SateLoc: A Virtual Fingerprinting Approach to Outdoor LoRa Localization using Satellite Images

Yuxiang Lin[1,2], Wei Dong[1,2], Yi Gao[1,2], and Tao Gu[3]

[1]College of Computer Science, Zhejiang University
[2]Alibaba-Zhejiang University Joint Institute of Frontier Technologies, China
[3]Computer Science and Information Technology, RMIT University
{linyx,dongw,gaoyi}@zju.edu.cn,tao.gu@rmit.edu.au

ABSTRACT

With the increasing relevance of the Internet of Things (IoT) and large-scale Location-Based Services (LBS), LoRa localization has been attractive due to its low cost, low power and long range properties. However, existing localization approaches based on Received Signal Strength Indicator (RSSI) are either easily affected by signal fading of different land-cover types or labor-intensive. In this work, we propose SateLoc, a LoRa localization system that utilizes satellite images to generate virtual fingerprints. Specifically, SateLoc first uses high-resolution satellite images to identify land-cover types. With the path loss parameters of each land-cover type, SateLoc can automatically generate a virtual fingerprinting map for each gateway. We then propose a novel multi-gateway combination strategy, which is weighted by the environment interference of each gateway, to produce a joint likelihood distribution for localization. We implement SateLoc with commercial LoRa devices without any hardware modification, and evaluate its performance in a 227,500m^2 urban area. Experimental results show that SateLoc achieves a median localization error of 47.1m, improving more than 40% compared to the state-of-the-art model-based approaches. More importantly, compared to the fingerprinting-based approach, SateLoc does not require the labor-intensive fingerprint acquisition process.

CCS CONCEPTS

• **Networks** → **Wide area networks**; **Location based services**;

KEYWORDS

LoRa localization, Virtual fingerprints, Land-cover information

1 INTRODUCTION

The vision of the Internet of Things (IoT) has been to communicate with everyday objects around us. It has been reported that the number of IoT devices will reach 75.44 billion by 2025 [37]. The location information of these IoT devices can be exploited in various city-scale applications, such as finding lost objects (e.g., keys), navigating and tracking pets. Over the years, various solutions have been proposed for such open-air localization. Traditional Global Positioning System (GPS) can provide accurate locations of the IoT devices. However, a GPS receiver typically consumes 30-50 mA power and needs additional power to forward data [16], which is incompatible with the low power consumption constraints of IoT systems [32]. Besides, GPS tends to lose connectivity in indoor environments. In order to remedy these defects, over the years, many localization approaches have been proposed based on wireless technologies such as Wi-Fi [39, 40], BLE [4, 41], and ZigBee [5, 38]. These approaches have been shown to be effective, however, they require dense deployment of Access Points (APs) due to short radio range, leading to high deployment and maintenance cost. Long-range radio-based localization has recently attracted attention from researchers.

Among existing long-range radios, LoRa has been widely deployed in many city-scale IoT networks [9]. Exploiting LoRa for localizing IoT devices over long distances has attracted increasing research interests [1, 11, 16, 18, 21, 32]. An important property of LoRa to enable remote sensing is its multipath and Doppler resistance [6, 30]. In addition, its low cost, low power, and good scalability make LoRa suitable for IoT localization.

While the research community has made significant advances in LoRa localization, many limitations exist. Existing LoRa localization algorithms rely on either calculating the Time Difference of Arrival (TDoA) [16, 32] or measuring the Received Signal Strength Indicator (RSSI) [1, 21]. TDoA-based approaches use the time differences obtained from different gateways to estimate node location. These approaches can only achieve sub-kilometer accuracy due to the limitation of the inherent time resolution of LoRa chips [14]. RSSI-based approaches can be further divided into two categories: 1) directly obtain the location by comparing the received RSSI values with an RSSI fingerprinting database (i.e., fingerprinting-based approaches [1, 10, 35]); 2) first estimate the distance from the signal source with a known signal propagation model, and then calculate the location using trilateration algorithms (i.e., model-based approaches [6, 12, 21, 22]). Fingerprinting-based approaches can usually achieve good accuracy. However, they are labor-intensive since they require a site survey to build a database, and update the database regularly to reflect changes in dynamic environments. As for the model-based approaches, one big issue is that they have not considered complex signal fading of different land-cover types (e.g., buildings, trees, and water) along LoRa link.

Signal fading on LoRa link may vary depending on different land-cover types. From our benchmark experiments in Section 3.2, we observe that different land-cover types indeed cause significantly different path loss to LoRa signal. The actual path loss of a LoRa link may deviate from its theoretical value at a scale of tens of dBs. This motivates us to investigate how to accurately quantify path loss with respect to various land-cover types in the real world. If we can obtain the actual path loss in any location, we can then use it as a

978-1-7281-5497-8/20/$31.00 ©2020 IEEE
DOI 10.1109/IPSN48710.2020.00-50

virtual fingerprint and further build a virtual fingerprint database for localization.

In this paper, we present SateLoc, a novel LoRa **Loc**alization system based on virtual fingerprints obtained from **Sate**llite images. Specifically, SateLoc first utilizes the high-resolution satellite image to identify the land-cover type of each pixel in the image. SateLoc then automatically estimates the path loss of an arbitrary LoRa link with the path loss parameters of the traversed land-cover types. We face two technical challenges in quantifying the path loss. The first challenge is to extract the LoRa link of an arbitrary slope from the image and further divide it into segments, each of which belongs to the same land-cover type. The second challenge is how to combine the path losses of these segments with their lengths and the corresponding path loss parameters. A simple sum-up of the path losses of all segments may not work because these segments are extracted at different distances from the gateway and will have different impacts on the overall path loss. To address the above challenges, we propose an algorithm that can automatically segment an arbitrary LoRa link and calculate its overall path loss (see details in Algorithm 1). For each gateway, SateLoc will generate a virtual fingerprinting map, which consists of the path loss between each pixel in the satellite image and the gateway. Obviously, this virtual fingerprinting map can be generated with little human effort. More importantly, SateLoc uses the Expected Signal Power (ESP) instead of RSSI for fingerprinting since ESP characterizes signal energy even when the signal power is below the noise floor after a long propagation or traveling through walls. Subsequently, we can combine the virtual fingerprinting maps of multiple gateways to get the final location. However, in reality, the measurements of different gateways may be suffered from different environment interference, making it inappropriate to utilize these measurements equally. To adaptively combining the fingerprinting maps of gateways, we propose a weighted combination strategy, which takes the environment interference of LoRa links as the weight. Finally, SateLoc will produce a joint location likelihood distribution to localize a LoRa node.

We implement SateLoc with commercial LoRa devices without any hardware modification. We evaluate its performance in a 350×650 m^2 urban area. Experimental results show that SateLoc achieves a median localization error of 47.1m, which is improved by 53.0% and 49.1%, respectively, compared to two state-of-the-art model-based approaches (i.e., standard log-normal model [6, 21, 22] and Okumura-Hata model [12]). The localization accuracy of SateLoc is also comparable to existing fingerprinting-based approaches [1], and importantly we do not require any labor-intensive fingerprint acquisition process.

In summary, we make the following key contributions:

- We propose SateLoc, which utilizes satellite images to achieve accurate LoRa localization with little human effort. SateLoc can automatically extract the land-cover information to generate virtual fingerprinting maps (i.e., ESP maps) for localization.
- We empirically analyze the model parameters in different land-cover types. With these parameters, we can adaptively get a more accurate path loss of an arbitrary LoRa link for fingerprinting.
- We propose a multi-gateway combination strategy, which quantifies the environment interference of each gateway as its weight, to jointly estimate node location.
- We implement SateLoc with commercial LoRa devices and conduct comprehensive experiments in the field. Results show that SateLoc achieves a median localization error of 47.1m in a 227,500m^2 area.

The rest of the paper is organized as follows. Section 2 reviews the related work of LoRa localization. Section 3 introduces the background and motivation of SateLoc. Section 4 describes an overview. In Section 5, we show how to automatically extract land-cover information. We present the ESP map extraction process and location estimation process in Section 6 and Section 7, respectively. We give the implementation details and evaluation results of SateLoc in Section 8. SateLoc is discussed in Section 9, and finally, Section 10 concludes this paper.

2 RELATED WORK

The related work of LoRa localization can be mainly divided into the following two categories: 1) TDoA-based localization, 2) RSSI-based localization.

2.1 TDoA-based Localization

TDoA-based approaches localize a LoRa node by the time differences when the same signal arrives at multiple gateways. In [16], a LoRaWAN network with TDoA capabilities was deployed to perform localization for stationary nodes. Carvalho et al. [8] evaluate the feasibility of using LoRaWAN for mobile applications. Podevijn et al. [32] further implement a TDoA system to evaluate the tracking performance in mobile scenarios.

However, TDoA-based approaches usually have poor localization performance because, in the current commodity LoRa hardware implementation, the resolution a timestamp can reach is one microsecond (μs) only [14, 18]. Such resolution can easily result in several hundred meters error in accuracy. In addition, these approaches require strict clock synchronization among gateways, making it more difficult for real deployment in the field.

A recent work μLocate in [27] designs a dedicated multi-band LoRa backscatter system to achieve a meter-level localization accuracy in a multi-room environment. In contrast, SateLoc aims to achieve accurate remote localization with off-the-shelf LoRa devices.

2.2 RSSI-based Localization

RSSI-based approaches are usually very efficient for remote LoRa localization with low cost since RSSI values are widely available in commodity LoRa devices.

Prior model-based approaches use RSSI measurements to estimate the distance from node to gateway using propagation models, which characterize the relationship between path loss and distance. Bor et al. [6] and Petajajarvi et al. [31] employ the standard log-normal model to describe path loss. Lam et al. [21, 22] further propose algorithms based on the log-normal model to eliminate noise interference for outdoor localization. Recently, Demetri et al. [12] first identify the type of surrounding environments (*urban* or

suburban) and then use the corresponding equation of the Okumura-Hata model to characterize path loss. The main drawback in these approaches is that most signal propagation models work better in theory, but work poorly in reality due to the complex signal fading. Specifically, they directly use unified models to characterize the path loss of a whole LoRa link and do not consider different land-cover types along the link, introducing deviation to the estimation of the overall path loss. Improving localization accuracy with existing propagation models remains a challenging task.

Fingerprinting-based approaches work more effectively in reality. In these approaches, RSSI values of known locations are manually collected into a database in an offline training phase, and then RSSI measurements will be compared with the database for online localization. Machine learning techniques, e.g., Bayesian inference method [10, 23], SVM [35] and k-Nearest-Neighbor (kNN) [1], are typically applied on collected RSSI measurements to estimate node location. Although fingerprinting-based approaches can be more accurate than model-based techniques, the construction of an effective fingerprint database usually requires a significant amount of manual efforts, especially for large-scale scenarios. Moreover, regular updates are essentially required to keep fingerprints up to date.

In SateLoc, we aim to combine the strength of the aforementioned two approaches, and on the other hand, address their drawbacks. Specifically, we utilize the land-cover information to enhance the existing propagation models to be more realistic and adaptive to real environments. With this enhanced model, we can then generate a more effective virtual map for fingerprint matching with little human effort.

3 BACKGROUND & MOTIVATION

In this section, we first give the background of LoRa, and then describe the motivation of using land-cover information to improve localization performance.

3.1 LoRa Background

LoRa is a physical layer technique that operates in sub-GHz ISM bands. LoRa adopts Chirp Spread Spectrum (CSS) modulation, in which a chirp is encoded using a linear variation of frequency over time. Different chirps ("0" and "1") differ from each other in the initial frequency. A typical LoRa system consists of LoRa gateways and LoRa nodes. Given a central frequency f_c, the frequency of a transmitted up chirp increases from $f_c - BW/2$ to $f_c + BW/2$. At the receiver side, an inverse down chirp whose frequency decreases from $f_c + BW/2$ to $f_c - BW/2$ can be utilized to decode the chirp. A LoRa receiver can decode LoRa transmissions up to 20dB below the noise floor, enabling very long communication distances. In addition, as LoRa uses its entire allocated bandwidth to transmit data, it is more robust to channel noise, multipath and Doppler effects. Therefore, LoRa can play an important role in remote localization.

3.2 Motivation

In practice, the deployment of LoRa may not be in *open space* only, but most likely in a combination of open space, forest, rangeland and etc. In this case, the Friis' law [12] cannot be applied to model its propagation fading. When traversing different land-cover types, LoRa signal propagation can be complicated and may exhibit different behaviors.

Table 1: Five representative land-cover types.

Type	Description
BUILT-UP	Built-up areas with human artifacts
FOREST	Trees
FIELD	Open space, Farms
WATER	Rivers, oceans, lakes
RANGELAND	Green land, grassland

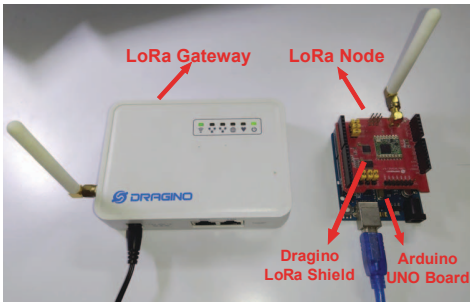


Figure 1: The commercial Dragino LG01 gateway and LoRa node. The LoRa node consists of a Dragino LoRa shield and an Arduino UNO microcontroller board [2].

We intend to study the influence of different land-cover types on LoRa link. We first use the log-distance path loss model [34] to describe the large-scale path loss $PL(d)$ in dB:

$$PL(d) = \overline{PL(d_0)} + 10n\log(\frac{d}{d_0}) + N_\sigma, \quad (1)$$

where d is the distance between LoRa node and gateway, $\overline{PL(d_0)}$ is the mean path loss at a known reference distance d_0, n is the path loss exponent of the corresponding environment, and N_σ is a zero-mean Gaussian random variable with a standard deviation of σ. The key advantage of this model is that it can model the path loss well in different environments with appropriate path loss parameters n and σ.

To present the impact of different land-cover types quantitatively, we conduct the following field experiments to identify the path loss parameters of each land-cover type.

Setup. In this experiment, we select five representative land-cover types as illustrated in Table 1. Each type may have different obstacles which cause different effects on LoRa signal propagation. Note that these types can be easily identified using multi-spectral satellite images [25].

We deploy a commercial Dragino LG01 gateway and a Dragino LoRa shield [15], as shown in Figure 1. Omnidirectional antennas are attached to both the gateway and the node. Besides, they are embedded with standard Semtech SX1278 LoRa RF-fronts [36]. In all land-cover types, both the gateway and the node are placed at 1m above the ground. In the case of the water type, the gateway and the node are placed at 1m above the water surface. The LoRa

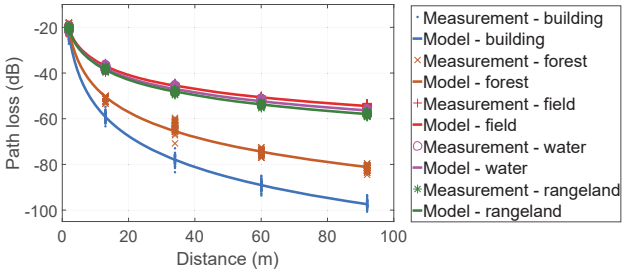


Figure 2: Path loss vs. distance in the five land-cover types. Dots are the measurements and solid lines show the theoretical values calculated from Equation 1.

node transmits packets periodically with the parameters configured as follows: Spreading Factor SF=12 [1], BandWidth BW=125 kHz, Coding Rate CR=4/8, and Transmission Power TP=13 dBm. We use a common reference distance $d_0 = 1$m to obtain a stable and accurate path loss reference $\overline{PL(d_0)}$ [17, 34].

Dataset. In each land-cover type, we deploy the LoRa node at five different distances to the gateway: 2m, 13m, 34m, 60m, and 92m, for packet collection. We set the maximum distance to 92m because a few walls in a built-up area or trees in a forest can make LoRa link disappear within 100m [6, 24]. In each location, we collect 50 packets, and record the RSSI and SNR of each packet for later analysis. In total, 1250 packets (50 packets $\times$ 5 locations $\times$ 5 types) are collected for parameter study.

Parameter Study. While LoRa can receive transmission with a signal power that is up to 20 dB below the noise floor, the common RSSI cannot capture the path loss of these extremely weak signals. LoRa uses the ESP to obtain the energy of the signal [12]:

$$ESP = RSSI + SNR - 10\log(1 + 10^{0.1SNR}). \quad (2)$$

For each packet, we use its SNR and RSSI to calculate the ESP and further the path loss [2]. With sufficient packets collected at the reference distance d_0, we can first obtain the reference path loss $\overline{PL(d_0)}$ of each land-cover type. Then for each type, we use the curve fitting function in MatLab to obtain its essential path loss parameters n and σ. Figure 2 shows both measured and theoretical path loss at different distances in the five land-cover types. Each fitting curve is represented with a unique color of the corresponding measurements. We use the corresponding n and set $\sigma = 0$ to generate these curves. As seen, the log-normal model can well capture the characteristics of LoRa links in different land-cover types. Taking a closer look at the figure, different types can lead to significantly different path loss. Both *built-up* and *forest* types will cause a relatively more severe path loss than the other types. As these land-cover types are commonly seen in an urban area, hence the path loss differences among these types cannot be ignored when estimating the path loss of a LoRa link.

[1] SF is the ratio between the symbol rate and chip rate and takes a value between 7 to 12. A higher value of SF results in a longer time for each symbol transmission and yields a longer communication range.

[2] To achieve the conversion between *ESP* and *path loss*, we have pre-measured the ESP value ESP_0 when the transceivers are placed together. Thus, we have: $PathLoss = ESP_0 - ESP$. In the following, these two terms can be used interchangeably.

Table 2: Reference path loss $\overline{PL(d_0)}$, path loss exponent n, standard deviation σ, and the average model fitting error in different land-cover types.

Type	$\overline{PL(d_0)}$	n	σ	Error (dB)
Built-up	23.091	4.499	1.887	2.351
Forest	21.254	3.616	0.854	1.531
Field	20.478	2.056	0.569	0.818
Water	20.709	2.158	0.599	0.882
Rangeland	20.392	2.274	0.622	0.987

Table 2 shows the reference path loss $\overline{PL(d_0)}$, path loss parameters n and σ of all land-cover types we have measured. The table also shows the average error between the modeled value and the measurements. As seen, path loss exponent n is larger in a more complex land-cover type, which contains more Non-Line-of-Sight (NLoS) paths and thus more severe fading. Standard deviation σ and the fitting error both depend on the path loss variation due to dynamic obstacles (e.g., pedestrians, vehicles) and static obstacles (e.g., buildings, trees) across the link [34]. Therefore, these two values are relatively higher in the *built-up* and *forest* types with more obstacles.

This experiment shows that different land-cover types can result in significantly different path loss effects. This strongly motivates us to utilize the land-cover information to improve LoRa localization accuracy. Our idea is first to identify the land-cover types from a satellite image, and then automatically generate an ESP map for localization.

4 OVERVIEW OF SATELOC

Figure 3 gives an overview of SateLoc. SateLoc integrates land-cover information from satellite images and location and signal information from gateways to localize a LoRa node. The workflow of SateLoc is described as follows.

(1) With the training image set, SateLoc first trains a Random Forest (RF) model, including a spectral feature extraction process and a parameter optimization process.

(2) When the RF model is trained, it can be applied to generate a land-cover map according to the satellite image of the testing area, which covers both the gateways and the node to be localized.

(3) With the land-cover map, SateLoc further produces an ESP map for each gateway based on its location in the testing area. The ESP map essentially serves as a virtual fingerprinting map.

(4) Every time when a LoRa node to be localized transmits a packet, SateLoc records all surrounding gateways that receive the packet. Using the signal information (i.e., RSSI and SNR) extracted from the LoRa packets, SateLoc can produce a location likelihood distribution for each gateway based on its ESP map.

(5) SateLoc uses a weighted combination strategy to combine the likelihood distributions of all gateways. Finally, a joint location likelihood distribution can be generated to determine node location.

The first three steps are offline processes that need to be performed only once. The last two online localization steps will be performed each time the LoRa signal transmitted to the gateways.

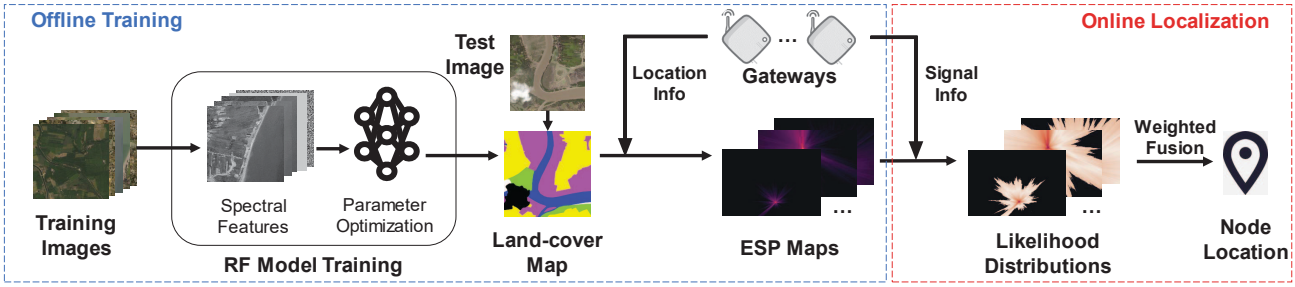


Figure 3: Overview of SateLoc and the workflow.

5 LAND-COVER CLASSIFICATION

In this section, we show how to identify land-cover types with the high-resolution satellite image datasets provided via DeepGlobe [13]. SateLoc uses Random Forest (RF) for the multi-class segmentation task. We select RF as our classification approach due to its capabilities of dealing with such large-scale datasets, modeling complex relationships among multiple image features, and avoiding overfitting.

5.1 Building an RF Model

The inputs to train the RF model are the 803 satellite images associated with labeled land-cover types at a spatial resolution of 50cm in the DeepGlobe dataset. We extract a feature vector for each pixel in satellite images. Each feature vector contains five typical spectral features, including the three raw RGB values, the Local Binary Patterns (LBP) [29], and the Normalized Different Vegetation Index (NDVI) [26]. During preprocessing, we first extract the same number of samples (i.e., 16,800 samples) for each land-cover type to eliminate the imbalance of sample sizes. Then we standardize all features to zero mean and unit variance to accelerate the convergence of the RF model.

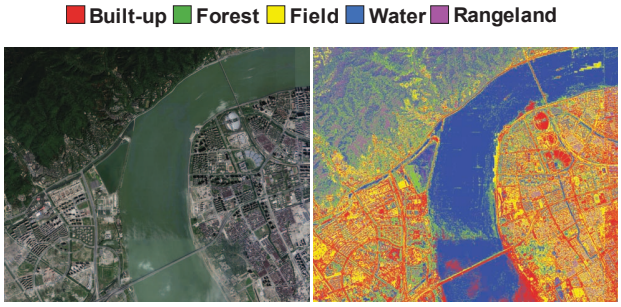
We use 70% of the feature vectors and their corresponding labels as the training set and the rest as the test set. We perform a grid-based search on the training set to automatically find the optimal hyper-parameters with a five-fold cross-validation strategy to maximize the classification accuracy. Specifically, we test the following key parameters in RF: decision tree number in [50,500] in increments of 50, max features per tree in [1,5] in increments of 1, and min samples split in [2,512] in exponentially increments. Finally, we find that RF performs the best when the decision tree number is equal to 400, max features per tree are equal to 1, and min samples split is equal to 32. We use the above configuration to construct the final RF model, which can be applied to the satellite image of the area of interest to predict the land-cover type of each pixel and further output a land-cover map.

5.2 Predicting Land-cover Type

To evaluate classification accuracy, we assess the experimental results with Overall Accuracy (OA), precision and recall. OA is the percentage of correctly classified pixels of all pixels, while precision and recall are at a granularity of each type. The values of OA, precision, and recall are in the range of 0 to 1, and the higher value indicates better classification performance.

Table 3: Classification accuracy of the RF model.

Type	Built-up	Forest	Field	Water	Rangeland
Precision	0.57	0.46	0.44	0.68	0.39
Recall	0.52	0.56	0.49	0.66	0.31
OA	0.51				



(a) Satellite image (b) Land-cover map

Figure 4: An example of land-cover type classification: (a) the ground truth satellite image in RGB, and (b) the corresponding land-cover map generated by SateLoc in a 5×5 km^2 urban area of Hangzhou, China.

Table 3 shows the classification accuracy of the trained RF model. The OA is 0.51. This is because the labels provided in the dataset are far from perfect [13]. In the dataset, many masks ignore terrain details and small structures not annotated in the ground truth. Incomplete and often inaccurate labeling presents a significant barrier for model development and evaluation. However, compared to the baseline CNN-based approach proposed in [13] with an OA of 0.433, our RF model still improves the accuracy by 17.8%.

Figure 4(a) shows an example satellite image of a portion of urban areas in Hangzhou, China. The image is directly sampled from Google Earth and its corresponding land-cover map is shown in Figure 4(b). We see that the classification and segmentation quality is good enough, except for some uncertainty between water (blue) and forest (green). However, these misclassified pixels are sparsely distributed with each covers an area of 0.6×0.6 m^2. These pixels are a small number of the long LoRa link and the induced path loss error is still acceptable. In addition, the land-cover map has a property that

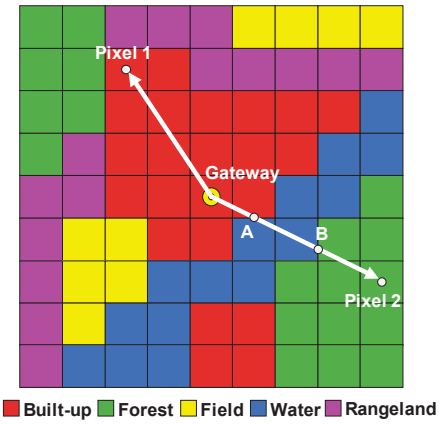


Figure 5: Illustration of ESP map generation with an example land-cover map. LoRa node at Pixel 1 transmits its packets in the same land-cover type (i.e. built-up) while node at Pixel 2 needs to transmit across multiple types.

each land-cover area is composed of a number of clustered pixels with clear boundaries. This property helps speed up the generation of the corresponding ESP map, which will be described in the next section.

6 GENERATING ESP MAPS

In this section, we describe how to utilize the land-cover map to generate the virtual fingerprinting map, i.e., ESP map, for each gateway.

Figure 5 shows an example land-cover map and each square represents a pixel in the map. We aim to generate the ESP map for the gateway (shown as the yellow circle). The basic idea of ESP map generation is to first select the pixels on the wireless link between the LoRa gateway and each candidate location, and then calculate the path loss with the land-cover types of these pixels. Take Pixel 1 as an example, we first connect it with the gateway using a straight line. Then we extract the coordinates and label of each pixel on the line. Since these pixels all belong to the *built-up* type, SateLoc directly uses the path loss model in Equation 1, along with the distance $d_{(Gateway,Pixel_1)}$ and the corresponding path loss exponent $n_{built-up}$, to calculate the ESP value at Pixel 1. It is worth noting that we set $\sigma_{built-up} = 0$ to eliminate the randomness of the generated fingerprinting maps.

In a more complicated case shown in Figure 5, the link between Pixel 2 and the LoRa gateway goes through several land-cover types. There will be two intersections A and B at the boundaries between different types. Starting from the gateway, we first calculate the path loss $PL_{(Gateway,A)}$ on segment $(Gateway,A)$ with Equation 1. For the next step, we replace the reference distance d_0 with the distance $d_{(Gateway,A)}$ between gateway and intersection A in the path loss model. The reference path loss $\overline{PL(d_0)}$ will also be replaced by $PL_{(Gateway,A)}$ correspondingly. Now we can calculate the path loss $PL_{(A,B)}$ on segment (A,B) with the updated path loss model and the path loss exponent n_{water}. We repeat the above steps until we obtain the path loss of the last segment, i.e., segment $(B,Pixel_2)$, of the link.

Algorithm 1 ESP map generation algorithm

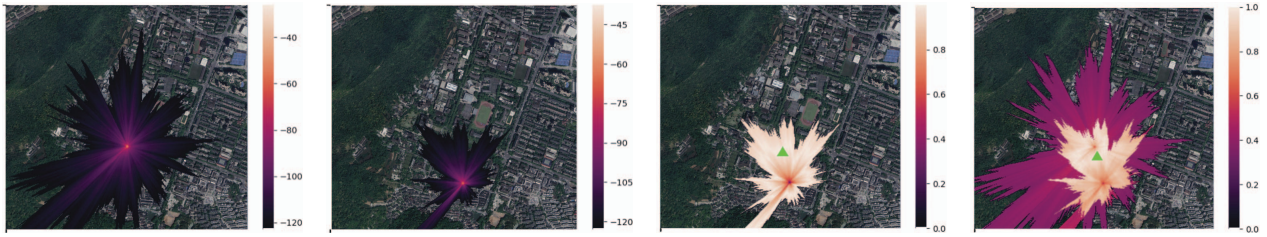
Input: Land-cover map LM with $X \times Y$ pixels; Gateway location (x_G, y_G); Basic ESP value ESP_0; Path loss exponents $n's$ of five land-cover types; Receiver sensitivity S; Pixel resolution R; The reference path loss $\overline{PL(d_0)}$ (set with Table 2 according to the land-cover type of the pixel where the gateway is placed) at the reference distance d_0

Output: ESP map $EM = \{EM(1,1),...,EM(X,Y)\}$

for i = 1 to X **do**
  for j = 1 to Y **do**
    // Get the pixel array P that consists of the coordinates $[x,y]$ and land-cover types c of all pixels on link $((x_G,y_G),(i,j)]$. The $getTraversedPixels$ function is designed based on the Bresenham's algorithm.
    $P = getTraversedPixels(x_G, y_G, i, j, LM)$
    // Initialize the link length L and the path loss PL using the first pixel P_1 in P. P_{1x}, P_{1y}, P_{1c} are the attributes x, y, c of P_1, respectively.
    $x_{end} = P_{1x}$, $y_{end} = P_{1y}$, $n = n_{P_{1c}}$
    $D = R\sqrt{(x_{end} - x_G)^2 + (y_{end} - y_G)^2}$
    $PL = \overline{PL(d_0)} + 10n \log \frac{D}{d_0}$
    $L = D$, $x_{start} = P_{1x}$, $y_{start} = P_{1y}$
    // Search for the segments whose pixels belong to the same type, and iteratively calculate their path losses with their distances to the gateway. P_{kx}, P_{ky}, P_{kc} indicate the attributes x, y, c of the k-th pixel in P, respectively. $|P|$ is the total pixel number of array P.
    for k = 2 to $|P| - 1$ **do**
      if $P_{kc} == P_{k+1c}$ **then**
        continue
      else
        $x_{end} = P_{kx}$, $y_{end} = P_{ky}$, $n = n_{P_{kc}}$
        $D = R\sqrt{(x_{end} - x_{start})^2 + (y_{end} - y_{start})^2}$
        $PL = PL + 10n \log \frac{L+D}{L}$
        $L = L + D$, $x_{start} = P_{kx}$, $y_{start} = P_{ky}$
    // Add the path loss of the last segment to get the overall path loss.
    $x_{end} = P_{|P|x}$, $y_{end} = P_{|P|y}$, $n = n_{P_{|P|c}}$
    $D = R\sqrt{(x_{end} - x_{start})^2 + (y_{end} - y_{start})^2}$
    $PL = PL + 10n \log \frac{L+D}{L}$
    $EM(i,j) = ESP_0 - PL$
    if $EM(i,j) < S$ **then**
      $EM(i,j) = -\infty$
return EM

After that, we sum up the path loss of all segments and get the ESP value at Pixel 2.

Algorithm 1 shows the details of our ESP map generation algorithm. For each pixel in the land-cover map, SateLoc first extracts all the traversed pixels from it to the gateway (not included) using a form of Bresenham's line drawing algorithm [7]. Then SateLoc divides these pixels into segments, each of which belongs to the same land-cover type. The path loss of each segment can be iteratively calculated with the corresponding path loss exponent and its distance to the gateway. Finally, SateLoc calculates the overall path loss and further the ESP value of the pixel. While pixels belonging to the same type are usually clustered in an area(illustrated in Section 5), there will be reasonable ESP calculations (each corresponds to a segment) for each pixel. Therefore, the overall map generation cost (including time cost and computing resource overhead) will be acceptable.



(a) ESP map in an area with sparse buildings nearby (in dBm) **(b) ESP map in an area with dense buildings nearby (in dBm)** **(c) Likelihood distribution** **(d) Joint likelihood distribution**

Figure 6: ESP to location in a 1.5×1.4 km^2 area of Yuquan Campus, Zhejiang University: (a) An example ESP map of $gateway_1$ placed in an area with sparse buildings. (b) An example ESP map of $gateway_2$ placed in an area with dense buildings. (c) An example likelihood distribution of $gateway_2$ when receiving an ESP value of ≈-116 dBm from the LoRa node. The while pixels show the most possible locations estimated using Equation 3. (d) The joint likelihood distribution of the LoRa node produced using Equation 6 when $gateway_1$ receives an ESP value of ≈-88 dBm and $gateway_2$ receives an ESP value of ≈-116 dBm. A brighter pixel indicates a higher likelihood. The green triangle marks the actual location of the node.

To make the map more accurate, we utilize the receiver sensitivity S of the gateway. Specifically, pixels with an ESP value less than S will not receive the LoRa signal from the gateway. Therefore, for pixels whose ESP value less than S, we just assign $-\infty$ to them in the ESP map. This will also help significantly reduce the computation overhead in the following likelihood distribution generation process. Taken from [36], we set $S = -136$ dBm according to the LoRa configuration in our experiments. Figure 6(a) and (b) show two example ESP maps over the 2D space when gateways are placed in Yuquan Campus, Zhejiang University with sparse and dense buildings nearby, respectively. As seen, the path loss does not spread uniformly from the gateway. The area with dense buildings nearby suffers from more severe signal fading and the corresponding LoRa coverage degrades more significantly.

7 FROM ESP MAP TO LOCATION

7.1 ESP Map to Likelihood Distribution

We have obtained the ESP map for each gateway, we now present our algorithm to infer the location of the node. As we record the RSSI and SNR of each transmitted packet from a LoRa node, the corresponding ESP_{node} can be calculated using Equation 2. Assuming there are M gateways in the testing area that may receive the LoRa signal, which means there are M ESP maps for localization. The likelihood $L_{m,i}$ of the node locating at the i-th pixel of the m-th ESP map can be computed as:

$$L_{m,i} = 1 - \frac{|\overline{ESP_{node}} - ESP_{m,i}|}{\max(|\overline{ESP_{node}} - ESP_m|) - \min(|\overline{ESP_{node}} - ESP_m|)}, \quad (3)$$

where $\overline{ESP_{node}}$ is the mean ESP value measured from the LoRa node. $ESP_{m,i}$ is the ESP value at the i-th pixel in the m-th ESP map. The denominator calculates the value range of ESP differences in the m-th ESP map, and normalizes the likelihood. For pixels with an ESP value of $-\infty$, we directly set their likelihoods as 0. Figure 6(c) shows an example likelihood distribution of $gateway_2$ over the 2D space when receiving a LoRa packet with an ESP value of ≈-116 dBm in our experiments. As seen, the shape of possible locations is not a circle since the LoRa links are affected by different land-cover types.

7.2 Likelihood Distribution Fusion

Figure 6(c) also shows that the pixel with the maximum likelihood can be far away from the ground truth when using only one gateway to localize LoRa node. To get a more accurate location, SateLoc usually requires at least two gateways (i.e., $M \geq 2$). The more gateways used, the higher the localization accuracy. In practice, however, there may not be many LoRa gateways covering the same area. To make a trade-off between deployment cost and localization accuracy, we set $M = 3$ in our experiments. The detailed performance with varying numbers of gateways will be evaluated in the evaluation.

A straightforward way to get the joint likelihood distribution is to equally add the likelihood obtained from all gateways. However, the performance of this method deteriorates due to different environment interference along different LoRa links. The reason is that each link contains various land-cover types, each of which has a specific level of environment interference. Based on Table 2, we use variance σ^2 to quantitatively represent the level of environment interference. This is because σ^2 can characterize the uncertainty of the path loss and can be directly summed up when the path loss is independent. A larger σ^2 indicates more severe environment interference. We hence develop a weighted combination algorithm to build the joint likelihood distribution and mitigate the effects of the environment interference. The algorithm involves the following three steps:

First, we estimate the environment interference of each pixel in the M likelihood distributions (one for each gateway). The environment interference $\Gamma_{m,i}$ of the LoRa link between the m-th gateway $Gateway_m$ and the i-th pixel $Pixel_i$ can be obtained by:

$$\Gamma_{m,i} = \sum_{k=1}^{|P_{m,i}|} \sigma_k^2, \quad (4)$$

where $P_{m,i}$ is the pixel array on link $(Gateway_m, Pixel_i)$, $|P_{m,i}|$ is the total pixel number in $P_{m,i}$, and σ_k^2 is the interference value at the k-th pixel in $P_{m,i}$. Obviously, a larger Γ indicates that we should assess less weight on the corresponding link.

Second, we calculate the corresponding weight $w_{m,i}$ at the i-th pixel of the m-th likelihood distribution:

$$w_{m,i} = \frac{\frac{\prod_{\theta=1}^{M} \Gamma_{\theta,i}}{\Gamma_{m,i}}}{\sum_{\theta=1}^{M} \frac{\prod_{\theta=1}^{M} \Gamma_{\theta,i}}{\Gamma_{m,i}}}. \quad (5)$$

For a specific pixel $Pixel_i$, a relatively larger environment interference $\Gamma_{m,i}$ among the M links (one for each gateway) will lead to a lower weight $w_{m,i}$. In addition, Equation 5 normalizes each weight $w_{m,i}$ to a range of 0 to 1. For each pixel, the sum of the M gateways' weights equals to 1.

Finally, SateLoc calculates the weighted sum of M likelihood distributions to form the joint likelihood distribution [3]. SateLoc just picks the pixel with the highest likelihood as the final location:

$$Location = \arg\max_{i} \sum_{m=1}^{M} w_{m,i} L_{m,i}. \quad (6)$$

Note that the first two steps are offline performed after getting the traversed pixels in Algorithm 1, while the last step needs to be performed in the online localization phase. Figure 6(d) shows an example joint likelihood distribution when $gateway_1$ receives an ESP value of ≈-88 dBm and $gateway_2$ receives an ESP value of ≈-116 dBm.

8 EVALUATION

In this section, we first present the implementation of SateLoc. We then describe our evaluation and present the experimental results in both localization accuracy and latency.

8.1 Implementation

We set the LoRa gateway shown in Figure 1 as the receiver, and the LoRa node as the transmitter. The RadioHead library [33] provides a LoRa communication library for embedded microprocessors and is installed in both the gateway and node. The transceivers use one channel in the 433 MHz band for transmitting or receiving LoRa signals. We carefully study the typical LoRa configuration parameters: TP, SF, BW, and CR. To ensure the LoRa transmission range and its resilience to noise, we select a set of parameters, i.e., TP = 13 dBm, SF = 12, BW = 125 kHz, CR = 4/8 in our experiments. Detailed localization performance of different parameters will also be evaluated. The node consistently broadcasts LoRa packets with unique IDs to all gateways at a period of five seconds. Note that we set such a short time period just for experimental analysis. In practice, the period can be much longer to provide a better lifetime. Each gateway is connected to a laptop through Ethernet to monitor and collect LoRa packets. For packets collected at gateways with the same ID, we record their RSSIs and SNRs. The core localization estimation algorithms of SateLoc are implemented on a desktop with an Intel i7 6700 CPU and 8GB RAM.

Figure 7 shows the site deployment of SateLoc in Yuquan Campus, Zhejiang University. The test satellite image (Figure 7 left, a 1.5 × 1.4 km^2 area) used in our experiments is directly extracted from Google Earth with a spatial resolution of 0.6m. Three LoRa gateways and one LoRa node are placed in a 350 × 650 m^2 area on a college campus (Figure 7 right). The advertised

[3] M can be an arbitrary positive integer.

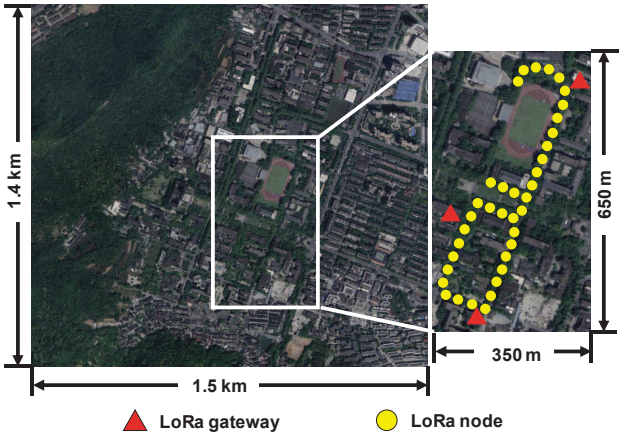


Figure 7: Deployment of SateLoc in Yuquan Campus, Zhejiang University. The left image is the overall satellite map (2816 × 3072 pixels) we used in SateLoc. The right image is the zoom-in map of a 350 × 650 m^2 urban area, which shows the deployment of the Dragino gateways (red triangles) and the measurement sites (yellow dots). The three gateways are placed at the 2D pixels of (1170, 2130), (1780, 1550), and (2240, 1660) in the satellite map.

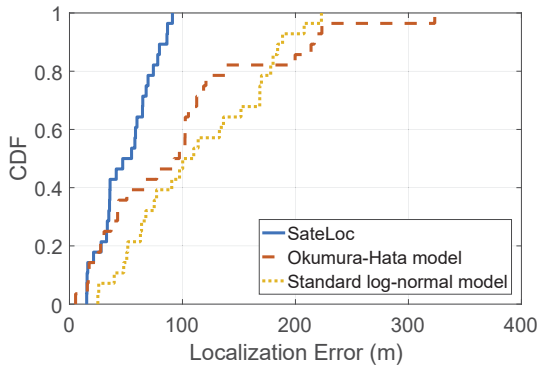
communication range of LoRa is expected to be more than 10km for suburban environments [24]. However, LoRa coverage can degrade significantly with the presence of obstacles in NLoS scenario. Our experimental area covers a large number of buildings and trees, and is full of NLoS paths. As already shown in Figure 6, the LoRa coverage in our experiments can be only a few hundred meters, which is consistent with the experimental results in [6, 20, 24, 28]. Therefore, it is reasonable to place three gateways in such complex outdoor scenario, which is also the deployment environments in urban areas, to maintain an acceptable LoRa link quality.

All LoRa gateways are placed at 10m above the ground in buildings. For each gateway, we record its location and pre-measure the path loss at the reference distance (i.e., 1m) for analysis. The LoRa node is powered by a 5V portable power supply, and is placed at a height of 1m when sending LoRa packets. We place the LoRa node at 38 test locations (shown in Figure 7) in the campus area. We manually label these locations in the form of 2D pixels in the satellite map as the ground truth. In each location, we collect the RSSIs and SNRs of 20 LoRa packets, which are aimed to calculate the average ESP to further mitigate the environment interference. We have also shown the performance with varying numbers of packets in our evaluation.

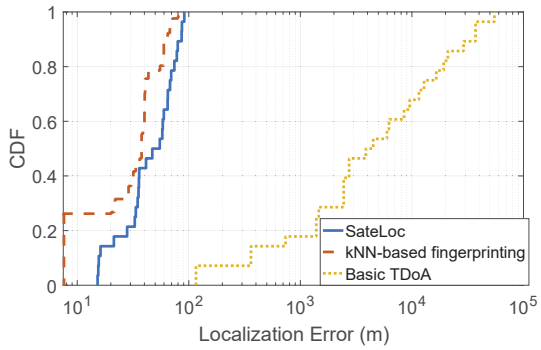
8.2 Performance Evaluation

As a target for precise and fast remote localization, the two most important metrics are used to testify SateLoc: accuracy and latency.

8.2.1 Accuracy. We first compare the performance of SateLoc with RSSI-based and TDoA-based approaches. Then, we show the effectiveness of the ESP map and our path loss model. Finally, we evaluate the effects of different LoRa configuration parameters.



(a) Comparison with model-based approaches



(b) Comparison with fingerprinting-based and TDoA-based approaches

Figure 8: Overall performance of SateLoc.

Comparison study. *Comparison with model-based approaches.* We compare the localization accuracy of SateLoc with two state-of-the-art approaches that use standard log-normal model [6, 21, 22] and Okumura-Hata model [12] to characterize the path loss. We use the parameters of the two models estimated in these works. In particular, the Okumura-Hata model is also related to the surrounding environment and consists of an urban equation and a suburban equation. We use its urban equation in the *built-up* and *forest* types and use the suburban equation elsewhere. Based on the estimated path losses, we employ a trilateration method to localize the node in these approaches. For a fair comparison, we use the same number of gateways (i.e., 3) and LoRa configuration parameters during benchmarking.

Figure 8(a) shows that the median localization error of SateLoc is 47.1m, while the median localization errors of the standard log-normal model and the Okumura-Hata model are 100.2m and 92.6m, respectively. Compared to the two model-based localization approaches, SateLoc improves the localization accuracy by 53.0% and 49.1%, respectively. Although there is still a gap compared to high-precision GPS, we believe that SateLoc can help achieve some prime applications such as parked vehicle tracking [11] with low power consumption. Specifically, the service provider may need to compute how much time each vehicle has been parked. GPS-based vehicle tracking is not widely deployed since GPS is power-hungry and requires a connection to the vehicle battery, which increases the complexity of vehicle installation [11]. In this case, the provider can use SateLoc to detect the start and end time of parking. Another appealing use case of LoRa localization is the anti-theft of a large number of stationary assets in construction sites, utility yards, airports, etc due to the low cost of LoRa nodes.

Comparison with fingerprinting-based approaches. We further compare the performance of SateLoc with a state-of-the-art fingerprinting-based approach [1], which uses kNN for localization. We randomly select 70% (i.e., 14) LoRa packets of each location as the training set and the rest as the test set. During training, we calculate the ESP value of each gateway in each location to construct the fingerprint database. In total, there are 532 fingerprints (14 packets $\times$ 38 locations), each consists of three ESP values of the corresponding gateways. We adopt Euclidean distance to measure fingerprint similarity. For each test packet, we extract candidate locations with the $k = 3$ smallest distances and use the centroid as its location. We get the final localization error of a test location by averaging the localization errors of packets collected there.

Figure 8(b) shows that SateLoc achieves comparable localization accuracy compared to the kNN-based approach (a median localization error of 37.8m). In addition, we find that the localization errors of the kNN-based approach are mainly distributed around 40m, which is roughly the distances between neighbor test locations in our experiments. This indicates that the accuracy of these fingerprinting-based approaches is highly dependent on the site deployment when collecting fingerprints. However, it is labor-intensive to construct and maintain an effective fingerprint database, especially in a large-scale area. Instead, SateLoc manages to remedy the defect using more effective virtual ESP fingerprints.

Comparison with TDoA-based approaches. We also make a comparison between SateLoc and a basic TDoA-based localization approach. As shown in Figure 8(b), TDoA-based approaches may not perform well due to the low clock precision of the low-cost commercial LoRa devices used in our experiments. Specifically, the clock error of our devices can reach up to tens of microseconds, leading to a localization error of thousands of meters. Two lessons learned from these comparisons to make good use of TDoA information are: 1) adding modules with high-precision clocks in the gateways, or 2) deploying relatively expensive devices with high clock precision.

Effectiveness of the ESP map and path loss model. The ESP map boundary of a gateway can be viewed as its signal coverage area. We move the LoRa node around the gateway and mark the furthest locations where LoRa signal can be received. Then we connect these locations to get the measured coverage boundary. Figure 9 shows the theoretical boundary of the ESP map and the measured LoRa coverage of a deployed gateway. As seen, the measured boundary can roughly match the automatically generated ESP map. Taking a closer look, the measured boundary is consistently smaller than the theoretical boundary. This is because the minimum measured ESP values (i.e., -123.02 dBm on average) are slightly above the theoretical ESP value (i.e., -136 dBm).

We also perform controlled experiments to validate our iterative aggregation path loss model and compare it with other state-of-the-art path loss models. We deploy the LoRa devices with different numbers of land-cover types in the middle and record the distances

Figure 9: Effectiveness of the ESP map of Yuquan Campus, Zhejiang University. The white lines show the measured signal coverage boundary.

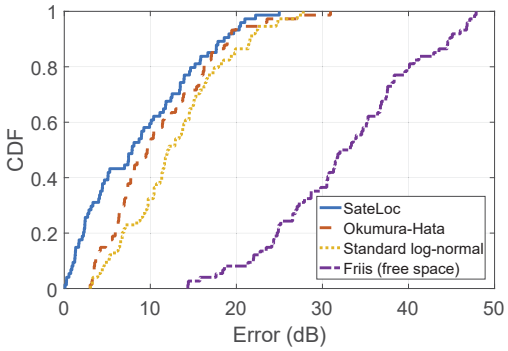


Figure 10: Path loss estimation errors of our iterative aggregation path loss model and existing models.

for each type. Note that it is difficult to get a complete ground truth ESP map since we need to take these samples site by site. Hence, we use per-sample values collected in our experiments as the ground truth. Figure 10 shows the detailed path loss estimation errors of SateLoc and other existing models in our experiments. As seen, SateLoc achieves a median path loss estimation error of 7.87 dB, reducing the error by 18.4%, 34.4% and, 75.4% compared to the Okumura-Hata model (i.e., 9.64 dB), standard log-normal model (i.e., 12.00 dB), and free space model (i.e., 32.05 dB), respectively. Results show that SateLoc can get more accurate path losses and further a more accurate ESP map with the novel map generation algorithm.

Impact factors. In the following, we evaluate the performance of SateLoc with different LoRa configurations by varying one of the parameters at a time during the packet collection process. We also evaluate the localization accuracy of SateLoc when using different numbers of gateways and packets, and in different weather conditions.

Impact of transmission power. Considering that different transmission power will lead to different transmission ranges, localization accuracy may be greatly affected. To evaluate the influence of transmission power, we set TP to: 13 dBm, 10 dBm, and 7 dBm, respectively. Figure 11(a) shows the localization error of different TPs. Results show that when the TP drops from 13 dBm to 7 dBm, the median localization error increases by 80.6%. Besides, the localization error in some locations can reach up to 300m. This is because some gateways cannot successfully receive packets from these locations with a limited transmission power, and thus cannot be used for localization.

Impact of spreading factor. Spreading factor can influence not only transmission range, but also the resistance to interference noise [3]. We evaluate the performance of SateLoc with three different SFs: 12, 10, 8. Figure 11(b) shows that the median localization errors are increased to 75.0m and 102.8m when the SFs are set to 10 and 8, respectively. Results show that a higher SF can significantly improve the localization performance of SateLoc, since it can reduce the interference of environmental noise.

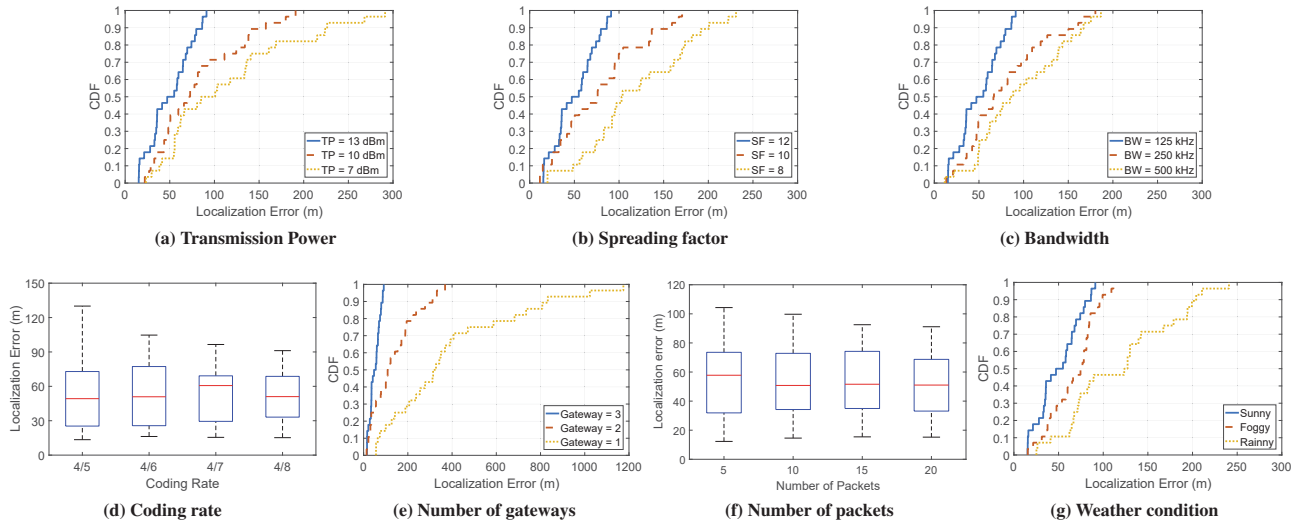
Impact of bandwidth. A higher bandwidth gives a higher data rate, but decreases the sensitivity due to additional noise [6]. Figure 11(c) shows the localization accuracy when the BWs are set to 125 kHz, 250 kHz, and 500 kHz, respectively. It is as expected that the localization errors increase with higher bandwidths. Compared to the 125 kHz BW, the localization errors of the 250 kHz BW and 500 kHz BW increase by 40.5% and 79.8%, respectively.

Impact of coding rate. Coding rate offers protection against the burstiness of interference and helps reduce the packet error rate. CR can be set to either 4/5, 4/6, 4/7 or 4/8. Figure 11(d) shows that the median localization errors of the four CRs are 48.4m, 47.8m, 59.6m, and 47.1m, respectively. Results show that SateLoc achieves consistent localization accuracy across different CRs. This is because changing the CR will not significantly influence both the RSSI and SNR of the received LoRa packets.

Impact of number of gateways. We also evaluate the performance of SateLoc when using the collected data from different numbers of gateways. As described in Section 7.2, we first use 1, 2, and 3 distributions (i.e., gateways) to form the joint likelihood distribution, respectively. Then we pick the pixel with the highest likelihood as the final location. The localization results are shown in Figure 11(e). As seen, a larger number of gateways will significantly decrease the localization error. Considering the practical coverage and the deployment overhead of LoRa gateways, we use 3 gateways in our experiments. We believe the localization accuracy can be further improved with more gateways.

Impact of number of packets. In our experiments, we collect 20 packets to calculate the average ESP values and localize the node around every two minutes (20 packets $\times$ 5 seconds/packet). In practice, we find that the ESP values of packets collected at the same location are stable. Besides, SateLoc can complete the online localization phase within three seconds (evaluated in Section 8.2.2). These indicate that we can use fewer packets to localize the node. Figure 11(f) shows that using fewer packets will not significantly decrease the localization accuracy. When the number of packets drops from 20 to 5, the median localization error only increases from 47.1m to 53.8m. For cases where response time is limited, SateLoc can still perform well using all available LoRa packets.

Impact of weather condition. Weather condition is one of the main influence factors in wireless communication. We collect experimental data on sunny, foggy, and rainy days. Figure 11(g) shows that bad weather conditions (i.e., more moisture in the air)



(a) Transmission Power (b) Spreading factor (c) Bandwidth

(d) Coding rate (e) Number of gateways (f) Number of packets (g) Weather condition

Figure 11: Localization accuracy of SateLoc with different factors.

can significantly reduce localization accuracy. This is because the path loss parameters in Table 2 are extracted on sunny days and cannot work well in other conditions. A straightforward solution is to extract the corresponding parameters in all weather conditions.

8.2.2 Latency. The two main phases that contribute to the latency of SateLoc in online localization are the likelihood distribution generation and the fusion phase.

In the generation phase, time delay mainly contributes to the likelihood calculation process. For each gateway, SateLoc takes about 5.85s to generate its likelihood distribution with the whole satellite image (2816 $\times$ 3072 pixels) we used in SateLoc. In the fusion phase, it takes around 0.38s to produce the joint likelihood distribution and get the maximum likelihood.

It is worth noting that time delay is highly relevant to image size. If we only use the pixels of the zoom-in deployment map in Figure 7, the generation time of each gateway can be further reduced to 0.75s and the fusion time can be reduced to 0.30s. Therefore, the overall time delay (0.75 s/gateway $\times$ 3 gateways + 0.30 s) of the online localization phase should be within 3s. In SateLoc, the time interval between adjacent packets is five seconds, which is sufficient for continuous localization. In a general LoRa network, the packet rate of a LoRa node can even be less than 30 packets/hour to prolong its lifetime [31]. As a result, SateLoc is compatible with existing LoRa networks.

9 DISCUSSION

In this section, we first discuss two possible ways to further improve the localization accuracy by improving land-cover classification accuracy and integrating TDoA information. Then we present a discussion on future directions of SateLoc in more environments and applications.

High-accuracy land-cover classification. Land-cover classification from high-resolution satellite images is known to be a challenging problem. The labels are also far from perfect due to the cost for annotating multi-class segmentation mask [13]. Quantifying the path loss of LoRa link accurately much depends on classification quality. While deep learning is a promising approach for accurate land-cover classification [19], we will investigate a more efficient deep learning approach in our future work. Besides, since the land-cover may change over time, it is necessary to update the test satellite image and re-classify the land-cover types periodically (although not frequently, such as updating every six months) to maintain a good localization accuracy.

Integrating TDoA information. Although SateLoc significantly reduces localization accuracy to around 50m, more efforts are required to apply this system for applications with high-accuracy localization requirements such as localizing goods in a warehouse. The localization accuracy can be further improved with multi-dimensional information available in off-the-shelf LoRa devices, such as TDoA. In our future work, we plan to investigate how to fuse the locations extracted from TDoA and RSSI.

More real-world evaluations. In this paper, we analyze the performance of SateLoc in an urban area of around 0.2 km^2. Our current evaluation includes land-cover types of buildings, fields, forest, and rangeland. In the future, we plan to deploy SateLoc in more environments to further evaluate the performance of SateLoc.

Large-scale tracking applications. SateLoc has the potential to help realize large-scale tracking using continuously estimated locations. However, tracking applications usually have higher requirements for system responsiveness. While SateLoc needs to calculate the likelihood of all pixels in a satellite image to estimate the start location, we can use the spatial information of adjacent locations to reduce the search space of subsequent locations. More importantly, the tracking accuracy can be further improved with road information extracted from satellite images.

10 CONCLUSION

In this paper, we propose SateLoc, which uses high-resolution satellite images for remote LoRa localization. We first analyze the path loss parameters in different land-cover types. With these parameters, we can capture the path loss of an arbitrary LoRa link. SateLoc consists of two phases: an offline training phase and an online localization phase. In the training phase, SateLoc trains an RF model for automatically generating a land-cover map according to the satellite image of the area of interest. Next, SateLoc produces a virtual fingerprinting map (i.e., ESP map) for each gateway without the labor-intensive fingerprint acquisition process. Then in the localization phase, SateLoc uses a novel weighted combination strategy to combine the fingerprinting maps of multiple gateways and output a joint location likelihood distribution for node localization. We implement SateLoc with commercial LoRa devices in a 227,500m^2 urban area. The experimental results show that SateLoc achieves a median localization error of 47.1m, while requiring little human effort.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers and the shepherd for their valuable comments and helpful suggestions. This work is supported by the National Science Foundation of China (No. 61772465 and No. 61872437), Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars under No. LR19F020001, and Australian Research Council (ARC) Discovery Project Grants, DP190101888 and DP180103932. Wei Dong is the corresponding author.

REFERENCES

[1] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn. Sigfox and LoRaWAN datasets for fingerprint localization in large urban and rural areas. *Data*, 3(2):13, 2018.
[2] Arduino. ARDUINO UNO REV3. https://store.arduino.cc/usa/arduino-uno-rev3, 2019.
[3] A. Augustin, J. Yi, T. Clausen, and W. Townsley. A study of LoRa: Long range & low power networks for the internet of things. *Sensors*, 16(9):1466, 2016.
[4] R. Ayyalasomayajula, D. Vasisht, and D. Bharadia. BLoc: CSI-based accurate localization for BLE tags. In *ACM CoNEXT*, pages 126–138. ACM, 2018.
[5] J. Blumenthal, R. Grossmann, F. Golatowski, and D. Timmermann. Weighted centroid localization in zigbee-based sensor networks. In *2007 IEEE international symposium on intelligent signal processing*, pages 1–6. IEEE, 2007.
[6] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso. Do LoRa low-power wide-area networks scale? In *ACM MSWiM*, pages 59–67. ACM, 2016.
[7] J. Bresenham. A linear algorithm for incremental digital display of circular arcs. *Communications of the ACM*, 20(2):100–106, 1977.
[8] D. F. Carvalho, A. Depari, P. Ferrari, A. Flammini, S. Rinaldi, and E. Sisinni. On the feasibility of mobile sensing and tracking applications based on lpwan. In *IEEE SAS*, pages 1–6. IEEE, 2018.
[9] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi. Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios. *IEEE Wireless Communications*, 23(5):60–67, 2016.
[10] W. Choi, Y.-S. Chang, Y. Jung, and J. Song. Low-Power LoRa Signal-Based Outdoor Positioning Using Fingerprint Algorithm. *ISPRS International Journal of Geo-Information*, 7(11):440, 2018.
[11] L. A. S. Committee et al. LoRaWAN Geolocation Whitepaper. *LoRa Alliance, January*, 2018.
[12] S. Demetri, M. Zúñiga, G. P. Picco, F. Kuipers, L. Bruzzone, and T. Telkamp. Automated Estimation of Link Quality for LoRa: A Remote Sensing Approach. In *ACM/IEEE IPSN*, 2019.
[13] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar. DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images. In *IEEE CVPR Workshops*, June 2018.
[14] A. Dongare, C. Hesling, K. Bhatia, A. Balanuta, R. L. Pereira, B. Iannucci, and A. Rowe. OpenChirp: A low-power wide-area networking architecture. In *IEEE PerCom Workshops*, pages 569–574. IEEE, 2017.
[15] Dragino. LoRa Wireless Node, Gateway. http://www.dragino.com/products/lora.html, 2017.
[16] B. C. Fargas and M. N. Petersen. GPS-free geolocation using LoRa in low-power WANs. In *IEEE GIoTS*, pages 1–6. IEEE, 2017.
[17] A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.
[18] C. Gu, L. Jiang, and R. Tan. LoRa-Based Localization: Opportunities and Challenges. *arXiv preprint arXiv:1812.11481*, 2018.
[19] F. Hu, G.-S. Xia, J. Hu, and L. Zhang. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11):14680–14707, 2015.
[20] S. Kartakis, B. D. Choudhary, A. D. Gluhak, L. Lambrinos, and J. A. McCann. Demystifying low-power wide-area communications for city IoT applications. In *ACM WiNTECH*, pages 2–8. ACM, 2016.
[21] K.-H. Lam, C.-C. Cheung, and W.-C. Lee. LoRa-based localization systems for noisy outdoor environment. In *IEEE WiMob*, pages 278–284. IEEE, 2017.
[22] K.-H. Lam, C.-C. Cheung, and W.-C. Lee. New rssi-based lora localization algorithms for very noisy outdoor environment. In *IEEE COMPSAC*, volume 2, pages 794–799. IEEE, 2018.
[23] Y. Li, Z. He, Y. Li, H. Xu, L. Pei, and Y. Zhang. Towards location enhanced IoT: Characterization of LoRa signal for wide area localization. In *IEEE UPINLBS*, pages 1–7. IEEE, 2018.
[24] J. Liando, A. Gamage, A. Tengourtius, and M. Li. Known and Unknown Facts of LoRa: Experiences from a Large-scale Measurement Study. In *ACM Transactions on Sensor Networks*, volume 15. ACM, 2019.
[25] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277–293, 2017.
[26] G. E. Meyer and J. C. Neto. Verification of color vegetation indices for automated crop imaging applications. *Computers and electronics in agriculture*, 63(2):282–293, 2008.
[27] R. Nandakumar, V. Iyer, and S. Gollakota. 3D Localization for Sub-Centimeter Sized Devices. In *ACM SenSys*, pages 108–119. ACM, 2018.
[28] P. Neumann, J. Montavont, and T. Noël. Indoor deployment of low-power wide area networks (LPWAN): A LoRaWAN case study. In *IEEE IPSN WiMob*, pages 1–8. IEEE, 2016.
[29] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):971–987, 2002.
[30] Y. Peng, L. Shangguan, Y. Hu, Y. Qian, X. Lin, X. Chen, D. Fang, and K. Jamieson. PLoRa: A passive long-range data network from ambient LoRa transmissions. In *ACM SIGCOMM*, pages 147–160. ACM, 2018.
[31] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo. On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology. In *IEEE ITST)*, pages 55–59. IEEE, 2015.
[32] N. Podevijn, D. Plets, J. Trogh, L. Martens, P. Suanet, K. Hendrikse, and W. Joseph. TDoA-based outdoor positioning with tracking algorithm in a public LoRa network. *WCMC*, 2018, 2018.
[33] RadioHead. RadioHead Packet Radio library for embedded microprocessors. https://github.com/dragino/RadioHead, 2019.
[34] T. S. Rappaport et al. *Wireless communications: principles and practice*, volume 2. prentice hall PTR New Jersey, 1996.
[35] H. Sallouha, A. Chiumento, and S. Pollin. Localization in long-range ultra narrow band IoT networks using RSSI. In *IEEE ICC*, pages 1–6. IEEE, 2017.
[36] Semtech. Semtech SX1278. https://www.semtech.com/uploads/documents/DS_SX1276-7-8-9_W_APP_V6.pdf, 2019.
[37] Statista. Internet of Things - number of connected devices worldwide 2015-2025. https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/#statisticContainer, 2019.
[38] M. Sugano, T. Kawazoe, Y. Ohta, and M. Murata. Indoor Localization System using RSSI Measurement of Wireless Sensor Network based on ZigBee Standard. *Wireless and Optical Communications*, 538, 2006.
[39] J. Wang, N. Tan, J. Luo, and S. J. Pan. WOLoc: WiFi-only outdoor localization using crowdsensed hotspot labels. In *IEEE INFOCOM*, pages 1–9. IEEE, 2017.
[40] C. Wu, J. Xu, Z. Yang, N. D. Lane, and Z. Yin. Gain without pain: Accurate WiFi-based localization using fingerprint spatial gradient. *ACM IMWUT*, 1(2):29, 2017.
[41] Y. Zhuang, J. Yang, Y. Li, L. Qi, and N. El-Sheimy. Smartphone-based indoor localization with bluetooth low energy beacons. *Sensors*, 16(5):596, 2016.