SCIENCE CHINA
Information Sciences

• LETTER •

February 2025, Vol. 68, Iss. 2, 129301:1–129301:2
https://doi.org/10.1007/s11432-024-4263-8

Localizing base stations with measured data: a concatenated image-based deep learning approach

Yang CHEN[1], Ziyuan LIU[1], Feng JIANG[2], Hong SHEN[1*], Wei XU[1*], Mengyu LI[2] & Ritao CHENG[2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*
[2]*China Mobile Group Design Institute Co., Ltd., Beijing 100080, China*

Received 11 July 2024/Revised 13 November 2024/Accepted 24 December 2024/Published online 9 January 2025

Citation Chen Y, Liu Z Y, Jiang F, et al. Localizing base stations with measured data: a concatenated image-based deep learning approach. Sci China Inf Sci, 2025, 68(2): 129301, https://doi.org/10.1007/s11432-024-4263-8

Localization has become an indispensable function of modern cellular communication systems [1]. In a cloud radio access network (C-RAN), the remote radio head (RRH) is the actual signal transmitter, whose location can assist in wireless network layout optimization and wireless resource management. The measurement of the RRH location can only be done manually on-site due to the lack of global positioning systems (GPS) at the RRH side, which is thus not a cost-effective solution for the widely used C-RAN with a large number of RRH units.

In general, there are three types of localization methods: two-step localization [2], direct positioning [3], and fingerprinting localization [4]. Different from both two-step and direct localization methods that need an accurate channel model, fingerprinting localization establishes the relationship between signal measurements, e.g., received signal strength indication (RSSI), and the actual location via offline learning without requiring the knowledge of the channel model. However, all the above studies focus on user equipment (UE) localization, which is different from the considered base station (BS) localization. In this study, by using the measured data provided by China Mobile, we develop a concatenated image-based deep learning (DL) method to deal with the challenging BS localization problem. We first propose a 2D sliding window-based data preprocessing approach. Then, we develop an enhanced U-Net called BSLocNet. Finally, the BS location estimate is obtained by performing an argmax-N approach on the output heatmap.

Data processing for BSLocNet. Considering a raw dataset including B BSs. The longitude and the latitude of the b-th BS ($1 \leqslant b \leqslant B$) are denoted by $x_{\rm BS}^b$ and $y_{\rm BS}^b$, respectively. There are L_b square grids in the coverage area of the b-th BS. The latitude, the longitude, and the average RSSI value of all grids in the b-th BS are denoted by $\mathbf{lon}_b \in \mathbb{R}^{L_b\times 1}$, $\mathbf{lat}_b \in \mathbb{R}^{L_b\times 1}$, and $\mathbf{RSSI}_b \in \mathbb{R}^{L_b\times 1}$, respectively. Our task is to estimate the BS coordinates based on the above dataset, which is expressed by

$$\left[\hat{x}_{\rm BS}^b, \hat{y}_{\rm BS}^b\right] = \mathscr{F}\left(\mathbf{lon}_b, \mathbf{lat}_b, \mathbf{RSSI}_b\right), 1 \leqslant b \leqslant B, \quad (1)$$

where $\left[\hat{x}_{\rm BS}^b, \hat{y}_{\rm BS}^b\right]$ denotes the estimated coordinate of the b-th BS and $\mathscr{F}(\cdot)$ represents a neural network.

We develop a snapshot-based RSSI image generation strategy based on the sliding window technique. Without loss of generality, let us consider the sliding window for the k-th snapshot of the b-th BS. We define $\mathcal{G}_k^b$ as the set of the grids located within the k-th sliding window for the b-th BS. The set of the horizontal and vertical coordinates of the i-th grid can be expressed by $\mathcal{X}_{b,i,k}$ and $\mathcal{Y}_{b,i,k}$, then we define $\mathcal{U}_{b,i,k} = \{[m,n] | \forall m \in \mathcal{X}_{b,i,k}, \forall n \in \mathcal{Y}_{b,i,k}\}$ as the set of pixel coordinates that corresponds to the area of the i-th grid in this snapshot. See Appendix A.1 for the process of generating the set of coordinates.

To facilitate the training of the neural network, we normalize the RSSI values as follows:

$$R_k^b(i) = \frac{\mathrm{RSSI}_b(i) - \min(\mathbf{RSSI})}{\max(\mathbf{RSSI}) - \min(\mathbf{RSSI})},\ i \in \mathcal{G}_k^b, \quad (2)$$

where $R_k^b(i)$ and $\mathrm{RSSI}_b(i), i \in \mathcal{G}_k^b$ represent the normalized and the original RSSIs of the i-th grid in the b-th BS locating in the k-th sliding window, respectively, and $\mathbf{RSSI}$ denotes the set of all RSSI values in the raw data.

Therefore, we can define the pixels of the RSSI image $\mathcal{RI}_k^b$ as follows:

$$\mathcal{RI}_k^b(m,n) = \begin{cases} R_k^b(i), & [m,n] \in \mathcal{U}_{b,i,k}, i \in \mathcal{G}_k^b, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The relationship between the BS location and the RSSI distribution depends on the geographic feature of the considered region. For the k-th window, we denote the corresponding GIS image as $\mathcal{GI}_k$.

Finally, for the b-th BS locating in the k-th window, the RSSI image $\mathcal{RI}_k^b$ and the GIS image $\mathcal{GI}_k$ are concatenated in channels as the network input, denoted as $\mathcal{CI}_k^b$. Note that there are total 4 channels since GIS images are colorful RGB images.

* Corresponding author (email: shhseu@seu.edu.cn, wxu@seu.edu.cn)

Table 1 Performance comparison of different methods.

Method	Mean error (m)		Pr{error ⩽ 50 m} (%)		Pr{error ⩾ 100 m} (%)	
	Train	Test	Train	Test	Train	Test
U-Net with coordinate output	231.17	233.65	7.37	7.46	74.75	74.60
Direct argmax-N	183.21	182.10	23.68	23.68	43.58	45.58
CNN	244.12	125.18	6.51	6.42	76.60	76.80
CoordAtt	39.16	45.16	74.26	72.34	7.28	8.10
Proposed method	11.11	18.36	98.84	98.46	0.26	0.35

We also propose a heatmap label that reflects every pixel's probability of being the actual BS location. For specific details, please refer to Appendix A.2.

Neural network design. U-Net [5] is employed as our backbone and refined by ResNet-style modules. The original U-Net has only 2 Conv layers in each downsample and upsample block, lacking the capability of providing a large perceptive field and deep feature extraction from high-resolution images. By replacing these 2 Conv layers with stacked residual modules, we not only enhance the network representation capability but also address the potential challenges resulting from a deeper network. Moreover, the same padding is adopted to avoid the interpolation operation for resizing when concatenating feature images in the U-Net pipeline. The specific architecture of the network can be found in Appendix B.

Firstly, we transform the actual coordinates into the pixel coordinates as follows:

$$x_k^b = \frac{x_{\mathrm{BS}}^b - \mathrm{lon}_l^k}{\mathrm{lon}_{\mathrm{pixel}}}, \tag{4}$$

$$y_k^b = \frac{y_{\mathrm{BS}}^b - \mathrm{lat}_l^k}{\mathrm{lat}_{\mathrm{pixel}}}, \tag{5}$$

where $\boldsymbol{\mu}_k^b = [x_k^b, y_k^b]$ is the pixel coordinate of the b-th BS, $[x_{\mathrm{BS}}^b, y_{\mathrm{BS}}^b]$ is the actual longitude and latitude coordinate of the b-th BS.

Then, we adopt a circular symmetric Gaussian distribution based label heatmap. The gray-scale of the pixel with coordinate $\boldsymbol{\mu} = [m, n]$ in the label is defined by

$$\boldsymbol{H}_k^b(m,n) = \exp\left(-\frac{1}{2}\left(\boldsymbol{\mu} - \boldsymbol{\mu}_k^b\right)^{\mathrm{T}} \boldsymbol{C}^{-1}\left(\boldsymbol{\mu} - \boldsymbol{\mu}_k^b\right)\right), \tag{6}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \delta_{xx} & \delta_{xy} \\ \delta_{yx} & \delta_{yy} \end{bmatrix} \tag{7}$$

represents the covariance matrix of the pixel coordinates along both the horizontal and the vertical axes. For simplicity, we assume that $\delta_{xy} = \delta_{yx} = 0$ and $\delta_{xx} = \delta_{yy} = \delta > 0$, where δ is the variance of the 2D Gaussian distribution. The MSE between the network output heatmap $\hat{\boldsymbol{H}}_k^b$ and label heatmap $\boldsymbol{H}_k^b$ is selected as the loss function L for supervised training, which takes the form

$$L = \frac{1}{|\boldsymbol{H}_k^b|} \sum_{m,n} \left(\boldsymbol{H}_k^b(m,n) - \hat{\boldsymbol{H}}_k^b(m,n)\right)^2, \tag{8}$$

where $|\boldsymbol{H}_k^b|$ is the total number of pixels in $\boldsymbol{H}_k^b$.

To sufficiently and robustly utilize the information of the predicted heatmap, we propose to select the brightest N pixels, record their coordinates in a set $\mathcal{A}_k^b$, and determine the predicted BS coordinate by

$$\hat{x}_k^b, \hat{y}_k^b t = \frac{\sum_{[m,n]\in\mathcal{A}_k^b} \hat{\boldsymbol{H}}_k^b(m,n) \times [m,n]}{\sum_{[m,n]\in\mathcal{A}_k^b} \hat{\boldsymbol{H}}_k^b(m,n)}. \tag{9}$$

Note that N is a hyperparameter determined beforehand (see Figure C2 in Appendix C). After estimating pixel coordinates, we need to perform the conversion from pixel coordinates to actual coordinates as follows:

$$\hat{x}_{\mathrm{BS}}^b = \hat{x}_k^b \mathrm{lon}_{\mathrm{pixel}} + \mathrm{lon}_l^k, \tag{10}$$

$$\hat{y}_{\mathrm{BS}}^b = \hat{y}_k^b \mathrm{lat}_{\mathrm{pixel}} + \mathrm{lat}_l^k, \tag{11}$$

where $[\hat{x}_k^b, \hat{y}_k^b]$ is the predicted pixel coordinate of the b-th BS from the network and $[\hat{x}_{\mathrm{BS}}^b, \hat{y}_{\mathrm{BS}}^b]$ is the predicted actual coordinate of the b-th BS.

Experiments based on measured data. We evaluate the performance of our proposed BS localization scheme on the measured data provided by China Mobile. We provide the performance comparison of different methods in Table 1. From Table 1, we can see that the proposed design outperforms all benchmark schemes. Compared to the U-Net with coordinate output, the residual modules and the heatmap output of our method can both help improve the network's learning ability and localization accuracy, as explained in the previous subsection. Moreover, the skip connection modules used by the proposed network can achieve feature fusion and thus endow the network with better performance than CNN. The evident performance gain over the argmax-N method is mainly attributed to the fact that this benchmark scheme can lead to a significant loss of spatial information when locating the BS. Finally, the coordinate attention mechanism used by CoordAtt is also less effective compared to the heatmap-based soft-label of our method. More details about the experiment can be found in Appendix C.

Acknowledgements This work was supported by joint research and development project with the Wireless Network Department of China Mobile Design Institute, which is entitled "Research on New Wireless Networking and Antenna Technology" (Grant No. R23101PE) and National Natural Science Foundation of China (Grant No. 61871108).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Xiao Z, Zeng Y. An overview on integrated localization and communication towards 6G. Sci China Inf Sci, 2022, 65: 131301
2 Shen H, Ding Z, Dasgupta S, et al. Multiple source localization in wireless sensor networks based on time of arrival measurement. IEEE Trans Signal Process, 2014, 62: 1938–1949
3 Weiss A J. Direct position determination of narrowband radio frequency transmitters. IEEE Signal Process Lett, 2004, 11: 513–516
4 Wang X, Gao L, Mao S, et al. CSI-based fingerprinting for indoor localization: a deep learning approach. IEEE Trans Veh Technol, 2017, 66: 763–776
5 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015. 5–9