Chapter 1

Introduction

Recent years have witnessed evolutionary growth in embedded systems and communication technologies, which helps in designing compact-sized mobile devices. Such devices generally operate over batteries and are powerful enough to perform most of the tasks the users can perform on a computer [1, 2]. Moreover, designing these compact devices incurs minimal costs, and they consume limited energy while executing a task. These low-cost, low-powered, and compact devices, like smartphones, smartwatches, smart glasses, have revolutionized in-situ task processing in real-time applications. Specifically, smartphones are widely adopted and the preferable choice for a significant number of users; in fact, most of the users only have smartphones as computing devices.

Besides the computational capacity of simultaneously processing multiple tasks, smartphones possess richer sensing capabilities. Smartphones have various onboard sensors, including accelerometer, gyroscope, magnetometer, touch, fingerprint sensor, heart rate, *etc* [3, 4]. These onboard sensors facilitate unprecedented opportunities to perform various sensing and monitoring activities. The sensing and monitoring activities are advantageous in different application domains such as health care, agriculture, activities monitoring, transport, localization and tracking, human safety, *etc* [1, 2, 4]. Figure 1.1 illustrates different applications of smartphones in smart transportation.

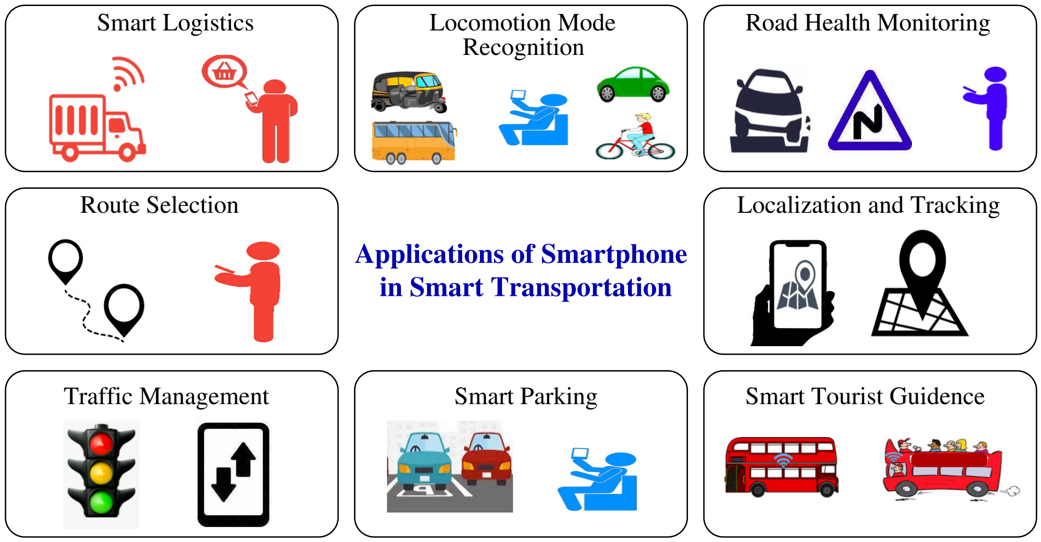

Figure 1.1: An illustration of smartphones applications in smart transport.

With the availability of various sensors in the smartphone, identifying a locomotion (or transportation) mode has become convenient and effortless in recent years. As the smartphone has become an integral part of our daily routine, which can capture crucial information about basic and complex locomotion activities; therefore, locomotion mode detection becomes a dominating area of research. It helps in estimating travel time, traffic management, journey planning, route selection, *etc* [5]. A smartphone user is no longer assumed to be static and can conveniently move outside, *i.e.,* walk, run or may use different transportation modes like bus, train, car, bike, bicycle, *etc* [6]. This movement of users generates a huge amount of sensory data on smartphones. In this thesis, we use "transportation" and "locomotion" as interchangeable terms.

A smartphone trains a recognition model on the collected and annotated sensory dataset to build a classifier, which learns the mapping between sensory instances and their corresponding class labels (or locomotion modes) during training. While testing, the class labels are predicted for a given sensory instance. However, collecting and annotating a large amount of data for all possible locomotion modes is cost-inefficient and time-consuming [7, 8]. A traditional classifier can recognize only seen locomotion modes that are given with training instances. In other words, such a classifier is in-capable to identify an unseen class that appears first time in testing. Such unseen class or locomotion mode can be identified by using the semantic information of the seen classes [9, 10]. Zero-shot learning [11] is a concept that extends the capability of the traditional classifier to identify the unseen classes. Figure 1.2(a) illustrates seen and unseen modes in the testing dataset. It shows the instances for auto-rickshaw and tractor are missing from training dataset; however, they appear in testing.

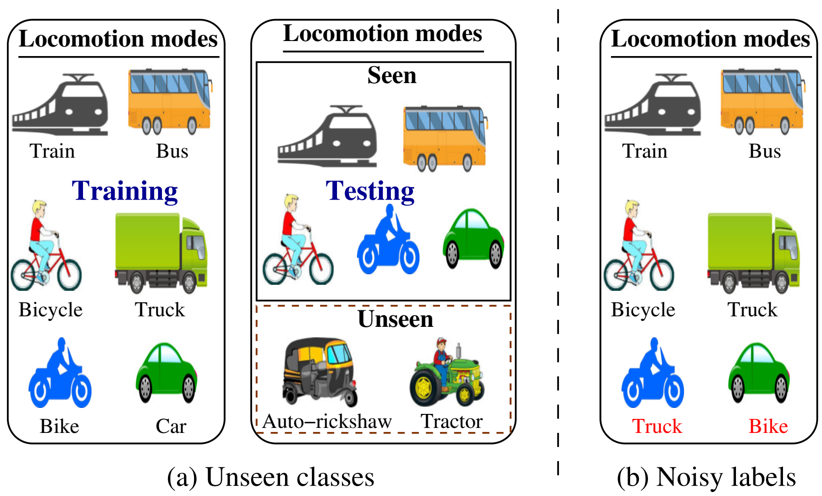

Figure 1.2: Example scenarios: (a) seen and unseen locomotion modes in the dataset and (b) locomotion modes with noisy labels in the dataset. The wrong labels of the car as bike and bike as truck (in red color) indicate noisy labels in training data.

Noisy labels in the dataset is another critical issue that arises while training a recognition model, which requires colossal training data with correctly annotated labels. Acquiring such annotated dataset is expensive and time-consuming. Therefore, the data analysts employ different annotation techniques such as labeling through crowdsourcing, web-based queries, and so on [12]. These annotation techniques may introduce noisy labels in the dataset as the labeling is carried out by non-expert volunteers. If noisy labels are present in the dataset, the classifier learns a wrong mapping which results in performance diminution [13]. In addition, the assumption of information about the concentration of the noisy labels in the training dataset is impractical. Figure 1.2(b) illustrates an example of locomotion modes with some wrong labels of car and bike as bike and truck (in red color), respectively.

Task processing on smartphones suffers from resource constraints, including processing capability, storage, and battery power. Such constraints hamper the successful execution or delay the processing. Cloud computing resolves the problem of resource constraints on smartphones. Initially, the smartphone transfers task to the Cloud, task executes on the Cloud, and result is propagated back to the smartphone. However, Cloud computing suffers from substantial communication delay and requires continuous long-range communication networks (like 4G or 5G). Such networks consume the colossal power of smartphones. Fog computing mitigates the shortcomings of Cloud computing and provides the mechanism of task execution using other devices (smartphones) in the proximity [14]. The Fog devices can interact with each other and provide parallel data processing [15–19]. The interaction among Fog devices overcomes the limitation of unequal storage and computation power among devices.

Federated Learning (FL) is an emerging paradigm that provides a distributed training framework, where data collection and model training are localized to the participants devices [20–22]. This localization preserves data privacy and reduces communication overhead while transmitting data to the server. Thus, FL proves to be a suitable technique for executing a task on smartphones. Each participant device in FL trains the model using the local dataset and sends the trained model's Weight Parameter Matrices (WPM) to the central server. In turn, the central server aggregates the WPM received from participant devices and sends them back the updated WPM. A participant device in FL uses its resources, such as memory and processing power (which may not be the same for all devices), to load the model and train them locally. The availability of resources at the participant devices depends on their type and other installed services. Such heterogeneity in device resources requires unequal time to train the model using local datasets. Moreover, a large-size model may not be successfully trained on a device with limited resources. Similarly, the time required to share the WPM between each participant device and the central server depends on the networking resources.

1.1 Motivation of the research work

Locomotion activity detection has been a field of great research interest that helps to understand the movement patterns of human [23–25]. Literature indicates several significant contributions towards building a pertinent approach for detecting locomotion activities [1, 2]. The smartphone trains a recognition model on the collected sensory dataset to detect transportation modes. The performance of the model heavily depends on labeled training instances. However, it is impractical to gather prior information (labeled instances) about all types of transportation modes. In addition, the inappropriate annotation of the collected sensory data generates abundant noisy labels. Thus, the recognition models must provide robustness against the uncertainty of noisy and unseen labels. Further, the task processing suffers from the resource constraints of smartphones, including processing capability, storage, and battery. Such constraints hamper the successful execution or delay the processing on smartphones. It generates the need for appropriate mechanisms to provide fast processing, reduce memory, and minimize energy consumption on smartphones.

The existing recognition approaches [5, 26–28] can identify only seen locomotion modes by extracting the deep features from the training instances. It is impractical to acquire prior knowledge (*i.e.,* labeled training instances) about each type of locomotion mode. Thus, the recognition model should be capable enough to identify a new locomotion mode without having any corresponding training instance. It indicates that the approaches [5, 26–28] can not be employed for identifying an unseen locomotion mode. In prior studies [29, 30], the authors employed machine learning-based recognition models to identify the locomotion modes, which heavily rely on the knowledge of domain-related features. Such dependency can be obviated by utilizing the automatic feature extraction capabilities of deep network models.

Next, the locomotion mode recognition models require a colossal amount of correctly annotated data instances, which compels to adopt automated annotation mechanisms or non-expert volunteers for annotating a large amount of data. Though it is cost-and time-effective but results in noisy labels in the annotated dataset. Additionally, assuming information about the concentration of noisy labels is intangible in the real-world scenario. The existing approaches [31–37] require prior information about the concentration, which helped in the easier handling of noisy labels by setting parameters accordingly. However, the approaches [31–41] do not provide any mechanism to determine the concentration. It motivates to develop a recognition approach that should be robust against noisy labels, utilizing no prior information about noisy labels. Some of the previous studies [33, 39, 40] can handle a low concentration. Therefore, in the real-world scenario, where the chances of noisy labels are high, we need a mechanism that can achieve adequate performance on a higher concentration.

Further, the existing literature on task partitioning and offloading in Fog computing suffer from the following limitations. The prior studies on task offloading [14–17, 42–46] presented hierarchical networks topology, comprising Edge, Fog, and Cloud layers for distributed tasks execution. However, the existing work did not consider the interaction among the Fog devices, *i.e.*, offloading from one Fog device to others. A few prior work [14, 42, 43, 47] considered heterogeneity of Fog devices limited to the transmission of the complete task; hence, utilizing available resources inefficiently. Later, the existing work [44–46] incorporated offloading of complete task from one device to another. However, task partitioning is requisite to ensure load balancing among different Fog devices while executing the task.

Finally, to address the heterogeneity of resources among the participants (smartphones) in FL, prior studies proposed mechanisms that discard slow processing participants, called *stragglers*, from the federation [48–50]. However, the removal of stragglers hampers effective utilization of the local dataset (on stragglers) and prohibits their performance improvement via FL. Some existing work [51] have considered a fixed size model for all the devices of heterogeneous resources. The fixed size model may not fully utilize the colossal resources devices. The existing work [52, 53] used Knowledge Distillation (KD) to resize and train the model that fit on the devices for FL. The KD is a student-teacher learning process. The training of the student model by the teacher model requires multiple epochs in KD; therefore, it delays the aggregation process at the central server. The existing work [54] considered unequal bandwidth issue in FL that faces weight staleness problem. Some devices update their parameters multiple times, while others may not participate in aggregation. In summary, the existing FL techniques in the presence of heterogeneous resources avoid the straggler devices, delay the aggression process, and/or reduce the number of aggregation rounds.

1.2 Contributions and organization of the thesis

In this thesis, we investigate the different challenges encountered while extending the capabilities of smartphone sensors for application in smart transport. We first consider the problem of unseen class (locomotion/transportation mode), *i.e.,* no training instance of the class in the training dataset; however, such instances may appear during testing. We employ the concept of zero-shot learning and deep learning features to construct a classifier, which can recognize both seen and unseen locomotion modes. We next highlight the performance deterioration of the built classifier in deep learning due to noisy labels in the training dataset. We propose a mechanism for handling noisy labels with no prior information about noise concentration. Further, we solve the problem of unavoidable delay due to task processing on smartphones having resource (storage/processing/battery) constraints. We propose a system for task partitioning and sub-tasks offloading among multiple Fog devices (or smartphones) to minimize delay and execution cost. Extending the problem of task execution delay and resource constraints on smartphones, we propose the technique to deploy optimal lightweight or large-size models on smartphones with heterogeneous device and networking resources. Specifically, this thesis investigates the following research problems:

- How to identify both seen and unseen locomotion (or transportation) modes using zero-shot learning and deep learning features of the labeled training instances?
- How to recognize a locomotion mode using deep learning models in the presence of noisy labels with no prior information about the noise concentration?
- How to split a smart transport task and offload the sub-tasks among Fog devices (smartphones), which ensures successful execution and incurs minimal cost?
- How to train an optimal lightweight or large-size model on the participants devices (smartphones) in FL with heterogeneous resources (unequal memory, processing power, and bandwidth) using local datasets?

The rest of the thesis is organized as follows:

Chapter 2: This chapter presents the preliminaries about the various techniques used in this thesis, including Fog computing, zero-shot learning, knowledge distillation, and federated learning. The chapter also presents state-of-the-art work covering the prior studies on task offloading in Fog computing, followed by the strategies to handle unseen classes and noisy labels in the dataset. Further, we discuss the overview of the federated learning techniques that solved the problem of resource constraints on the participating devices (smartphones).

Chapter 3: In this chapter, we consider the problem of unseen locomotion modes in the given dataset. During the construction of the classifier, we incorporate the concept of zero-shot learning to enhance the capability of the traditional classifier such that it can recognize both seen and unseen locomotion modes. *We address the problem of identifying seen and unseen locomotion modes using semantic information and deep learning features of labeled training instances.* To solve this problem, this work proposes a sensors based *Deep* learning model by incorporating the concept of *Zero*-shot learning for identifying the unseen locomotion modes. We call this model as DeepZero. It first extracts the features from the training data using a sequential combination of convolutional neural network and long short term memory. Later, the model builds a classifier by learning a mapping between the extracted features and semantic information of the class labels. Finally, we develop an android application to collect a Locomotion Mode Recognition (LMR) dataset using acceleration, gyroscope, and magnetometer sensors of the smartphone. This chapter conducts various experiments to evaluate the performance of the DeepZero model using the LMR dataset along with an existing Sussex-Huawei Locomotion (SHL) dataset [55].

Chapter 4: In this chapter, we consider the noisy labels in the sensory dataset of locomotion modes with no prior information about noise concentration. Our target is to build a deep learning model that is robust towards the noisy labels in the dataset. *We address the problem of recognizing a locomotion mode using deep learning models in the presence of noisy labels.* To solve the problem, this work proposes a deep learning based approach called as Locomotion mode Recognition with Noisy Labels (LRNL). The approach builds an ensemble model by incorporating three different models *i.e.,* conventional, noise adaptive, and noise corrective. The conventional model simultaneously extracts spatial and temporal features using deep learning techniques. It is robust against a low concentration of noisy labels and the baseline for other models. For handling a moderate concentration, the noise adaptive model introduces a new loss function. Later, the noise corrective model uses the low-rank estimate for handling higher concentrations. Finally, we conduct experiments to evaluate the performance of the LRNL approach using the collected LMR dataset along with two existing datasets, *i.e.*, SHL [55] and Transportation Mode Detection (TMD) [56].

Chapter 5: In this chapter, we consider the problem of task execution on the smartphone, which minimizes delay and execution cost. *Specifically, we address the problem of effectively splitting a smart transport task and offloading the sub-tasks among Fog* *devices (smartphones), ensuring successful execution and incurring the minimal cost.* This chapter presents a Fog computing based transportation system. The system uses multiple Fog devices (smartphones) to assist the passengers. The passengers and vehicles work as end-users and the Edge devices in the system, respectively. A passenger generates a task and forwards it to the Fog devices for further processing using the Edge device. Selected Fog devices parallel process the fraction of the task such that the complete task processes within the given time constraint. We present a Knapsack-based task offloading algorithm, which helps to utilize the resources of Fog devices. We also present a competitive game model and near Nash Equilibrium solution for estimating the optimal value of the fraction of the task process at Fog devices. Finally, we develop a prototype and present results to investigate the performance of the system.

Chapter 6: In this chapter, we consider the heterogeneity of device and networking resources while allocating and training optimal lightweight or large-size models on smartphones using FL. We present a fast federated learning technique to train the deep learning models for prediction or classification tasks at the participant devices using the local datasets. *We address the problem of successfully training a model on the participant devices (smartphones) with heterogeneous resources (unequal memory, processing power, and bandwidth) using local datasets.* The proposed technique starts with the collection of available resource information of participant devices and selects a generic model which directly works on most of the devices. We next propose a knowledge distillation-based early-halting approach for devices where the generic model does not directly fit. The early halting in the fast federated learning technique speeds up the training of the model at the participant devices. Finally, we perform a real-world study to evaluate the feasibility and performance of the technique and compare it with state-of-the-art federated learning techniques on the pervasive domain of locomotion mode recognition with smartphones.

Chapter 7: The chapter summarizes the principal findings of the research work presented in this thesis. We also cover the promising directions of futuristic research in the pervasive domain of smartphone-based sensing and computation for smart transport.

Technical reports, research papers and textbooks are listed in References. The publications of the research work presented in this thesis are listed in the List of Publications.