Localization through Deep Learning in New and Low Sampling Rate Environments

Thanh Dat Le and Yan Huang

University of North Texas, Denton TX 76205, USA
{thanhle5}@my.unt.edu, {yan.huang}@unt.edu

Abstract. Source localization in wireless networks is essential for spectrum utilization optimization. Traditional methods often require extensive transmitter information while existing deep learning approaches perform poorly in new and low sampling rate environments. We introduce LocNet, a deep learning approach that overcomes these limitations using a compact UNet-like architecture incorporating environmental maps. Unlike other deep learning strategies, LocNet adopts loss functions designed explicitly for imbalanced data, moving beyond the conventional mean-square error loss. Our comparative analysis reveals that LocNet outperforms other deep learning models by more than a factor of two. This advancement underscores LocNet's suitability for real-world deployment across diverse operational contexts.

Keywords: Deep Learning · Source Localization · Imbalanced Data.

1 Introduction

Source localization, a foundational problem with decades of research, has broad applications including forest fire tracking [16], search and rescue [1] and pinpointing radiation sources [12]. In the wireless sensor network domain, particularly for spectrum monitoring, source localization optimizes spectrum utilization by detecting and removing unauthorized transmitters [20].

Traditional transmitter localization methods include Time-Of-Arrival (TOA), Angle-Of-Arrival (AOA) and Received Signal Strength (RSS). While TOA and AOA are effective; they require information about the transmitter's structure and power as well as costly hardware to acquire the measurements [22]. On the other hand, RSS which measures the raw signal power from a sensor at a receiving location, does not rely on transmitter information.

Given the advancements in artificial intelligence, particularly in deep learning, a data-driven technique is able to leverage the RSS to localize transmitters. Prior works employed deep learning [21, 18, 20, 22, 10, 8] to localize a randomly placed transmitter in an study area and did outperform the traditional methods [3, 6, 7, 13]. However, there are two major challenges that were not addressed: (1) Low sampling rate: existing methods require a large number of sensors to capture extensive RSS values and do not perform well in low sampling rate environments; and (2) Unknown environment: performance of existing methods diminish in large and unknown environments that were not covered during training [18, 21]. We will address these two challenges and make the following contributions:

- **Compact Deep Learning Model**: We propose LocNet, a compact UNet based regression model, optimized for stable training and robust performance for localization. The model is orders of magnitudes smaller than many State-Of-The-Art (SOTA) deep learning models.
- **High Performance in Low Sampling Rate and Unknown Environments**: We employ a loss function that tailors to the imbalanced dataset, moving beyond the Mean-Square Error (MSE) loss approach in training. Additionally, we integrate environment maps into the framework to improve localization performance. This integration boosts the model's effectiveness in completely unknown environments that are not included in the training dataset.
- **Extensive Experiments**: To test the robustness of our model, we compare LocNet with four SOTA deep learning based models across various unseen environments with sampling rates ranging from 0.01% to 0.1%. Results show that LocNet consistently outperforms the SOTA models. We conducted an ablation study to show the effectiveness of our design choices and provide further visual analysis of the best performing and worst performing environments to understand the impact of the environments and sample distribution.

In Section 2, we define the problem of transmitter localization. Section 3 contains literature reviews of the prior works in this area. Section 4 defines our methodology; including the base architecture and comparison with Unet. Section 5 describes the dataset, presents the experiment design and discusses the experimental results. Section 6 provides the directions of future research.

2 Problem Definition

Consider a geographic Region Of Interest (ROI) discretized into a two-dimensional grid of $\mathrm{H} \times \mathrm{W}$; where H and W are the height and width of the grid respectively with $(\mathrm{H}, \mathrm{W} \in \mathbb{Z}^+)$. Let S be a matrix with a size of $\mathrm{H} \times \mathrm{W}$ where S_{ij} represents a reading from a sensor if it is non-zero and no information nor buildings present. Additionally, a matrix E of $\mathrm{H} \times \mathrm{W}$ represents the environment mask where buildings are represented by -1, sensors are represented as 1 and vacancy (neither 1 nor -1) is represented as 0. The objective is to localize the transmitter location with S and the environmental mask E.

3 Related Works

Deep learning techniques [21, 18, 20, 22, 10, 8] have achieved better localization performance than traditional methods [3, 6, 7, 13]. These methods typically yield either the coordinate-based or the heatmap-based result for transmitter localization.

3.1 Coordinate Based Localization

The coordinate-based approach determines the (x, y) coordinates of the transmitters on a 2D map as the output of the localization process.

Zhang et al. [21] pioneered integrating deep learning in localizing a transmitter via RSS. They proposed a novel hybrid architecture that combines a Hidden Markov Model (HMM) with Multiple Layers of Perceptrons (MLP). The MLP is a feature extractor in this design, while HMM serves as a localizer to predict the transmitter's (x, y) coordinates. For data collection, Zhang et al. partitioned the ROI into multiple cells; each cell covered a small area and contained five (5) randomly placed signal sensors. The model subsequently learns from this data. The method works well in known and small-scale training environments.

Wang et al. [18] took a similar approach. Instead of randomly placing sensors in each cell, Wang et al. uniformly deployed 5 to 12 sensors across the ROI. The authors introduced a Multi-Task Gated Convolution Neural Network (MT-GCNN), combining the Gated 1D Convolution Neural Net with skip connections (GCNN) and two MLP heads. These MLP heads were responsible for finding the transmitters' containment cells and estimating the distance between the transmitters and the cells' center. A separate linear regression calculates the transmitters' location using the model's outputs.

DeepTxFinder [22] can localize transmitters with any sensor deployment across the ROI; where each is detected by training a separate deep learning model that shares a similar architecture. Each model had several Convolution Layers for feature extractions and an MLP for prediction. The system first predicts the number of transmitters in ROI and then selects the model to pinpoint them. The model performs well at high sampling rate environments.

3.2 Heatmap Based Localization

The heatmap-based deep learning technique converts sample readings into a 2D heatmap; with each pixel representing the probability of a transmitter's presence. Post-processing techniques, such as thresholding or the argmax functions, are used to identify the transmitter location.

DeepMTL [20] uses a modified object detection algorithm YOLOv3 [14] to generate a heatmap for transmitters' location detection. The authors demonstrated that YOLOv3 outperformed manual tuning via argmax or thresholding in localization. DeepMTL achieved remarkable performance in scalability and localization as compared to DeepTxFinder.

MSLocNet [8] and TL;DL [10] employ autoencoders to address the low sampling rate problem. An autoencoder comprises an encoder for extracting essential input features (latent spaces) and a decoder for generating a localized heatmap. Lin et al. implemented ResNet blocks [4] within the encoder. A ResNet block is characterized by multiple stacked convolution layers that are interspersed with skip connections that bypass selected layers. Furthermore, they innovated by introducing fusion blocks in the decoder. These blocks are a combination of bilinear upsampling and convolution layers that are specifically designed for effective heatmap generation. Mitchell et al. employed the original UNet architecture

[15]. Both models used manual thresholding for localization and excelled at low sampling rates.

Among the six deep learning models above, the model proposed by Zhang et al. [21] is specifically designed for the same fixed training and testing environment while the model by Wang et al. [18] requires the sensors to be uniformly distributed. We will focus on localization with randomly distributed sensors where the sampling rate can be very low and the training and testing environments are disjoint. We compare with DeepTxFinder [22], DeepMTL [20], MSLocNet [8], TL;DL. [10]; which suffer performance degradation when dealing with such environments.

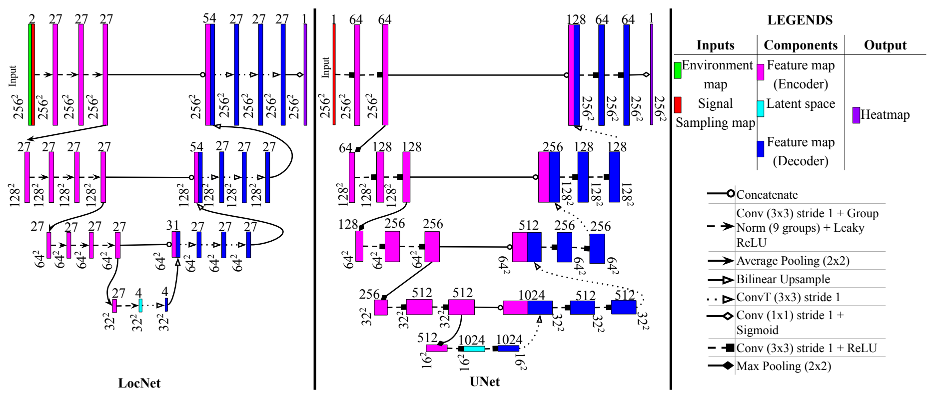


Fig. 1: Our proposed LocNet (128K parameters) in comparison with original UNet (31M parameters). Top number represents the number of channels and bottom number represents the channels's dimension. Bar width is proportional to number of channels.

4 Method

Deep Learning Autoencoder Autoencoder is a widely-used deep learning architecture for various tasks. It comprises two primary components: Encoder and Decoder. The encoder compresses the input(s) (or downsampling high-dimension features) into lower-dimension features; which is achieved by using different deep learning layers such as convolutional layers, pooling layers or fully-connected layers. In contrast, the Decoder utilizes the encoder's output to decompress and attempt to reconstruct the original input. Depending on the training data and ground truths, the decoder can learn tasks like image restoration, generation, and more.

UNet The UNet architecture, part of the deep learning autoencoder family, was first introduced by [15]. UNet's essential feature is the skip-connections, which link corresponding encoder-decoder layers between pooling and upsampling layers. These connections enhance information retention, facilitate the fusion of high-level features and improve gradient flow; thus aiding generalization for high-level encoders. Frost et al [10]. adopted this architecture in radio localization and designated it as TL;DL (as shown in the right of Fig. 1).

The Proposed LocNet Drawing from studies [10] and [8], we hypothesize that an architecture similar to AutoEncoder is well-suited for localization tasks. Our proposed LocNet model, as shown in Fig. 1 , is inspired by AutoEncoder architectures of Tenganya et al. [17] and Locke et al. [9] initially designed for generating radio propagation maps. We adapt UNet architecture as well as the features from the inspired models. UNet's skip connections are beneficial in handling highly sparse sampling maps by retaining and transferring information from encoder to decoder for better utilization. Moreover, LocNet utilizes both an environment map and a sampling map for localization to aid the model in generalizing to new environments for localizing a transmitter.

Distinct from the TL;DL, which follows the original UNet [15], LocNet is characterized by its compactness by reducing the number of channels of convolution as well as deconvolution layers. This design avoids unnecessary feature reductions in both the encoder and decoder. Additionally, we employ convolution transposes ("deconvolution layers") following bilinear upsampling to align with the decoder's purpose. This allows recovery of granularity from the latent space; serving as a "reversing" encoder. Group Normalization is integrated before the activation layer of each convolution and deconvolution layer. This technique groups the correlated features, helps convolution adjust the filters for better extraction and improves training stability. A sigmoid activation is applied at the end of the decoder to generate a probability map. To pinpoint the transmitter's (x, y) coordinate, we applied argmax function on the decoder's output. The detailed illustration of LocNet can be found in Fig. 1

Focal loss. Previous works $[10, 20, 8]$ have employed Mean-Squre Error (MSE) as a loss function for the localization problem. However, localization is a problem with highly imbalanced data and MSE applies equal weighting to all samples' losses. There are loss functions specifically designed to handle imbalances in image processing domain including Balanced Binary Cross-Entropy (BBCE), Focal Loss (FL) and Focal Tversky loss (FT). We introduce these losses first and provide insights on why FL is chosen for LocNet for the problem of localization.

Balanced Binary Cross Entropy (BBCE) : BBCE is a modified version of Binary Cross Entropy (BCE). BCE calculates the log predictions of each location for each class, as illustrated in Eq. 1. BBCE introduces a balancing coefficient, denoted as α, which adjusts the weight of the loss contributions from each class to address the class imbalance problem, as illustrated in Eq. 2.

$$BCE(p,y) = \begin{cases} -log(p) & \text{if } y = 1 \\ -log(1-p) & \text{otherwise} \end{cases} \quad (1)$$

$$BBCE(p,y) = \begin{cases} - * \alpha * log(p) & \text{if } y = 1 \\ -(1-\alpha) * log(1-p) & \text{otherwise} \end{cases} \quad (2)$$

Focal Tversky (FT) : FT loss is an enhanced variant of Tversky loss, as illustrated in Eq. 3, incorporating a gamma (γ) exponent term to prioritize the learning of the minority class. Tversky loss is a function that introduces a trade-off between false positives and false negatives; thus facilitating a controlled balance (β) when assessing the similarity between predicted and actual samples.

$$Tversky(p,y) = \frac{1 + p * y}{1 + y * p + \beta * (1-y) * p + (1-\beta) * y * (1-y)} \quad (3)$$

$$FT(p,y) = (1 - Tversky(p,y))^{\gamma} \quad (4)$$

Focal loss (FL) : FL loss function addresses the imbalance between classes and directs a deep learning model to focus on the minority class; as illustrated in Eq. 5. It has two tuning parameters: γ and α. The γ parameter regulates the loss by down-weighting the "easy" or majority class samples; thus allowing the model to concentrate on the "hard" or minority class samples. Meanwhile, α provides an additional weighting for balancing the loss contributions between the two classes.

$$FL(p,y) = \begin{cases} -\alpha * log(p) * (1-p)^{\gamma} & \text{if } y = 1 \\ -(1-\alpha) * log(1-p) * p^{\gamma} & \text{otherwise} \end{cases} \quad (5)$$

All three of the above loss functions are designed to handle imbalanced data and have the ability to direct LocNet's focus toward localizing a transmitter; which is a minor class in our dataset. We choose FL as the loss function to for LocNet based on the following observations: BBCE solves the imbalanced problem but lacks a mechanism to focus on a minority region such as FL. FT may not effectively work for our task since it is primary used for segmentation tasks and addresses data imbalance indirectly by tuning false positives and false negatives based on predictions.

5 Experimental Design and Results

In the experiments, we compare our LocNet model with four other SOTA models:

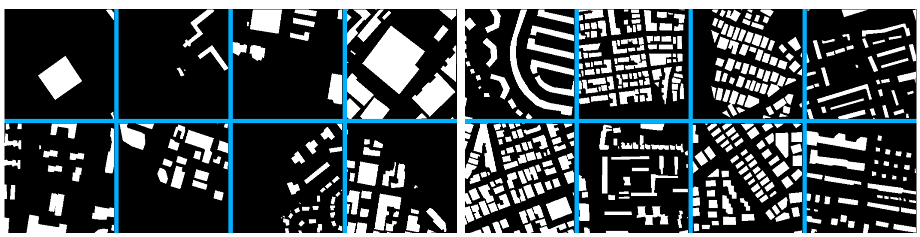
- *LocNet*: Our model.
- *TL;DL*: Inspired from UNet architecture [10].
- *MSLocNet*: Adopting AutoEncoder architecture with Residual Block [8].

– *DeepMTL*: A Two-stages localizer; combination between heatmap generator model with YOLOv3 [20].
– *DeepTxFinder*: coordinate (x, y) based regression that utilizing Convolution with Fully Connected Neural Network [2].

We adopt Root Mean Square Error of euclidean distance between two points (RMSE) extensively used in the SOTA models as the metric to compare localization performance between models.

$$RMSE = \sqrt{\frac{1}{N} * \sum_{n=1}^{N} Euclidean\_Distance(p_n, t_n)^2} \tag{6}$$

where N is number of testing sampling maps, p_n is the predicted transmitter location, t_n the ground truth transmitter location for map n, and $Euclidean\_Distance(p_n, t_n)$ is the Euclidean distance between the two locations.



(a) Simple environments that LocNet has the lowest localization error (increasing RMSE in row major order).

(b) Challenging environments that LocNet has highest localization error (decreasing RMSE in row major order).

Fig. 2: Test environment visualization.

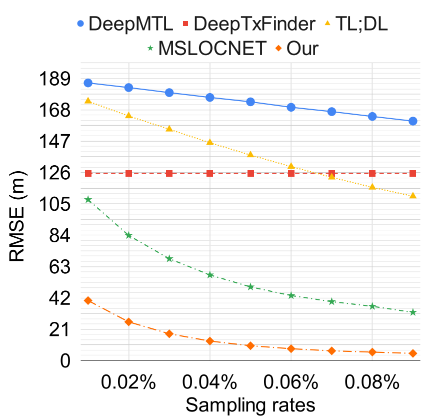
5.1 Dataset

The RadioMapSeer dataset [19] is used to create the low-sampling dataset for benchmarking our proposal against SOTA methods in outdoor localization. It comprises 701 maps, which are taken from OpenStreetMap [11] in different areas in major cities such as Ankara, Berlin, Glasgow, Ljubljana, London and Tel Aviv. Each map is a 2D binary image, where a value of 1 indicates a building is present and 0 indicates a building is not present. For each map there are 80 transmitter locations randomly distributed. Each transmitter location has a generated signal propagation map by using the Dominant Path Model (DPM) method via WinProp program[5] resulting in $701 \times 80 = 56,080$ signal maps. The transmitter, receivers, and buildings have heights of $1.5m$, $1.5m$, and $25m$ [19]. The simulations are stored as 2D grayscale images of 256 x 256 pixels; in which

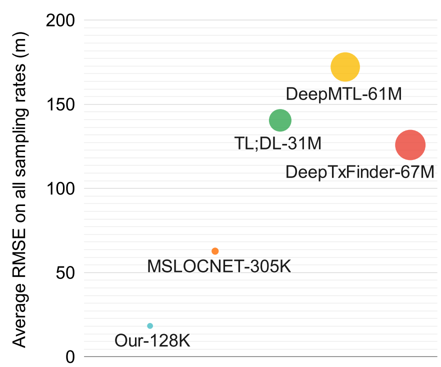
1 pixel covers an area of $1m^2$[19]. Fig. 2 shows 16 test environment maps with 8 easy and 8 most challenging environments based on our model performance.

For our benchmark, we sought to create a low-sampling dataset derived from the RadioMapSeer dataset. To this end, for each environment map and transmitter pair, we generate 5 sampling maps by using uniform sampling method with the sampling rate between 0.01% to 0.1% of the total pixels represented. Each sampling is represented by two matrices: (1) a signal map with sampling readings on sampled locations and zero on other locations; (2) a companion environment mask map with 1 representing sampled locations, 0 indicating "masked-out" areas, and -1 as building locations. This methodology yields 280,400 low-sampling maps from the 56,080 full signal maps. Based on the environments, We divide this dataset into training, validation, and testing sets (501/100/100). We ensure that the environments in all sets are entirely separate to avoid data leakage.

Implementation: We trained our model for 100 epochs with batches size of 64 maps. We employed the AdamW optimizer with a learning rate of 5e-4. In comparing between losses, we set $\alpha = 0.75$ for both BBCE and FL, $\gamma = 3$ for FL and FT, and $\beta = 0.7$ for FT loss. We reconstructed SOTA models based on the methodologies described in their research publications.



(a) Performance of our LocNet model and SOTAs.

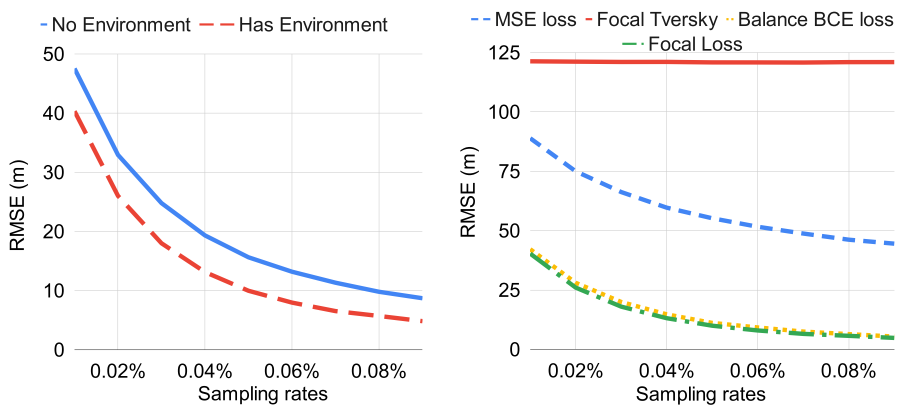


(b) Performance and model sizes (diameter and numbers representing numbers of trainable parameters)

Fig. 3: Performance Comparison with SOTAs

5.2 Comparing with SOTAs

As shown in Figure 3a, LocNet outperforms other SOTA models consistently across all sampling rates in the test sets where environments are not seen in training data, reducing the localization error by more than a factor of two. Moreover, all models' performances increase as the sampling rate increases, which reveals all models benefit from more samples. Among the SOTAs, MSLocNet consistently has the smallest performance gap from LocNet. Fig. 3b shows that LocNet is three orders of magnitudes smaller than DeepTxFinder and less than half the size of MSLocNet in training parameters, which enables edge device deployment.



(a) LocNet Performance w.r.t Environment Map

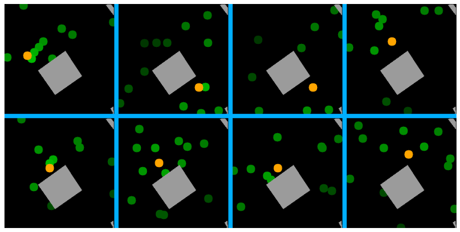
(b) LocNet Performance w.r.t. Loss Functions

Fig. 4: Ablation study on environment map and loss functions.

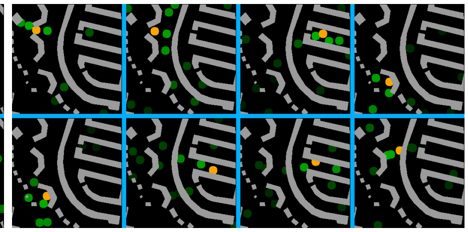
5.3 Ablation Study

Environment vs no environment: As shown in Fig 4a, fusing the environment map with the sampling map helps improve the localization performance. On average, the localization error is reduced by at least 15% for each sampling rate.

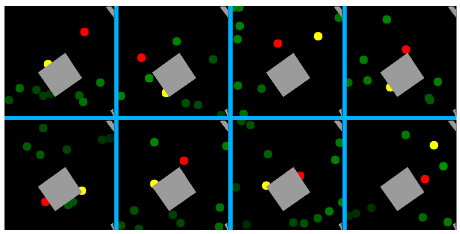
Comparing between losses: As shown in Fig 4b, FT and MSE result in the poorest performance. With MSE, our model typically estimates locations to be approximately 44 meters less accurate than using FL or BBCE. FL and BBCE demonstrated similar effectiveness. We believe the reason can be that both FL and BBCE employ log-form error calculation after performing a sigmoid function on predictions, which can result in significant loss values for incorrect predictions. These high loss values can lead to a high changing rate in gradients, thus enhancing the model's learning capability and likely avoiding saddle points. On the contrary, MSE and FT do not exhibit similar rates of gradient movement, which potentially causes the model's learning to hit a plateau.



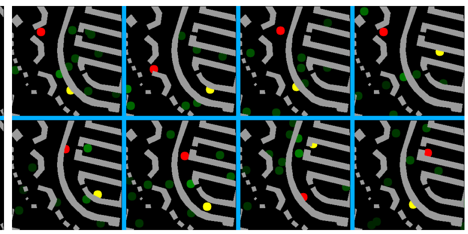
(a) Best cases for the simplest environment in Fig. 2a



(b) Best cases for the most challenging environment in Fig. 2a



(c) Worst cases for simplest environment in Fig. 2a



(d) Worst cases for the most challenging environment in Fig. 2b

Sampling Predicted Ground truth Overlapping predicted and ground truth

Fig. 5: Visualization of best and worse cases (dot sizes enlarged for visualization purpose).

5.4 Visual Analysis

In testing, LocNet performed better in environments with fewer buildings and more open spaces, as shown in Fig 2a, likely due to simpler signal propagation estimation in such settings. In contrast, dense urban areas in Fig. 2b cause signal attenuation, which increases the complexity of signal propagation estimation and leads to a reduction in localization performance. LocNet becomes more effective when more sensors are closer to the transmitter as shown between Fig. 5a and 5c and between Fig. 5b and 5d. Having more sensors near the transmitter simplifies signal propagation's estimation and assists the model in focusing on a specific area for localization, thus increasing the performance.

6 Conclusion and Future Work

We have presented LocNet, a specialized model for localizing a transmitter in sparse sampling maps. We employ UNet architecture, enhanced with Group Normalization, and utilize the FL function to deal with imbalanced and sparse data issues. Additionally, experimental evidence supports the notion that providing an environmental map of the area of interest further refines localization efficiency. We plan to extend our model's capabilities to localize a transmitter with different power strengths and improve our architecture to achieve greater performance in localization tasks.

Acknowledgement

The research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-23-2-0014. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, not withstanding any copyright notation herein.

References

1. Atif, M., Ahmad, R., Ahmad, W., Zhao, L., Rodrigues, J.J.P.C.: Uav-assisted wireless localization for search and rescue. IEEE Systems Journal **15**(3), 3261–3272 (2021)
2. Bizon, I., Nimr, A., Schulz, P., Chafii, M., Fettweis, G.P.: Blind transmitter localization using deep learning: A scalability study. In: IEEE Wireless Communications and Networking Conference (WCNC) (2023)
3. Destino, G., Abreu, G.: On the maximum likelihood approach for source and network localization. IEEE Transactions on Signal Processing **59**(10), 4954–4970 (2011)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
5. Hoppe, R., Wölfle, G., Jakobus, U.: Wave propagation and radio network planning software winprop added to the electromagnetic solver package feko. In: International Applied Computational Electromagnetics Society Symposium - Italy (ACES). pp. 1–2 (2017)
6. Khaledi, M., Khaledi, M., Sarkar, S., Kasera, S., Patwari, N., Derr, K., Ramirez, S.: Simultaneous power-based localization of transmitters for crowdsourced spectrum monitoring. In: Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking. p. 235–247 (2017)
7. Lin, L., So, H., Chan, Y.: Accurate and simple source localization using differential received signal strength. Digital Signal Processing **23**(3), 736–743 (2013)
8. Lin, M., Huang, Y., Li, B., Huang, Z., Zhang, Z., Zhao, W.: Deep learning-based multiple co-channel sources localization using bernoulli heatmap. Electronics **11**(10) (2022)
9. Locke IV, W.A.: Deep learning approaches to radio map estimation. In: Master thesis. UNT Digital Library, University of North Texas (2023)
10. Mitchell, F., Baset, A., Patwari, N., Kasera, S.K., Bhaskara, A.: Deep learning-based localization in limited data regimes. In: Proceedings of the ACM Workshop on Wireless Security and Machine Learning. p. 15–20 (2022)
11. OpenStreetMap: (2023), https://www.openstreetmap.org. Last accessed 10 Oct 2023

12. Pinto, L.R., Vale, A., Brouwer, Y., Borbinha, J., Corisco, J., Ventura, R., Silva, A.M., Mourato, A., Marques, G., Romanets, Y., Sargento, S., Gonçalves, B.: Radiological scouting, monitoring and inspection using drones. Sensors **21**(9) (2021)
13. Rahman, M.Z., Habibi, D., Ahmad, I.: Source localisation in wireless sensor networks based on optimised maximum likelihood. In: Australasian Telecommunication Networks and Applications Conference (2008)
14. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint:1804.02767 (2018)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing (2015)
16. Sharma, A., Singh, P.K., Kumar, Y.: An integrated fire detection system using iot and image processing technique for smart cities. Sustainable Cities and Society **61**, 102332 (2020)
17. Teganya, Y., Romero, D.: Deep completion autoencoders for radio map estimation. IEEE Transactions on Wireless Communications **21**(3), 1710–1724 (2022)
18. Wang, W., Zhu, L., Huang, Z., Li, B., Yu, L., Cheng, K.: Mt-gcnn: Multi-task learning with gated convolution for multiple transmitters localization in urban scenarios. Sensors **22**(22) (2022)
19. Yapar, Ç., Levie, R., Kutyniok, G., Caire, G.: Dataset of pathloss and ToA radio maps with localization application. arXiv preprint:2212.11777 (2022)
20. Zhan, C., Ghaderibaneh, M., Sahu, P., Gupta, H.: Deepmtl: Deep learning based multiple transmitter localization. In: IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) (2021)
21. Zhang, W., Liu, K., Zhang, W., Zhang, Y., Gu, J.: Deep neural networks for wireless localization in indoor and outdoor environments. Neurocomputing **194**, 279–287 (2016)
22. Zubow, A., Bayhan, S., Gawłowicz, P., Dressler, F.: Deeptxfinder: Multiple transmitter localization by deep learning in crowdsourced spectrum sensing. In: 2020 29th International Conference on Computer Communications and Networks (ICCCN) (2020)