Learning Solutions for LWR-type Traffic Flow Models

Bilal Thonnam Thodi*[1,2] and Saif Eddin Jabari[3,4]

[1]Ph.D. Candidate, New York University Tandon School of Engineering, USA

[2]Graduate Research Assistant, Engineering Division, New York University Abu Dhabi, UAE

[3]Associate Professor, Civil and Urban Engineering, New York University Abu Dhabi, UAE

[4]Global Network Associate Professor, New York University Tandon School of Engineering, USA

SHORT SUMMARY

First-order macroscopic traffic flow models in the form of partial differential equations are conventionally solved using numerical schemes which are grid-dependent. We propose a kernel-based method for learning solutions of first-order traffic flow models. The solution kernels are approximated by Fourier Neural Operators - a variant of deep neural networks. Unlike the conventional schemes, our method learns solutions to arbitrary initial and boundary conditions. This avoids resolving the problem for every new instance of input conditions, thereby lowering the computational cost. We apply this method for learning traffic density solutions of the Lighthill-Witham-Richards (LWR) traffic flow model. Numerical experiments to show the neural network solution's accuracy, grid-independence, robustness, and computational complexity are included.

Keywords: Deep Learning, Learning Traffic Dynamics, LWR Traffic Flow Model.

1. INTRODUCTION

Macroscopic descriptions of traffic flow assume a continuum fluid approximation and model the evolution of aggregated variables such as traffic density, flow and speed over space and time. Of several models available, the class of Lighthill-Witham-Richards (LWR) models (Lighthill & Whitham, 1955; Richards, 1956) is widely used for modeling freeway and urban traffic. The LWR-type models are non-linear hyperbolic partial differential equations (PDE), for which there exist discontinuous solutions even with smooth initial densities (LeVeque, 1992; Whitham, 1999).

The LWR PDE is conventionally solved using first-order numerical schemes such as the Godunov scheme (Lebacque, 1996) or the minimum supply-demand method (Daganzo, 1994). These methods discretize the space-time domain to computational cells and approximate an average PDE solution in each cell. The size of these computational cells is restricted by the Courant-Friedrich-Lewy (CFL) condition for numerical stability. The solutions are approximate and improve with finer discretizations, i.e., close to the CFL limit. These methods fall under finite-volume based schemes (Kessels, 2019). There also exist finite-difference based numerical schemes, where a viscous form of LWR PDE is solved; viscosity is added for well-posedness. In this case, the solutions are defined at discrete points in the computational domain. However, adding viscosity leads to unphysical solutions. The computational cost for these numerical schemes is proportional to the grid resolution.

Solution for general initial and boundary conditions, measurable over the space-time domain, and that obtained at lower computational cost are still missing in the literature. To this end, we propose

a methodology for learning LWR PDE solutions using deep learning (DL) models. DL-based solutions for PDEs have attained recent attention due to their good generalization power and lower computational cost (Raissi, Yazdani, & Karniadakis, 2020). Some recent studies include (Li et al., 2020; Raissi, Perdikaris, & Karniadakis, 2019; Rudy, Brunton, Proctor, & Kutz, 2017). These studies exploited the representation power of deep neural networks in approximating complex dynamics. In the traffic flow literature, (Shi, Mo, Huang, Di, & Du, 2021) attempt to solve a viscous form of LWR PDE using physics informed neural networks (Raissi et al., 2019). But, the neural network model is trained to solve an instance of boundary condition and requires re-training for new instance of boundary condition, which is computationally costly. We overcome this limitation in our proposed solution.

Unlike classical numerical techniques, our solution method is data-driven. The idea is to write the LWR PDE solution as an integral of kernel convolution with input conditions and then learn the kernel function from data. In this study, we approximate the kernel function using a Fourier Neural Operator (FNO) (Li et al., 2020) − a variant of deep neural networks, recently proposed to solve general PDEs. The FNO model learns density solutions for arbitrary input conditions and is grid resolution-invariant. The latter property allows one to train the solution kernels at a lower grid resolution (which is computationally less expensive to generate) and test them at higher grid resolution. Towards the end of the paper, we conduct experiments to assess the generalization performance, resolution-invariance and computational aspects of the proposed neural network solution.

In the remainder of the paper, we present the LWR model of traffic flow and formulate the Fourier Neural Operator solution. Finally, solutions for a few selected problem instances is presented, followed by a brief discussion on the features and limitations of the proposed method.

2. METHODOLOGY

LWR Traffic Flow Model

Consider a road section with boundaries $x = x_{\mathrm{l}}$ and $x = x_{\mathrm{u}}$, and assume a finite time period $t \in [0, t_{\mathrm{m}}]$. Denote $\rho(x,t)$ as the traffic density and $q(x,t)$ as the traffic flow at (x,t). The LWR model of traffic density dynamics is,

$$\begin{aligned} &\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial Q(\rho(x,t))}{\partial x} = 0; && x \in [x_{\mathrm{l}},\, x_{\mathrm{u}}], \quad t \in [0,\, t_{\mathrm{m}}] \\ &\rho(x_{\mathrm{l}},t) = \bar{\rho}_{\mathrm{l}}(t), \;\; \rho(x_{\mathrm{u}},t) = \bar{\rho}_{\mathrm{u}}(t); && t \in (0,\, t_m] \\ &\rho(x,0) = \bar{\rho}_0(x); && x \in (x_l,\, x_u) \end{aligned} \tag{1}$$

where $Q(\rho(x,t)) = q(x,t)$ is the equilibrium flow-density relation, $\bar{\rho}_0(x)$ is the initial density profile, and $\bar{\rho}_{\mathrm{l}}(t)$, $\bar{\rho}_{\mathrm{u}}(t)$ are the prescribed boundary densities. The system (1) is an initial-boundary valued problem. Next, we formulate an approximate solution for (1).

General Solution for Traffic Density

We assume that the traffic density $\rho(x,t)$ is approximated using a convolutional integral as,

$$\rho(x,t) := \int_{\xi=x_u}^{\xi=x_l} \int_{\tau=0}^{\tau=t_m} g(x-\xi, t-\tau)\, f(\xi,\tau)\, d\xi\, d\tau = f * g\,(x,t). \tag{2}$$

Here f is an input function that encodes the prescribed initial and boundary densities, g is an unknown kernel function, and $*$ is the convolution operator. The solution (2) implies that the traffic density at any point in the space-time plane is obtained by convolving the prescribed densities from initial and boundary points using the kernel function g.

The solution (2) is motivated from the Green's function for solving linear PDEs. For instance, the general solution to the famous heat equation can be written in similar form as (2) with g as a Gaussian kernel function (Evans, 2010). We assume there exist similar kernel function g for the non-linear traffic density dynamics. We rewrite the convolutional integral in (2) using the Fourier Transforms as follows:

$$\begin{aligned}\rho(x,t) &= f * g\ (x,t) \\ &= \mathscr{F}^{-1}\left[\mathscr{F}\left(f * g\right)\right]\ (x,t) \\ &= \mathscr{F}^{-1}\left[\mathscr{F}\left(f\right) \times G\right]\ (x,t),\end{aligned} \tag{3}$$

where $\mathscr{F}(.)$ and $\mathscr{F}^{-1}(.)$ are the Fourier Transform and the Inverse Fourier Transform, and $G := \mathscr{F}(g)$ is the kernel function defined in the Fourier space. The second equality in (3) is the Fourier decomposition. The third equality in (3) follows from the convolution theorem. We formulate (3) to a neural network model, which is described next.

Fourier Neural Operator (FNO) Model

The solution (3) is a global operation, i.e., the traffic density at (x,t) depends on the inputs from whole space-time domain because of the Fourier decomposition. We add an additional local term $wf(x,t)$ to (3) to account for local correlations in traffic density. Together with a non-linear function $\sigma(a) := \max(0,a)$, we have the density at (x,t) as,

$$\rho(x,t) = \sigma\left[w \times f + \mathscr{F}^{-1}\left(\mathscr{F}\left(f\right) \times G\right)\right]\ (x,t)\ . \tag{4}$$

Eq. (4) defines the traffic density solution at a single point (x,t). We rewrite (4) into a matrix form to define the traffic density in the entire space-time domain,

$$\rho = \sigma\left[W \times \rho^{(0)} + \mathscr{F}^{-1}\left(\mathscr{F}\left(\rho^{(0)}\right) \times G\right)\right], \tag{5}$$

where $\mathscr{D} = \{(x,t)\}^{n \times m}$ is the discretized space-time domain, $\rho := \{\rho(x,t)\}_{\mathscr{D}} \in \mathbb{R}^{n \times m}_{[0,k_{\text{jam}}]}$ is the density matrix defined over $\mathscr{D}$, and $\rho^{(0)} := f$ is the input matrix (prescribed initial and boundary densities) over $\mathscr{D}$.

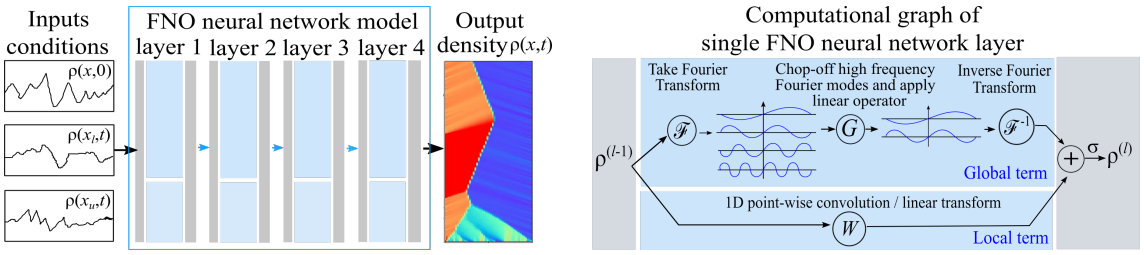
In (5), $W : \mathbb{R}^{n \times m \times k_1} \to \mathbb{R}^{n \times m \times k_2}$ is a point-wise linear operator (one-dimensional convolution) that maps feature vector from latent dimension k_1 to k_2. $G \in \mathbb{C}^{n \times m \times k_1 \times 2}$ is the kernel parameter in the Fourier space. $\mathscr{F}(\rho) : \mathbb{R}^{n \times m \times k_1} \to \mathbb{C}^{n \times m \times k_2 \times 2}$ is the two-dimensional Fourier Transform. The non-linearity $\sigma(.)$ is applied element-wise.

Eq. (5) is the parametric neural network approximation of the traffic density for (1), with trainable parameters W and G. It takes an input matrix $\rho^{(0)}$ and outputs the traffic density over $\mathscr{D}$. This forms a single layer of the Fourier Neural Operator (FNO) model proposed in (Li et al., 2020). We stack deep layers of (5) to improve generalization. The forward propagation for an L layer FNO model is written as,

$$\rho^{(l)} = \sigma\left[W^{(l)} * \rho^{(l-1)} + \mathscr{F}^{-1}\left(\mathscr{F}\left(\rho^{(l-1)}\right) \times G^{(l)}\right)\right], \quad l = 0, \ldots, L, \tag{6}$$

where $W^{(l)}$ and $G^{(l)}$ are the trainable parameters in each layer l. A 4-layer FNO model architecture is shown in Figure 1.

The forward computation of (6) consist of two operations: a local convolution and a global Fourier decomposition. This is depicted in Figure 1b. The first term extracts local information. The second term extracts global information and handle sparse inputs by retaining low-order information from the Fourier Transform. In other words, the high frequency Fourier coefficients are removed before



(a) A 4-layer FNO model (b) A single FNO layer

Figure 1: Fourier Neural Operator model (Adapted from (Li et al., 2020)).

the Inverse Fourier Transform. This improves the generalization and avoids over-fitting, especially to noisy data (Li et al., 2020).

The solution (6) is a numerically differentiable function. Hence, the parameters $\left\{\left(W^{(l)}, G^{(l)}\right) : l = 0, \ldots, L\right\}$ can be optimized using gradient descent-based algorithms in a supervised learning setting. Solution for a new instance of initial and boundary conditions (i.e., testing) is obtained from a single forward pass of (6). The computational cost of (6) is limited by the 2-D Fourier Transforms, which is of order $\mathbb{O}(n \log m)$ with Fast Fourier Transform techniques.

The solution (6) is also grid resolution-invariant, i.e., measurable at all points in $\mathscr{D}$, since we approximate the density in the Fourier space. In the experiments, we investigate this by training (6) at a lower grid resolution and testing at higher resolution.

3. RESULTS AND DISCUSSION

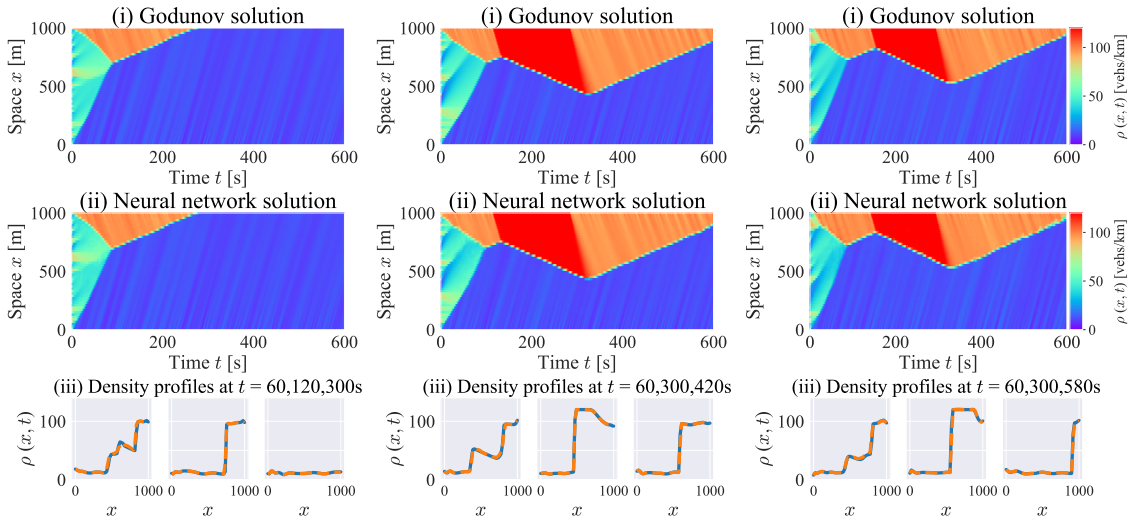
We empirically assess the quality of the FNO solution (6). The data for training (6) is generated from the Godunov numerical scheme (Kessels, 2019) with random input conditions: initial densities $\rho_0 \sim \mathrm{U}[0, 100]$ (vehs/km), boundary fluxes, $q_{\text{in}} \sim \mathrm{U}[300, 1500]$ (vehs/hr) and $q_{\text{out}} \sim \mathrm{U}[800, 1500]$ (vehs/hr). For some instances, we set the outflow $q_{\text{out}} = 0$ vehs/hr for an arbitrary time period to replicate traffic signals.

Other parameters of the training data are: length of road section $x_u - x_l = 1000$ m, time period $t_m = 600$ secs, and cell size dimensions 20 m $\times$ 1 sec. We assume the Greenshield's fundamental relation with jam density $k_{\text{jam}} = 120$ vehs/km and free-flow speed $v_{\text{free}} = 60$ km/hr. The training set consist of 7500+ samples. The number of layers of the FNO model (6) is $l = 4$. The testing results are discussed below.

Learned Density Solution

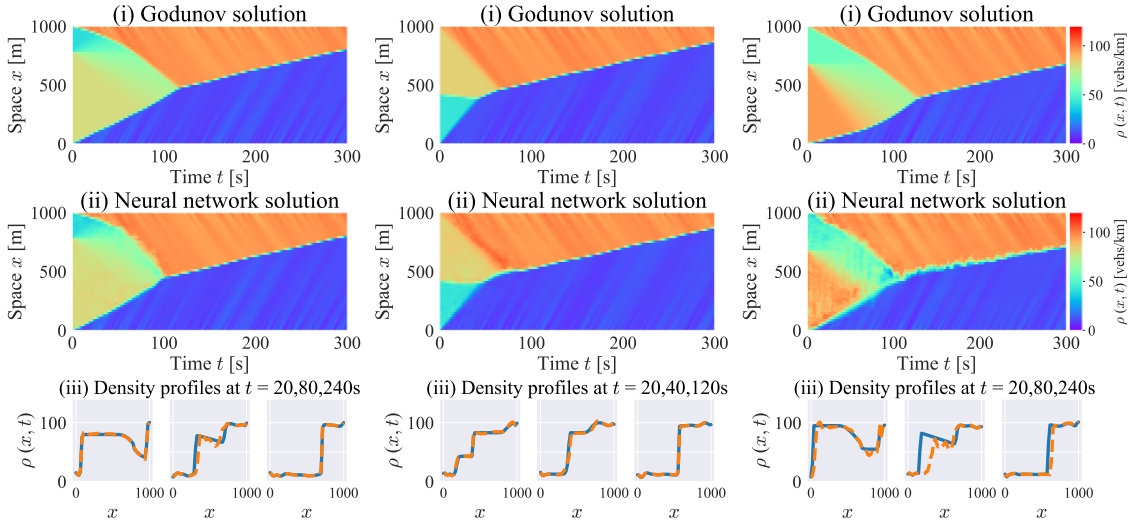
The traffic density solutions for a few selected problem instances are shown in Figure 2 through Figure 4. The first row shows the Godunov solution, the second row shows the FNO solution (6), and the third row shows the density profiles at various time instants. The first two columns represent examples of good approximation, and the third column shows an example of poor approximation. Overall, we observe an average approximation error of $\leq 5\%$, which is acceptable.

Figure 2 shows sample solutions from a validation dataset, i.e., for different input conditions but with traffic scenarios similar to the training data. We see that the shockwaves separating different regimes are correctly reconstructed. The model also accurately predicts the minor waves emanating from the boundaries. The density profiles show a perfect approximation.



(a) without traffic signal (b) one traffic signal (c) one traffic signal

Figure 2: Comparison of solutions for three problem instances from the validation dataset (Val).



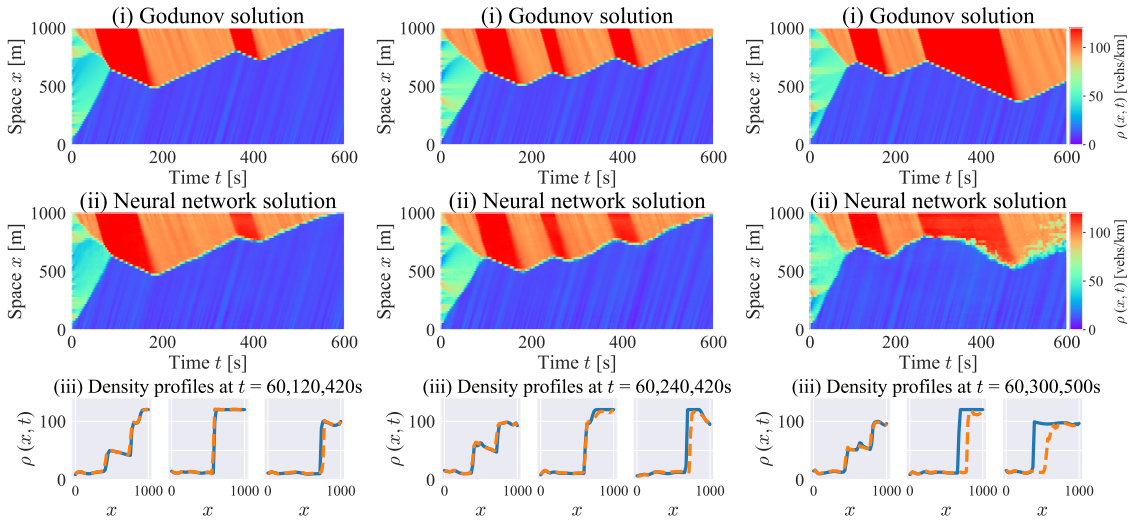
(a) rarefaction wave (b) shockwave (c) rarefaction wave

Figure 3: Comparison of solutions for three problem instances from the testing dataset - different initial conditions (Test-IC).

Figure 3 and Figure 4 shows sample solutions from a testing dataset, i.e., for the traffic scenarios and input conditions not seen during training. Figure 3 present solutions for jump initial conditions; the sub-figures are (a) rarefaction wave and (b) shockwave. The shockwave reconstruction is accurate, but the rarefaction waves incur minor errors. An example with a bad rarefaction wave is shown in sub-figure (c). The training data lack scenarios with rarefaction waves of longer periods.

Figure 4 present solutions for (a) two and (b) three traffic signals - not seen during training. The queuing dynamics predicted by the FNO are accurate, except in a few test cases. An example of a poor approximation is shown in the sub-figure (c), for which the FNO solution predicts queue dissipation earlier than that in the true solution. Nevertheless, the average errors are within the acceptable range, given that these scenarios are not observed in the training data.

The results presented above are promising. However, the test cases presented here to assess the generalization error of the FNO model are far from complete. This is only a preliminary experi-



(a) two traffic signals (b) three traffic signals (c) two traffic signals

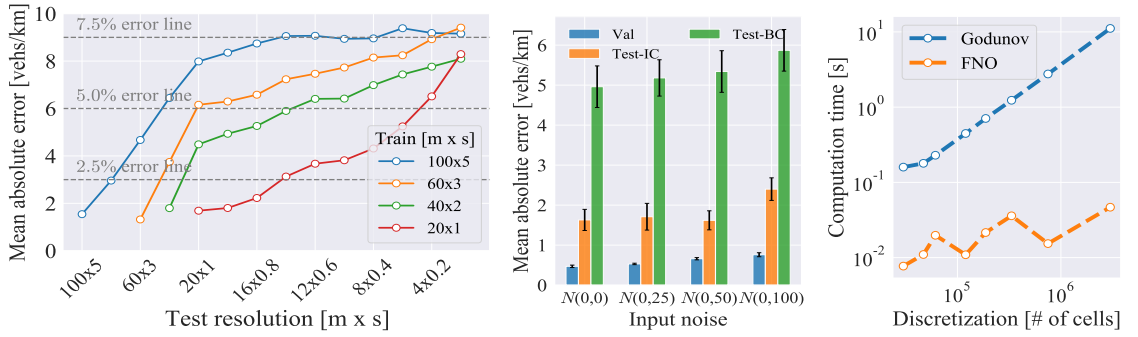
Figure 4: Comparison of solutions for three problem instances from the testing dataset - different boundary conditions (Test-BC).

ment, where the model is trained with a small set of random numerical solutions. More detailed study incorporating diverse traffic scenarios can improve the generalization further.

Properties of Learned Solutions

We next discuss a few properties of the FNO solution (6).

Resolution Invariance. As mentioned before, the solution (6) is grid resolution-invariant, i.e., defined at all (x,t). This feature allows one to train models using solutions of coarser-resolution and generalize them to finer grid resolutions. This is beneficial since obtaining coarser resolution data is computationally cheaper. To study the limit of this resolution-invariance trade-off, we train four models at lower grid resolutions and test them at their respective higher resolutions. The average density error is summarized in Figure 5a.



(a) Resolution-invariance (b) Noise robustness (c) Time complexity

Figure 5: Properties of the learned density solution.

We observe that most test cases fall below the 7.5% error rate. The models perform well near the resolutions they are trained on, but the performance declines at a finer resolution. The declination rate depends on the train resolution. For instance, see the curve corresponding to 100 m $\times$ 5 sec and 20 m $\times$ 1 sec in Figure 5a. To conclude, assuming an acceptable error rate of $\leq 5\%$, the FNO solution (6) can generalize to resolutions $\times 4$ higher than the train resolution.

Robustness. We further assess the FNO model (6) performance for noisy inputs, i.e., for weak initial and boundary conditions. The density errors for different noise levels (white Gaussian) in the boundary flows are shown in Figure 5b. We see a negligible increase in the density errors at higher noise levels, implying that the FNO model can handle noisy inputs. This also affirms the stability of the FNO solutions to input perturbations.

Computational time. An important, appealing feature of the proposed method is the computational complexity. As noted before, the prime computational bottleneck of the solution (6) is the Fourier transform, which of order $\mathbb{O}(n \log m)$, whereas the Godunov scheme is $\mathbb{O}(nm)$. We compare the average compute time of (6) and the Godunov scheme in Figure 5c for different grid size. Note that the FNO model is evaluated on a GPU (to exploit the parallel computation in the neural network model), and the Godunov scheme is evaluated on a CPU. Clearly, the FNO model incurs a negligible increase in the compute time for large problem sizes compared to the Godunov scheme.

Discussion on Learning-based Solution

The results show a promising direction for solving macroscopic traffic flow models using data-driven learning-based techniques. How well these techniques compare against traditional numerical solvers is still an open research question. This depends on the generalization performance (out-of-sample error) of neural network models, and we assessed them empirically in this study. The choice of training dataset also affects the generalization error and is an open research question.

A limitation of the proposed learning-based solution (6) is that it is agnostic to traffic physics. This partly explains the poor approximation seen in a few test cases, for e.g., Figure 2c, Figure 3c and Figure 4c. We expect the solution (6) to learn the governing physics from the training data, but the physics is not guaranteed to hold while testing. The neural network solutions that explicitly impose traffic physical constraints is important for reliable applications (Thodi, Khan, Jabari, & Menendez, 2021a, 2021b), and is a potential future direction.

4. CONCLUSIONS

We propose a data-driven method for learning the traffic density solutions of the LWR traffic flow model. We approximate the traffic density using the Fourier Neural Operator (FNO) model – a variant of the deep neural network. The proposed method learns density solutions for arbitrary initial and boundary conditions, unlike classical numerical schemes. This avoids the need for resolving the problem for every new instance of input conditions, and thereby lower computational cost.

Our empirical analysis has shown that the average approximation error of the FNO model is $\leq$ 5%, which is within the acceptable range. The model trained with random input conditions can sufficiently generalize to selected traffic scenarios. However, we have seen a few test cases where the FNO model failed to produce the correct solution. Arbitrary generalization warrants a diverse training dataset and an explicit constraining of physical dynamics. We also found that the FNO solutions can be transferred to resolutions $\times 4$ higher than the train resolution, robust to noisy inputs and incur lower computational costs than traditional solvers.

The ultimate research question we try to answer here is whether we can learn traffic density dynamics using parametric models with a lower computational burden than numerical solvers. This study shows promising results. Our efforts continue to build on this research direction with a more explicit constraining of traffic physics on the learning model.

ACKNOWLEDGMENT

This work was supported by the NYU Abu Dhabi (NYUAD) Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001.

REFERENCES

Daganzo, C. F. (1994, August). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28*(4), 269–287.

Evans, L. C. (2010). *Partial differential equations*. Providence, R.I.: American Mathematical Society.

Kessels, F. (2019). *Traffic Flow Modelling: Introduction to Traffic Flow Theory Through a Genealogy of Models*. Cham: Springer International Publishing.

Lebacque, J.-P. (1996). The godunov scheme and what it means for first order traffic flow models. In *Internaional symposium on transportation and traffic theory* (pp. 647–677).

LeVeque, R. (1992). *Numerical methods for conservation laws* (Vol. 132). Springer.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). *Fourier neural operator for parametric partial differential equations.*

Lighthill, M., & Whitham, G. (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. In *Royal society of london. series a, mathematical and physical sciences* (Vol. 229, pp. 317–345).

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019, February). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Raissi, M., Yazdani, A., & Karniadakis, G. E. (2020, February). Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, *367*(6481), 1026–1030.

Richards, P. I. (1956, February). Shock Waves on the Highway. *Operations Research*, *4*(1), 42–51.

Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017, April). Data-driven discovery of partial differential equations. *Science Advances*, *3*(4), e1602614.

Shi, R., Mo, Z., Huang, K., Di, X., & Du, Q. (2021). Physics-informed deep learning for traffic state estimation. *arXiv preprint arXiv:2101.06580*.

Thodi, B. T., Khan, Z. S., Jabari, S. E., & Menendez, M. (2021a). *Incorporating kinematic wave theory into a deep learning method for high-resolution traffic speed estimation.*

Thodi, B. T., Khan, Z. S., Jabari, S. E., & Menendez, M. (2021b). Learning traffic speed dynamics from visualizations. In *2021 ieee international intelligent transportation systems conference (itsc)* (p. 1239-1244).

Whitham, G. B. (1999). *Linear and Nonlinear Waves: Whitham/Linear*. Hoboken, NJ, USA: John Wiley & Sons, Inc.