the free-flow speed. For example, upstream arrival curves combined with stochastic free-flow speeds are employed to obtain the pseudo arrival curves at the downstream section (i.e., $x = x_2$) as shown in Fig. 1(b), which is used to calculate the queue profile at intersection [8].

For the sake of simplicity, we hide the spatial index of the arrival and departure curves, since the symbol contains spatial information. For example, the lane-based arrival curve at the upstream sections $A_l(t, x = x_1)$ is simply denoted as $A_l(t)$.

B. Cumulative Departure Curves and Matched Vehicles

According to the LPR records, we can obtain the accurate and complete timestamp sequences for all vehicles that depart from the stop bar on each monitored lane of both upstream and downstream intersections. By ordering the vehicle departing timestamp t_i , the corresponding cumulative departure index D_{l,t_i}^* of each vehicle can be determined. The cumulative departure curve $D_{l'}^*(t)$ can be obtained by connecting departure observations (t_i, D_{l,t_i}^*) , $i = 1, 2, \dots, D$.

The matched vehicles could be obtained through vehicle license plate matching between upstream and downstream lanes using vehicle departure records, as shown in Fig. 1(b). Specifically, a vehicle is regarded as a matched vehicle m if it is both recognized by the LPR camera to depart from the upstream lane at time t_m and depart from the downstream lane at time t'_m . Under FIFO rules, the lane-based cumulative arrival index of the matched vehicle is equal to the lane-based cumulative departure index, i.e.,

$$A_{l,t_m}^* = D_{l',t'_m}^*, \quad (3)$$

where t'_m is the timestamp of matched vehicles when passing the downstream approaching lane.

C. Historical and Real-Time ACR Task

In the process of estimating the lane-based arrival curves, we utilize License Plate Recognition (LPR) data from both upstream and downstream locations, dividing the task into two distinct types: historical and real-time Arrival Curve Reconstruction (ACR). For historical lane-based ACR, there is an inherent time lag because once a vehicle is detected by the upstream LPR camera and then enters the target link, it takes a link travel time before being detected by the downstream LPR camera, as illustrated in Fig. 1(b). This time lag results in delayed updates to the historical ACR. In contrast, the real-time lane-based ACR aims to reconstruct the arrival curve instantaneously, regardless of this time lag. This makes real-time ACR more challenging than the historical ACR but particularly valuable for applications that require immediate traffic management responses.

IV. METHODOLOGY

In this section, we formulate the lane-based arrival curve reconstruction problem based on lane choice learning in subsection IV-A. Then, a Bayesian arrival curve learner is designed to learn the lane choice patterns and uncertainties in subsection IV-B. Finally, the historical and real-time arrival curve reconstruction procedures are introduced in subsection IV-C, respectively.

A. Lane Choice Learning-Based Arrival Curve Reconstruction

The reconstruction of the lane-based arrival curve is equivalent to estimating the lane-based cumulative arrival index of any arriving vehicle. The existing approaches typically employed interpolation methods such as the Gaussian Process to infer the lane-based cumulative arrival indices of unmatched vehicles [22], [23] based on cumulative indices of matched vehicles, which are extracted through vehicle license plate matching using LPR data upstream and downstream intersections, as detailed in subsection III-B. While effective in historical arrival curve reconstruction, these methods depend on historical matched vehicles when applying to real-time reconstruction, weakening their reliability in real-time applications.

In this study, we design a lane choice learning-based arrival curve reconstruction, which enables the utilization of additional real-time link-based arrival observations from LPR data in real-time reconstruction. Specifically, the lane-based cumulative arrival index of any arriving vehicle can be expressed as

$$A_{l,t} = A_{l,t_m}^* + \Delta a_{l,m,t}, \quad (4)$$

where A_{l,t_m}^* is the lane-based cumulative arrivals of the previous matched vehicle arriving at t_m , and $\Delta a_{l,m,t}$ is the lane-based arrival accumulations during the interval $(t_m, t]$. The lane-based arrival accumulations can be determined by utilizing the link-based arrival accumulations from LPR data and a lane choice proportion, i.e.,

$$\Delta a_{l,m,t} = \alpha_l \cdot \Delta s_{m,t}^*, \quad (5)$$

where $\Delta s_{m,t}^*$ is the observed link-based arrival accumulations during the interval $(t_m, t]$, α_l is a lane choice proportion that distributes the link-based arrivals to the downstream lane l . However, in practical scenarios, the link-based arrivals are partially observed due to the incomplete LPR systems monitoring of upstream approaching lanes, as demonstrated in Fig. 1(a). Therefore, instead of using the linear proportion, this study introduces a non-linear lane choice mapping function to capture the relationship between the partially observed link-based arrivals and lane-based arrivals, i.e.,

$$\Delta a_{l,m,t} = f(\Delta s_{m,t}^*), \quad (6)$$

where f is the non-linear lane choice mapping function, $\Delta s_{m,t}^*$ is the partially observed link-based arrival accumulations, which come from multiple upstream lanes, i.e.,

$$\Delta s_{m,t}^* = \sum_{l'} \Delta s_{l',m,t}^*, \quad (7)$$

where $\Delta s_{l',m,t}^*$ is the link-based arrival accumulations coming from upstream approaching lane l' . According to Assumption 1, the link-based arrival accumulations from the upstream lane are equal to the upstream lane-based departure accumulations during $(t_m, t]$ i.e.,

$$\Delta s_{l',m,t}^* = \Delta d_{l',m,t}^* = D_{l',t}^* - D_{l',t_m}^*, \quad (8)$$

where $D_{l',t}^*$ and D_{l',t_m}^* is the observable cumulative departure index of upstream lane-based departure curves at time t and t_m .

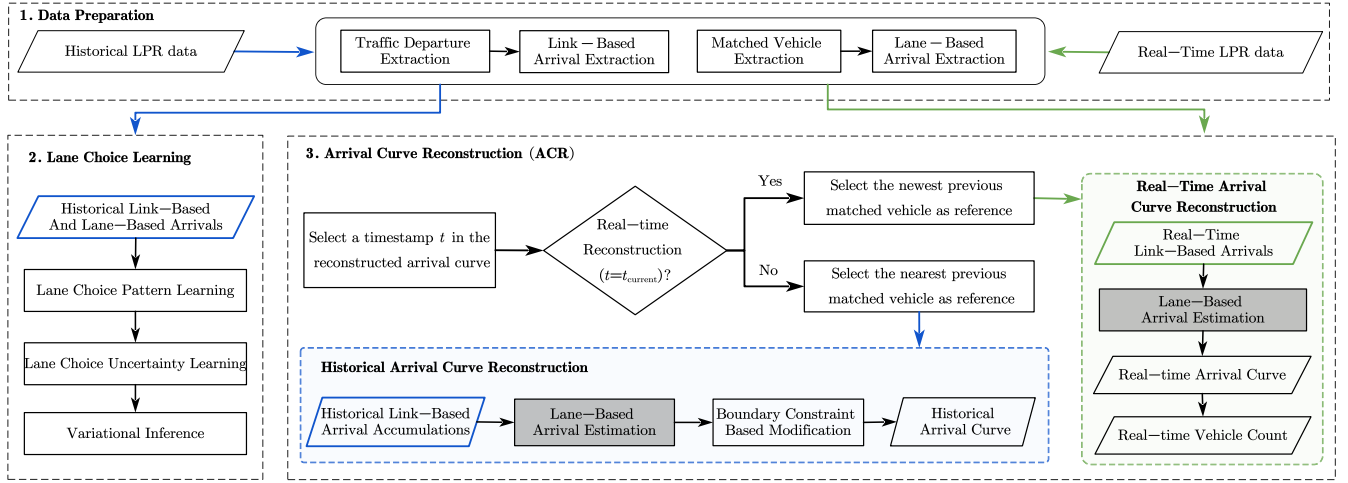


Fig. 2. Overview of the proposed lane choice learning-based arrival curve reconstruction framework.

The lane choice patterns of link-based arrivals from different upstream lanes are distinguishing. Therefore, to capture refined lane choices and avoid information loss, the Eq. (6) is rewritten as

$$\Delta a_{l,m,t} = f\left(\Delta s_{l',m,t}^*\right), \quad (9)$$

where $\Delta s_{l',m,t}^*$ is a vector consisting of multiple observed upstream link-based arrival accumulations from different upstream lanes.

In order to better capture lane choice patterns, we introduce additional upstream signal phase information because different upstream signal phases could correspond to distinguishing lane choice patterns. For example, the upstream through phase probably contributes the most arrivals of downstream through lane. Therefore, we introduce two additional features, the time in the signal cycle and the estimation span, and reformulate the Eq. (6) as

$$\Delta a_{l,m,t} = f\left(\Delta s_{l',m,t}^*, t_m^c, \delta\right), \quad (10)$$

where t_m^c is the time in signal cycles of the referenced matched vehicles m , $\delta = t - t_m$ is the estimation span, i.e., the time difference between the current estimation and previous matched vehicles. In addition to LPR data, the proposed framework is also compatible with the emerging CAV data, which can provide additional observations, e.g., link-based arrivals, about mid-block merging traffic, and right-turning traffic at intersections, which are usually not monitored by LPR cameras.

B. Lane Choice Learning via Bayesian Deep Learning

In this subsection, we propose a Bayesian Arrival Curve Learner (BACL) to capture the lane choice patterns and quantify the lane choice uncertainties. Specifically, the lane choice pattern learning using BACL is illustrated in subsection IV-B1. We specify how the lane choice uncertainties are quantified in subsection IV-B2.

1) *Learning Lane Choice Patterns*: To learn lane choice patterns, we employ a Bayesian deep learning approach termed BACL to serve as the lane choice mapping function f , which is trained utilizing historical link-based and lane-based arrival observations. Based on historical matched vehicles, The historical lane-based arrivals can be calculated as

$$\Delta a_{l,m}^* = A_{l,t_{m+1}}^* - A_{l,t_m}^*, \quad (11)$$

where $A_{l,t_{m+1}}^*$ and A_{l,t_m}^* denote the cumulative arrival index of m and $m+1$ matched vehicles. The historical link-based arrivals from upstream lane l' can be calculated based on observed lane-based departure curves, i.e.,

$$\Delta s_{l',m}^* = \Delta d_{l',m}^* = D_{l',t_{m+1}}^* - D_{l',t_m}^*, \quad (12)$$

where $D_{l',t_{m+1}}^*$ and D_{l',t_m}^* is the cumulative departure index of upstream lane-based departure curves at time t_{m+1} and t_m . Given the historical lane-based and link-based arrival observations, the lane choice pattern learning using BACL is reformulated as follows

$$\begin{aligned} \min \sum_{l=1}^L \sum_{m=1}^{M_l-1} F(\Delta a_{l,m}, \Delta a_{l,m}^*), \\ \text{s.t. } \Delta a_{l,m} = f^\omega(\Delta s_{l',m}^*, t_m^c, \delta), \end{aligned} \quad (13)$$

where

$\Delta a_{l,m}$ is the estimated lane-based arrival accumulation during the time interval $(t_m, t_{m+1}]$, where t_m and t_{m+1} denote the arrival timestamp of m^{th} and $(m+1)^{th}$ matched vehicles,

$\Delta a_{l,m}^*$ is the observed upstream lane-based arrival accumulation during the time interval $(t_m, t_{m+1}]$,

$\Delta s_{l',m}^*$ is a vector consisting of multiple observed link-based arrival accumulations from different upstream lanes l' , during the time interval $(t_m, t_{m+1}]$,

$F(\cdot)$ is a probabilistic loss function,

$f^\omega(\cdot)$ is the BACL-based lane choice mapping function,

M_l is the number of matched vehicles in lane l ,

L is the number of lanes.

2) *Learning Lane Choice Uncertainty*: This subsection further specifies how lane choice uncertainties are characterized. To quantify the lane choice uncertainties, instead of using fixed parameter values, we put a prior distribution $p(\omega)$ over the space of model parameters ω of the BACL-based lane choice mapping function $f^\omega(\cdot)$, which represents our prior belief as to which parameters are likely to have mapped link-based arrivals to lane-based arrivals before we observe any matched vehicle data, as illustrated in Fig. 3. In this study, we assume the Gaussian distribution as the prior and posterior distribution of model weights and the predictive distribution of estimated lane-based arrivals, given its computation simplicity, variational inference efficiency, and expressive capability.

For the sake of simplicity, we denote the link-based arrival accumulations, time in the signal cycle, and estimation span as model input, i.e., $\mathbf{x} = \{\Delta s_{l',m}^*, t_m^c, \delta\}$, and the corresponding lane-based arrival accumulations as model output, i.e., $y = \Delta a_{l,m}$. The historical dataset can be expressed as $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and $\mathbf{y} = \{y_1, \dots, y_N\}$, and the posterior distribution over the space of parameters can be given by Bayes's theorem

$$p(\omega|\mathbf{X}, \mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{X}, \omega) p(\omega)}{p(\mathbf{y}|\mathbf{X})}, \quad (14)$$

where $p(\mathbf{y}|\mathbf{X}, \omega)$ denotes the likelihood distribution by which the model inputs, e.g., link-based arrival accumulations, generate the lane-based arrivals given parameter setting ω . Then, we can estimate lane-based arrival accumulation given new model inputs by integrating all likelihood generated from the posterior distribution, i.e.,

$$p(\hat{y}|\hat{\mathbf{x}}, \mathbf{X}, \mathbf{y}) = \int p(\hat{y}|\hat{\mathbf{x}}, \omega) p(\omega|\mathbf{X}, \mathbf{y}) d\omega, \quad (15)$$

where $\hat{\mathbf{x}}$ denotes the new model inputs, $\hat{y}$ represents the estimated lane-based arrival accumulation, and $p(\hat{y}|\hat{\mathbf{x}}, \omega)$ is the likelihood.

3) *Variational Inference and Parameter Update*: It is intractable to solve the posterior distribution of weights $p(\omega|\mathbf{X}, \mathbf{y})$ directly using Bayes theorem in Eq. (14). Therefore, we use a variational weight distribution $q_\theta(\omega)$ to approximate the actual posterior. The parameters of the variational weight distribution can be estimated by minimizing the KL divergence between the variational distribution $q_\theta(\omega)$ and the actual posterior $p(\omega|\mathbf{X}, \mathbf{y})$, i.e.,

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \text{KL}(q_\theta(\omega) || p(\omega|\mathbf{X}, \mathbf{y})) \\ &= \arg \min_{\theta} \int q_\theta(\omega) \log \frac{q_\theta(\omega)}{p(\omega) p(\mathbf{y}|\mathbf{X}, \omega)} d\omega \\ &= \arg \min_{\theta} \text{KL}(q_\theta(\omega) || p(\omega)) - \mathbb{E}_{q_\theta(\omega)} [\log p(\mathbf{y}|\mathbf{X}, \omega)], \end{aligned} \quad (16)$$

which is equivalent to minimizing negative Evidence Lower Bound (ELBO)

$$\text{KL}[q_\theta(\omega) || p(\omega)] - \sum_{i=1}^N \int q_\theta(\omega) \log p(y_i | f^\omega(\mathbf{x}_i)) d\omega, \quad (17)$$

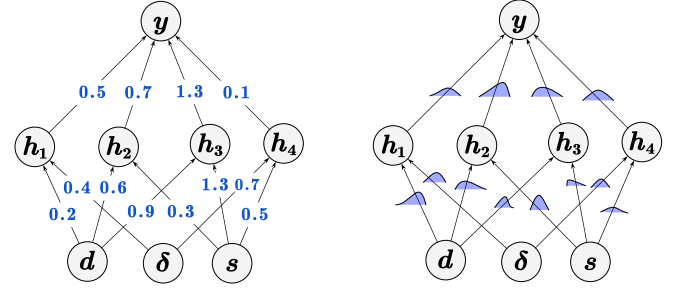


Fig. 3. Left: Classical neural network, each weight has a fixed value. Right: Bayesian neural network, each weight is represented by a distribution.

where $\mathbf{x}_i$ and y_i denote the i sample in the training dataset, and N represents the number of samples. In forward arrival estimation, directly sampling the weights from the variational posterior distribution $q_\theta(\omega)$ will lose the corresponding gradients that are essential in the backward parameter estimation process. Therefore, to keep the gradient and reserve the uncertainties of arrival estimation, we use the Gaussian reparameterization trick. Specifically, the sampling process is removed from the model structure, and the posterior weight ω is generated using a parameter-free noise distribution $\epsilon \sim \mathcal{N}(0, I)$, as shown in Fig. 4. The corresponding variational posterior parameters $\theta = (\mu, \rho)$ is expressed as

$$\omega = t(\theta, \epsilon) = \mu + \log(1 + \exp(\rho)) \circ \epsilon, \quad (18)$$

where $\circ$ is point-wise multiplication. According to the Lemma 1 in [32], the gradients with respect to the mean value and standard deviation parameter of the variational distribution are given as

$$\Delta_\mu = \frac{\partial f(\omega, \theta)}{\partial \omega} + \frac{\partial f(\omega, \theta)}{\partial \mu}, \quad (19)$$

$$\Delta_\rho = \frac{\partial f(\omega, \theta)}{\partial \omega} \frac{\epsilon}{1 + \exp(-\rho)} + \frac{\partial f(\omega, \theta)}{\partial \rho}, \quad (20)$$

where $f(\omega, \theta) = \log q_\theta(\omega) - \log p(\omega) p(\mathbf{y}|\mathbf{X}, \omega)$. The parameters of the variational distribution are updated as

$$\mu \leftarrow \mu - \alpha \Delta_\mu, \quad (21)$$

$$\rho \leftarrow \rho - \alpha \Delta_\rho, \quad (22)$$

where α denotes a learning rate.

4) *Arrival Accumulation Estimation and Uncertainty Quantification*: Given the estimated variational posterior distribution, the mean value and variance of estimated arrival accumulation are derived by generating weight samples from the variational distribution, i.e.,

$$\mathbb{E}[\hat{y}] = \int p(\hat{y}|\hat{\mathbf{x}}, \omega) q_\theta(\omega) d\omega \approx \frac{1}{M} \sum_{m=1}^M f^{\hat{\omega}_m}(\hat{\mathbf{x}}), \quad (23)$$

$$\begin{aligned} \text{Var}[\hat{y}] &= \underbrace{\frac{1}{M} \sum_{m=1}^M [f^{\hat{\omega}_m}(\hat{\mathbf{x}})]^2}_{\text{epistemic lane choice uncertainty}} - \mathbb{E}[\hat{y}]^2 \\ &\quad + \underbrace{\frac{1}{M} \sum_{m=1}^M \sigma_n(\hat{\mathbf{x}})}_{\text{aleatoric arrival uncertainty}}, \end{aligned} \quad (24)$$

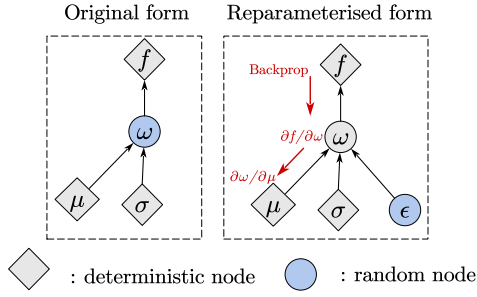


Fig. 4. Illustration of the reparameterization trick.

where $f^{\hat{\omega}_m}(\hat{\mathbf{x}})$ is the estimated arrivals of BACL model for the input $\hat{\mathbf{x}}$ using the generated weight $\hat{\omega}_m$ sample from the variational posterior distribution $q_{\theta}(\omega)$, M is the number of samples, $\hat{y} = \Delta a_{l,m,t}$, $\hat{\mathbf{x}} = (\Delta s_{l',m,t}^*, t_m^c, \delta)$.

C. Historical and Real-Time Arrival Curve Reconstruction

The historical Arrival Curve Reconstruction (ACR) aims to estimate the lane-based cumulative arrival index of historical unmatched vehicles, while the real-time ACR seeks to estimate the lane-based cumulative arrival index of real-time arriving vehicles.

1) *Historical ACR*: The first step in historical ACR is to determine the time range of reconstructed curves, which is defined as the range between the arriving time of two adjacent matched vehicles, i.e.,

$$T_m = (t_m, t_{m+1}], m = 1, 2, \dots, M_l - 1, \quad (25)$$

where T_m is the m reconstructed time range of the arrival curves. Based on Eq. (23), we can obtain the mean estimation of arrival accumulations $\Delta \hat{a}_{l,m,t}$. The two adjacent matched vehicles provide boundary constraints that limit the up value of estimated arrival accumulations. The estimated lane-based arrival accumulations within the range T_m must equal the observed one, which is calculated in Eq. (11). For each reconstructed area T_m , we introduced a modified factor to incorporate the boundary constraint into the estimation of arrival accumulations, i.e.,

$$\lambda_m = \frac{\Delta a_{l,m}^*}{\Delta \hat{a}_{l,m,t}}, \quad (26)$$

where $\Delta \hat{a}_{l,m,t}$ is the estimated arrival accumulation. The Eq. (4) can be written as

$$A_{l,t} = A_{l,t_m}^* + \lambda_m \cdot \Delta \hat{a}_{l,m,t}. \quad (27)$$

The pseudocode of historical ACR is presented in Algorithm 1.

2) *Real-Time ACR*: In real-time reconstruction, the real-time link-based arrival accumulations after the newest matched vehicle M can be obtained from observed lane-based departure curves using real-time LPR data

$$\Delta s_{l',M,t}^* = D_{l',t}^* - D_{l',t_M}^*, \quad (28)$$

where $D_{l',t}^*$ and D_{l',t_M}^* are the observed cumulative departure indices of lane-based departure curves at time t and t_M . Based on the real-time link-based arrivals, we can estimate the mean and variance of real-time lane-based arrival accumulations

Algorithm 1 Historical Lane-Based Arrival Curve Reconstruction

Input: The historical LPR data set, signal timing;

Output: The historical arrival curve $A_l(t)$, $t < t_M$;

- 1: Extract lane-based cumulative arrival index of matched vehicle observations $A_{l,m}^*$, $m = 1, 2, \dots, M_l - 1$ using Eq. (3)
- 2: **for** $m=1:M_l-1$ **do**
- 3: Select the m matched vehicle as a reference;
- 4: **for** t in range $(t_m, t_{m+1}]$ **do**
- 5: Calculate the reconstructed time span $\delta = t - t_m$;
- 6: Calculate the link-based arrival accumulations merged from different upstream lanes via Eq. (8);
- 7: Estimate the mean value and variance of the estimated lane-based arrival accumulations $\Delta \hat{a}_{l,m,t}$ via Eq. (10), Eq. (23), and Eq. (24);
- 8: Modify the mean estimated arrival accumulation and estimate the lane-based cumulative arrival index of unmatched vehicles via Eq. (27);
- 9: **end for**
- 10: **end for**

using the trained BACL via Eq. (23) and Eq. (24), and the real-time arrival curve reconstruction can be expressed as

$$A_{l,t} = A_{l,t_M}^* + \Delta \hat{a}_{l,M,t}, \quad (29)$$

where A_{l,t_M}^* is the lane-based cumulative arrival index of the newest matched vehicle, $\Delta \hat{a}_{l,M,t}$ is the mean estimation of real-time lane-based arrival accumulation using trained BACL. The pseudocode of real-time ACR using real-time LPR data is presented in Algorithm 2.

For practical applications, the real-time cumulative arrival curve can be applied to real-time lane-based vehicle count estimation. Specifically, given the estimated real-time lane-based arrival curve $A_l(t)$, the real-time lane-based vehicle count $G_l(t)$ can be determined as

$$G_l(t) = A_l(t) - D_l^*(t), \quad (30)$$

where $D_l^*(t)$ is the lane-based downstream departure curve. We present the estimation examples of real-time lane-based vehicle count estimation in subsection V-C.

V. EXPERIMENTS

In this section, we evaluate our proposed method on a real-world LPR dataset. We introduce the test site, baseline methods, and evaluation metrics in subsection V-A. The historical arrival curve reconstruction model performances are evaluated in subsection V-B. The real-time estimation performances of lane-based arrival curves and vehicle counts are assessed in subsection V-C. Finally, we inspect the model performance using different utilization levels of LPR data in subsection V-D.

A. Experimental Settings

1) *Test Site Description*: In field experiments, we select two adjacent intersections and the connecting link on Qianjing

Algorithm 2 Real-Time Lane-Based Arrival Curve Reconstruction

Input: The real-time LPR data set, signal timing;

Output: The real-time arrival curve $A_l(t)$, $t > t_M$;

- 1: Extract the lane-based cumulative arrival index of the newest matched vehicle $A_{l,M}^*$ using Eq. (3);
- 2: Select the newest matched vehicle M as a reference;
- 3: **while** No new matched vehicle update **do**
- 4: Select the reconstructed timestamp $t = t_{current}$;
- 5: Calculate the reconstructed time span $\delta = t - t_M$;
- 6: Calculate the link-based arrival accumulations merged from different upstream lanes via Eq. (28);
- 7: Estimate the mean value and variance of the estimated lane-based arrival accumulations $\Delta \hat{a}_{l,M,t}$ via Eq. (10), Eq. (23), and Eq. (24);
- 8: Obtain the real-time lane-based arrival curve $A_l(t)$ via Eq. (29);
- 9: **end while**

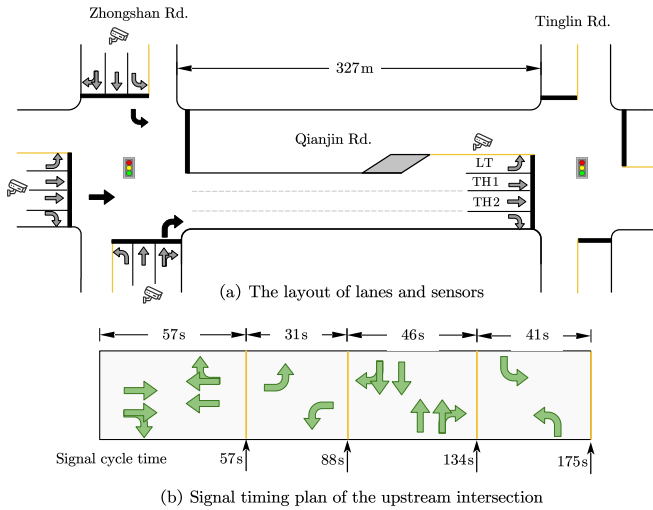


Fig. 5. Visualization of the test site layout and signal timing.

Road in Kunshan, Jiangsu Province, China as our test sites. As shown in Fig. 5, the LPR data are collected at the upstream and downstream intersections from June 4th to June 8th, 2018, spanning a total of five days. The evaluation includes three traffic scenarios, each lasting 15 minutes: the morning peak (AP) from 8:15 to 8:30 AM, the off-peak (OP) from 2:15 to 2:30 PM, and the evening peak (PP) from 5:15 to 5:30 PM. Three of four downstream lanes are selected as test lanes: the left-turn lane (LT), through lane 1 (TH1), and lane 2 (TH2). Three of the five days' data are used to learn the arrival curves, while the data from the remaining two days are served for evaluating the reconstruction of arrival curves. The LPR sensors at the test site did not work well and the matching rate of LPR data is 61.6% on average. Additional video data are also collected to fix unrecognized vehicle license plates, improving the matching rate to 97.8%. This corrected dataset is used to provide ground truth for model validation.

2) *Benchmark Method:* We compared the proposed Bayesian method with the State-Of-The-Art (SOTA) probabilistic and deterministic methods.

- **AF-GP** (Arrival Function-based Gaussian Process [22], [23]): A customized Gaussian process for arrival curve reconstruction, in which a customized mean function is calibrated by two-stage cyclic arrival rates to capture arrival patterns and a covariance function is developed to characterize arrival uncertainties. The arrival curves are reconstructed by inferring the cumulative arrival indices of unmatched vehicles using the AF-GP model based on matched vehicles.
- **LC-NN** (Lane Choice-based Feedforward Neural Network): A plain neural network that reconstructs the arrivals curves based on lane choice learning, using matched vehicles and real-time link-base arrivals. The LC-NN is used for comparison to demonstrate the superior performance of the Bayesian approach in modeling lane choice uncertainty.

3) *Evaluation Metrics:* To assess the arrival curve reconstruction performance, we employ a deterministic metric Root Mean Square Error (RMSE) and a probabilistic metric Continuous Ranked Probability Score (CRPS). The metric RMSE is used to evaluate the mean estimation of cumulative arrivals

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i^* - a_i)^2}, \quad (31)$$

where a_i^* and a_i represent the actual value and estimated mean value of the i^{th} unobserved cumulative arrivals, respectively.

In addition to the deterministic RMSE, the CRPS is employed to serve as a probabilistic evaluation of the estimated arrival distribution, reflecting both the accuracy of the mean estimation and the appropriateness of the uncertainty quantification

$$\text{CRPS} = \int_{-\infty}^{\infty} (F(a) - H(a - a^*))^2 da, \quad (32)$$

where $F(a)$ is the cumulative distribution function (CDF) of the estimated distribution, a^* is the observed value, H is the Heaviside step function, i.e., 0 for $a < a^*$ and 1 for $a \geq a^*$. A lower CRPS value indicates that the distribution estimate is closer to the actual observed value.

B. Historical Arrival Curve Reconstruction Results

To evaluate the model performance of historical arrival curve reconstruction (ACR) from both deterministic and probabilistic perspectives, we summarize the average RMSEs and CRPSs (including standard deviations) across various LPR data matching rates in Tab. I and Tab. II. The metrics are calculated using testing data spanning nine scenarios (3 testing lanes $\times$ 3 testing periods). The proposed BACL model consistently outperforms the baseline methods across diverse matching rates. Tab. I shows that the RMSEs for interpolation-based AF-GP remain relatively stable when matching rates (MRs) exceed 40%, but they increase significantly, with larger standard deviations, at MRs below 20%.

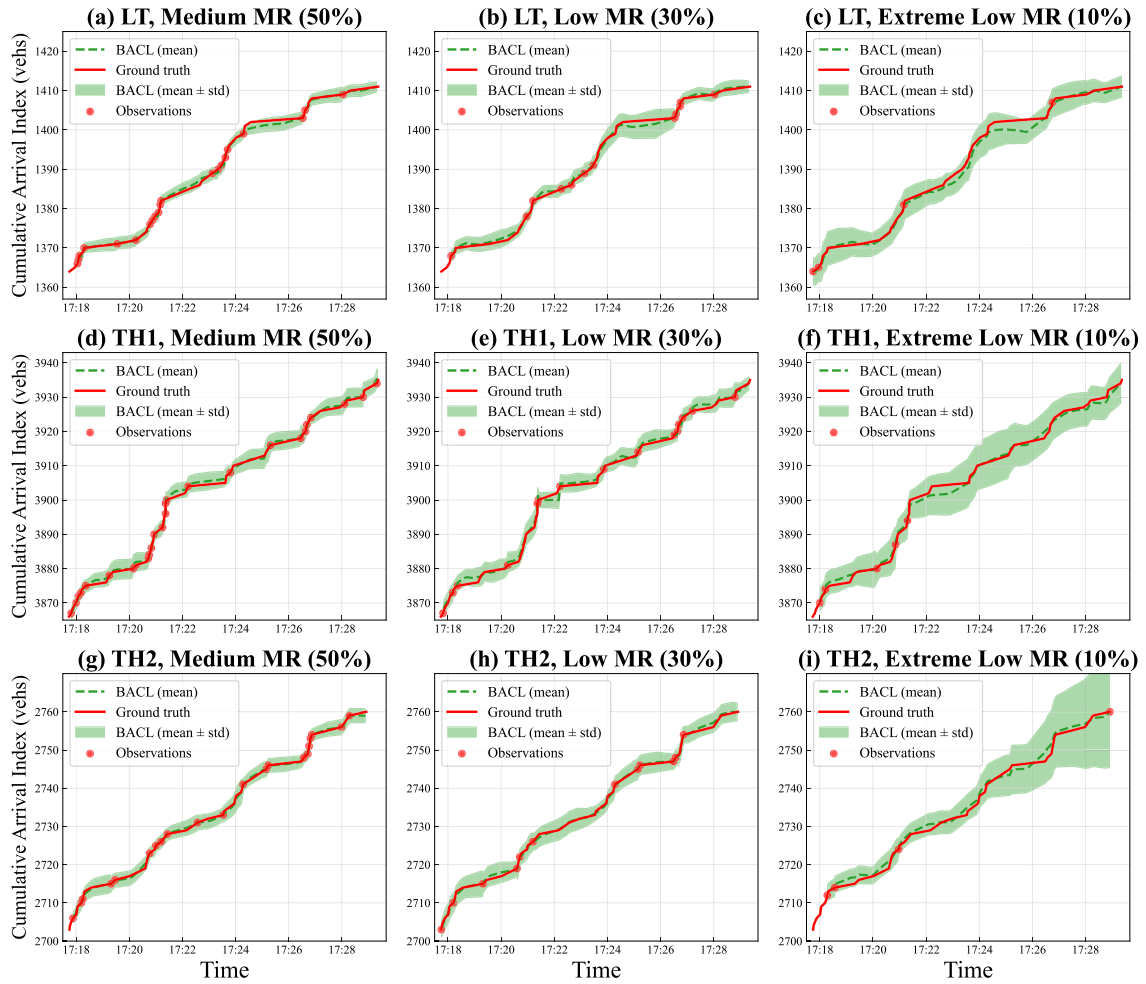


Fig. 6. Historical arrival curve reconstruction examples with mean estimates and uncertainty characterization.

In contrast, the lane choice-based neural network (LC-NN) and proposed BACL, which utilize link-based arrival information, display stable RMSEs with small variance even at low MRs. This stability suggests the effectiveness of the lane choice-based learning mechanism in reconstructing arrival curves. By comparing the BACL with LC-NN, it can be seen that while the LC-NN performs competitively with the BACL at high matching rates, the LC-NN's RMSE increases dramatically in low matching rates scenarios where the uncertainties of lane choices are even higher, highlighting the value of proposed Bayesian deep learning approach.

By comparing the probabilistic CRPSs of the baseline and proposed method in Tab. II, we can observe that the CRPSs of AF-GP increase sharply in low MRs, while those for BACL remain robust with smaller variance. This consistency indicates that the estimated arrival distributions from BACL more closely align with the actual values, demonstrating the superiority of the proposed Bayesian deep learning approach in capturing lane choice and arrival uncertainties.

To vividly demonstrate the ACR performance, we visualize nine examples from three downstream lanes during PM peak hours across three matching rates (MRs) of 50%, 30%, and 10%, shown in Fig. 6. The green dashed lines and filled areas represent the BACL's estimated arrival mean values

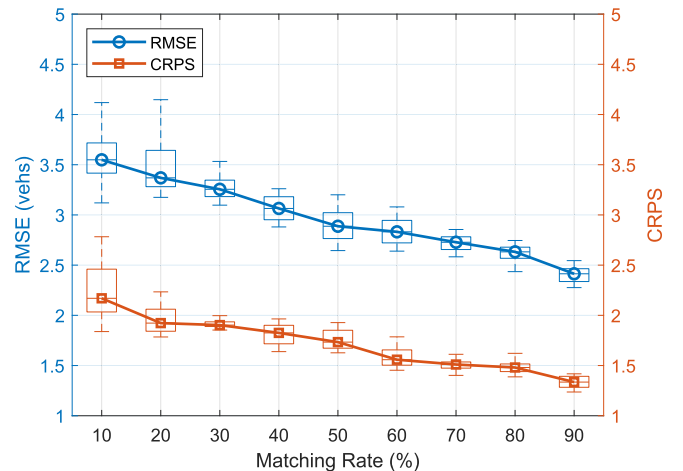


Fig. 7. The RMSE and CRPS distribution of real-time arrival curve reconstruction under various matching rates.

and uncertainties, respectively. It can be seen that the BACL consistently offers accurate and reliable reconstruction results with reasonable uncertainty characterization. In the medium and low MR scenarios, e.g., Fig. 6(a), (d), (g), the BACL's mean estimates closely align with the ground truth arrival

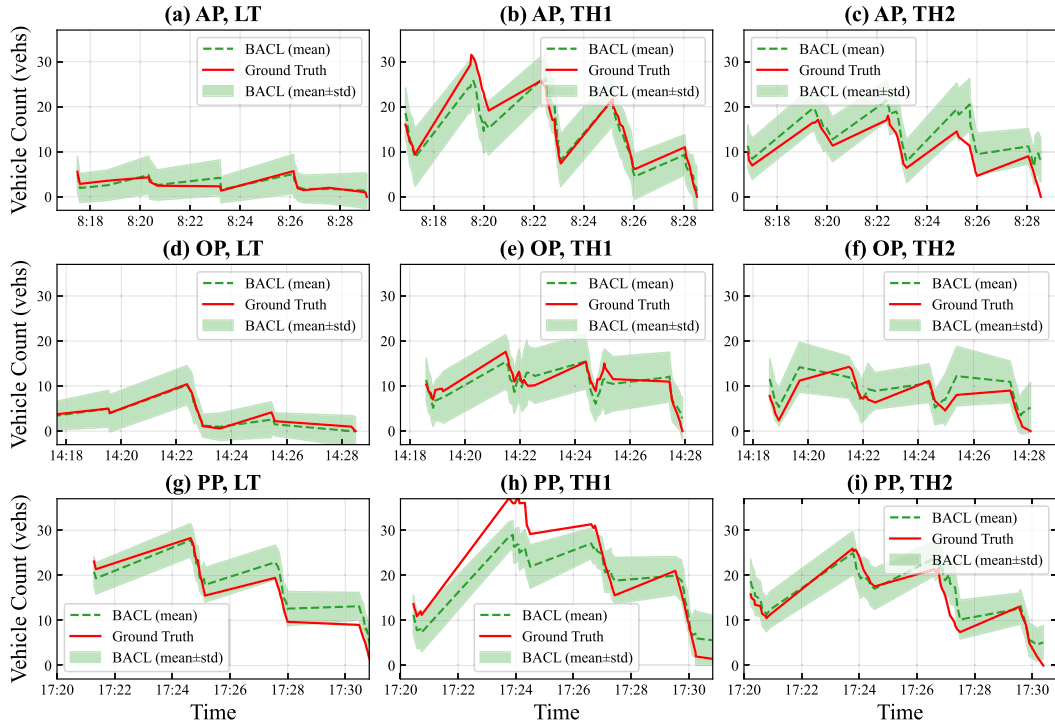


Fig. 8. Real-time lane-based vehicle count estimation examples.

TABLE I
AVERAGE RMSES WITH STANDARD DEVIATIONS OF HISTORICAL ACR
UNDER VARIOUS MATCHING RATES

Matching Rate	Model		
	AF-GP	LC-NN	BACL
10	5.903±2.415	2.201±0.281	1.607±0.200
20	3.485±1.971	1.830±0.175	1.359±0.106
30	2.907±1.764	1.445±0.071	1.224±0.170
40	2.203±0.899	1.374±0.157	0.926±0.439
50	1.886±0.901	1.136±0.053	0.736±0.165
60	1.552±0.419	1.063±0.048	0.734±0.152
70	1.422±0.529	0.901±0.035	0.615±0.216
80	1.138±0.280	0.807±0.028	0.531±0.135
90	1.109±0.340	0.681±0.017	0.504±0.152

TABLE II
AVERAGE CRPSs WITH STANDARD DEVIATIONS OF HISTORICAL ACR
UNDER VARIOUS MATCHING RATES

Matching Rate	Model	
	ACR-GP	BACL
10	5.054±0.997	0.866±0.101
20	2.670±0.767	0.624±0.045
30	1.830±0.292	0.609±0.078
40	1.289±0.155	0.503±0.077
50	0.992±0.173	0.500±0.053
60	0.876±0.113	0.423±0.053
70	0.754±0.101	0.357±0.053
80	0.692±0.049	0.354±0.059
90	0.632±0.048	0.336±0.061

curves. With relatively sufficient observations, the associated uncertainty estimates are small and trend consistently with the mean estimates. In scenarios with extremely low MRs, e.g.,

Fig. 6(c), (f), (i), the mean estimates may slightly diverge from the ground truth. However, the accompanying uncertainty (primarily lane choice uncertainty) estimates enhance reliability, ensuring the results remain robust even with limited data.

C. Real-Time Arrival Curve Reconstruction Results

To assess the real-time reconstruction performance, we display the distributions of RMSE and CRPS under various matching rates, as shown in Fig. 7. For each matching rate, experiments are repeated 20 times. The links connecting the boxplot centers represent the median values of RMSEs and CRPSs. Compared to the historical reconstruction results in Tab. I and Tab. II, the RMSEs and CRPSs for real-time reconstruction are even higher. This increase is attributed to the historical reconstructions benefiting from bilateral boundary constraints, e.g., the adjacent matched vehicles illustrated in Fig. 1(b). Also, the estimation span is usually smaller than the update time of matching vehicles, e.g., 31s~141s. In contrast, the real-time reconstruction faces more stringent challenges with only unilateral boundary constraints (e.g., the newest matched vehicle), and the estimation span is typically larger than the vehicle link travel time, e.g., 265s~437s. Despite numerical differences, the RMSE and CRPS metrics exhibit consistent trends across varying matching rates. As the matching rates increase, the values of both metrics improve, showing decreased values and smaller variances. This improvement is due to higher matching rates bringing the estimation span closer to the vehicle link travel time, therefore reducing the lane choice uncertainties associated with link-based arrivals. Compared to the deterministic RMSE, the variances of probabilistic CRPSs are even lower at the same matching rates,

TABLE III
SIGNIFICANCE OF INPUT FEATURES

Matching Rate	Model		
	w/o link arrivals	w/o lane vector	BACL
10	2.582±0.094	1.479± 0.264	1.607±0.200
20	1.884±0.068	1.138±0.152	1.359±0.106
30	1.622±0.188	1.018±0.279	1.224±0.170
40	1.189±0.135	0.940±0.248	0.926±0.439
50	1.018±0.137	0.769±0.195	0.736±0.165
60	1.020±0.107	0.790±0.149	0.734±0.152
70	0.983±0.132	0.655±0.200	0.615±0.216
80	0.877±0.120	0.614±0.351	0.531±0.135
90	0.878±0.136	0.541±0.215	0.504±0.152

indicating the reliability of our approach in quantifying lane choice uncertainties.

To further show the real-time arrival reconstruction performance, we visualize nine real-time lane-based vehicle count estimation examples from three lanes during three testing periods under 50% matching rates, as shown in Fig. 8. Overall, the BACL provides accurate lane-based vehicle count estimates with reliable uncertainties, indicating the superiority of the proposed BACL in estimating real-time operational metrics.

D. Significance of Different Input Features

To examine the significance of input features, two additional BACL models are configured using different levels of LPR data features in Tab. III. First, in the w/o link arrivals setting, the link-based arrivals input is removed in the BACL model, making it challenging to learn the lane choices from link-based arrivals to lane-based arrivals. As a result, the model degrades sharply across various matching rates, implying the importance of incorporating link-based arrival data. Second, the w/o lane vector setting uses aggregated link-based arrivals instead of a detailed vector input that contains upstream lane information. Interestingly, in scenarios with low MRs, this simplification results in lower RMSEs. The reduced complexity of this model variant requires fewer parameters, making it more suitable for conditions with limited data.

VI. CONCLUSION

In this study, a Bayesian deep learning approach is developed to reconstruct real-time lane-based traffic arrival curves using LPR data. Targeting real-time reconstruction, a lane choice learning process is designed to effectively capture the relationship between link-based arrivals and lane-specific arrivals. Moreover, the lane choice uncertainties are characterized using Bayesian parameter inference techniques, minimizing the arrival curve reconstruction uncertainties. Extensive experiment results demonstrate the superiority and necessity of lane choice modeling in reconstructing historical and real-time lane-based arrival curves.

There are several future directions to advance this study. First, the same matching rates of LPR data are used in the arrival curve learning and reconstruction stages. It is practically meaningful to test the model performance with different matching rates at the two stages because the matching rate in reality is not stationary due to equipment aging or maintenance. Second, it is possible to extend the proposed

link-level model to a network-scale traffic arrival acquisition model, in which the proposed model can serve as a sub-module to obtain link-level traffic demand.

REFERENCES

- [1] D. Ni, *Traffic Flow Theory: Characteristics, Experimental Methods, and Numerical Techniques*. Amsterdam, The Netherlands: Butterworth-Heinemann, 2015.
- [2] C. An, X. Guo, R. Hong, Z. Lu, and J. Xia, "Lane-based traffic arrival pattern estimation using license plate recognition data," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 4, pp. 133–144, Jul. 2022.
- [3] M. Li, J. Tang, Q. Chen, and Y. Liu, "Traffic arrival pattern estimation at urban intersection using license plate recognition data," *Phys. A, Stat. Mech. Appl.*, vol. 625, Sep. 2023, Art. no. 128995.
- [4] C. An, Y. He, J. Lu, Z. Lu, and J. Xia, "One-stage estimation of cyclic arrival rates using license plate recognition data," *J. Intell. Transp. Syst.*, pp. 1–14, Aug. 2024, doi: [10.1080/15472450.2024.2392720](https://doi.org/10.1080/15472450.2024.2392720).
- [5] J. Zheng and H. X. Liu, "Estimating traffic volumes for signalized intersections using connected vehicle data," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 347–362, Jun. 2017.
- [6] C. An, Y.-J. Wu, J. Xia, and W. Huang, "Real-time queue length estimation using event-based advance detector data," *J. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 277–290, Jul. 2018.
- [7] S. Lee, K. Xie, D. Ngoduy, and M. Keyvan-Ekbatani, "An advanced deep learning approach to real-time estimation of lane-based queue lengths at a signalized junction," *Transp. Res. C, Emerg. Technol.*, vol. 109, pp. 117–136, Dec. 2019.
- [8] H. Wu, L. Luo, T. Oguchi, K. Tang, and H. Zhu, "Stochastic queue profile estimation using license plate recognition data," *Phys. A, Stat. Mech. Appl.*, vol. 643, Jun. 2024, Art. no. 129790.
- [9] L. Li, W. Huang, and H. K. Lo, "Adaptive coordinated traffic control for stochastic demand," *Transp. Res. C, Emerg. Technol.*, vol. 88, pp. 31–51, Mar. 2018.
- [10] Z. Yao, L. Shen, R. Liu, Y. Jiang, and X. Yang, "A dynamic predictive traffic signal control framework in a cross-sectional vehicle infrastructure integration environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1455–1466, Apr. 2020.
- [11] Z. Xu et al., "Trajectory optimization for a connected automated traffic stream: Comparison between an exact model and fast heuristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2969–2978, May 2021.
- [12] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1841–1851, Mar. 2022.
- [13] E. Amini, A. Omidvar, and L. Eleftheriadou, "Optimizing operations at freeway weaves with connected and automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 126, May 2021, Art. no. 103072.
- [14] N. Dobrota, A. Stevanovic, and N. Mitrovic, "A novel model to jointly estimate delay and arrival patterns by using high-resolution signal and detection data," *Transportmetrica A, Transp. Sci.*, vol. 20, no. 1, pp. 1–33, Jan. 2024.
- [15] C. Tan, J. Yao, X. Ban, and K. Tang, "Cumulative flow diagram estimation and prediction based on sampled vehicle trajectories at signalized intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11325–11337, Aug. 2022.
- [16] X. Luo, D. Ma, S. Jin, Y. Gong, and D. Wang, "Queue length estimation for signalized intersections using license plate recognition data," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 209–220, Fall 2019.
- [17] K. Tang, H. Wu, J. Yao, C. Tan, and Y. Ji, "Lane-based queue length estimation at signalized intersections using single-section license plate recognition data," *Transportmetrica B, Transp. Dyn.*, vol. 10, no. 1, pp. 293–311, Dec. 2022.
- [18] W. Shao and L. Chen, "License plate recognition data-based traffic volume estimation using collaborative tensor decomposition," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3439–3448, Nov. 2018.
- [19] W. Rao, Y.-J. Wu, J. Xia, J. Ou, and R. Kluger, "Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 29–46, Oct. 2018.
- [20] B. Mo, R. Li, and J. Dai, "Estimating dynamic origin-destination demand: A hybrid framework using license plate recognition data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 7, pp. 734–752, 2020.
- [21] W. Yao, M. Zhang, S. Jin, and D. Ma, "Understanding vehicles commuting pattern based on license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103142.

- [22] X. Zhan, R. Li, and S. V. Ukkusuri, "Lane-based real-time queue length estimation using license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 57, pp. 85–102, Aug. 2015.
- [23] B. Mo, R. Li, and X. Zhan, "Speed profile estimation using license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 82, pp. 358–378, Sep. 2017.
- [24] C. M. Day et al., "Evaluation of arterial signal coordination: Methodologies for visualizing high-resolution event data and measuring travel time," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2192, no. 1, pp. 37–49, Jan. 2010.
- [25] J. Zheng, H. X. Liu, S. Misgen, K. Schwartz, B. Green, and M. Anderson, "Use of event-based traffic data in generating time–space diagrams for evaluation of signal coordination," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2439, no. 1, pp. 94–104, Jan. 2014.
- [26] X. Zhan, Y. Zheng, X. Yi, and S. V. Ukkusuri, "Citywide traffic volume estimation using trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 272–285, Feb. 2017.
- [27] Y. Zhao, J. Zheng, W. Wong, X. Wang, Y. Meng, and H. X. Liu, "Various methods for queue length and traffic volume estimation using probe vehicle trajectories," *Transp. Res. C, Emerg. Technol.*, vol. 107, pp. 70–91, Oct. 2019.
- [28] P. Hao, Z. Sun, X. J. Ban, D. Guo, and Q. Ji, "Vehicle index estimation for signalized intersections using sample travel times," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 513–529, Nov. 2013.
- [29] H. Zhang, H. X. Liu, P. Chen, G. Yu, and Y. Wang, "Cycle-based end of queue estimation at signalized intersections using low-penetration-rate vehicle trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3257–3272, Aug. 2020.
- [30] F. Li, K. Tang, J. Yao, and K. Li, "Real-time queue length estimation for signalized intersections using vehicle trajectory data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2623, no. 1, pp. 49–59, Jan. 2017.
- [31] Z. Wen and X. Weng, "Inferring the number of vehicles between trajectory-observed vehicles," *J. Intell. Transp. Syst.*, vol. 28, no. 6, pp. 816–829, Nov. 2024.
- [32] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.



Yang He received the B.E. degree in traffic engineering from Chang'an University, Xi'an, China, in 2020. He is currently pursuing the Ph.D. degree with the Intelligent Transportation System Research Center, Southeast University. His current research interests include transportation network modeling, traffic state estimation, and low-rank modeling.



Chengchuan An received the Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2019. From 2014 to 2016, he was a Visiting Scholar with the Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, USA. Since 2020, he has been a Post-Doctoral Researcher with the Intelligent Transportation System Research Center, Southeast University. His current research interests include intelligent traffic signal control systems and traffic data mining.



Jiawei Lu received the Ph.D. degree in transportation engineering from Arizona State University in 2022. He is currently a Post-Doctoral Fellow with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology. His research interests include transportation network modeling, traffic simulation, and distributed traffic control.



Yao-Jan Wu is a Professor in transportation engineering with the Civil and Architectural Engineering and Mechanics Department and the Executive (Founding) Director of the Center for Applied Transportation Sciences (CATS), The University of Arizona (UA). He has authored or co-authored over 160 refereed publications, including more than 80 journal articles, and has presented his research findings at more than 100 national and international conferences and invited speaker events. His research focuses on a strong connection between information technology (IT) and traditional transportation research.



Zhenbo Lu received the Ph.D. degree in traffic information engineering and control from Southeast University, Nanjing, China, in 2011. He is currently an Associate Professor with the Intelligent Transportation System Research Center, Southeast University. His research interests include transportation planning, traffic simulation, and intelligent transportation systems.



Jingxin Xia received the Ph.D. degree in transportation engineering from the University of Kentucky, USA, in 2006. He is currently a Professor with the Intelligent Transportation System Research Center, Southeast University, Nanjing, China. He has published more than 40 peer-reviewed articles so far. His research interests include traffic flow theory, transportation network modeling, traffic signal control, and intelligent transportation systems.