Deep Learning for Spatio-Temporal Data Mining: A Survey

Senzhang Wang, Jiannong Cao, *Fellow, IEEE*, Philip S. Yu, *Fellow, IEEE*,

Abstract—**With the fast development of various positioning techniques such as Global Position System (GPS), mobile devices and remote sensing, spatio-temporal data has become increasingly available nowadays. Mining valuable knowledge from spatio-temporal data is critically important to many real world applications including human mobility understanding, smart transportation, urban planning, public safety, health care and environmental management. As the number, volume and resolution of spatio-temporal datasets increase rapidly, traditional data mining methods, especially statistics based methods for dealing with such data are becoming overwhelmed. Recently, with the advances of deep learning techniques, deep leaning models such as convolutional neural network (CNN) and recurrent neural network (RNN) have enjoyed considerable success in various machine learning tasks due to their powerful hierarchical feature learning ability in both spatial and temporal domains, and have been widely applied in various spatio-temporal data mining (STDM) tasks such as predictive learning, representation learning, anomaly detection and classification. In this paper, we provide a comprehensive survey on recent progress in applying deep learning techniques for STDM. We first categorize the types of spatio-temporal data and briefly introduce the popular deep learning models that are used in STDM. Then a framework is introduced to show a general pipeline of the utilization of deep learning models for STDM. Next we classify existing literatures based on the types of ST data, the data mining tasks, and the deep learning models, followed by the applications of deep learning for STDM in different domains including transportation, climate science, human mobility, location based social network, crime analysis, and neuroscience. Finally, we conclude the limitations of current research and point out future research directions.**

Index Terms—**Deep learning, Spatio-temporal data, data mining**

I. Introduction

Spatio-temporal data mining (STDM) is becoming growingly important in the big data era with the increasing availability and importance of large spatio-temporal datasets such as maps, virtual globes, remote-sensing images, the decennial census and GPS trajectories. STDM has broad applications in various domains including environment and climate (e.g. wind prediction and precipitation forecasting), public safety (e.g. crime prediction), intelligent transportation (e.g. traffic flow prediction), human mobility (e.g. human trajectory pattern

S.Z. Wang (szwang@nuaa.edu.cn) is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China, and also with the Department of Computing, The Hong Kong Polytechnic University, HongKong, China.

J.N. Cao is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China.

P.S. Yu is with the Department of Computer Science, University of Illinois at Chicago, Chicago, USA, and also with Institute for Data Science, Tsinghua University, Beijing, China.

mining), etc. Classical data mining techniques that are used to deal with transaction data or graph data often perform poorly when applied to spatio-temporal datasets because of many reasons. First, ST data are usually embedded in continuous space, whereas classical datasets such as transactions and graphs are often discrete. Second, patterns of ST data usually present both spatial and temporal properties, which is more complex and the data correlations are hard to capture by traditional methods. Finally, one of the common assumptions in traditional statistical based data mining methods is that data samples are independently generated. When it comes to the analysis of spatio-temporal data, however, the assumption about the independence of samples usually does not hold because ST data tends to be highly self correlated.

Although STDM has been widely studied in the last several decades, a common issue is that traditional methods largely rely on feature engineering. In other words, conventional machine learning and data mining techniques for STDM are limited in their ability to process natural ST data in their raw form. For example, to analyze human's brain activity from fMRI data, usually careful feature engineering and considerable domain expertise are needed to design a feature extractor that transforms the raw data (e.g. the pixel values of the scanned fMRI images) into a suitable internal representation or feature vector. Recently, with the prevalence of deep learning, various deep leaning models such as convolutional neural network (CNN) and recurrent neural network (RNN) have enjoyed considerable success in various machine learning tasks due to their powerful hierarchical feature learning ability, and have been widely applied in many areas including computer vision, natural language processing, recommendation, time series data prediction, and STDM. Compared with traditional methods, the advantages of deep learning models for STDM are as follows.

- **Automatic feature representation learning** Significantly different from traditional machine learning methods that require hand-crafted features, deep learning models can automatically learn hierarchical feature representations from the raw ST data. In STDM, the spatial proximity and the long-term temporal correlations of the data are usually complex and hard to be captured. With the multi-layer convolution operation in CNN and the recurrent structure of RNN, such spatial proximity and temporal correlations in ST data can be automatically and effectively learned from the raw data directly.

- **Powerful function approximation ability** Theoretically, deep learning can approximate any complex non-linear
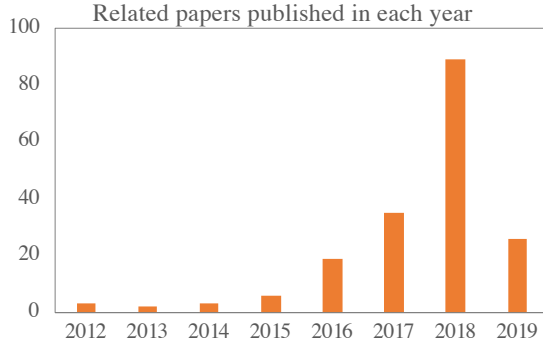
Fig. 1. Number of papers that explore deep learning techniques for STDM published in recent years.

functions and can fit any curves as long as it has enough layers and neuros. Deep learning models usually consist of multiple layers, and each layer can be considered as a simple but non-linear module with pooling, dropout, and activation functions so that it transforms the feature representation at one level into a representation at a higher and more abstract level. With the composition of enough such transformations, very complex functions can be learned to perform more difficult STDM tasks with more complex ST data.

Figure 1 shows the number of yearly published papers that explore deep learning techniques for various STDM tasks. One can see that there is a significant increase trend of the paper number in the last three years. Only less than 10 related papers are published each year from 2012 to 2015. From 2016 on, the number increases rapidly and many researchers try different deep learning models for different types of ST data in different applications domains. In 2018, there are about 90 related papers published, which is a large number. The complete number for 2019 is unavailable now, but we believe that the growing trend will keep on this year and also the future several years. Given the richness of problems and the variety of real applications, there is an urgent need for overviewing the existing works that explore deep learning techniques in the rapidly advancing field of STDM due to the following reasons. It can highlight the similarities and differences of using different deep learning models for addressing STDM problems of diverse applications domains. This can enable the cross-pollination of ideas across disparate research areas and application domains, by making it possible to see how deep learning model (e.g., CNN and RNN) developed for a certain problem in a particular domain (e.g., traffic flow prediction in transportation) can be useful for solving a different problem in another domain (e.g., crime prediction in crime analysis).

Related surveys on STDM There are a few recent surveys that have reviewed the literature on STDM in certain contexts from different perspectives. [9] and [143] discussed the computational issues for STDM algorithms in the era of "big data" for application domains such as remote sensing, climate science, and social media analysis. [87] focused on frequent pattern mining from spatio-temporal data. It stated the challenges of pattern discovery from ST data and classified the patterns into three categories: individual periodic pattern; pairwise movement pattern and aggregative patterns over multiple trajectories. [18] reviewed the state-of-the-art in STDM research and applications, with emphasis placed on the data mining tasks of prediction, clustering and visualization for spatio-temporal data. [130] reviewed STDM from the computational perspective, and emphasized the statistical foundations of STDM. [112] reviewed the methods and applications for trajectory data mining, which is an important type of ST data. [75] provided a comprehensive survey on ST data clustering. [4] discussed different types of ST data and the relevant data mining questions that arise in the context of analyzing each type of data. They classify literature on STDM into six major categories: clustering, predictive learning, change detection, frequent pattern mining, anomaly detection, and relationship mining. However, all these works reviewed STDM from the perspective of traditional methods rather than deep learning methods. [114] and [157] provided a survey that specially focused on the utilization of deep learning models for analyzing traffic data to enhance the intelligence level of transportation systems. There still lacks of a broad and systematic survey on exploring deep learning techniques for STDM in general.

Our Contributions Compared with existing works, our paper makes notable contributions summarized as follows:

- **First survey** To our knowledge, this is the first survey that reviews recent works exploring deep learning techniques for STDM. In light of the increasing number of studies on deep learning for spatio-temporal data analytics in the last several years, we first categorize spatio-temporal data types, and present the popular deep learning models that are widely used in STDM. We also summarize the data representations for different data types, and summarize which deep learning models are suitable to handle which types of data representations of ST data.
- **General framework** We present a general framework for deep learning based STDM, which consists of the following major steps: data instance construction, data representation, deep learning model selection and addressing STDM problems. Under the guidance of the framework, given a particular STDM task, one can better use the proper data representations and select or design the suitable deep learning models for the task under study.
- **Comprehensive survey** This survey provides a comprehensive overview on recent advances of using deep learning techniques for different STDM problems including predictive learning, representation learning, classification, estimation and inference, anomaly detection, and others. For each task, we provide detailed descriptions on the representative works and models for different types of ST data, and make necessary comparison and discussion. We also categorize and summarize current works based on the application domains including transportation, climate science, human mobility, location based social network, crime analysis, and neuroscience.
- **Future research directions** This survey also highlights

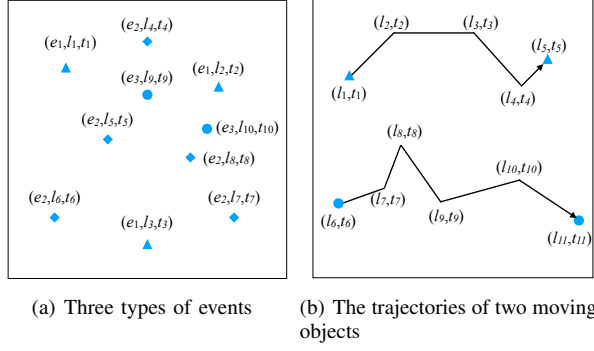(a) Three types of events (b) The trajectories of two moving objects

Fig. 2. Illustration of event and trajectory data types

several open problems that are not well studied currently, and points out possible research directions in the future.

Organization of This Survey The rest of this survey is organized as follows. Section II introduces the categorization of the ST data, and briefly introduces the deep learning models that are widely used for STDM. Section III provides a general framework for using deep learning for STDM. Section IV overviews various STDM tasks addressed by deep learning models. Section V presents a gallery of applications across various domains. Section VI discusses the limitations of existing works and suggests future directions. We finally conclude this paper in Section VII.

II. Categorization of Spatio-Temporal Data

A. Data Types

There are various types of ST data that differs in the way of data collection and representation in different real-world applications. Different application scenarios and ST data types lead to different categories of data mining tasks and problem formulations. Different deep learning models usually have different preferences to the types of ST data and have different requirements for the input data format. For example, CNN model is designed to process image-like data, while RNN is usually used to process sequential data. Thus it is important to first summarize the general types of ST data and represent them properly. We follow and extend the categorization in [4], and classify the ST data into the following types: event data, trajectory data, point reference data, raster data, and videos.

Event data. Event data comprise of discrete events occurring at point locations and times (e.g., crime events in the city and traffic accident events in a transportation network). An event can generally be characterized by a point location and time, which denotes where and when the event occurred, respectively. For example, a crime event can be characterized as such a tuple (e_i, l_i, t_i), where e_i is the crime type, l_i is the location where the crime occurs and t_i is the time when it occurs. Fig. 1(a) shows an illustration of the event data. It shows three types of events denoted by different shapes of the symbol. ST event data are common in real-world applications such as criminology (incidence of crime and related events), epidemiology (disease outbreak events), transportation (car accident), and social network (social event and trending topics).

Trajectory data. Trajectories denote the paths traced by bodies moving in space over time. (e.g., the moving route of a bike trip or taxi trip). Trajectory data are usually collected by the sensors deployed on the moving objects that can periodically transmit the location of the object over time, such as GPS on a taxi. Fig. 1(b) shows an illustration of two trajectories. Each trajectory can be usually characterized as such a sequence $\{(l_1, t_1), (l_2, t_2)...(l_n, t_n)\}$, where l_i is the location (e.g. latitude and longitude) and t_i is the time when the moving object passes this location. Trajectory data such as human trajectory, urban traffic trajectory and location based social networks are becoming ubiquitous with the development of Mobile applications and IoT techniques.

Point reference data. Point reference data consist of measurements of a continuous ST field such as temperature, vegetation, or population over a set of moving reference points in space and time. For example, meteorological data such as temperature and humidity are commonly measured using weather balloons floating in space, which continuously record weather observations. Point reference data can be usually represented as a set of tuples as follows $\{(r_1, l_1, t_1), (r_2, l_2, t_2)...(r_n, l_n, t_n)\}$. Each tuple (r_i, l_i, t_i) denotes the measurement of a sensor r_i at the location l_i of the ST filed at time t_i. Fig. 3 shows an example of the point reference data (e.g. sea surface temperature) in a continuous ST field at two time stamps. They are measured by the sensors at reference locations (shown as while circles) on the two time stamps. Note that the locations of the temperature sensors change over time.

Raster data. Raster data are the measurements of a continuous or discrete ST field that are recorded at fixed locations in space and at fixed time points. The major difference between point reference data and raster data is that the locations of the point reference data keep changing while the locations of the raster data are fixed. The locations and times for measuring the ST field can be regularly or irregularly distributed. Given m fixed locations $S = \{s_1, s_2, ...s_m\}$ and n time stamps $T = \{t_1, t_2, ...t_n\}$, the raster data can be represented as a matrix $R^{m \times n}$, where each entry r_{ij} is the measurement at location s_i at time stamp t_j. Raster data are also quite common in real-world applications such as transportation, climate science, and neuroscience. For example, the air quality data (e.g. PM2.5) can be collected by the sensors deployed at fixed locations of a city, and the data collected in a continuous time period form the air quality raster data. In neuroscience, functional magnetic resonance imaging or functional MRI (fMRI) measures brain activity by detecting changes associated with blood flow. The scanned fMRI signals also form the raster data for analyzing the brain activity and identifying some diseases. Fig. 4 shows an example of the traffic flow raster data of a transportation network. Each road is deployed a traffic sensor to collect real time traffic flow data. The traffic flow data of all the road sensors in a whole day (24 hours) form a raster data.

Video. A video that consists of a sequence of images can be also considered as a type of ST data. In the spatial domain, the neighbor pixels usually have similar RGB values and thus present high spatial correlations. In the temporal domain, the images of consecutive frames usually change smoothly and
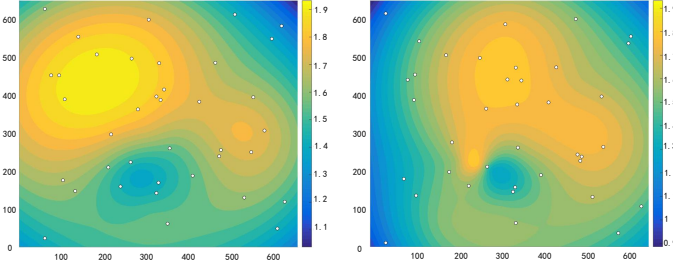
Fig. 3. Illustration of ST reference point data in two time stamps. The while circles are the locations of the sensors that record the readings of the ST field. The color bars show the distribution of the ST field.
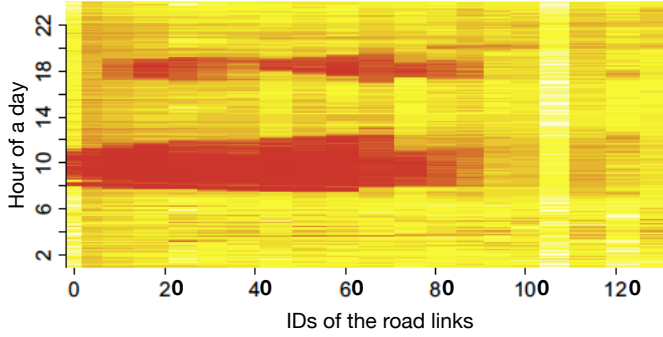

Fig. 4. Illustration of raster data collected from traffic flow sensors. The x-axis is the ID of the road links in a transportation network, and the y-axis is the hour of a day. Different colors denote different traffic flows on the road links captured by the road sensors deployed at fixed locations.

present high temporal dependency. A video can be generally represented as a three dimensional tensor with one dimension representing time t and the other two representing an image. Actually, video data can be also considered as a special raster data if we assume that there is a "sensor" deployed at each pixel and at each frame the "sensors" will collect the RGB values. Deep learning based video data analysis is extremely hot and a large number of papers are published in recent years. Although we categorize videos as a type of ST data, we focus on reviewing related works from the perspective of data mining and video data analysis falls into the research areas of computer vision and pattern recognition. Thus in this survey we do not cover the ST data type of videos.

B. Data Instances and Representations

The basic unit of data that a data mining algorithm operates upon is called a data instance. For a classical data mining setting, a data instance can be usually represented as a set of features with a label for supervised learning or without labels for unsupervised learning. In the ST data mining scenario, there are different types of data instances for different ST data types. For different data instances, there are several types of data representations that are used to formulate the data for further mining by the deep learning models.

Data instances. In general, the ST data can be summarized into the following data instances: points, trajectories, time series, spatial maps and ST raster as shown in the left part of Fig. 5. A ST point can be represented as a tuple containing
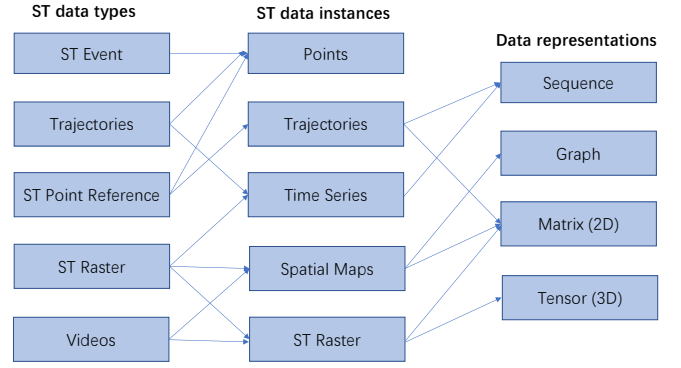

Fig. 5. Data instances and representations of different ST data types

the spatial and temporal information as well as some additional features of an observation such as the types of crimes or traffic accidents. Besides ST events, trajectories and ST point reference can also be formed as points. For example, one can break a trajectory into several discrete points to count how many trajectories have passed a particular region in a particular time slot. Besides formed as points and trajectories, trajectories can be also formed as time series in some applications. If we fix the location and count the number of trajectories traversing the location, it forms a time series data. The data instance of spatial maps contains the data observations of all the sensors in the entire ST filed at each time stamp. For example, the traffic speed readings of all the loop sensors deployed at the expressway at time t form a spatial map data. The data instance of the ST raster data contains the measurements spanning the entire set of locations and time stamps. That is, a ST raster comprises of a set of spatial maps.

Different data instances can be extracted from ST raster as time series, spatial maps or ST raster itself, depending on different applications and analytic requirements. First, we can consider the measurements at a particular ST grid of the ST field as a time series for some time series mining tasks. Second, for each time stamp the measurements of an ST raster can be considered as a spatial map. Third, one can also consider all the measurements spanning all the locations and time stamps as a whole for analysis. In such a case, ST raster itself can be a data instance.

Data representations. For the above mentioned five types of ST data instances, four types of data representations are generally utilized to represent them as the input of various deep learning models, sequence, graph, 2-dimensional matrix and 3-dimensional tensor as shown in the right part of Fig. 5. Different deep learning models require different types of data representations as input. Thus how to represent the ST data instances relies on the data mining task under study and the selected deep learning model.

Trajectories and time series can be both represented as sequences. Note that trajectories sometime are also represented as a matrix whose two dimensions are the row and column ids of grid ST field. Each entry value of the matrix denotes whether the trajectory traverses the corresponding grid region. Such a data representation is usually used to
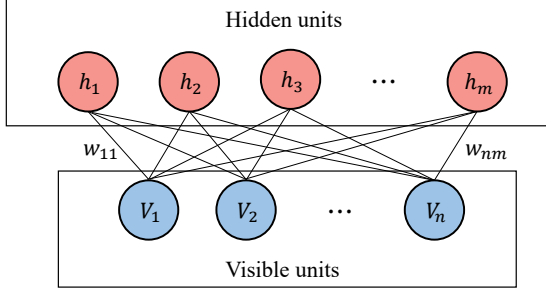
Fig. 6. Structure of the RBM model.

Fig. 7. Structure of the CNN model.

Fig. 8. Structure of GraphCNN model.

facilitate the utilization of CNN models [67], [118], [142]. Although graph can be also represented as a matrix, here we categorize graph and image matrix as two different types of data representations. This is because graph nodes does not follow the Euclidean distance as the image matrix does, and thus the way to deal with graphs and image matrices are totally different. We will discuss more details on the methods to handle the two types of data representations later. Spatial maps can be both represented as graphs and matrices, depending on different applications. For example, in urban traffic flow prediction, the traffic data of a urban transportation network can be represented as a traffic flow graph [85], [155] or cell region-level traffic flow matrix [121], [137]. Raster data are usually represented as 2D matrices or 3D tensors. For the case of matrix, the two dimensions are locations and time steps, and for the case of tensor, the three dimensions are row region cell id, column region id and time stamp. Matrix is a simpler data representation format compared with tensor, but it loses the spatial correlation information among the locations. Both are widely used to represent raster data. For example, in wind forecasting, the wind speed time series data of multiple anemometers deployed in different locations are usually merged as a matrix, and then is feed into a CNN or RNN model for future wind speed prediction [96], [200]. In neuroscience, one's fMRI data are a sequence of scanned fMRI brain images, and thus can be represented as a tensor like a video. Many works use the fMRI images tensor as the input of CNN model for feature learning to detect the brain activity [66], [76] and diagnose diseases [116], [158].

C. *Preliminary* of Deep Learning Models

In this subsection, we briefly introduce several deep learning models that are widely used for STDM, including RBM, CNN, GraphCNN, RNN, LSTM, AE/SAE, and Seq2Seq.

Restricted Boltzmann Machines (RBM). A Restricted Boltzmann Machine is a two-layer stochastic neural network [53] which can be used for dimensionality reduction, classification, feature learning and collaborative filtering. As shown in Fig. 6, the first layer of the RBM is called the visible, or input layer with the neuron nodes $\{v_1, v_2, ...v_n\}$, and the second is the hidden layer with the neuron nodes $\{h_1, h_2, ...h_m\}$. As a fully-connected bipartite undirected graph, all nodes in RBM are connected to each other across layers by undirected weight edges $\{w_{11}, ...w_{nm}\}$, but no two nodes of the same layer are
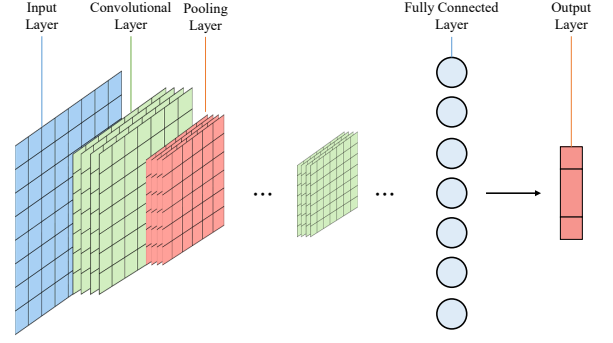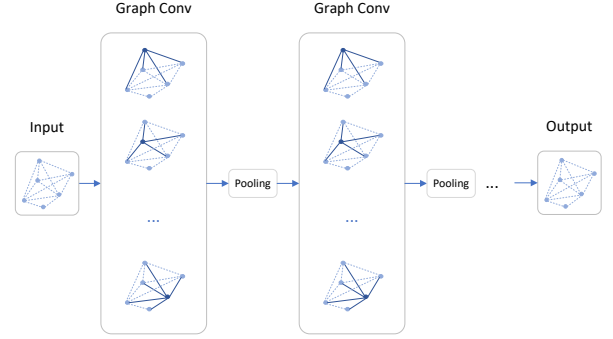
linked. The standard type of RBM has a binary-valued nodes and also bias weights. RBM tries to learn a binary code or representation of the input, and depending on the particular task, RBM can be trained in either supervised or unsupervised ways. RBM is usually used for learning features.

CNN. Convolutional neural networks (CNN) is a class of deep, feed-forward artificial neural networks that are applied to analyze visual imagery. A typical CNN model usually contains the following layers as shown in Fig. 7: the input layer, the convolutional layer, the pooling layer, the fully-connected layer and the output layer. The convolutional layer will determine the output of neurons of which are connected to local regions of the input through the calculation of the scalar product between their weights and the region connected to the input volume. The pooling layer will then simply perform downsampling along the spatial dimensionality of the given input to reduce the number of parameters. The fully-connected layers will connect every neuron in one layer to every neuron in the next layer to learn the final feature vectors for classification. It is in principle the same as the traditional multi-layer perceptron neural network (MLP). Compared with traditional MLPs, CNNs have the following distinguishing features that make them achieve much generalization on vision problems: 3D volumes of neurons, local connectivity and shared weights. CNN is designed to process image data. Due to its powerful ability in capturing the correlations in the spatial domain, it is now widely used in mining ST data, especially the spatial maps and ST rasters.

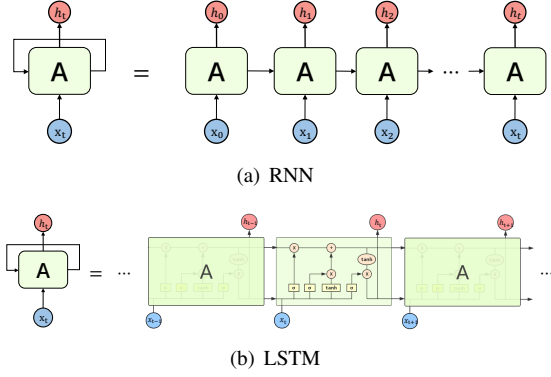**GraphCNN.** CNN is designed to process images which can

(a) RNN

(b) LSTM

Fig. 9. Structure of the RNN and LSTM models

Fig. 10. Structure of Seq2Seq model.

Fig. 11. Structure of the one-layer AE model.

be represented as a regular grid in the Euclidean space. However, there are a lot of applications where data are generated from the non-Euclidean domain such as graphs. GraphCNN is recently widely studied to generalize CNN to graph structured data [160]. Fig. 8 shows an structure illustration of a GraphCNN model. The graph convolution operation applies the convolutional transformation to the neighbors of each node, followed by pooling operation. By stacking multiple graph convolution layers, the latent embedding of each node can contain more information from neighbors which are multi-hops away. After the generation of the latent embedding of the nodes in the graph, one can either easily feed the latent embeddings to feed-forward networks to achieve node classification of regression goals, or aggregate all the node embeddings to represent the whole graph and then perform graph classification and regression. Due to its powerful ability in capturing the node correlations as well as the node features, it is now widely used in mining graph structured ST data such as network-scale traffic flow data and brain network data.

RNN and LSTM. A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. RNN is designed to recognize the sequential characteristics and use patterns to predict the next likely scenario. They are widely used in the applications of speech recognition and natural language processing. Fig. 9(a) shows the general structure of a RNN model, where X_t is the input data, A is the parameters of the network and h_t is the learned hidden state. One can see the output (hidden state) of the previous time step $t-1$ is input into the neural of the next time step t. Thus the historical information can be stored and passed to the future.

A major issue of standard RNN is that it only has short-term memory due to the issue of vanishing gradients. Long Short-Term Memory (LSTM) network is an extension for recurrent neural networks, which is capable of learning long-term dependencies of the input data. LSTM enables RNN to remember their inputs over a long period of time due to the special memory unit as shown in the middle part of Fig. 9(b). An LSTM unit is composed of three gates: input, forget and output gate. These gates determine whether or not to let new input in (input gate), delete the information because it is not important (forget gate) or to let it impact the output at the current time step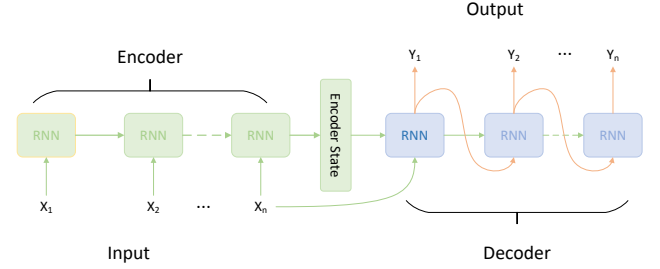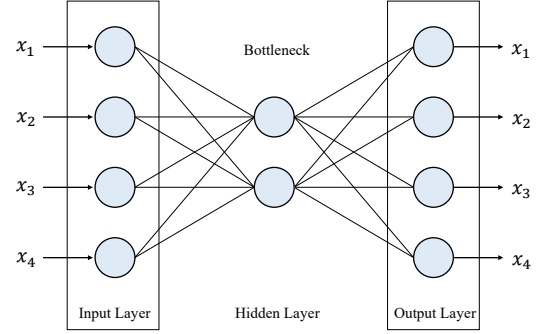 (output gate). Both RNN and LSTM are widely used to deal with sequence and time serious data for learning the temporal dependency of the ST data.

Seq2Seq. A sequence to sequence (Seq2Seq) model aims to map a fixed length input with a fixed length output where the length of the input and output may differ [138]. It is widely used to various NLP tasks such as machine translation, speech recognition and online chatbot. Although it is initially proposed to address NLP tasks, Seq2Seq is general framework and can be used to any sequence-based problem. As shown in Fig. 10, a Seq2Seq model generally consists of 3 parts: encoder, intermediate (encoder) vector and decoder. Due to the powerful ability in capturing the dependencies among the sequence data, Seq2Seq model is widely used in ST prediction tasks where the ST data present high temporal correlations such as urban crowd flow data and traffic data.

Autoencoder (AE) and Stacked AE. An autoencoder is a type of artificial neural network that aims to learn efficient data codings in an unsupervised manner [53]. As shown in Fig. 11, it features an encoder function to create a hidden layer (or multiple layers) which contains a code to describe the input. There is then a decoder which creates a reconstruction of the input from the hidden layer. An autoencoder creates a compressed representation of the data in the hidden layer or bottleneck layer by learning correlations in the data, which can be considered as a way for dimensionality reduction. As an effective unsupervised feature representation learning technique, AE facilitates various down stream data mining and machine learning tasks such as classification and clustering. A stacked autoencoder (SAE) is a neural network consisting of multiple layers of sparse autoencoders in which the outputs of each layer is wired to the inputs of the successive layer [7].
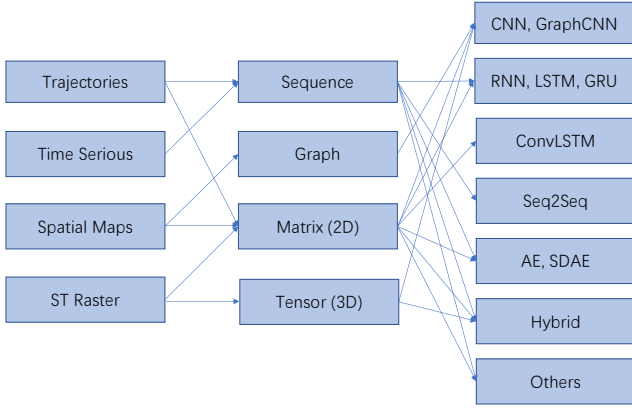
Fig. 12. Data representation for different DL models

III. FRAMEWORK

In this section, we will introduce how to use deep learning models for addressing STDM problems in general. First, we will give a framework that describes the pipeline which contains ST data instance construction, ST data representation, deep learning model section & design, and finally addressing the problem. Next we will introduce these major steps in detail.

A general pipeline for using deep learning models for ST data mining is shown in Fig. 13. Given the raw ST data collected from various location sensors, including the event data, trajectory data, point reference data and raster data, data instances are first constructed for data storage. As we discussed before, the ST data instances can be point, time series, spatial maps, trajectory and ST raster. To apply deep learning models for various mining tasks, the ST data instances need to be further represented as a particular data format to fit the deep learning models. The ST data instances can be represented as sequence data, 2D matrix, 3D tensors and graphs. Then for different data representations, different deep learning models are suitable to process them. RNN and LSTM models are good at handling sequence data with short-term or long-term temporal correlation, while CNN models are effective to capture the spatial correlation in the image like matrices. The hybrid model that combines RNN and CNN can capture both the spatial and temporal correlations of a tensor representation of the ST raster data. Finally, the selected deep learning models are used to address various STDM tasks such as prediction, classification, representation learning, etc.

A. ST Data Preprocessing

ST data preprocessing aims to represent ST data instances as a proper data representation format that the deep learning model can handle. Usually the input data format of a deep learning model can be a vector, a matrix or a tensor depending on different models. Fig. 12 shows the ST data instances and their corresponding data representations. One can see that usually one type of ST data instance corresponds to one typical data representations. Trajectory and time series data can be naturally represented as sequence data. Spatial map data can

be represented as a 2D matrix. ST raster can be represented as a 2D matrix or 3D tensor.

However, it is not always the case. For example, trajectory data sometimes are represented as a matrix, and then CNN model is applied to better capture the spatial features [24], [67], [103], [117], [150]. The ST field where the trajectories are measured such as a city is first partitioned into grid cell regions. Then the ST field can be modeled as a matrix with each cell region representing an entry. If a trajectory paths over the cell region, the corresponding entry value is set to 1; otherwise it is set to 0. In this way, a trajectory data can be represented as a matrix and thus CNN can be applied. Sometimes a spatial map is represented as a graph. For example, the sensors deployed in the express ways are usually modeled as a graph where the nodes are the sensors and the edges denote the road segments between two neighbor sensors. In such a case, GraphCNN models are usually utilized to process the sensor graph data and predict the future traffic (volume, speed, etc.) for all the nodes [22], [85]. ST raster data can be both represented as 2D matrices or 3D tensors, depending on the data types and applications. For example, a series of fMRI brain image data can be represented as a tensor and input into a 3D-CNN model for diseases classification [78], [116], and it can be also represented as a matrix by extracting the time series correlations between pair-wise regions of the brain for brain activity analysis [48], [113].

B. Deep Learning Model Selection & Design

With the data representations of the ST data instances, the next step is to feed them into the selected or designed deep learning models for different STDM tasks. As shown in the right part of Fig. 12, there are different deep learning model options for each type of data representation. Sequence data can be used as the input of the models including RNN, LSTM, GRU, Seq2Seq, AE, hybrid models and others. RNN, LSTM and GRU are all recurrent neural networks that are suitable to predict the sequence data. Sequence data can also be processed by Seq2Seq model. For example, in multi-step traffic prediction, a Seq2Seq model which consists a set of LSTM units in the encoder layer and a set of LSTM units in the decoder layer is usually applied to predict the traffic speed or volume in the next several time slots simultaneously [89], [90]. As a feature learning model, AE or SAE can be used to various data representations to learn a low-dimensional feature coding. Sequence data can also be encoded as a low-dimensional feature with AE or SAE. GraphCNN is particularly designed to process the graph data to capture the spatial correlations among the neighbor nodes. If the input is a single matrix, usually CNN model is applied, and if the input is a sequence of matrices, RNN models, ConvLSTM and hybrid models can be applied depending on the problems under study. If the goal is only for feature learning, AE and SAE models can be applied. For tensor data, usually it is handled by a 3D-CNN or the combination of 3D-CNN with RNN models.

Table I summarizes the works using deep learning models to handle different types of ST data. As shown in the table, CNN, RNN and their variants (e.g. GraphCNN and ConvLSTM) are
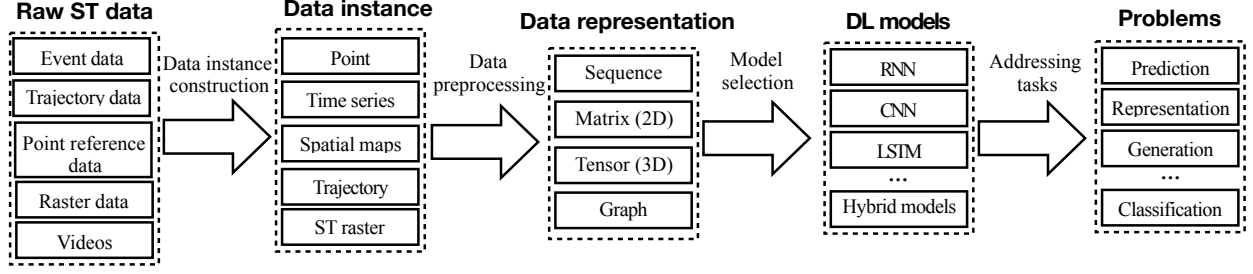
Fig. 13.   A general pipeline for using DL models for ST data mining

two most widely used deep learning models for STDM. CNN model is mostly used to process the spatial maps and ST raster. Some works also used CNN to handle trajectory data, but currently there is no work using CNN for time series data learning. GraphCNN model is specially designed to handle graph data, which can be categorized into spatial maps. RNN models including LSTM and GRU can be broadly applied in dealing with trajectories, time series, and the sequences of spatial maps. ConvLSTM can be considered as a hybrid model which combines RNN and CNN, and are usually used to handle spatial maps. AE and SDAE are mostly used to learn features from time series, trajectories and spatial maps. Seq2Seq model is generally designed for sequential data, and thus only used to handle time series and trajectories. The hybrid models are also common for STDM. For example, CNN and RNN can be stacked to learn the spatial features first, and then capture the temporal correlations among the historical ST data. Hybrid models can be designed to fit all the four types of data representations. Other models such as network embedding [164], multi-layer perceptron (MLP) [57], [186], generative adversarial nets (GAN) [49], [93], Residual Nets [78], [89], deep reinforcement learning [50], etc. are also used in recent works.

C. Addressing STDM Problems

Finally, the selected or designed deep learning models are used to address various STDM tasks such as classification, predictive learning, representation learning and anomaly detection. Note that usually how to select or design a deep learning model depends on the particular data mining task and the input data. However, to show the pipeline of the framework we first show the deep learning model and then the data mining tasks. In next section, we will categorize different STDM problems and review the works based on the problems and ST data types in detail.

IV. DEEP LEARNING MODELS FOR ADDRESSING DIFFERENT STDM PROBLEMS

In this section, we will categorize the STDM problems, and introduce the corresponding deep learning models proposed to address them. Fig. 14 shows the distribution of various STDM problems addressed by deep learning models, including prediction, representation learning, detection, classification, inference/estimation, recommendation and others. One can

see the largest category of the studied STDM problems is prediction. More than 70% related papers focus on studying the ST data prediction problem. This is mainly because an accurate prediction largely relies on high quality features, while deep learning models are especially powerful in feature learning. The second largest problem category is representation learning, which aims to learning feature representations for various ST data in an unsupervised or semi-supervised way. Deep learning models are also used in other STDM tasks including classification, detection, inference/estimation, recommendation, etc. Next we will introduce the major STDM problems in detail and summarize the corresponding deep learning based solutions.
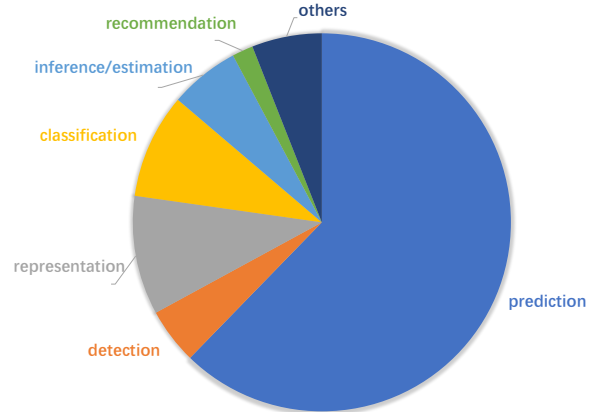
Fig. 14.   Distributions of the STDM problems addressed by deep learning models

A. Predictive Learning

The basic objective of predictive learning is to predict the future observations of the ST data based on its historical data. For different applications, both the input and output variables can belong to different types of ST data instances, resulting in a variety of predictive learning problem formulations. In the following, we will introduce the predictive problems based on the types of ST data instance as the model input.

Points. Points are usually merged in temporal or spatial domains to form time series or spatial maps such as crimes [31], [57], [145], [56], traffic accidents [201] and social events [43], so that deep learning models can be applied. [145] adapted ST-ResNet model to predict crime distribution over

TABLE I
DIFFERENT DL MODELS FOR PROCESSING FOUR TYPES OF ST DATA.

	Trajectories	Time Series	Spatial Maps (Image-like data & Graphs)	ST Raster
CNN	[24], [67], [103], [117], [150]		[11], [154], [199], [152], [100], [31], [139], [148], [184], [80], [69], [15], [72], [200], [113], [54], [68]	[188], [12], [123], [141], [106], [74], [131], [149], [116], [128], [128], [76], [78]
GraphCNN			[85], [155], [94], [111], [144], [22], [92], [175], [44], [8], [85], [155]	
RNN(LSTM,GRU)	[42], [77], [165], [99], [91], [163], [35], [159], [64], [38], [135], [181], [88], [81], [190], [37], [169], [166], [41], [65], [192]	[126], [27], [177], [90], [23], [89], [178], [17], [179], [101], [97], [14], [34]	[125], [107], [156], [2], [3], [39], [62], [162]	[23]
ConvLSTM			[1], [98], [161], [198], [151], [73], [201], [70], [147]	
AE/SDAE	[115], [197], [13]	[55], [167], [104]	[32], [16], [191], [48], [52], [182]	
RBM/DBN	[117]	[136]		[140], [58], [66]
Seq2Seq	[82], [170], [20], [171]	[90], [89]		
Hybrid	[164], [142], [108]	[96], [59]	[189], [30], [19], [6], [174], [187], [84], [109], [134], [49], [176]	[105], [127]
Others	[36], [10], [46], [195], [26], [193], [168]	[124], [93]	[133], [145], [202], [21], [183], [146], [79], [43], [185], [186], [132]	[122], [63], [71]

the Los Angeles area. Their models contains two staged. First, they transformed the raw crime point data as image-like crime heat maps by merging all the crime events happened in the same time slot and region of the city. Then, they adapted hierarchical structures of residual convolutional units to train a crime prediction model with the crime heat maps as input. Similarly, [57] proposed to use GRU model to predict the crime of a city. [201] studied the traffic accident prediction problem using the Convolutional Long Short-Term Memory (ConvLSTM) neural network model. They also first merged the point data of traffic accidents and modeled the traffic accident count in a spatio-temporal field as a 3-D tensor. Each entry (i, j, t) of the tensor represents the traffic accident count at the grid cell (i, j) in time slot t. The historical traffic accident tensors are input into CovnLSTM for prediction. [43] proposed a spatial incomplete multi-task deep learning framework to effectively forecast the subtypes of future events happened at different locations.

Time series. In road-level traffic prediction, the traffic flow data on a road or freeway can be modeled as a time series. Recently, many works tried various deep learning models for road-level traffic prediction [104], [136], [191]. [104] for the first time utilized stacked autoencoder to learn features from the traffic flow time series data for road-segment level traffic flow prediction. [136] considered the traffic flow data at a freeway as time series and proposed to use Deep Belief Networks (DBNs) to predict the future traffic flow based on the previous traffic flow observations. [126] studied the problem of taxi demand forecasting, and modeled the taxi demand at a particular area as a time series. A deep learning model with fully-connected layers is proposed to learn features from the

historical time series of taxi demand, and then the features are integrated with other context features such as weathers and social media texts to predict the future demand.

RNN and LSTM are widely used for time series ST data prediction. [90] integrated LSTM and sequence to sequence model to predict the traffic speed of a road segment. Besides the traffic speed information, their model also considered other external features including the geographical structure of roads, public social events such as national celebrations, and online crowd queries for travel information. The weather variables such as wind speed are also typically modeled as time series and then RNN/LSTM models are applied for future weather forecasting [14], [17], [55], [97], [124], [179]. For example, [17] proposed an ensemble model for probabilistic wind speed forecasting. The model integrated traditional wind speed prediction models including wavelet threshold denoising (WTD) and adaptive neuro fuzzy inference system (ANFIS) with recurrent neural network (RNN). In the area of fMRI data analysis, fMRI time series data are usually used to study the functional brain network and diagnose disease. [34] proposed to use LSTM model for classification of individuals with autism spectrum disorders (ASD) and typical controls directly from the resting-state fMRI time-series. [59] developed a deep convolutional auto-encoder model named DCAE for learning mid-level and high-level features from complex, large-scale tfMRI time series in an unsupervised manner. The time series data usually do not contain the spatial information, and thus the spatial correlations among the data are not explicitly considered in deep learning based prediction models.

Spatial maps. The spatial maps can be usually represented as image-like matrices, and thus are suitable to be processed

with CNN models for predictive learning [69], [80], [184], [200]. [184] proposed a CNN based prediction model to capture the spatial features in urban crow flow prediction. A real-time crowd flow forecasting system called UrbanFlow is built, and the crowd flow spatial maps are as its input. For forecasting the supply-demand in ride-sourcing services, [69] proposed the hexagon-based convolutional neural networks (H-CNN), where the input and output are both numerous local hexagon maps. In contrast to the previous studies that partitioned a city area into numerous square lattices, they proposed to partition the city area into various regular hexagon lattices because hexagonal segmentation has an unambiguous neighborhood definition, smaller edge-to-area ratio, and isotropy. Wind speed data of one monitoring site can be modeled as time series, while the data of multiple sites can be represented as spatial maps. CNN models can be also applied to predict wind speed of multiple sites simultaneously [200].

Given a sequence of spatial maps, to capture both the temporal and spatial correlations many works tried to combine CNN with RNN for the prediction. [161] proposed a convolutional LSTM (ConvLSTM) and used it to build an end-to-end trainable model for the precipitation nowcasting problem. This work combined the convolutional structure in CNN and the LSTM unites to predict the spatio-temporal sequences under a sequence-to-sequence learning framework. ConvLSTM is a sequence-to-sequence prediction model, whose each layer is a ConvLSTM unit that has convolutional structures in both the input-to-state and state-to-state transitions. The input and output of the model are both spatial map matrices. Following this work, many works tried to apply ConvLSTM to other spatial map prediction tasks of different domains [1], [6], [28], [70], [73], [98], [151], [198]. [151] proposed a novel cross-city transfer learning method for deep spatio-temporal prediction, called *RegionTrans*. *RegionTrans* contained multiple ConvLSTM layers to catch the spatio-temporal patterns hidden in the data. [73] applied ConvLSTM network to predict precipitation by using multi-channel radar data. [198] proposed an end-to-end deep neural network for predicting the passenger pickup/dropoff demands in mobility-on-demand (MOD) service. A encoder-decoder framework based on convolutional and ConvLSTM units was employed to identify complex features that capture spatio-temporal influences and pickup-dropoff interactions on citywide passenger demands. The passenger demands in the cell regions of a city was modeled as a spatial map and represented as a matrix. Similarly, [1] proposed a FCL-Net model which fused ConvLSTM layers, standard LSTM layers and convolutional layers for forecasting of passenger demand under on-demand ride services. [98] proposed a unified neural network module called Attentive Crowd Flow Machine (ACFM). ACFM is able to infer the evolution of the crowd flow by learning dynamic representations of temporally-varying data with an attention mechanism. ACFM is composed of two progressive ConvLSTM units connected with a convolutional layer for spatial weight prediction.

Some other models can be also used for predicting spatial maps, such as GraphCNN [8], [22], [92], [144], ResNet [146], [183], [185], and hybrid methods [49], [109], [177]. Note in this paper we consider that spatial maps contain both

image-like data and graph data. Although graphs are also represented as matrices, they require totally different technique such as GraphCNN or GraphRNN. In road network-scale traffic prediction, the transportation network can be naturally modeled as a graph, and then GraphCNN or GraphRNN is applied. [85] proposed to model the traffic flow on a transportation network as a diffusion process on a directed graph and introduced Diffusion Convolutional Recurrent Neural Network (DCRNN) for traffic forecasting. It incorporated both spatial and temporal dependency in the traffic flow of the entire road network. Specifically, DCRNN captures the spatial dependency using bidirectional random walks on the graph, and the temporal dependency using the encoder-decoder architecture with scheduled sampling. [155] proposed a new topological framework called Linkage Network to model the road networks and presented the propagation patterns of traffic flow. Based on the Linkage Network model, a novel online predictor, named Graph Recurrent Neural Network (GRNN), is designed to learn the propagation patterns in the graph. It simultaneously predicts traffic flow for all road segments based on the information gathered from the whole graph. [144] introduced an ST weighted graph (STWG) to represent the sparse spatio-temporal data. Then to perform micro-scale forecasting of the ST data, they built a scalable graph structured RNN (GSRNN) on the STWG.

Trajectories. Currently, two types of deep learning models, RNN and CNN are used for trajectory prediction depending on the data representations of the trajectories. First, trajectories can be represented as the sequence of locations as shown in Fig. 12. In such a case, RNN and LSTM models can be applied [38], [64], [77], [88], [135], [163], [165]. [163] proposed Collision-Free LSTM, which extended the classical LSTM by adding Repulsion pooling layer to share hidden-states of neighboring pedestrians for human trajectory prediction. Collision-Free LSTM can generate the future sequence based on pedestrian past positions. [64] studied the urban human mobility prediction problem, which given a few steps of observed mobility from one person, tries to predict where he/her will go next in a city. They proposed a deep-sequence learning model with RNN to effectively predict urban human mobility. [135] proposed a model named DeepTransport to predict the transportation mode such as walk, taking train, taking bus, etc, from a set of individual peoples GPS trajectories. Four LSTM layers are used to constructed DeepTransport to predict a user's transportation mode in the future.

Trajectories can be also represented as a matrix. In such a case, CNN models can be applied to better capture the spatial correlations [67], [103], [142]. [67] proposed a CNN-based approach for representing semantic trajectories and predicting future locations. In a semantic trajectory, each visited location is associated with a semantic meaning such as *home*, *work*, *shoppint*, etc. They modeled the semantic trajectories as a matrix whose two dimensions are semantic meanings and trajectory ID. The matrix is input into a CNN with multiple convolutional layers to learn the latent features for next visited semantic location prediction. [103] modeled trajectories as two-dimensional images, where each pixel of the image represented whether the corresponding location was visited in

the trajectory. Then multi-layer convolutional neural networks were adopted to combine multi-scale trajectory patterns for destination prediction of taxi trajectories. Modeling trajectories as image-like matrix is also utilized in other tasks such anomaly detection and inference [111], [150], which will be introduced in detail later.

ST raster. As we discussed before, ST raster data can be represented as matrices whose two dimensions are location and time, or tensors whose three dimensions are cell region ID, cell region ID, and time. Usually for ST raster data prediction, 2D-CNN (matrices) and 3D-CNN (tensors) are applied, and sometimes they are also combined with RNN. [188] proposed a multi-channel 3D-cube successive convolution network named 3D-SCN to nowcast storm initiation, growth, and advection from the 3D radar data. [121] modeled the traffic speed data at multiple locations of a road in successive time slots as a ST raster matrix, and then input it into a deep neural network for traffic flows prediction. [106] explored the similar idea as [121] for traffic prediction on a large transportation network. [12] proposed a 3D Convolutional neural networks for citywide vehicle flow prediction. Instead of predicting traffic on a road, they tried to predict vehicle flows in each cell region of a city. So they modeled the citywide vehicle flow data in successive time slots as ST rasters and input them into the proposed 3D-CNN model. Similarly, [131] modeled the mobility events of passengers in a city in different time slots as a 3D tensor, and then used 3D-CNN model to predict the supply and demand of the passengers for transportation. Note that the major difference between ST raster and spatial maps is that ST raster is the merged ST field measurements of multiple time slots, while spatial map is the ST field measurement in only one time slot. Thus the same type of ST data sometimes can be represented as both spatial maps and ST raster depending on the real application scenarios and the purposes of data analysis.

B. Representation Learning

Representation learning aims to learn the abstract and useful representations of the input data to facilitate downstream data mining or machine learning tasks, and the representations are formed by composition of multiple linear or non-linear transformations of the input data. Most existing works on representation learning for ST data focused on studying the data types of trajectories and spatial maps.

Trajectories. Trajectories are ubiquitous in location-based social networks (LBSNs) and various mobility services, and RNN and CNN models are both widely used to learn the trajectory representations. [82] proposed a seq2seq-based model to learn trajectory representations, for the fundamental research problem of trajectory similarity computation. The trajectory similarity based on the learned representations is robust to non-uniform, low sampling rates and noisy sample points. Simiarly, [170], [171] proposed to transform a trajectory into a feature sequence to describe object movements, and then employed a sequencetosequence autoencoder to learn fixedlength deep representations for clustering. Location-based social network (LBSN) data usually contain two important aspects, i.e., the mobile trajectory data and the social network

of users. To model the two aspects and mine their correlations, [164] proposed a neural network model to jointly learn the social network representation and the users' mobility trajectory representation. RNN and GRU models are used to capture the sequential relatedness in mobile trajectories at the short or long term levels. [10] proposed a content-aware POI embedding model named CAPE for POI recommendation. In CAPE, the embedding vectors of POIs in a user's check-in sequence are trained to be close to each other. [26] proposed a geographical convolutional neural tensor network named GeoCNTN to learn the embeddings of the locations in LBSNs. [41] proposed to use RNN and Autoencoder to learn the user check-in embedding and trajectory embedding, and used the embeddings for user social circle inference in LBSNs.

Spatial maps. There are several works that study how to learn representations of the spatial maps. [21] proposed a convolutional neural network architecture for learning spatiotemporal features from raw spatial maps of the sensor data. [153] formulated the problem of learning urban community structures as a spatial representation learning task. A collective embedding learning framework was presented to learn urban community structures by unifying both static POIs data and dynamic human mobility graph spatial map data. [182] studied how to learn nonlinear representations of brain connectivity patterns from neuroimage data to inform an understanding of neurological and neuropsychiatric disorders. A deep learning architecture named Multi-side-View guided AutoEncoder (MVAE) is proposed to learn the representations of the input brain connectome data derived from fMRI and DTI images.

C. Classification

The classification task is mostly studied in analyzing fMRI data. Recently, brain imaging technology has become a hot topic within the field of neuroscience, including functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG), and Magnetoencephalography (MEG) [120]. Particularly, fMRI combined with deep learning methods, has been widely used in the study of neuroscience for various classification tasks such as disease classification, brain function network classification and brain activation classification when watching words or images [158]. Various types of ST data can be extracted from the raw fMRI data depending on different classification tasks. [34] proposed the use of recurrent neural networks with long short-term memory (LSTMs) for classification of individuals with autism spectrum disorders (ASD) and typical controls directly from the resting-state fMRI time-series data generated from different brain regions. [48], [52], [54], [71], [113], [132] modeled the fMRI data as spatial maps, and then used them as the input of the classification models. Instead of using each individual resting-state fMRI time-series data directly, [48] and [52] calculated the whole-brain functional connectivity matrix based on the Pearson correlation coefficient between each pair of resting-state fMRI time-series data. Then the correlation matrix can be considered as a spatial map, and is input to a DNN model for ASD classification. [113] proposed a more general convolutional neural network architecture for functional connectome classification called connectome-convolutional neural

network (CCNN). CCNN is able to combine information from diverse functional connectivity metrics, and thus can be easily adapted to a wide range of connectome based classification or regression tasks, by varying which connectivity descriptor combinations are used to train the network.

Some works also directly use the 3D structural MRI brain scanned images as the ST raster data, and then 3D-CNN model is usually applied to learn features from the ST raster for classification [63], [66], [78], [116], [128], [194]. [78] proposed two 3D convolutional network architectures for brain MRI classification, which are the modifications of a plain and residual convolutional neural networks. Their models can be applied to 3D MRI images without intermediate handcrafted feature extraction. [194] also designed a deep 3D-CNN framework for automatic, effective, and accurate classification and recognition of large number of functional brain networks reconstructed by sparse 3D representation of whole-brain fMRI signals.

D. Estimation and Inference

Current works on ST data estimation and inference mainly focus on the data types of spatial maps and trajectories.

Spatial maps. While monitoring stations have been established to collect pollutant statistics, the number of stations is very limited due to the high cost. Thus, inferring fine-grained urban air quality information is becoming an essential issue for both government and people. [19] studied the problem of air quality inference for any location based on the air pollutant of some monitoring stations. They proposed a deep neural network model named ADAIN for modeling the heterogeneous data and learning the complex feature interactions. In general, ADAIN combines two kinds of neural networks: i.e., feed-forward neural networks to model static data and recurrent neural networks to model sequential data, followed by hidden layers to capture feature interactions. [139] investigated the application of deep neural networks to precipitation estimation from remotely sensed information. A stacked denoising auto-encoder is used to automatically extract features from the infrared cloud images and estimate the amount of precipitation. Estimating the duration of a potential trip given the origin location, destination location as well as the departure time is a crucial task in intelligent transportation systems. To address this issue, [83] proposed a deep multi-task representation learning model for arrival time estimation. This model produces meaningful representation that preserves various trip properties and at the same time leverages the underlying road network and the spatiotemporal prior knowledge.

Trajectories [147], [181] tried to estimate the travel time of a path from the mobility trajectory data. [181] proposed a RNN based deep model named DEEPTRAVEL which can learn from the historical trajectories to estimate the travel time. [147] proposed an end-to-end Deep learning framework for Travel Time Estimation called DeepTTE that estimated the travel time of the whole path directly rather than first estimating the travel times of individual road segments or sub-paths and then summing up them. [111] studied the problem of inferring the purpose of a users visit at a certain location from trajectory data. They proposed a graph convolutional neural networks (GCNs) for the inference of activity types (i.e., trip purposes) from GPS trajectory data generated by personal smartphones. The mobility graphs of a user is constructed based on all his/her activity areas and connectivities based on the trajectory data, and then the spatio-temporal activity graphs are fed into GCNs for activity types inference. [42] studied the problem of Trajectory-User Linking (TUL), which aims to identify and link trajectories to users who generate them in the LBSNs. A Recurrent Neural Networks (RNN) based model called TULER is proposed to address the TUL problem by combining the check-in trajectory embedding model and stacked LSTM. Identifying the distribution of users transportation modes, e.g. bike, train, walk etc., is an essential part of travel demand analysis and transportation planning [24], [148]. [24] proposed a CNN model to infer travel modes based on only raw GPS trajectories, where the modes are labeled as walk, bike, bus, driving, and train.

E. Anomaly Detection

Anomaly detection or outlier detection aims to identify the rare items, events or observations which raise suspicions by differing significantly from the majority of the data. Current works on anomaly detection for ST data mainly focus on the data types of events and spatial maps.

Events. [137] tried to detect the non-recurring traffic congestions caused by temporal disruptions such as accidents, sports games, adverse weathers, etc. A convolutional neural network (CNN) is proposed to identify non-recurring traffic anomalies that are caused by events. [189] studied how to detect traffic accidents from social media data. They first thoroughly investigated the 1-year over 3 million tweet contents in Northern Virginia and New York City, and then two deep learning methods: Deep Belief Network (DBN) and Long Short-Term Memory (LSTM) were implemented to identify the traffic accident related tweets. [199] proposed to utilize Convolutional Neural Networks (CNN) for automatic detection of traffic incidents in urban networks by using traffic flow data. [16] collected big and heterogeneous data including human mobility data and traffic accident data to understand how human mobility will affect traffic accident risk. A deep model of Stack denoise Autoencoder was proposed to learn hierarchical feature representation of human mobility, and these features were used for efficient prediction of traffic accident risk level.

Spatial maps. [100] presented the first application of Deep Learning techniques as alternative methodology for climate extreme events detection such as hurricanes and heat waves. The model was trained to classify tropical cyclone, weather front and atmospheric river with the climate image data as the input. [72] studied how to detect and localize extreme climate events in very coarse climate data. The proposed framework is based on two deep neural network models, (1) Convolutional Neural Networks (CNNs) to detect and localize extreme climate events, and (2) Pixel recursive recursive super resolution model to reconstruct high resolution climate data from low resolution climate data. To address the issue
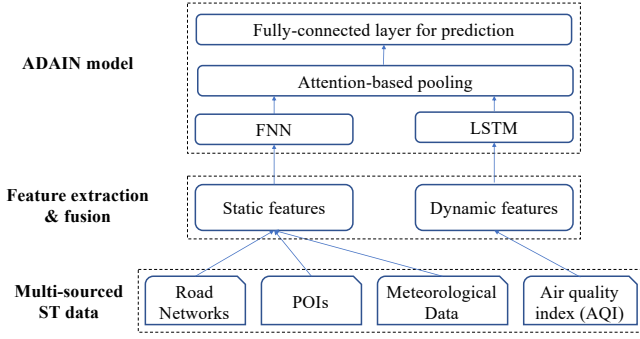
Fig. 15. ADAIN model framework.

of limited labeled extreme climate events, [123] presented a multichannel spatiotemporal CNN architecture for semi-supervised bounding box prediction. The approach proposed in [123] is able to leverage temporal information and unlabeled data to improve the localization of extreme weathers.

F. Other tasks.

Besides the problems we discussed above, deep learning models are also applied in other STDM tasks include recommendation [10], [81], [193], pattern mining [118], relation mining [197], etc. [10] proposed a content-aware hierarchical POI embedding model CAPE for POI recommendation. From text contents, CAPE captures not only the geographical influence of POIs, but also the characteristics of POIs. [193] also proposed to exploit the embedding learning technique to capture the contextual check-in information for POI recommendation. [118] proposed a deep-structure model called DeepSpace to mine the human mobility patterns through analyzing the mobile data of human trajectories. [197] studied the problem of Trajectory-User Linking (TUL), which aims to link trajectories to users who generate them from the geo-tagged social media data. A semi-supervised trajectory-user relation learning framework called TULVAE (TUL via Variational AutoEncoder) is proposed to learn the human mobility in a neural generative architecture with stochastic latent variables that span hidden states in RNN.

G. Fusing Multi-Sourced Data

Besides the ST data under study, there are usually some other types of data that are highly correlated to the ST data. Fusing such data together with the ST data can usually improve the performance of various STDM tasks. For example, the urban traffic flow data can be significantly affected by some external factors such as weather, social events, and holidays. Some recent works try to fuse the ST data and other types of data into a deep learning architecture for jointly learning features and capturing the correlations among them [16], [19], [89], [174], [178], [188], [201]. Generally, there are two popular ways to fuse the multi-sourced data in applying deep learning models for STDM, raw data-level fusion and latent feature-level fusion.

Raw data-level fusion. For the raw data-level fusion, the multi-sourced data are integrated first and then input into

the deep learning model for feature learning. [201] studied the traffic accident prediction problem by using the Convolutional Long Short-Term Memory (ConvLSTM) neural network model. First, the entire studied area is partitioned into grid cells. Then a number of fine-grained urban and environmental features such as traffic volume, road condition, rainfall, temperature, and satellite images are collected and map-matched with each grid cell. Given the number of accidents as well as the external features at each location mentioned above as the model input, a Hetero-ConvLSTM model to predict the number of accidents that will occur in each grid cell in future time slots is proposed. [19] proposed the ADAIN model which fused both the urban air quality information from monitoring stations and the urban data that are closely related to air quality, including POIs, road networks and meteorology for inferring fine-grained urban air quality of a city. The framework of ADAIN model is shown in Fig 15. Features are first manually extracted from the multi-sourced data including road networks, POIs, meteorological data and urban air quality index data. Then all the features are fused together and then fed into FNN and RNN models for feature learning.

Latent feature-level fusion. For the latent feature-level fusion, different types of raw features are input into different deep learning models first, and then a latent feature fusion component is used to fuse different types of latent features. [89] proposed a deep-learning-based approach called ST-ResNet, which is based on the residual neural network framework to collectively forecast the inflow and outflow of crowds in each region of a city. As shown in Fig 16, ST-ResNet handles two types of data, the ST crowd flow data sequences in a city and the external features including the weather and holiday events. Two components are designed to learn the latent features of the external features and the crowd flow data features seperately, and then a feature fusing function $tanh$ is used to integrate the two types of learned latent features. [174] proposed a Deep Multi-View Spatial-Temporal Network (DMVST-Net) framework to combine multi-view data for taxi demand prediction. DMVST-Net consists of three views: temporal view, spatial view and semantic view. CNN is used to learn features from the spatial view, LSTM is used to learn features from the temporal view and network embedding is applied to learn the correlations among regions. Finally, a fully connected neural network is applied to fuse all the latent features of the three views for taxi demand prediction.

H. Attention Mechanism

Attention is a mechanism that was developed to improve the performance of the Encoder-Decoder RNN on machine translation [5]. A major limitation of the Encoder-Decoder RNN is that it encodes the input sequence to a fixed length internal representation, which results in worse performance for long input sequences. To address this issue, attention allows the model to learn which encoded words in the source sequence to pay attention to and to what degree during the prediction of each word in the target sequence. Although attention is initially proposed in machine translation with the word sequence data as the input, it actually can be applied
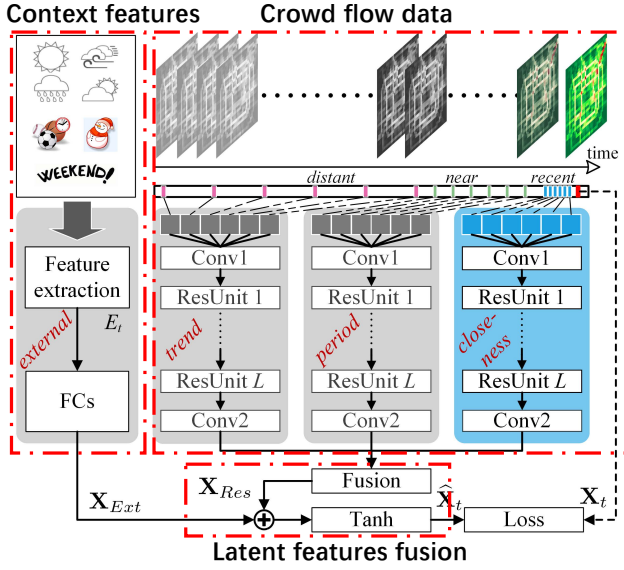
Fig. 16. ST-ResNet architecture [89].

to any kind of inputs such as images, which is called visual attention. As many ST data can be represented as sequential data (time series and trajectories) and image-like spatial maps, attention can also be incorporated into deep learning model for improving the performance of various STDM tasks [19], [38], [39], [57], [81], [88], [98], [142], [198].

The neural attention mechanism used in STDM can be generally categorized into spatial domain attention [19], [39] and temporal domain attention [38], [57], [81], [198]. Some works use both spatial and temporal domain attentions [88], [98], [142]. [39] proposed a combined attention model in the spatial domain. It utilizes both "soft attention" as well as "hard-wired" attention in order to map the trajectory information from the local neighborhood to the future positions of the pedestrian of interest. [57] proposed an attentive hierarchical recurrent network model named DeepCrime for crime prediction. The temporal domain attention mechanism is applied to capture the relevance of crime patterns learned from previous time slots in assisting the prediction of future crime occurrences, and automatically assign the importance weights to the learned hidden states at different time frames. In the proposed attention mechanism, the importance of crime occurrence in the past time slots is estimated by deriving a normalized importance weight via a softmax function. [88] proposed a multi-level attention network for predicting the geo-sensory time series that are generated by sensors deployed in different geospatial locations to continuously and cooperatively monitor the surrounding environment, such as air quality. Specifically, in the first level attention, a spatial attention mechanism consisting of local spatial attention and global spatial attention is proposed to capture the complex spatial correlations between different sensors time series. In the second level attention, a temporal attention is applied to model the dynamic temporal correlations between different time intervals in a time series.

V. APPLICATIONS

Large volumes of ST data are generated from various application domains such as transportation, on-demand service, climate & weather, human mobility, location-based social network (LBSN), crime analysis, and neuroscience. Table II shows the related works of the application domains mentioned above. One can see that the largest proportion of the works fall into transportation and human mobility due to the increasing availability of the urban traffic data and human mobility data. In this section, we will describe the applications of deep learning techniques used for STDM in different applications.

A. Transportation

With the increasing availability of transportation data collected from various sensors like loop detector, road camera, and GPS, there is an urgent need to utilize deep learning methods to learn the complex and highly non-linear spatio-temporal correlations among the traffic data to facilitate various tasks such as traffic flow prediction [30], [60], [90], [121], [136], [167], traffic incident detection [125], [189], [199] and traffic congestion prediction [108], [137]. Such transportation related ST data usually contain information of the traffic speed, volume, or traffic incidents, the locations of the road segments of regions, and the time. Transportation data can be modeled as time series, spatial maps and ST raster in different application scenarios. For example, in road network-scale traffic flow prediction the traffic flow data collected from multiple road loop sensors can be modeled as a raster matrix where one dimension is the locations of the sensors and the other is the time slots [106]. The loop sensors can be also connected as a sensor graph based on the connections among the road links where the sensors are deployed, and the the traffic data of a road network can be modeled as a graph spatial map so that GraphCNN models can be applied [85], [175]. While in road-level traffic prediction, the historical traffic flow data on each single road is modeled as a time series, and then RNN or other deep learning models are used for traffic prediction of a single road [60], [90], [167].

B. On-Demand Service

In recent years, various on-demand services such as Uber, Mobike, DiDi, GoGoVan have become increasingly popular due to the wide use of mobile phones. The on-demand services have taken over the traditional businesses by serving people with what and where they want. Many on-demand services produce a large number of ST data which involve the locations of the customers and the required service time. For example, Uber and DiDi are two popular ride-sharing on-demand service providers in USA and China, respectively. They both provide services including taxi hailing, private car hailing, and social ride-sharing to users via a smartphone application. To better meet customers' demand and improve the service, a crucial problem is how to accurately predict the demand and supply of the service at different locations and time. Deep learning methods for STDM in the application of on-demand service mostly focus on predicting the demand

TABLE II
RELATED WORKS IN DIFFERENT APPLICATION DOMAINS

Application domains	Related works
Transportation	[189], [199], [125], [32], [30], [6], [33], [136], [60], [121], [27], [114], [177], [90], [23], [73], [117], [135], [89], [148], [85], [137], [155], [157], [79], [107], [201], [22], [24], [108], [105], [16], [106], [93], [15], [175], [176], [104], [149], [102], [51], [40], [173], [47], [25]
On-demand Service	[1], [8], [126], [174], [146], [80], [69], [84], [198], [70], [44], [191], [50], [45], [86]
Climate & Weather	[19], [11], [100], [188], [161], [139], [17], [79], [94], [46], [70], [186], [96], [97], [129], [14], [127], [200]
Human Mobility	[67], [164], [154], [152], [98], [36], [163], [151], [82], [64], [183], [38], [118], [135], [181], [115], [195], [111], [142], [42], [170], [153], [159], [83], [190], [37], [202], [185], [35], [99], [20], [169], [186], [49], [3], [131], [103], [191], [171], [41], [65], [197], [13], [147], [180], [95]
Location Based Social Network	[164], [10], [26], [193], [77], [81], [99], [165], [166], [197], [168], [192]
Crime Analysis	[31], [133], [145], [57], [123], [55], [124], [72], [179]
Neuroscience	[120], [158], [110], [34], [48], [52], [58], [59], [63], [66], [71], [113], [116], [128], [54], [76], [68], [132], [78], [182]

and supply. [1] proposed to apply deep learning methods to forecast the demand-supply distributions of the dockless bike-sharing system. [92] proposed a graph CNN model to predict the station-level hourly demand in a large-scale bike-sharing network. [126], [174] proposed to use LSTM model to predict the taxi demand in different areas. [146] applied ResNet model to predict the supply-demand for online car-hailing services. The historical demand-supply in different regions of the city under study is usually modeled as spatial maps or raster tensors, so that CNN, RNN and combined models are applied to predict the future.

C. Climate & Weather

Climate science is the scientific study of climate, scientifically defined as weather conditions averaged over a period of time. The weather data usually contain the atmospheric and oceanic conditions (e.g., temperature, pressure, wind-flow, and humidity) that are collected by various climate sensors deployed at fixed or floating locations. As the climate data of different locations usually present high spatio-temporal correlations, STDM techniques are widely used for short-term and long-term weather forecasting. Especially, with the recent advances of deep learning techniques, many works tried to incorporate deep learning models for analyzing various weather and environment data [79], [129], such as air quality inference [19], [94], precipitation prediction [100], [161], wind speed prediction [96], [200], and extreme weather detection [100]. The data related to climate and weather can be spatial maps (e.g. radar reflectivity images) [188], time series (e.g. wind speed) [17], and events (e.g. extreme weather events) [100]. [19] proposed a neural attention model to predict the urban air quality data of different monitoring stations. [100] proposed to use CNN model for detecting extreme weather in climate databases. CNN model can be also used to estimate the precipitation from the remote sensing images [100].

D. Human Mobility

With the wide use of mobile devices, recent years have witnessed an explosion of extensive geolocated datasets related to human mobility. The large volume of human mobility data enable us to quantitatively study individual and collective human mobility patterns, and to generate models that can capture and reproduce the spatiotemporal structures and regularities in human trajectories. The study of human mobility is especially important for applications such as estimating migratory flows, traffic forecasting, urban planning, human behavior analysis, and personalized recommendation. Deep learning techniques applied on human mobility data mostly focus on human trajectory data mining such as trajectory classification [36], trajectory prediction [38], [64], [163], trajectory representation learning [82], [170], mobility pattern mining [118], and human transportation mode inference from trajectories [24], [42]. Based on different application scenarios and analytic purposes, human trajectories can be modeled as different types of ST data types and data representation so that different deep learning models can be applied. The most widely used models for human trajectory data mining are RNN an CNN models, and sometimes the two types of models are combined to capture both the spatial and temporal correlations among the human mobility data.

E. Location Based Social Network (LBSN)

Location-based social networks such as Foursquare and Flickr are social networks that use GPS features to locate the users and let the users broadcast their locations and other contents from their mobile device [196]. A LBSN does not only mean adding a location to an existing social network so that people can share location-embedded information, but also consists of the new social structure made up of individuals connected by their locations in the physical world as well as their location-tagged media content. LBSN data contain a large number of user check-in data which consists of the instant location of an individual at a given timestamp. Currently, deep learning methods have been used in analyzing the user generated ST data in LBSN, and the studied tasks include next check-in location prediction [67], user representation learning in LBSN [164], geographical feature extraction [26] and user check-in time prediction [165].

F. Crime Analysis

Law enforcement agencies store information about reported crimes in many cities and make the crime data publicly available for research purposes. The crime event data typically

has the type of crime (e.g., arson, assault, burglary, robbery, theft, and vandalism), as well as the time and location of the crime. Patterns in crime and the effect of law enforcement policies on the amount of crime in a region can be studied using this data with the goal of reducing crime [4]. As the crimes happened at different regions of a city usually present high spatial and temporal correlations, deep learning models can be used with the crime account heat map of a city as the input to capture such complex correlations [31], [57], [145]. For example, [31] proposed a Spatiotemporal Crime Network based on CNN to forecast the crime risk of each region in the urban area for the next day. [145] proposed to utilize ST-ResNet model to collectively predict crime distribution over the Los Angeles area. [57] developed a new crime prediction framework–DeepCrime, which is a deep neural network architecture that uncovers dynamic crime patterns and explores the evolving inter-dependencies between crimes and other ubiquitous data in urban space. As we discussed before, crime data are typical ST event data, but are usually represented as spatial maps through merging the data in spatial and temporal domains so that deep learning models can be applied for analytics.

G. Neuroscience

In recent years, brain imaging technology has become a hot topic within the field of neuroscience. Such technology includes functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG), Magnetoencephalography (MEG), and functional Near Infrared Spectroscopy (fNIRS). The spatial and temporal resolutions of neural activity measured by these technologies is quite different from another. fMRI measures the neural activity from millions of locations, while it is only measured from tens of locations for EEG data. fMRI typically measures activity for every two seconds, while the temporal resolution of EEG data is is typically 1 millisecond. Because of its space resolving power, fMRI and EEG combined with deep learning methods, has been widely used in the study of neuroscience [34], [63], [113], [128]. As we discussed before, deep learning models are mostly used for the classification task in neuroscience by using the fMRI data or EEG data such as disease classification [34], brain function network classification [113] and brain activation classification [63]. For example, Long-Short Term Memory network (LSTM) was used to identify Autism Spectrum Disorder (ASD) [34], Convolutional Neural Networks (CNN) were used to diagnose amnestic Mild Cognitive Impairment (aMCI) [113] and Feedforward Neural Networks (FNN) were used to classify Schizophrenia [119].

VI. OPEN PROBLEMS

Though many deep learning methods have been proposed and widely used for STDM in diverse application domains discussed above, challenges still exist due to the highly complex, large volume, and fast increasing ST data. In this section, we provide some open problems that have not been well addressed by current works and need further studies in the future.

Interpretable models. Current deep learning models for STDM are mostly considered as *black-boxs* which lack of interpretability. Interpretability gives deep learning models the ability to explain or to present the model behaviors in understandable terms to humans, and it is an indispensable part for machine learning models in order to better serve people and bring benefits to society [29]. Considering the complex data types and representations of ST data, it is more challenging to design interpretable deep learning models compared with other types of data such as images and word tokens. Although attention mechanisms are used in some previous works to increase the model interpretability such as periodicity and local spatial dependency [19], [57], [88], how to build a more interpretable deep learning model for STDM tasks is still not well studied and remains an open problem.

Deep learning model selection. For a given STDM task, sometimes multiple types of related ST data can be collected and different data representations can be choosen. How to properly select the ST data representations and the corresponding deep learning modes is not well studied. For example, in traffic flow prediction, some works model the traffic flow data of each road as a time series so that RNN, DNN or SAE are used for prediction [104], [136]; some works model the traffic flow data of multiple road links as spatial maps so that CNN is applied for prediction [184]; and some works model the traffic flow data of a road network as a graph so that GraphCNN is adopted [85]. There lacks of deeper studies on how to properly select deep learning models and data representations of the ST data for better addressing the STDM task under study.

Broader applications to more STDM tasks. Although deep learning models have been widely used in various STDM tasks discussed above, there are some tasks that have not been addressed by deep learning models such as frequent pattern mining and relationship mining [4], [87]. The major advantage of deep learning is its powerful feature learning ability, which is essential to some STDM tasks such as predictive learning and classification that largely rely on high quality features. However, for some STDM tasks like frequent pattern mining and relationship mining, learning high quality features may not be that helpful because these tasks do not require features. Based on our review, currently there are very few or even no works that utilize deep learning models for the tasks mentioned above. So it remains an open problem that how deep learning models along or the integration of deep learning models with traditional models such as frequent pattern mining and graphical models can be extented to broader applications to more STDM tasks.

Fusing multi-modal ST datasets. In big data era, multi-modal ST datasets are increasingly available in many domains such as neuroimaging, climate science and urban transportation. For example, in neuroimaging, fMRI and DTI can both capture the imaging data of the brain activity with different technologies that provide different spatiotemporal resolutions [61]. How to use deep learning models to effectively fuse them together to better perform the tasks of disease classification and brain activity recognition is less studied. The multi-modal transportation data including taxi trajectory data, bike-sharing trip data and public transport check-in/out data of a city can

all reflect the mobility of urban crowd flow from different perspectives [30]. Fusing and analyzing them together rather than separately can more comprehensively capture the underlying mobility patterns and make more accurate predictions. Although there are recent attempts that tried to apply deep learning models for transferring knowledge from the crowd flow data among different cities [151], [172], how to fuse multi-modal ST datasets with deep learning models is still not well studied and needs more research attention in the future.

VII. CONCLUSION

In this paper, we conduct a comprehensive overview of recent advances in exploring deep learning techniques for STDM. We first categorize the different data types and representations of ST data and briefly introduce the popular deep learning models used to STDM. For different types of ST data and their representations, we show the corresponding deep learning models that are suitable to handle them. Then we give a general framework showing the pipeline of utilizing deep learning models for addressing STDM tasks. Under the framework, we overview current works based on the categorization of the ST data types and the STDM tasks including predictive learning, representation learning, classification, estimation and inference, anomaly detection, and others. Next we summarize the applications of deep learning techniques for STDM in different domains including transportation, on-demand service, climate & weather, human mobility, location-based social network (LBSN), crime analysis, and neuroscience. Finally, we list some open problems and point out the future research directions for this fast growing research filed.

REFERENCES

[1] Y. Ai, Z. Li, M. Gan, Y. Zhang, D. Yu, W. Chen, and Y. Ju. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications*, pages 1–13, 2018.

[2] A. Akbari Asanjan, T. Yang, K. Hsu, S. Sorooshian, J. Lin, and Q. Peng. Short-term precipitation forecast based on the persiann system and lstm recurrent neural networks. *Journal of Geophysical Research: Atmospheres*, 123(22):12–543, 2018.

[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 961–971, 2016.

[4] G. Atluri, A. Karpatne, and V. Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 51(4):83, 2018.

[5] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In *In Proceedings of International Conference on Learning Representations 2015*, 2015.

[6] J. Bao, P. Liu, and S. V. Ukkusuri. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention*, 122:239–254, 2019.

[7] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In *In Proceedings of Advances in Neural Information Processing Systems*, 2006.

[8] D. Chai, L. Wang, and Q. Yang. Bike flow prediction with multi-graph convolutional networks. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 397–400. ACM, 2018.

[9] V. Chandola, R. R. Vatsavai, D. Kumar, and A. Ganguly. Analyzing big spatial and big spatiotemporal data: a case study of methods and applications. *Big Data Analytics*, 33(239), 2015.

[10] B. Chang, Y. Park, D. Park, S. Kim, and J. Kang. Content-aware hierarchical point-of-interest embedding model for successive poi recommendation. In *IJCAI*, pages 3301–3307, 2018.

[11] A. Chattopadhyay, P. Hassanzadeh, and S. Pasha. A test case for application of convolutional neural networks to spatio-temporal climate data: Re-identifying clustered weather patterns. *arXiv preprint arXiv:1811.04817*, 2018.

[12] C. Chen, K. Li, S. G. Teo, G. Chen, X. Zou, X. Yang, R. C. Vijay, J. Feng, and Z. Zeng. Exploiting spatio-temporal correlations with multiple 3d convolutional neural networks for citywide vehicle flow prediction. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 893–898. IEEE, 2018.

[13] C. Chen, C. Liao, X. Xie, Y. Wang, and J. Zhao. Trip2vec: a deep embedding approach for clustering and profiling taxi trip purposes. *Personal and Ubiquitous Computing*, pages 1–14, 2018.

[14] M. Chen, J. M. Davis, C. Liu, Z. Sun, M. M. Zempila, and W. Gao. Using deep recurrent neural network for direct beam solar irradiance cloud screening. In *Remote Sensing and Modeling of Ecosystems for Sustainability XIV*, volume 10405, page 1040503. International Society for Optics and Photonics, 2017.

[15] M. Chen, X. Yu, and Y. Liu. Pcnn: Deep convolutional networks for short-term traffic congestion prediction. *IEEE Transactions on Intelligent Transportation Systems*, (99):1–10, 2018.

[16] Q. Chen, X. Song, H. Yamada, and R. Shibasaki. Learning deep representation from big and heterogeneous data for traffic accident inference. In *AAAI*, pages 338–344, 2016.

[17] L. Cheng, H. Zang, T. Ding, R. Sun, M. Wang, Z. Wei, and G. Sun. Ensemble recurrent neural network based probabilistic wind speed forecasting approach. *Energies*, 11(8):1958, 2018.

[18] T. Cheng, J. H. B. Anbaroglu, and G. Tanakaranond. Spatiotemporal data mining. *Handbookd of Regional Science*, pages 1173–1193, 2014.

[19] W. Cheng, Y. Shen, Y. Zhu, and L. Huang. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. In *AAAI*, 2018.

[20] K.-H. Chow, A. Hiranandani, Y. Zhang, and S.-H. G. Chan. Representation learning of pedestrian trajectories using actor-critic sequence-to-sequence autoencoder. *arXiv preprint arXiv:1811.08069*, 2018.

[21] O. Costilla-Reyes, P. Scully, and K. B. Ozanyan. Deep neural networks for learning spatio-temporal features from tomography sensors. *IEEE Transactions on Industrial Electronics*, 65(1):645–653, 2018.

[22] Z. Cui, K. Henrickson, R. Ke, and Y. Wang. High-order graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *arXiv preprint arXiv:1802.07007*, 2018.

[23] Z. Cui, R. Ke, and Y. Wang. Deep stacked bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. In *6th International Workshop on Urban Computing (UrbComp 2017)*, 2016.

[24] S. Dabiri and K. Heaslip. Inferring transportation modes from gps trajectories using a convolutional neural network. *Transportation research part C: emerging technologies*, 86:360–371, 2018.

[25] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, and S. He. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[26] D. Ding, M. Zhang, X. Pan, D. Wu, and P. Pu. Geographical feature extraction for entities in location-based social networks. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 833–842. International World Wide Web Conferences Steering Committee, 2018.

[27] M. F. Dixon, N. G. Polson, and V. O. Sokolov. Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading. *Applied Stochastic Models in Business and Industry*, 2017.

[28] B. Du, H. Peng, S. Wang, M. Bhuiyan, L. Wang, Q. Gong, L. Liu, and J. Li. Deep irregular convolutional residual lstm for urban traffic passenger flows prediction. *IEEE Transactions on Intelligent Transportation Systems*, 99:1–14, 2019.

[29] M. Du, N. Liu, and X. Hu. Techniques for interpretable machine learning. *arXiv:1808.00033 [cs.LG]*, 2018.

[30] S. Du, T. Li, X. Gong, Z. Yu, and S.-J. Horng. A hybrid method for traffic flow forecasting using multimodal deep learning. *arXiv preprint arXiv:1803.02099*, 2018.

[31] L. Duan, T. Hu, E. Cheng, J. Zhu, and C. Gao. Deep convolutional neural networks for spatiotemporal crime prediction. In *2017 International Conference on Information and Knowledge Engineering (IKE)*, pages 61–67, 2017.

[32] Y. Duan, Y. Lv, W. Kang, and Y. Zhao. A deep learning based approach for traffic data imputation. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pages 912–917. IEEE, 2014.

[33] Y. Duan, Y. Lv, Y.-L. Liu, and F.-Y. Wang. An efficient realization of deep learning for traffic data imputation. *Transportation research part C: emerging technologies*, 72:168–181, 2016.

[34] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan. Identifying autism from resting-state fmri using long short-term memory networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 362–370. Springer, 2017.

[35] Y. Endo, K. Nishida, H. Toda, and H. Sawada. Predicting destinations from partial trajectories using recurrent neural network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 160–172. Springer, 2017.

[36] Y. Endo, H. Toda, K. Nishida, and J. Ikedo. Classifying spatial trajectories using representation learning. *International Journal of Data Science and Analytics*, 2(3-4):107–117, 2016.

[37] Z. Fan, X. Song, T. Xia, R. Jiang, R. Shibasaki, and R. Sakuramachi. Online deep ensemble learning for predicting citywide human mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):105, 2018.

[38] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 1459–1468. International World Wide Web Conferences Steering Committee, 2018.

[39] T. Fernando, S. Denman, S. Sridharan, and C. Fookes. Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection. *Neural networks*, 108:466–478, 2018.

[40] T. G. C. N. for Traffic Speed Prediction Considering External Factors. Liang ge and hang li and junling liu and aoli zhou. In *In Proceedings of The 20th International Conference on Mobile Data Management*, 2019.

[41] Q. Gao, G. Trajcevski, F. Zhou, K. Zhang, T. Zhong, and F. Zhang. Trajectory-based social circle inference. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 369–378. ACM, 2018.

[42] Q. Gao, F. Zhou, K. Zhang, G. Trajcevski, X. Luo, and F. Zhang. Identifying human mobility via trajectory embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1689–1695. AAAI Press, 2017.

[43] Y. Gao, L. Zhao, L. Wu, Y. Ye, H. Xiong, and C. Yang. Incomplete label multi-task deep learning for spatio-temporal event subtype forecasting. 2019.

[44] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. 2019.

[45] X. Geng, Y. Li, L. Wang, L. Zhang, J. Ye, Y. Liu, and Q. Yang. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[46] S. Giffard-Roisin, M. Yang, G. Charpiat, B. Kégl, and C. Monteleoni. Fused deep learning for hurricane forecast fused deep learning for hurricane track forecast from reanalysis data. In *Climate Informatics Workshop Proceedings 2018*, 2018.

[47] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[48] X. Guo, K. C. Dominick, A. A. Minai, H. Li, C. A. Erickson, and L. J. Lu. Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. *Frontiers in neuroscience*, 11:460, 2017.

[49] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, number CONF, 2018.

[50] S. He and K. G. Shin. Spatio-temporal capsule-based reinforcement learning for mobility-on-demand network coordination. In *In proceeding of the 27th International conference on World Wide Web*, 2019.

[51] Z. He, C.-Y. Chow, and J.-D. Zhang. Stcnn: A spatio-temporal convolutional neural network for long-term traffic prediction. In *In Proceedings of The 20th International Conference on Mobile Data Management*, 2019.

[52] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, 17:16–23, 2018.

[53] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2013.

[54] T. Horikawa and Y. Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8:15037, 2017.

[55] M. Hossain, B. Rekabdar, S. J. Louis, and S. Dascalu. Forecasting the weather of nevada: A deep learning approach. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–6. IEEE, 2015.

[56] C. Huang, C. Zhang, J. Zhao, X. Wu, N. Chawla, and D. Yin. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *In proceeding of the 27th International conference on World Wide Web*, 2019.

[57] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla. Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1423–1432. ACM, 2018.

[58] H. Huang, X. Hu, J. Han, J. Lv, N. Liu, L. Guo, and T. Liu. Latent source mining in fmri data via deep neural network. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 638–641. IEEE, 2016.

[59] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, S. Zhao, L. Guo, and T. Liu. Modeling task fmri data via deep convolutional autoencoder. *IEEE transactions on medical imaging*, 37(7):1551–1561, 2018.

[60] W. Huang, G. Song, H. Hong, and K. Xie. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning. *IEEE Trans. Intelligent Transportation Systems*, 15(5):2191–2201, 2014.

[61] R. J. Huster, S. Debener, T. Eichele, and C. S. Herrmann. Methods for simultaneous eeg-fmri: An introductory review. *Journal of Neuroscience*, 32(18):6053–6060, May 2012.

[62] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena. Structural-rnn: Deep learning on spatio-temporal graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5308–5317, 2016.

[63] H. Jang, S. M. Plis, V. D. Calhoun, and J.-H. Lee. Task-specific feature extraction and classification of fmri volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *NeuroImage*, 145:314–328, 2017.

[64] R. Jiang, X. Song, Z. Fan, T. Xia, Q. Chen, Q. Chen, and R. Shibasaki. Deep roi-based modeling for urban human mobility prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):14, 2018.

[65] X. Jiang, E. N. de Souza, A. Pesaranghader, B. Hu, D. L. Silver, and S. Matwin. Trajectorynet: An embedded gps trajectory representation for point-based classification using recurrent neural networks. *arXiv preprint arXiv:1705.02636*, 2017.

[66] M. Jin, T. Curran, and D. Cordes. Classification of amnestic mild cognitive impairment using fmri. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 29–32. IEEE, 2014.

[67] A. Karatzoglou, N. Schnell, and M. Beigl. A convolutional neural network approach for modeling semantic trajectories and predicting future locations. In *International Conference on Artificial Neural Networks*, pages 61–72. Springer, 2018.

[68] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh. Brainnetcnn: convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

[69] J. Ke, H. Yang, H. Zheng, X. Chen, Y. Jia, P. Gong, and J. Ye. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[70] J. Ke, H. Zheng, H. Yang, and X. M. Chen. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85:591–608, 2017.

[71] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*, 124:127–146, 2016.

[72] S. Kim, S. Ames, J. Lee, C. Zhang, A. C. Wilson, and D. Williams. Resolution reconstruction of climate data with pixel recursive model. In *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on*, pages 313–321. IEEE, 2017.

[73] S. Kim, S. Hong, M. Joh, and S.-k. Song. Deeprain: Convlstm network for precipitation prediction using multichannel radar data. *arXiv preprint arXiv:1711.02316*, 2017.

[74] Z. Kira, W. Li, R. Allen, and A. R. Wagner. Leveraging deep learning for spatio-temporal understanding of everyday environments. In *IJCAI Workshop on Deep Learning and Artificial Intelligence*, 2016.

[75] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo. Spatio-temporal clustering: a survey. *Data Mining and Knowledge Discovery Handbook, Springer*, 2015.

[76] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller. Deep mri brain extraction: a 3d convolutional neural network for skull stripping. *NeuroImage*, 129:460–469, 2016.

[77] D. Kong and F. Wu. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In *IJCAI*, pages 2341–2347, 2018.

[78] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova. Residual and plain convolutional neural networks for 3d brain mri classification. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 835–838. IEEE, 2017.

[79] T. Kurth, S. Treichler, J. Romero, M. Mudigonda, N. Luehr, E. Phillips, A. Mahesh, M. Matheson, J. Deslippe, M. Fatica, et al. Exascale deep learning for climate analytics. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*, page 51. IEEE Press, 2018.

[80] D. Lee, S. Jung, Y. Cheon, D. Kim, and S. You. Forecasting taxi demands with fully convolutional networks and temporal guided embedding. 2018.

[81] R. Li, Y. Shen, and Y. Zhu. Next point-of-interest recommendation with temporal and multi-level context attention. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1110–1115. IEEE, 2018.

[82] X. Li, K. Zhao, G. Cong, C. S. Jensen, and W. Wei. Deep representation learning for trajectory similarity computation. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 617–628. IEEE, 2018.

[83] Y. Li, K. Fu, Z. Wang, C. Shahabi, J. Ye, and Y. Liu. Multi-task representation learning for travel time estimation. In *International Conference on Knowledge Discovery and Data Mining,(KDD)*, 2018.

[84] Y. Li and B. Shuai. Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms. *Multimedia Tools and Applications*, pages 1–12, 2018.

[85] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. 2018.

[86] Y. Li, Z. Zhu, D. Kong, M. Xu, and Y. Zhao. Learning heterogeneous spatial-temporal representation for bike-sharing demand prediction. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[87] Z. Li. Spatiotemporal pattern mining: Algorithms and applications. *Frequent Pattern Mining*, 2014.

[88] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng. Geoman: Multi-level attention networks for geo-sensory time series prediction. In *IJCAI*, pages 3428–3434, 2018.

[89] B. Liao, J. Zhang, M. Cai, S. Tang, Y. Gao, C. Wu, S. Yang, W. Zhu, Y. Guo, and F. Wu. Dest-resnet: A deep spatiotemporal residual network for hotspot traffic speed prediction. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 1883–1891. ACM, 2018.

[90] B. Liao, J. Zhang, C. Wu, D. McIlwraith, T. Chen, S. Yang, Y. Guo, and F. Wu. Deep sequence learning with auxiliary information for traffic prediction. *arXiv preprint arXiv:1806.07380*, 2018.

[91] D. Liao, W. Liu, Y. Zhong, J. Li, and G. Wang. Predicting activity and location with multi-task context aware recurrent neural network. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pages 3435–3441, 2018.

[92] L. Lin, Z. He, and S. Peeta. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, 97:258–276, 2018.

[93] Y. Lin, X. Dai, L. Li, and F.-Y. Wang. Pattern sensitive prediction of traffic flow based on generative adversarial framework. *IEEE Transactions on Intelligent Transportation Systems*, (99):1–6, 2018.

[94] Y. Lin, N. Mago, Y. Gao, Y. Li, Y.-Y. Chiang, C. Shahabi, and J. L. Ambite. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 359–368. ACM, 2018.

[95] Z. Lin, J. Feng, Z. Lu, Y. Li, and D. Jin. Deepstn+: Contextaware spatial temporal neural network for crowd flow prediction in metropolis. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[96] H. Liu, X. Mi, and Y. Li. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Conversion and Management*, 166:120–131, 2018.

[97] H. Liu, X. Mi, and Y. Li. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, lstm network and elm. *Energy Conversion and Management*, 159:54–64, 2018.

[98] L. Liu, R. Zhang, J. Peng, G. Li, B. Du, and L. Lin. Attentive crowd flow machines. *arXiv preprint arXiv:1809.00101*, 2018.

[99] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *AAAI*, pages 194–200, 2016.

[100] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*, 2016.

[101] Y. Liu, Y. Wang, X. Yang, and L. Zhang. Short-term travel time prediction by deep learning: A comparison of different lstm-dnn models. In *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*, pages 1–8. IEEE, 2017.

[102] Z. Liu, Z. Li, K. Wu, and M. Li. Urban traffic prediction from mobility data using deep learning. *IEEE Network*, 32(4):40–46, 2018.

[103] J. Lv, Q. Li, Q. Sun, and X. Wang. T-conv: A convolutional neural network for multi-scale taxi trajectory prediction. In *Big Data and Smart Computing (BigComp), 2018 IEEE International Conference on*, pages 82–89. IEEE, 2018.

[104] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang, et al. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intelligent Transportation Systems*, 16(2):865–873, 2015.

[105] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou. Lc-rnn: A deep learning model for traffic speed prediction. In *IJCAI*, pages 3470–3476, 2018.

[106] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4):818, 2017.

[107] X. Ma, Y. Li, Z. Cui, and Y. Wang. Forecasting transportation network speed using deep capsule networks with nested lstm models. *arXiv preprint arXiv:1811.04745*, 2018.

[108] X. Ma, H. Yu, Y. Wang, and Y. Wang. Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS one*, 10(3):e0119044, 2015.

[109] X. Ma, J. Zhang, B. Du, C. Ding, and L. Sun. Parallel architecture of convolutional bi-directional lstm neural networks for network-wide metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[110] A. H. Marblestone, G. Wayne, and K. P. Kording. Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience*, 10:94, 2016.

[111] H. Martin, D. Bucher, E. Suel, P. Zhao, F. Perez-Cruz, and M. Raubal. Graph convolutional neural networks for human activity purpose imputation from gps-based trajectory data. 2018.

[112] J. D. Mazimpaka and S. Timpf. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, (13):61–99, 2016.

[113] R. J. Meszlényi, K. Buza, and Z. Vidnyánszky. Resting state fmri functional connectivity-based classification using a convolutional neural network architecture. *Frontiers in neuroinformatics*, 11:61, 2017.

[114] H. Nguyen, L.-M. Kieu, T. Wen, and C. Cai. Deep learning methods in transportation domain: a review. *IET Intelligent Transport Systems*, 12(9):998–1004, 2018.

[115] N. T. Nguyen, Y. Wang, H. Li, X. Liu, and Z. Han. Extracting typical users' moving patterns using deep learning. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 5410–5414. IEEE, 2012.

[116] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen. 3d deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 212–220. Springer, 2016.

[117] X. Niu, Y. Zhu, and X. Zhang. Deepsense: A novel learning mechanism for traffic prediction with taxi gps traces. In *Global Communications Conference (GLOBECOM), 2014 IEEE*, pages 2745–2750. IEEE, 2014.

[118] X. Ouyang, C. Zhang, P. Zhou, and H. Jiang. Deepspace: An online deep learning framework for mobile big data to understand human mobility patterns. *arXiv preprint arXiv:1610.07009*, 2016.

[119] P. Patel, P. Aggarwal, and A. Gupta. Classification of schizophrenia versus normal subjects using deep learning. In *In Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, 2016.

[120] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, 8:229, 2014.

[121] N. G. Polson and V. O. Sokolov. Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79:1–17, 2017.

[122] Z. Qiu, T. Yao, and T. Mei. Learning spatio-temporal representation with pseudo-3d residual networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5534–5542. IEEE, 2017.

[123] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, and C. Pal. Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Advances in Neural Information Processing Systems*, pages 3402–3413, 2017.

[124] S. Rasp and S. Lerch. Neural networks for post-processing ensemble weather forecasts. *arXiv preprint arXiv:1805.09091*, 2018.

[125] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei. A deep learning approach to the citywide traffic accident risk prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3346–3351. IEEE, 2018.

[126] F. Rodrigues, I. Markou, and F. C. Pereira. Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach. *Information Fusion*, 49:120–129, 2019.

[127] I. Roesch and T. Günther. Visualization of neural network predictions for weather forecasting. In *Computer Graphics Forum*. Wiley Online Library, 2017.

[128] S. Sarraf and G. Tofighi. Deep learning-based pipeline to recognize alzheimer's disease using fmri data. In *Future Technologies Conference (FTC)*, pages 816–820. IEEE, 2016.

[129] S. Scher. Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, 45(22):12–616, 2018.

[130] S. Shekhar, Z. Jiang, R. Y. Ali, E. Eftelioglu, X. Tang, V. M. V. Gunturi, and X. Zhou. Spatiotemporal data mining: A computational perspective. *ISPRS International Journal of Geo-Information*, 4:2306–2338, 2015.

[131] B. Shen, X. Liang, Y. Ouyang, M. Liu, W. Zheng, and K. M. Carley. Stepdeep: A novel spatial-temporal mobility event prediction framework based on deep neural network. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 724–733. ACM, 2018.

[132] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE journal of biomedical and health informatics*, 22(1):173–183, 2018.

[133] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems*, pages 5617–5627, 2017.

[134] Y. Shi, Y. Tian, Y. Wang, and T. Huang. Sequential deep trajectory descriptor for action recognition with three-stream cnn. *IEEE Transactions on Multimedia*, 19(7):1510–1520, 2017.

[135] X. Song, H. Kanasugi, and R. Shibasaki. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In *IJCAI*, volume 16, pages 2618–2624, 2016.

[136] R. Soua, A. Koesdwiady, and F. Karray. Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, pages 3195–3202. IEEE, 2016.

[137] F. Sun, A. Dubey, and J. White. Dxnat-deep neural networks for explaining non-recurring traffic congestion. In *In Proceedings of IEEE International Conference on Big Data*, 2017.

[138] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014.

[139] Y. Tao, X. Gao, A. Ihler, K. Hsu, and S. Sorooshian. Deep neural networks for precipitation estimation from remotely sensed information. In *Evolutionary Computation (CEC), 2016 IEEE Congress on*, pages 1349–1355. IEEE, 2016.

[140] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatio-temporal features. In *European conference on computer vision*, pages 140–153. Springer, 2010.

[141] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

[142] D. Varshneya and G. Srinivasaraghavan. Human trajectory prediction using spatially aware deep attention models. *arXiv preprint arXiv:1705.09436*, 2017.

[143] R. R. Vatsavai, V. Chandola, S. Klasky, A. Ganguly, A. Stefanidis, and S. Shekhar. Spatiotemporal data mining in the era of big spatial data: algorithms and applications. In *In SIGSPATIAL international workshop on analytics for big geospatial data*, 2012.

[144] B. Wang, X. Luo, F. Zhang, B. Yuan, A. L. Bertozzi, and P. J. Brantingham. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. *arXiv preprint arXiv:1804.00684*, 2018.

[145] B. Wang, D. Zhang, D. Zhang, P. J. Brantingham, and A. L. Bertozzi. Deep learning for real time crime forecasting. *arXiv preprint arXiv:1707.03340*, 2017.

[146] D. Wang, W. Cao, J. Li, and J. Ye. Deepsd: supply-demand prediction for online car-hailing services using deep neural networks. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 243–254. IEEE, 2017.

[147] D. Wang, J. Zhang, W. Cao, J. Li, and Y. Zheng. When will you arrive? estimating travel time based on deep neural networks. AAAI, 2018.

[148] H. Wang, G. Liu, J. Duan, and L. Zhang. Detecting transportation modes using deep neural network. *IEICE TRANSACTIONS on Information and Systems*, 100(5):1132–1135, 2017.

[149] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong. Traffic speed prediction and congestion source exploration: A deep learning method. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 499–508. IEEE, 2016.

[150] L. Wang, X. Geng, J. Ke, C. Peng, X. Ma, D. Zhang, and Q. Yang. Ridesourcing car detection by transfer learning. *arXiv preprint arXiv:1705.08409*, 2017.

[151] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang. Crowd flow prediction by deep spatio-temporal transfer learning. *CoRR*, abs/1802.00386, 2018.

[152] L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4305–4314, 2015.

[153] P. Wang, Y. Fu, J. Zhang, X. Li, and D. Lin. Learning urban community structures: A collective embedding perspective with periodic spatial-temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(6):63, 2018.

[154] P. Wang, Z. Li, Y. Hou, and W. Li. Action recognition based on joint trajectory maps using convolutional neural networks. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 102–106. ACM, 2016.

[155] X. Wang, C. Chen, Y. Min, J. He, B. Yang, and Y. Zhang. Efficient metropolitan traffic prediction based on graph recurrent neural network. *arXiv preprint arXiv:1811.00740*, 2018.

[156] Y. Wang, M. Long, J. Wang, Z. Gao, and S. Y. Philip. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Advances in Neural Information Processing Systems*, pages 879–888, 2017.

[157] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee. Enhancing transportation systems via deep learning: A survey. *Transportation Research Part C: Emerging Technologies*, 2018.

[158] D. Wen, Z. Wei, Y. Zhou, G. Li, X. Zhang, and W. Han. Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: A brief overview and our opinion. *Frontiers in neuroinformatics*, 12:23, 2018.

[159] H. Wu, Z. Chen, W. Sun, B. Zheng, and W. Wang. Modeling trajectories with recurrent neural networks. IJCAI, 2017.

[160] Z. Wu, S. Pan, F. Chen, G. Long, and P. S. Y. Chengqi Zhang. A comprehensive survey on graph neural networks. *arXiv: 1901.00596v2*, 2019.

[161] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.

[162] C. Xu, J. Ji, and P. Liu. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies*, 95:47–60, 2018.

[163] K. Xu, Z. Qin, G. Wang, K. Huang, S. Ye, and H. Zhang. Collision-free lstm for human trajectory prediction. In *International Conference on Multimedia Modeling*, pages 106–116. Springer, 2018.

[164] C. Yang, M. Sun, W. X. Zhao, Z. Liu, and E. Y. Chang. A neural network approach to jointly modeling social networks and mobile tra-

jectories. *ACM Transactions on Information Systems (TOIS)*, 35(4):36, 2017.

[165] G. Yang, Y. Cai, and C. K. Reddy. Recurrent spatio-temporal point process for check-in time prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2203–2211. ACM, 2018.

[166] G. Yang, Y. Cai, and C. K. Reddy. Spatio-temporal check-in time prediction with recurrent neural network based survival analysis. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[167] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen. Optimized structure of the traffic flow forecasting model with a deep learning approach. *IEEE transactions on neural networks and learning systems*, 28(10):2371–2381, 2017.

[168] J. Yang and C. Eickhoff. Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling. *ACM Transactions on Information Systems (TOIS)*, 36(3):32, 2018.

[169] D. Yao, C. Zhang, J. Huang, and J. Bi. Serm: A recurrent model for next location prediction in semantic trajectories. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2411–2414. ACM, 2017.

[170] D. Yao, C. Zhang, Z. Zhu, Q. Hu, Z. Wang, J. Huang, and J. Bi. Learning deep representation for trajectory clustering. *Expert Systems*, 35(2):e12252, 2018.

[171] D. Yao, C. Zhang, Z. Zhu, J. Huang, and J. Bi. Trajectory clustering via deep representation learning. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 3880–3887. IEEE, 2017.

[172] H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In *In proceeding of the 27th International conference on World Wide Web*, 2019.

[173] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *In Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 2019.

[174] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, and J. Ye. Deep multi-view spatial-temporal network for taxi demand prediction. *arXiv preprint arXiv:1802.08714*, 2018.

[175] B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting.

[176] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 17(7):1501, 2017.

[177] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu. Deep learning: A generic approach for extreme condition traffic forecasting. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 777–785. SIAM, 2017.

[178] R. Yuecheng, X. Zhimian, Y. Ruibo, and M. Xu. Du-parking: Spatio-temporal big data tells you realtime parking availability. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 646–654. ACM, 2018.

[179] M. A. Zaytar and C. El Amrani. Sequence to sequence weather forecasting with long short term memory recurrent neural networks. *Int J Comput Appl*, 143(11), 2016.

[180] G. Zhan. A semantic sequential correlation based lstm model for next poi recommendation. In *In Proceedings of The 20th International Conference on Mobile Data Management*, 2019.

[181] H. Zhang, H. Wu, W. Sun, and B. Zheng. Deeptravel: a neural network based travel time estimation model with auxiliary supervision. *arXiv preprint arXiv:1802.02147*, 2018.

[182] J. Zhang, B. Cao, S. Xie, C.-T. Lu, P. S. Yu, and A. B. Ragin. Identifying connectivity patterns for brain diseases via multi-side-view guided deep architectures. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 36–44. SIAM, 2016.

[183] J. Zhang, Y. Zheng, and D. Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pages 1655–1661, 2017.

[184] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi. Dnn-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 92. ACM, 2016.

[185] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence*, 259:147–166, 2018.

[186] P. Zhang, Y. Jia, J. Gao, W. Song, and H. K. Leung. Short-term rainfall forecasting using multi-layer perceptron. *IEEE Transactions on Big Data*, 2018.

[187] S. Zhang, G. Wu, J. P. Costeira, and J. M. Moura. Fcn-rlstm: Deep spatio-temporal neural networks for vehicle counting in city cameras. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 3687–3696. IEEE, 2017.

[188] W. Zhang, L. Han, J. Sun, H. Guo, and J. Dai. Application of multi-channel 3d-cube successive convolution network for convective storm nowcasting. *arXiv preprint arXiv:1702.04517*, 2017.

[189] Z. Zhang, Q. He, J. Gao, and M. Ni. A deep learning approach for detecting traffic accidents from social media data. *Transportation research part C: emerging technologies*, 86:580–596, 2018.

[190] J. Zhao, J. Xu, R. Zhou, P. Zhao, C. Liu, and F. Zhu. On prediction of user destination by sub-trajectory understanding: A deep learning based approach. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1413–1422. ACM, 2018.

[191] L. Zhao, Y. Song, M. Deng, and H. Li. Temporal graph convolutional network for urban traffic flow prediction method. *arXiv preprint arXiv:1811.05320*, 2018.

[192] P. Zhao, H. Zhu, Y. Liu, Z. Li, J. Xu, and V. S. Sheng. Where to go next: A spatio-temporal lstm model for next poi recommendation. *arXiv preprint arXiv:1806.06671*, 2018.

[193] S. Zhao, T. Zhao, I. King, and M. R. Lyu. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th international conference on world wide web companion*, pages 153–162. International World Wide Web Conferences Steering Committee, 2017.

[194] Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, L. Guo, X. Hu, J. Han, and T. Liu. Automatic recognition of fmri-derived functional networks using 3-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1975–1984, 2018.

[195] S. Zheng, Y. Yue, and J. Hobbs. Generating long-term trajectories using deep hierarchical networks. In *Advances in Neural Information Processing Systems*, pages 1543–1551, 2016.

[196] Y. Zheng. Tutorial on location-based social networks. In *In proceeding of the 21st International conference on World Wide Web*, 2012.

[197] F. Zhou, Q. Gao, G. Trajcevski, K. Zhang, T. Zhong, and F. Zhang. Trajectory-user linking via variational autoencoder. In *IJCAI*, pages 3212–3218, 2018.

[198] X. Zhou, Y. Shen, Y. Zhu, and L. Huang. Predicting multi-step citywide passenger demands using attention-based neural networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 736–744. ACM, 2018.

[199] L. Zhu, F. Guo, R. Krishnan, and J. W. Polak. A deep learning approach for traffic incident detection in urban networks. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 1011–1016. IEEE, 2018.

[200] Q. Zhu, J. Chen, L. Zhu, X. Duan, and Y. Liu. Wind speed prediction with spatio–temporal correlation: A deep learning approach. *Energies*, 11(4):705, 2018.

[201] Y. Zhuoning, Z. Xun, and Y. Tianbao. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 984–992. ACM, 2018.

[202] A. Zonoozi, J.-j. Kim, X.-L. Li, and G. Cong. Periodic-crn: A convolutional recurrent model for crowd density prediction with recurring periodic patterns. In *IJCAI*, pages 3732–3738, 2018.