DCCP: Deep Convolutional Neural Networks for Cellular Network Positioning

Yu Lin[1], Yao Tong[2], Qinkun Zhong[2], Ruipeng Gao[2]*, Buyi Yin[1], Lei Liu[1], Li Ma[1], and Hua Chai[1]
Didi Chuxing, Beijing, China[1]
Beijing Jiaotong University, Beijing, China[2]
{whulinyu,yinbuyi,liuleifrey,malimarey,chaihua}@didiglobal.com[1], {tongyao,rpgao,qkzhong}@bjtu.edu.cn[2]

* Corresponding author

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685658

***Abstract*—Although location awareness is prevalent outdoors due to the GPS, we get confused and disoriented in many blocked environments such as in urban canyons and under multi-level flyovers. A straightforward solution is to employ cellular signals for positioning, but the cellular signatures are always sparse and uneven in vast region, and vary among different devices and postures. In this paper, we propose DCCP, a novel cellular network positioning approach that transforms the localization problem into a corresponding object recognition task in geographic space. Specially, we elicit the receptive region of each cellular station via crowdsourced user queries, and exploit neighbour base stations to derive a multi-dimensional feature map. We also devise a CNN model to learn local correlations among nearby map grids, and employ it for cellular positioning. Extensive experiments on two real-world traffic datasets from the DiDi platform have demonstrated our effectiveness compared with the state-of-the-art. This is the first approach to use only user queries instead of RF signatures for cellular network positioning, and our system meets requirements of the E911.**

***Index Terms*—cellular network positioning, user query, multi-dimensional feature map, CNN**

I. Introduction

Thanks to the explosion of GPS systems and devices, drivers are always conscious of their positions based on the popularity of location-based services (LBS) [1]. The location awareness has become an essential infrastructure in current sharing economy, especially for the ride-hailing platforms such as Uber, Lyft, and DiDi. However, whenever we drive into GPS blocked environments such as in urban canyons and under multi-level flyovers, we lack sufficient GPS receptions and easily get confused in such maze-like road structures.

This perennial challenge has sparked decades of research and yielded various Cellular Network Positioning (CNP) techniques [2]. A naive method is to use the Cell ID (CID) mandatory, i.e., estimating users' approximate locations based on their connected cellular base station, but the localization errors are often larger than 150 meters due to the wide coverage of 4G cellular towers [3]. A dedicated device [4] is deployed to scan surrounding cellular signals and produce corresponding distances towards nearby base stations via a particular radio propagation model, but it is not extensible for commercial smartphones. Recently, a series of efforts [4], [5] investigate various Radio-Frequency (RF) features for cellular network positioning. They always collect Received Signal Strength Indicator (RSSI), Reference Signal Receiving Power (RSRP), and Reference Signal Receiving Quality (RSRQ) cellular signatures at each location to construct a fine-grained fingerprint database, thus reducing localization errors caused by the non-line-of-sight signal propagation and multi-path fading effects. However, due to the arbitrary postures of smartphones and various quality of on-device antennas, the distribution of RF features is always broad and there exist similar signal strengths even in two far-away positions (Figure 1), resulted in extreme localization errors for more than 80 meters based on the state-of-the-art [6].

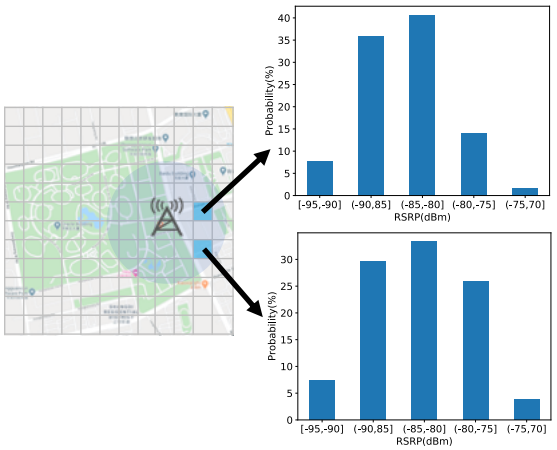


Fig. 1. Unreliable RSRP measurements for cellular network positioning. The signal strength varies widely among different devices and postures, and two far-away areas may share similar signal strength distributions.

In this paper, we propose *DCCP*, a novel Deep Convolutional neural network for Cellular network Positioning. Our intuition is to rank geographic grids based on the quantity and update time of user queries, instead of using the signal strength which varies among different devices and postures. The overview of our method is shown in Figure 2. First, we extract the receptive region for each base station based on the quantity and update time of user queries. Next, we construct a multi-dimensional feature map via nearby base stations, and explore a CNN model to infer the user's relative position towards the center of the reception region.

Specifically, we make the following contributions:

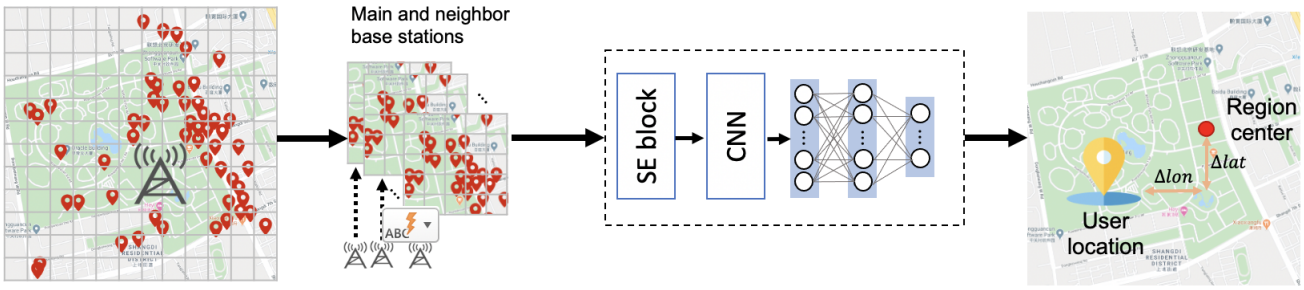


Fig. 2. The overview of our cellular network position method. (a) For each base station, we take user queries and map grids as input, and identify its reception region based on the quantity and update time of previous queries. (b) We exploit nearby base stations to construct a multi-dimensional feature map, and (c) explore a CNN model to produce user's relative position towards the center of reception region.

- We elicit the receptive region of each cellular station only based on user queries, thus preclude the interference from the broad and uneven distributed RF signatures.
- We extract features of nearby cellular stations, and devise a multi-dimensional feature map as model inputs, instead of directly predicting the missing fingerprints.
- We propose a novel CNN model to extract the local correlation among adjacent map grids, and infer user's relative locations in the receptive region.
- We conduct extensive experiments on real-world datasets collected by the DiDi platform. Results have shown our effectiveness compared with the state-of-the-art.

II. Related Work

Received Signal Strength in UMD. UMD (User Measurement Data) is reported by LTE smartphones regularly. It contains the RF measurements such as reference signal receiving power (RSRP), received signal strength indicator (RSSI) and reference signal receiving quality (RSRQ) [2]. RSRP is a crucial indicator of LTE, which is the average value of the signal power received on all REs carrying the reference signal in the symbol. RSSI is the average value of the power of all signals (including pilot signals and data signals, neighboring interference signals, noise signals, etc.). RSRQ is the ratio of RSRP and RSSI, i.e., $RSRQ = N * RSRP/RSSI$ where N denotes the signal bandwidth. Although we can build a wireless signal propagation model to transform these channel measurement values into distances from/to base stations, there are too many occlusions in urban cities, thus causing huge localization errors [1].

Cellular Network Positioning. Cellular network positioning is an indispensable component of location-based services (LBS). It has been developed and applied for decades. The most frequently used at an early age is the cell-id based positioning method [3], [7]. Then some hardware-specific approaches are proposed, e.g., AOA [8] and TOA [9] methods. However, this kind of technology is highly dependent on the density of base stations, and the information in the practical UMD is always insufficient for TOA and TDOA. The fingerprint-based method [10] came out since cellular stations are densely deployed. Cellsense [11] proposed a cellular RSSI fingerprint positioning method. Margolies *et al.* [12] constructed a wide-area radio map and developed a fingerprint-based test platform for cellular network positioning. In order to track users, Ray *et al.* [5] exploited the predicted fingerprints to match the UMD time series with the physical route, thus deriving continuous trajectories. Chakraborty *et al.* [4] proposed a geo-tag method based on the Gaussian mixture model (GMM), which formulates the characteristics of the RF signal as a Gaussian distributed random variable. Tian *et al.* [13] redesigned the subspace identification mechanism and took full use of the internal relations of RF fingerprints. They also proposed a large-scale fingerprint prediction method to fill up the missing fingerprints, but the cellular localization accuracy is still insufficient, especially for arbitrary postures and devices.

Machine Learning for Positioning. With the maturity of machine learning algorithms, all the collected information can be fully utilized as features for learning. Industries are often employing a geographic grid ranking method for cellular positioning. They score each map grid based on various observations in UMD, which can not capture the local correlation of adjacent grids in space, resulting in insufficient accuracy with different postures and devices. CNN has achieved great success in the field of computer vision [14]. It has specific characteristics including local area connection, weight sharing, down-sampling, etc. R-CNN [15] can derive the candidate area and produce accurate prediction results via the local correlation of adjacent image pixels. It inspired us that CNN has the potential ability to extract the local correlation of adjacent grids in map space and improve the accuracy of cellular network positioning.

III. Methodology

In this paper, we propose our cellular network position method *DCCP*, which only utilizes the quantity and updating time of user queries as inputs, instead of measuring RF features. It extracts the receptive region of each base station for

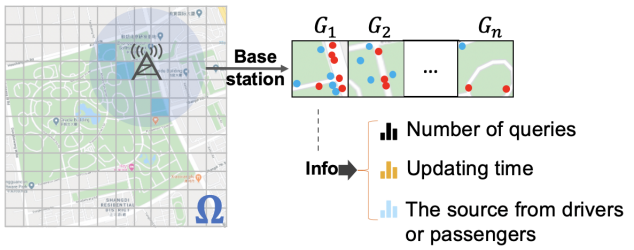
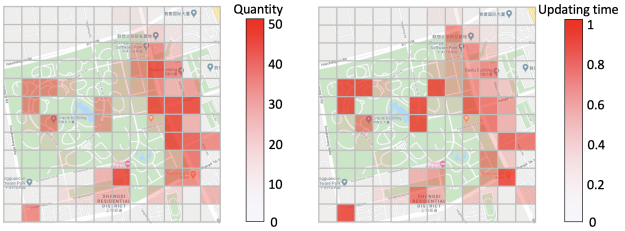


Fig. 3. One base station covers multiple map grids, and we calculate the amount, updating time, and source over all queries in each grid.



(a) Collection numbers heat (b) Update time heat

Fig. 4. Heat maps on queries. The gradation of color indicates the quantity and updating time of queries.

better attention, involves adjacent base stations to extend the dimension of feature maps, and devises a deep CNN model to infer the user's relative location based on the local correlation of map grids.

A. Query Database Construction

Thanks to the explosion of ride-hailing platforms, full-fledged crowdsourcing mechanisms have been deployed to collect user queries among tens of millions of smartphones. One typical query record can be denoted as: $Q = \{t, l, d, p, C_{1:m}, R_{1:m}\}$, where t represents the time stamp, l is the current location of user, d and p represent whether the record comes from a driver or a passenger, m is the number of base stations which are scanned at the current location, $C_{1:m}$ denotes the Cell ID of each base station, and $R_{1:m}$ is the corresponding signal strength RSRP.

Next, we randomly select 80% of the collected records in the past one month to establish a query database (shown in Figure 3). We divide the urban area into small grids with the same size (e.g., $11m \times 11m$ in our project), and classify grids according to the coverage of each cellular station. Since there are always coverage overlaps among different cellular stations, a grid may belong to multiple stations. We then count the number of queries α, the latest updating time as β, and the source (d for driver and p for passenger) to construct the database.

B. Receptive Region Detection

Since crowdsourcing queries are always sparse and uneven in vast areas, we propose a receptive region detection algorithm for better attention. As elucidated in Figure 4, the quantity and updating time of queries profoundly infer user locations at grid levels. Thus, we exploit both of them to elicit a small and inclusive receptive region of each base station.

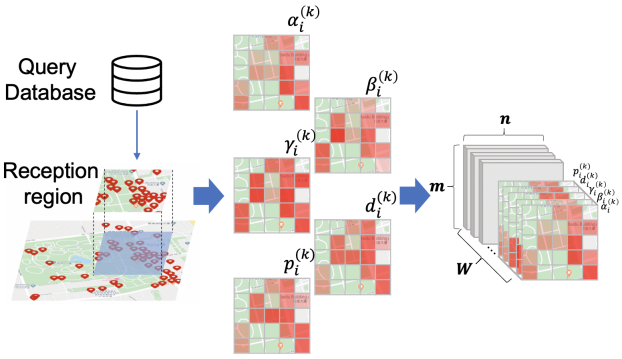


Fig. 5. Receptive region detection and feature map construction.

We devise a time attenuation factor to combine queries' quantity and updating time, and calculate the weight for each grid, i.e.,

$$\gamma_i = \alpha_i \exp\{-(\mathrm{t} - \beta_i)\} \tag{1}$$

where α_i represents the number of queries from all cellular stations at the grid G_i, β_i is the last updating time among all queries, and t is the current time.

Based on this weighting mechanism, we score all map grids for each cellular station, select the top n grids and derive their center as the original point for localization. Finally, we produce the receptive region as an $m \times n$ rectangular area according to the same center (shown in Figure 5).

C. Input Features

Existing cellular network positioning techniques always fill up the missing fingerprints based on sparse and uneven RF observations, thus the localization accuracy is easily impeded by arbitrary postures and qualities of smartphone antennas. Instead of predicting a fine-grained fingerprint database, we observe that cellular signals from some neighbor base stations are still strong due to the wide coverage of LTE networks (shown in Figure 7). Thus, our intuition is to utilize neighbor cellular stations to extend the dimension of our feature maps. Based on our experiments, we exploit one main station and six neighboring stations in urban areas.

For each grid G_i, we devise a five-dimensional feature map of each base station k, including:

- The number of queries $\alpha_i^{(k)}$;
- The last updating time $\beta_i^{(k)}$;
- The time attenuation factor $\gamma_i^{(k)}$;
- The number of drivers $d_i^{(k)}$;
- The number of passengers $p_i^{(k)}$.

Figure 6 contrasts the workflow of feature map construction. For each station, we elicit the reception region and produce a 5-dimensional feature map. Finally we stack W feature maps among all base stations as the input of our CNN model.

D. Model

Our CNN model exploits the multi-dimensional feature maps F as input, and predicts the deviation $(\Delta x, \Delta y)$ of

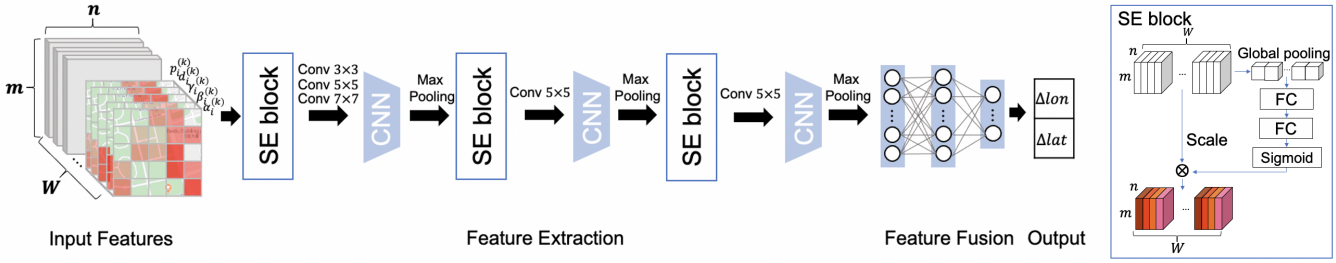


Fig. 6. DCCP structure. It consists of three convolutional layers and three fully connected layers. Before each layer of convolution, there is a SE block as the attention mechanism, including global pooling, fully connection and sigmoid activation.

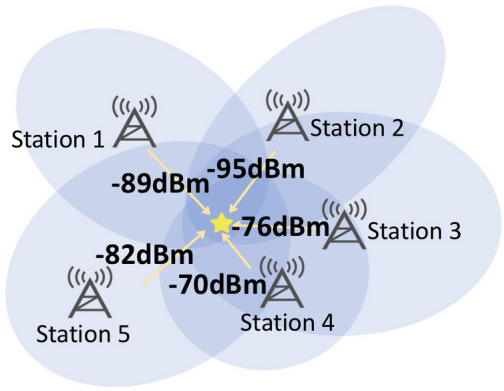


Fig. 7. A user receives multiple cellular signals from different base stations, and their RSRP signatures are slightly different.

user's relative position towards the original point x_0, y_0) in the receptive region, i.e.,

$$(\Delta x, \Delta y) = \mathrm{CNN}(F) \tag{2}$$

Structure. Our model is composed of three convolutional layers and three fully connected layers, as shown in Figure 6. We also add an SE module [16] before each convolutional layer as the attention mechanism among different channels. The SE module consists of global pooling, fully connection and sigmoid activation. We use different sizes of convolution kernels (3×3, 5×5, and 7×7) to extract features in the first convolutional layer, and employ a 5×5 convolution kernel for the last two convolutional layers. After each convolution, we use a 2×2 maximum pool layer to reduce network parameters.

Loss function. The user's predicted global position $(\hat{x}, \hat{y})$ is computed as:

$$(\hat{x}, \hat{y}) = (x_0 + \Delta x, y_0 + \Delta y) \tag{3}$$

where (x_0, y_0) is the coordinates of the center in reception region. We leverage the harversine formula as our loss function, which calculates the distance between the predicted position and its ground truth:

$$\mathrm{L} = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\hat{y} - y}{2}\right) + \cos y \cos \hat{y} \sin^2\left(\frac{\hat{x} - x}{2}\right)}\right) \tag{4}$$

where r is the radius of earth and (x, y) is the ground truth (gathered via opportunistic GPS observations).

Implementation. Our models are implemented in TensorFlow, and trained for 30000 epochs using the Adam optimizer with learning rate of $5e-4$. The ratio for training/validation/test split is $0.64 : 0.18 : 0.2$. We set the batch size of 512 and dropout of 0.5. The training and validation process is shown in Figure 8.

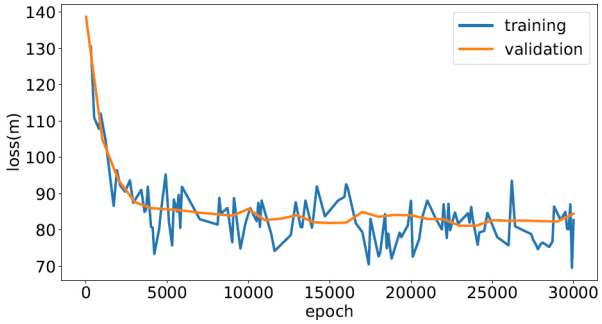


Fig. 8. Training and validation process.

IV. Evaluation

A. Experiment setting

Data collection. Based on the DiDi ride-hailing platform, We collected one million crowdsourcing queries within one month, and constructed a query database covering a small area (Shelchi with 2.6 km^2) and a large city (Beijing with 16410 km^2). Each query includes its timestamp, a cell ID of the connected cellular station, local indicators of six neighbouring cellular stations, a driver/passenger tag, and opportunistic GPS locations as the ground truth. An example application for data record is shown in Figure 9.

Experiment overview. 1) We use different sizes of windows to generate the receptive regions, and test which one has the largest coverage probability to involve the correct query. 2) In order to verify the contribution of neighbor base stations for cellular network positioning, we adopt a comparison experiment. 3) The feature ablation experiments are used to confirm the validity of each feature we devise. 4) In order to test whether our accuracy and reliability can meet the requirements of E911, we measure localization errors with CDF curves. 5) Finally, we compare with some state-of-the-art methods

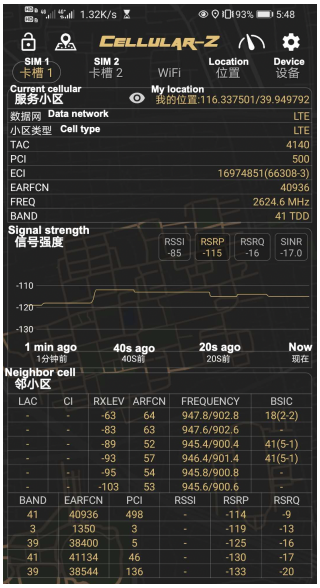


Fig. 9. An application to record the information of cellular stations. EARFCN denotes the local indicator of each neighbor station.

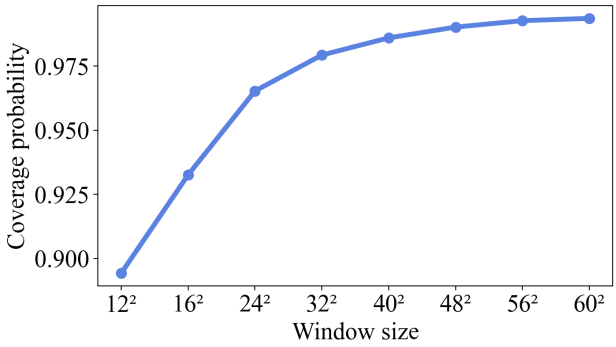


Fig. 10. Coverage probability with different sizes of reception region.

including CID, GMM, SSOA, and geographic grid ranking algorithm, in both small area and large city.

B. Receptive region coverage

In order to find the window size of the receptive region with the best performance, we adjust the window size and evaluate the coverage probability by:

$$C = \frac{\sum_{N}^{v=1} I_v}{N}, I_v = \begin{cases} 0, & t_v \in G_r \\ 1, & t_v \notin G_r \end{cases} \quad (5)$$

where t_v is the ground truth location of this query, G_r is the current grid, I_v represents whether it is in the receptive region, and C denotes the coverage probability. In addition, the broader coverage of a receptive region indicates the better performance.

The results are shown in Figure 10. For systematic performance impact, e.g., ensuring online operations, we choose 32×32 as the window size for receptive region detection.

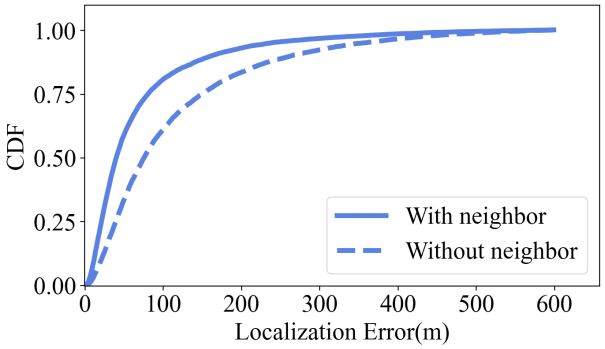


Fig. 11. CDF of localization errors with/out neighbor base stations.

TABLE I
FEATURE ABLATION.

Method	MSE	MAE	MAPE
DCCP-all	**29537.0118**	**81.5595**	**1.7084**
-no number of fingerprint	29747.2639	81.8877	1.7152
-no last collected time	29423.2092	81.8405	1.7143
-no drivers	30308.4701	83.5592	1.75
-no passengers	30203.5582	83.7155	1.7539

C. Effect of neighbor base stations

Our feature map consists of features from both main base station and its neighbor base stations. To verify the effectiveness of neighbor base stations for cellular network positioning, we compare localization errors (Euclidean distance between the predicted location and the ground truth) with/out neighbor base stations. As Figure 11 shows, the localization error is almost reduced to the half with a neighbor base station.

D. Feature ablation

We divide features produced into 4 categories and confirm the validity of each feature, as illustrated in TABLE I. When we remove any features, all three errors (MSE, MAE, and MAPE) are increased compared to the full-feature model. This indicates the effectiveness of our features.

E. Comparison with others

We compare the performance of our DCCP with CID [17], Gaussian mixture model [18], SSOA [13] and Geo-grid ranking algorithm. We conduct localization at 30000 test samples in Shelchi and Beijing, respectively.

In Figure 12, we observe that the 80-percentile localization error of DCCP is within $100m$, while the errors are larger than $190m$ for the other four methods, more than $2x$ errors of our method. For SSOA, it transforms the fingerprint prediction problem into a subspace recognition problem via matrix completion. However, it involves more errors between the predicted result and the ground truth. For CID, its performance is impacted by the uneven distribution of base stations. For GMM, the signal strength received at a given location is not always a Gaussian distributed random variable. Therefore, DCCP achieves the best performance to produce more accurate and robust locations in urban environments.

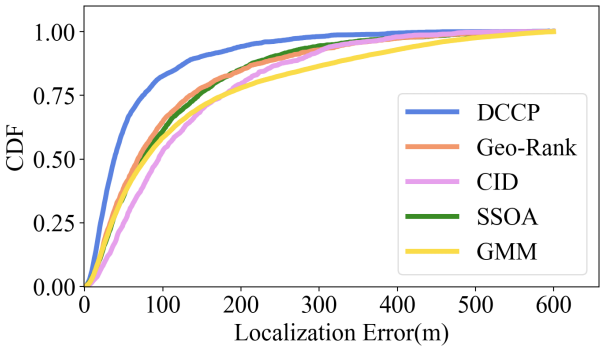


Fig. 12. CDF of localization errors in Shelchi.

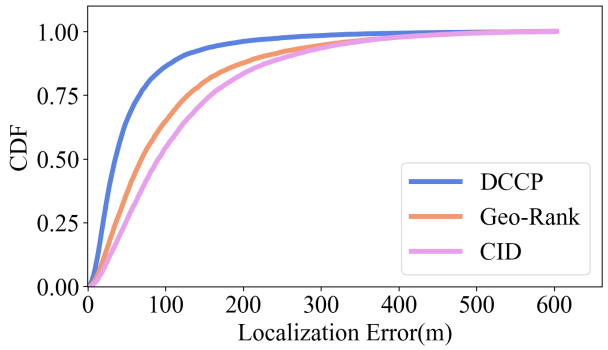


Fig. 13. CDF of localization errors in Beijing. SSOA and GMM are discarded due to their poor training efficiency in the large city.

We also use the E911's localization requirement as a benchmark, which announces the probability of $100m$ errors should be larger than 67-percentile and probability of $300m$ errors larger than 90-percentile. The localization error for our DCCP is within $100m$ at 81.2-percentile and within $300m$ at 98-percentile, for SSOA is within $100m$ at 63-percentile and within $300m$ at 96-percentile, for Geo-Rank is within $100m$ at 62-percentile and within $300m$ at 93-percentile, for CID is within $100m$ at 51-percentile and within $300m$ at 91-percentile, and for GMM is within $100m$ at 59-percentile and within $300m$ at 83-percentile. Thus, only our method satisfies the E911's localization requirements.

In Figure 13, we find DCCP can still achieve better performance than Geo-Rank and CID in large city. SSOA and GMM are discarded due to their poor training efficiency in the large city with massive data. Compared with the small region, its predictive ability has not decreased. Therefore, DCCP is consistently prior to all the state-of-the-art approaches.

V. Conclusion

This paper proposes DCCP, a novel approach that transforms the localization problem into a corresponding object recognition task in geographic space. Instead of gathering RF fingerprints for cellular network positioning, we only use user queries and devise an end-to-end CNN model for localization. Extensive experiments have shown our effectiveness compared with the state-of-the-art, and our reliability perfectly meets the requirements of E911. With the densely deployed 5G base stations at present, our method has broad applicability in various applications such as VR/AR games, mobile robotics and autonomous driving.

Acknowledgments

This work is supported in part by Beijing NSF L192004, NSFC 62072029, DiDi Research Collaboration Plan, and CCF-Tencent Open Fund.

References

[1] W. Zhang, X. Cui, D. Li, D. Yuan, and M. Wang, "The location privacy protection research in location-based service," in *Proceedings of IEEE International Conference on Geoinformatics*, 2010.

[2] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1g to 5g," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[3] E. Trevisani and A. Vitaletti, "Cell-id location technique, limits and benefits: an experimental study," in *Proceedings of IEEE workshop on mobile computing systems and applications*, 2004, pp. 51–60.

[4] A. Chakraborty, L. E. Ortiz, and S. R. Das, "Network-side positioning of cellular-band devices with minimal effort," in *Proceedings of IEEE INFOCOM*, 2015, pp. 2767–2775.

[5] A. Ray, S. Deb, and P. Monogioudis, "Localization of lte measurement records with missing information," in *Proceedings of IEEE INFOCOM*, 2016.

[6] E. Alimpertis, A. Markopoulou, C. Butts, and K. Psounis, "City-wide signal strength maps: Prediction with random forests," in *Proceedings of the World Wide Web Conference (WWW)*, 2019, p. 2536–2542.

[7] T. Kos, M. Grgic, and G. Sisul, "Mobile user positioning in gsm/umts cellular networks," in *Proceedings IEEE ELMAR*, 2006, pp. 185–188.

[8] D. Niculescu and B. Nath, "Ad hoc positioning system (aps) using aoa," in *Proceedings of IEEE INFOCOM*, vol. 3, 2003, pp. 1734–1743.

[9] M. T. Simsim, N. M. Khan, R. Ramer, and P. B. Rapajic, "Time of arrival statistics in cellular environments," in *Proceedings of IEEE VTC*, vol. 6, 2006, pp. 2666–2670.

[10] Q. D. Vo and P. De, "A survey of fingerprint-based outdoor localization," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 491–506, 2015.

[11] M. Ibrahim and M. Youssef, "Cellsense: An accurate energy-efficient gsm positioning system," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 286–296, 2011.

[12] R. Margolies, R. Becker, S. Byers, S. Deb, R. Jana, S. Urbanek, and C. Volinsky, "Can you find me now? evaluation of network-based localization in a 4g lte network," in *Proceedings of IEEE INFOCOM*, 2017.

[13] X. Tian, X. Wu, H. Li, and X. Wang, "Rf fingerprints prediction for cellular network positioning: A subspace identification approach," *IEEE Transactions on Mobile Computing*, vol. 19, no. 2, pp. 450–465, 2019.

[14] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE international symposium on circuits and systems*. IEEE, 2010, pp. 253–256.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of IEEE CVPR*, 2014, pp. 580–587.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[17] C. I. based localization: [Online], https://combain.com/about/about-positioning/cell-id-positioning/, 2017.

[18] A. Chakraborty, L. E. Ortiz, and S. R. Das, "Network-side positioning of cellular-band devices with minimal effort," in *Proceedings of IEEE INFOCOM*, 2015, pp. 2767–2775.