

# PROBABILISTIC FINE-GRAINED URBAN FLOW INFERENCE WITH NORMALIZING FLOWS

Ting Zhong, Haoyang Yu, Rongfan Li, Xovee Xu\*, Xucheng Luo, and Fan Zhou

University of Electronic Science and Technology of China, Chengdu, China  
{zhongting, xucheng, fan.zhou}@uestc.edu.cn, haoyang.yu417@outlook.com, xovee@ieee.org

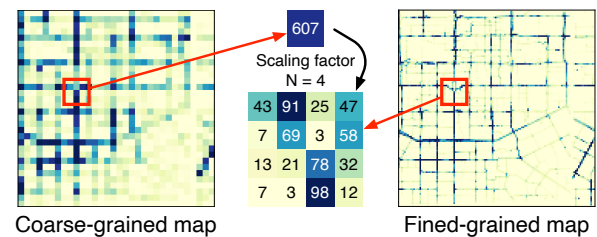
## ABSTRACT

Fine-grained urban flow inference (FUFI) aims at enhancing the resolution of traffic flow, which plays an important role in intelligent traffic management. Existing FUFI methods are mainly based on techniques from image super-resolution (SR) models, which cannot fully capture the influence of external factors and face the ill-posed problem in SR tasks. In this paper, we propose UFI-Flow – Urban Flow Inference via normalizing Flow, a novel model for addressing the FUFI problem in a principled manner by using a single probabilistic loss. UFI-Flow therefore directly accounts for the ill-posed nature of the problem and learns spatial correlations between urban flow maps. In addition, an augmented distribution fusion mechanism is further proposed to reinforce the influence of external factors in the joint distribution inference. We conduct comprehensive experiments on real-world datasets to show the superiority of the proposed model compared to the state-of-the-art baseline approaches.

**Index Terms**— Spatial-temporal data mining, urban flow inference, normalizing flow, mobile sensing

## 1. INTRODUCTION

Urban traffic flow forecasting systems are crucial for various applications such as urban planning, public safety, and daily travel [1]. On one hand, it can provide insights for decision making, risk assessment, and traffic management for governments and entrepreneurs [2, 3]. On the other hand, it may help improve the travel plans in advance, stagger the peak and save commute time [4]. Conventionally, urban traffic flow inferences and predictions rely on a massive number of sensors deployed citywide. However, these data are usually coarse and fuzzy due to the limitation of the number of sensors. Toward this end, the Fine-Grained Urban Flow Inference (FUFI) problem has been studied recently [5, 6], which aims to infer the fine-grained (FG) flow map from its corresponding coarse-grained (CG) one. As a result, FUFI models



**Fig. 1:** Structural constraints between coarse- and fine-grained map in Xi'an, China. The sum of the traffic flows in the lower side subregions is strictly equal to the traffic flow in the upper side superregion.

can achieve a satisfactory performance within only a few sensors, saving electricity, deployment, maintenance, and labor for efficient urban computing.

FUFI problem is similar to traditional single image super-resolution (SR) task [7, 8], but it has several intrinsic differences that pose new challenges: (1) strict structural constraints, the sum of traffic flow within a certain region (subregions) in the inferred FG map is strictly equal to the sum of traffic flow in the corresponding superregion of the original CG map (cf. Fig. 1); and (2) the FG inference results are greatly affected by various external factors, e.g., time, traffic and weather conditions.

**Current solutions.** Existing methods have proposed  $N^2$ -Normalization [5] and  $AN^2$ -Normalization [9] mechanisms to solve the structural constraints and use sub-network to fuse the external factors for final inference. However, these methods still face several notable shortcomings: (1) They train a deterministic mapping using  $MSE$  or  $MAE$  losses. Despite the impressive results they achieved, the ill-posed nature of the FUFI tasks is ignored, resulting the generated flow maps tend to be blurry [10, 11]. The employed reconstruction losses favor the prediction of an average traffic flow over the plausible FUFI solution, which leads to low perceptual quality and significant reduction of details in high-frequency traffic flow; and (2) They ignore the external factors after the structural constraint process, which may result in important information loss during the inference procedure and, as a result, significant performance degradation.

**Present work.** To address the aforementioned problems in FUFI, we propose a novel approach for Urban Flow Inference

\* Corresponding author

This work was supported by National Natural Science Foundation of China (Grant No. 62176043 and No. 62072077), and Sichuan Science and Technology Program (Grant No. 2020YFG0053).

via normalizing Flow (UFI-Flow). Specifically, we propose an invertible normalizing flow based network to generate latent variables by constructing a mapping between CG-FG map pairs, where the latent variables represent the conditional probability distribution of the FG map [12, 13]. Then we reconstruct the FG map from the learned latent variables via deterministic mapping from the corresponding CG map. Finally, the UFI-Flow with expressive distributions is optimized by maximizing the log-likelihood of the generated FG maps. UFI-Flow is capable of learning and generating FG flow maps that are consistent with the input CG maps, without any additional constraints or losses. We further propose an augmented distribution fusion (ADF) mechanism to address the distribution variables neglected problem for further performance improvement.

## 2. METHODOLOGY

### 2.1. Problem Definition

Given an urban flow map  $\mathcal{M}$  of interest, we divide it into  $H \times W$  spatial grid-cells based on geographic locations. We denote the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column of  $\mathcal{M}$  as  $g_{ij}$ . The flow volume in  $g_{ij}$  is denoted as  $x_{ij} \in \mathbb{R}_+$ , where  $\mathbf{X} \in \mathbb{R}_+^{H \times W}$  is the flow map at a given time-span. Given a map pair of a coarse-grained map  $\mathbf{X}^c$  and a fined-grained map  $\mathbf{X}^f$ , an up-scaling factor  $N$ , and external factor set  $E$ , the FUPI problem can be defined as follows:

$$\tilde{\mathbf{X}}^f = \mathcal{F}(\mathbf{X}^c | E, \mathbf{X}^f, N; \Theta), \quad (1)$$

where  $\mathbf{X}^c \in \mathbb{R}_+^{H \times W}$  and  $\mathbf{X}^f \in \mathbb{R}_+^{NH \times NW}$ .

### 2.2. Overview

Figure 2 depicts the framework of our proposed UFI-Flow model. First, we apply an invertible normalizing flow to generate unconstrained flow maps. We then address the structural constraint problem and reinforce the influence of external factors using ADF. Besides, we use external factor fusion (EFF) to enhance the inference performance.

### 2.3. Probabilistic Urban Flow Inference Network

Prior FUPI methods mainly learn a deterministic mapping by using reconstruction losses [11, 14]. However, they ignore the ill-posed nature of FUPI tasks. We address this drawback via estimating a full conditional distribution  $P_{\mathbf{X}^f | \mathbf{X}^c}(\mathbf{X}^f | \mathbf{X}^c; \theta)$  of CG map. The true posterior is usually intractable, thus we propose to approximate the complex urban flow distributions by transforming a simple base distribution via a series of invertible transformations [13, 12, 15, 16].

We first extract features of  $E$  from EFF module using fully connected layers. Then we concatenate extracted features  $E'$  with CG map  $\mathbf{X}^c$  as the input of normalizing flow

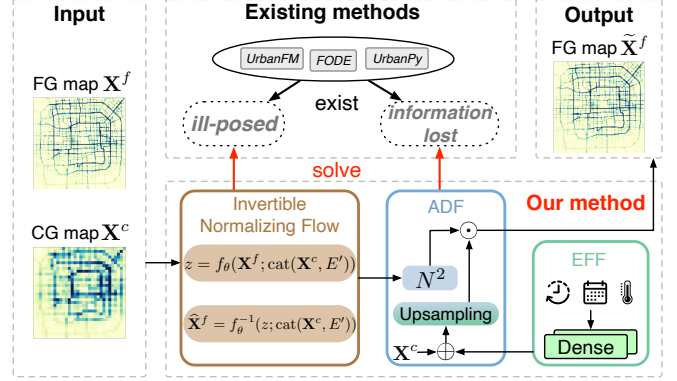


Fig. 2: Overview of the proposed UFI-Flow model.

(NF) to obtain the latent variable  $z$  that has a simple distribution (e.g., Gaussian)  $z = f_\theta(\mathbf{X}^f; \text{cat}(\mathbf{X}^c, E'))$ . Note that NFs are invertible, the unconstrained flow map  $\tilde{\mathbf{X}}^f$  can be exactly reconstructed by  $\tilde{\mathbf{X}}^f = f_\theta^{-1}(z; \text{cat}(\mathbf{X}^c, E'))$ . In this way, the inference network can be expressed as  $\mathbf{X}^c \leftrightarrow z \leftrightarrow \mathbf{X}^f$ , where the forward process is flow map inference, and the backward process is flow map estimation. In practice, we define the probability  $P_{\mathbf{X}^f | \mathbf{X}^c}(\mathbf{X}^f | \mathbf{X}^c, E'; \theta)$  by mapping samples  $z \sim q_z$  to the unconstrained maps as follows:

$$P_{\mathbf{X}^f | \mathbf{X}^c}(\mathbf{X}^f | \mathbf{X}^c, E'; \theta) = q_z(f_\theta(\mathbf{X}^f; \mathbf{X}^c)) \left| \det \frac{\partial f_\theta}{\partial \mathbf{X}^f} \right|. \quad (2)$$

Here mapping function  $f_\theta$  consists of  $N$  transformations and has an affine coupling layer structure [16]. The determinant of the Jacobian is derived as  $P_{\mathbf{X}^f | \mathbf{X}^c}(\mathbf{X}^f | \mathbf{X}^c, E'; \theta)$ , where  $\mathbf{h}^{n+1} = f_\theta^n(\mathbf{h}^n; \text{cat}(\mathbf{X}^c, E'))$  is the latent state of the  $n$ -th transformation. We let  $\mathbf{h}^0 = \mathbf{X}^f$  and  $\mathbf{h}^N = z$ . The determinants can be calculated easily and efficiently with the special designed Jacobian structure. Then we reconstruct the unconstrained flow maps via the inverse of normalizing flow  $\tilde{\mathbf{X}}^f = f_\theta^{-1}(z; \text{cat}(\mathbf{X}^c, E'))$ .

At last, we estimate the density of unconstrained flow maps and optimize  $f_\theta$  by performing a negative log-likelihood loss [17, 18] defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{NLL}}(\theta; \mathbf{X}^c, \mathbf{X}^f) &= -\log P_{\mathbf{X}^f | \mathbf{X}^c}(\mathbf{X}^f | \mathbf{X}^c, E'; \theta) \\ &= -\log q_z(z) - \sum_{n=0}^{N-1} \log \left| \det \frac{f_\theta^n}{\mathbf{h}^n} \right|. \end{aligned} \quad (3)$$

### 2.4. Augmented Distribution Fusion

Existing models proposed several normalization and distributional upsampling techniques to address the strict structural constraint of FUPI problem [5, 9]. More specifically, the flows in CG and FG maps obey the following constraint (cf. Fig. 1):

$$x_{ij}^c = \sum_{i'j'} x_{i'j'}^f, \quad \text{s.t.} \left\lfloor \frac{i'}{N} \right\rfloor = i, \left\lfloor \frac{j'}{N} \right\rfloor = j, \quad (4)$$

where  $x_{ij}^c$  is the flow volume in the CG map and  $x_{i'j'}^f$  is the corresponding flow volume in the superregions of the FG map. Existing solutions apply  $N^2$ -Normalization to the generated unconstrained maps, ignoring the influence of external factors [5]. In [9] the authors proposed  $AN^2$ -Normalization to consider the external factors, they still neglect the impact of external factors in the inference stage and only fuse the external factors during the constraining phase. In a nutshell, current studies obtain the final fine-grained flow map  $\tilde{\mathbf{X}}^f$  by upsampling from the learned joint distribution as follows:

$$\tilde{\mathbf{X}}^f = \mathbf{X}_{up}^c \odot \mathcal{D}^f, \quad (5)$$

where  $\mathcal{D}^f$  is generated by structural constraints  $A/N^2$ -Normalization and  $\mathbf{X}_{up}^c$  is the upsamplings of the CG map.

To address this issue, we propose to enhance the FG map inference via emphasizing the distribution of the external factors. We design a new fusing mechanism to sample the fusion CG map which is concatenated by the pixel-level feature maps  $E'$  and CG map  $\mathbf{X}^c$  as follows:

$$\mathbf{X}_{up}^c = \mathcal{U}(\text{cat}(\mathbf{X}^c, E'); N), \quad (6)$$

where  $\mathcal{U}$  represents the nearest-neighbor upsampling with scaling factor  $N$ . Finally, the dynamic spatial dependency between  $\mathbf{X}_{up}^c$  and  $\mathcal{D}^f$  is captured to infer the final fine-grained flow map  $\tilde{\mathbf{X}}^f$  according to Eq. (5).

### 3. EXPERIMENT

#### 3.1. Experimental Settings

**Datasets.** To evaluate the effectiveness of our method, we conduct extensive experiments on two large-scaled real-world datasets<sup>1</sup>: DiDi-Xi'an<sup>2</sup> and Taxi-BJ [19], both are collected from two metropolises in China (Xi'an and Beijing). As mentioned in previous studies [5, 9, 20], flow distribution is affected by various external factors. We collect meteorology factors of Xi'an from Wold Climate Data<sup>3</sup> and the weather condition includes 9 categories (e.g., rainy, sunny and cloudy), then we digitize these categories into ordinal values. We also include humidity and atmospheric pressure data for each flow map. Temperature, wind speed, humidity and pressure are scaled into the range  $[0, 1]$  with a min-max linear normalization. Time and date factors (e.g., hour of the day and day of the week) are transformed into ordinal values.

The statistics of two datasets are described in Table 1.

**Baselines.** We compare our method with following six baselines: (1) **SRCNN** [21]: is a successful convolutional neural network for SR problem; (2) **VDSR** [22]: uses residual network structure [23] and achieves better results than SRCNN; (3) **SRResNet** [24]: is a ResNet-based variant of the

<sup>1</sup><https://github.com/PattonYu/fufi-didi-xian>

<sup>2</sup>[outreach.didichuxing.com/research/opendata](https://outreach.didichuxing.com/research/opendata)

<sup>3</sup><https://en.tutiempo.net>

**Table 1:** Dataset description.

Dataset	DiDi-Xi'an	Taxi-BJ
Time range	10/1/2016-10/31/2016	7/1/2013-10/31/2013
Time interval	10 minutes	30 minutes
Coarse-grained size	32 × 32	32 × 32
Fine-grained size	128 × 128	128 × 128
Upscaling factor ( $N$ )	4	4
Latitude range	34.20°N - 34.28°N	39.82°N - 39.99°N
Longitude range	108.92°E - 109.01°E	116.26°E - 116.49°E
<b>External Factors (meteorology, time, etc.)</b>		
Temperature / °C	[11.0,32.0]	[-24.6,41.0]
Wind speed / mph	[0,9.3]	[0,48.6]
Weather conditions	9 types (e.g., Rainy,Sunny)	16 types
Humidity	[36%,100%]	\
Pressure /hPa	[1004,1026]	\
Holidays	7	18

VDSR model, which allows stacking more network layers and deepening the network depth; (4) **UrbanFM** [5]: is the first method for FUFU problem and it proposes  $N^2$ -normalization to solve the structural constraint; (5) **FODE** [9]: extends neural ODE [25] that allows more accurate and flexible solution with less memory cost; (6) **UrbanPy** [20]: is an extension of UrbanFM and solves the insufficiency problem when inferring flow map at higher upscaling rates by decomposing the original task into multiple sub-tasks.

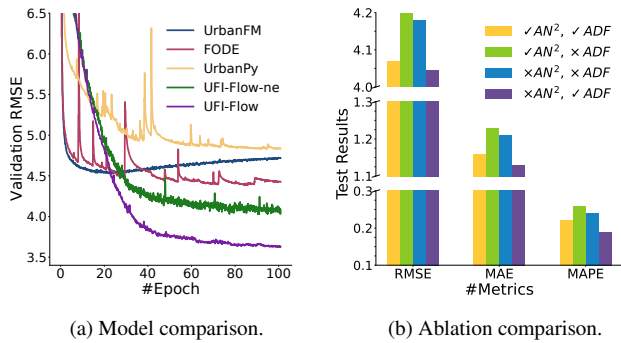
**Evaluation metrics and parameter settings.** We evaluate our model and baselines via three common metrics that are widely used for urban flow data: root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). We divide each dataset into training, validation, and test sets in a ratio of 2:1:1. The scaling factor  $N = 4$ . The training phase was performed using the Adam optimizer [26] with a learning rate  $2.5e^{-4}$ . The first dense layer for feature extraction has 128 hidden units and the second layer has 32x32 hidden units.

**Table 2:** Performance comparison. The best performances are in bold and the second best performances are underlined.

Dataset	DiDi-Xi'an			Taxi-BJ		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SRCNN	6.312	3.380	2.099	4.297	2.491	0.741
VDSR	5.182	2.702	1.741	4.159	2.213	0.467
SRResNet	5.063	1.799	0.922	4.164	2.457	0.713
UrbanFM	5.038	1.346	0.350	3.950	2.011	0.327
FODE	<u>4.860</u>	1.413	0.464	<u>3.860</u>	<u>1.963</u>	<u>0.313</u>
UrbanPy	5.210	<u>1.321</u>	<u>0.293</u>	3.950	1.995	0.329
<b>UFI-Flow</b>	<b>4.091</b>	<b>1.134</b>	<b>0.192</b>	<b>3.845</b>	<b>1.927</b>	<b>0.285</b>

#### 3.2. Results Analysis

**Model comparison.** Table 2 shows the experimental results of all models on two datasets. We have the following remarks. First, UFI-Flow significantly outperforms all baselines in non-trivial margins. It has better performance on



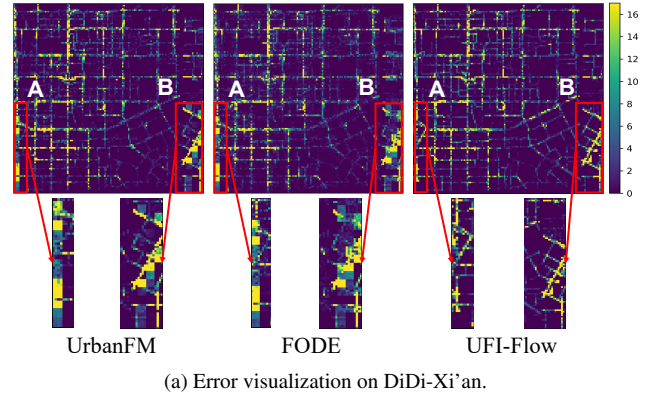
**Fig. 3:** (a) Convergence speed of our model and baselines; (b) The performance comparison of ablation studies, note that UFI-Flow is equivalent to  $[\times AN^2, \checkmark ADF]$ .

DiDi-Xi'an dataset with improvements of 15.8%, 14.1%, and 34.4% in terms of RMSE, MAE, and MAPE, respectively, compared to the best performances of baselines (underlined). Conventional image SR methods (SRCNN, VDSR, and SR-ResNet) perform poorly compared to FUFU methods (UrbanFM, FODE, UrbanPy, and UFI-Flow), largely due to their lack of considering the structural constraint and the influence of external factors. In contrast, our proposed UFI-Flow is superior than baselines, which verifies its effectiveness on FUFU problem.

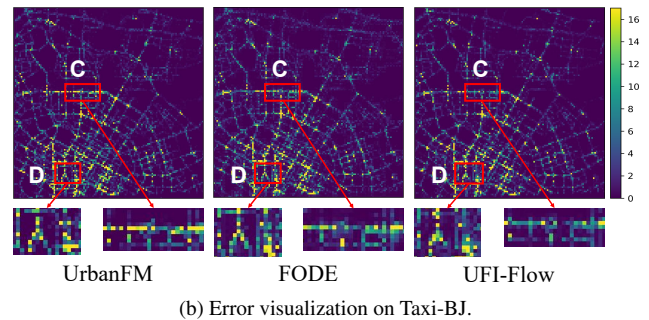
In Fig. 3a, we investigate the RMSE training process of FUFU methods on DiDi-Xi'an dataset, where UFI-Flow-ne removes the *EFF* block and ignores the influence of external factors. We can see that UFI-Flow achieves the optimal performance and has a stable training process. UFI-Flow-ne has a similar training speed as UFI-Flow and its slightly performance decrease demonstrates the benefit of our proposed external factor fusion. FODE and UrbanFM converge faster but their performances are less effective than UFI-Flow and UFI-Flow-ne, this shows the importance of learning spatial correlation of urban flows in both fine- and coarse-granularities. As an extension of UrbanFM, UrbanPy has more model parameters designed for inferring flow map with large scaling factors, which, leads to the slower convergence speed and worse performance with smaller scaling factors.

**Ablation study on *ADF* mechanism.** We conduct ablation studies to investigate the individual contributions of  $AN^2$  and *ADF*, the results are shown in Fig. 3b. Unexpectedly, using the  $AN^2$  mechanism decreases the model performance. UFI-Flow – which uses *ADF* but not  $AN^2$  – performs the best among four combinations. This verifies the effectiveness of our proposed augmented distribution fusion.

**Visualization analysis.** Figure 4 shows the inference error between the generated FG map and the ground truth on two datasets. The size of the error map is  $128 \times 128$  and we color the map cells with absolute errors  $|\mathbf{X}^f - \tilde{\mathbf{X}}^f|$ . Brighter col-



(a) Error visualization on DiDi-Xi'an.



(b) Error visualization on Taxi-BJ.

**Fig. 4:** Inference error visualization comparison between baselines and UFI-Flow on two datasets.

ors indicate larger errors compared to ground truth. To better visualize the inference errors, we focus on several selected areas, which are suburbs marked as A and B in Fig. 4a and congested areas marked as C and D in Fig. 4b. It is observed that UFI-Flow has much less bright pixels than baselines (UrbanFM and FODE), whether on populous or suburb areas. This visualization again verifies our model can generate finer and more accurate flow maps due to its capability of better fusing the external factors and learning the spatial interactions between CG and FG map pairs.

#### 4. CONCLUSIONS

We propose a novel method UFI-Flow for fine-grained urban flow inference, which infers the FG maps by calculating the conditional distribution of the FG map and optimizing the training process by a probabilistic log-likelihood loss between coarse- and fine-grained flow map pairs. Besides, we introduce *ADF*, a new mechanism to enhance the inferred flow map by fusing the distributions of external factors in inference stage. Extensive experiments show that UFI-Flow achieves superior performance on two real-world datasets compared to the six strong baselines. Our future work will focus on designing more powerful normalizing flow structures that are suitable for flow maps and effective for the FUFU problem.

## 5. REFERENCES

- [1] Bin Lu, Xiaoying Gan, Haiming Jin, Luoyi Fu, and Haisong Zhang, “Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting,” in *CIKM*, 2020, pp. 1025–1034.
- [2] Yuxuan Liang, Kun Ouyang, Junkai Sun, Yiwei Wang, Junbo Zhang, Yu Zheng, David S. Rosenblum, and Roger Zimmermann, “Fine-grained urban flow prediction,” in *WWW*, 2021, pp. 1833–1845.
- [3] Fan Zhou, Pengyu Wang, Xovee Xu, Wenxin Tai, and Goce Trajcevski, “Contrastive trajectory learning for tour recommendation,” *ACM TIST*, vol. 13, no. 1, 2022.
- [4] Peng Xie, Tianrui Li, Jia Liu, Shengdong Du, Xin Yang, and Junbo Zhang, “Urban flow prediction from spatiotemporal data using machine learning: A survey,” *Information Fusion*, vol. 59, pp. 1–12, 2020.
- [5] Yuxuan Liang, Kun Ouyang, Lin Jing, Sijie Ruan, Ye Liu, Junbo Zhang, David S. Rosenblum, and Yu Zheng, “UrbanFM: Inferring fine-grained urban flows,” in *SIGKDD*, 2019, pp. 3132–3142.
- [6] Haoyang Yu, Xovee Xu, Ting Zhong, and Fan Zhou, “Fine-grained urban flow inference via normalizing flows (student abstract),” in *AAAI*, 2022.
- [7] Jianrui Cai, Shuhang Gu, Radu Timofte, Lei Zhang, and et al., “Ntire 2019 challenge on real image super-resolution: Methods and results,” in *CVPR workshop*, 2019, pp. 2211–2223.
- [8] Zhihao Wang, Jian Chen, and Steven C.H. Hoi, “Deep learning for image super-resolution: A survey,” *TPAMI*, 2020.
- [9] Fan Zhou, Liang Li, Ting Zhong, Goce Trajcevski, Kunpeng Zhang, and Jiahao Wang, “Enhancing urban flow maps via neural odes,” in *IJCAI*, 2020, pp. 1295–1302.
- [10] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte, “SRFlow: Learning the super-resolution space with normalizing flow,” in *ECCV*, 2020, vol. 12350, pp. 715–732.
- [11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in *CVPR*, 2016, pp. 1646–1654.
- [12] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio, “Density estimation using real NVP,” in *ICLR*, 2017.
- [13] Danilo Jimenez Rezende and Shakir Mohamed, “Variational inference with normalizing flows,” in *ICML*, 2015, vol. 37, pp. 1530–1538.
- [14] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *CVPR*, 2017, pp. 5835–5843.
- [15] Laurent Dinh, David Krueger, and Yoshua Bengio, “NICE: non-linear independent components estimation,” in *ICLR*, 2015.
- [16] Diederik P. Kingma and Prafulla Dhariwal, “Glow: Generative flow with invertible 1x1 convolutions,” in *NeurIPS*, 2018, pp. 10236–10245.
- [17] Christina Winkler, Daniel E. Worrall, Emiel Hoogeboom, and Max Welling, “Learning likelihoods with conditional normalizing flows,” *arXiv:1912.00042*, 2019.
- [18] Lynton Ardizzone, Carsten L uth, Jakob Kruse, Carsten Rother, and Ullrich K othe, “Guided image generation with conditional invertible neural networks,” *arXiv:1907.02392*, 2019.
- [19] Junbo Zhang, Yu Zheng, and Dekang Qi, “Deep spatiotemporal residual networks for citywide crowd flows prediction,” in *AAAI*, 2017, pp. 1655–1661.
- [20] Kun Ouyang, Yuxuan Liang, Ye Liu, Zekun Tong, Sijie Ruan, David Rosenblum, and Yu Zheng, “Fine-grained urban flow inference,” *TKDE*, 2020.
- [21] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Learning a deep convolutional network for image super-resolution,” in *ECCV*, 2014, vol. 8692, pp. 184–199.
- [22] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in *CVPR*, 2016, pp. 1646–1654.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *CVPR*, 2016, pp. 770–778.
- [24] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *CVPR*, 2017, pp. 105–114.
- [25] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud, “Neural ordinary differential equations,” in *NeurIPS*, 2018, pp. 6572–6583.
- [26] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in *ICLR*, 2015.