

Received October 18, 2021, accepted November 4, 2021, date of publication November 8, 2021, date of current version November 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3126775

# Autonomous 3-D UAV Localization Using Cellular Networks: Deep Supervised Learning Versus Reinforcement Learning Approaches

**GHADA AFIFI<sup>1</sup>** AND **YASSER GADALLAH<sup>1</sup>**, (Senior Member, IEEE)

Department of Electronics and Communications Engineering, The American University in Cairo, New Cairo 11835, Egypt

Corresponding author: Ghada Afifi (ghadasameh@aucegypt.edu)

**ABSTRACT** Unmanned aerial vehicles (UAVs) are becoming an integral part of numerous commercial and military applications. In many of these applications, the UAV is required to self-navigate in highly dynamic urban environments. This means that the UAV must have the ability to determine its location in an autonomous and real time manner. Existing localization techniques rely mainly on the Global Positioning System (GPS) and do not provide a reliable real time localization solution, particularly in dense urban environments. Our objective is to propose an effective alternative solution to enable the UAV to autonomously determine its location independent of the GPS and without message exchanges. We therefore propose utilizing the existing 5G cellular infrastructure to enable the UAV to determine its 3-D location without the need to interact with the cellular network. We formulate the UAV localization problem to minimize the error of the RSSI measurements from the surrounding cellular base stations. While exact optimization techniques can be applied to accurately solve such a problem, they cannot provide the real time calculation that is needed in such dynamic applications. Machine learning based techniques are strong candidates to provide an attractive alternative to provide a near-optimal localization solution with the needed practical real-time calculation. Accordingly, we propose two machine learning-based approaches, namely, deep neural network and reinforcement learning based approaches, to solve the formulated UAV localization problem in real time. We then provide a detailed comparative analysis for each of the proposed localization techniques along with a comparison with the optimization-based techniques as well as other techniques from the literature.

**INDEX TERMS** 5G, UAV 3-D autonomous localization, optimization, deep learning, reinforcement learning, neural networks, Q-learning.

## I. INTRODUCTION

Unmanned aerial vehicles (UAVs) are increasingly demonstrating their potential for use in various military and civilian applications. UAVs have the advantages of a generally small size, convenient use and strong maneuverability which makes them suitable for such applications. The applications of UAVs in the military field include intelligence reconnaissance, emergency response, and geographic survey, while the civilian field applications include package delivery, aerial photography, infrastructure inspection, electric power inspection, agricultural plant protection, search and rescue and environmental monitoring [1]. However, there are several challenges for the UAV to deliver on the forementioned applications.

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang<sup>1</sup>.

One of the most important enabling requirements for the UAV to deliver on such applications is the ability to determine its location at any given point in time. The localization technique used currently by most commercial UAVs is a combination of the GPS and the Inertial Measurement Unit (IMU). However, in dense urban environments and indoor environments, the global navigation satellite systems (GNSS) signals do not provide an accurate and reliable localization solution due to reflections by high-rise structures and line of-sight (LOS) blockage. Also, the dependence on detectable transmissible signals may comprise the success of certain missions. Hence, it is necessary to design an alternative high-precision positioning method that is not based on GPS or other detectable signals.

As an alternative to the global positioning system (GPS), the radio received signal strength index (RSSI) measurements

from the surrounding cellular infrastructure can be used for localization purposes due to its simplicity and cost-effectiveness. Since the cellular infrastructure is widely deployed in urban areas worldwide, the reliance on cellular signals can offer an attractive alternative to GPS for UAV localization applications. We focus our attention on the 5G technology in this paper since it is expected to be widely deployed around the world. According to the 3rd Generation Partnership Project (3GPP), the 5G technology offers attractive features such as dense small cell deployments, millimeter wave (mmWave) communications, and device-to-device (D2D) communications [28]. The 5G technology addresses the current cellular network challenges, including ultra-low latency and higher reliability and capacity requirements, by optimizing the network operations to guarantee, in real-time, the QoS needs of emerging wireless and IoT services and consequently is expected to be widely deployed [29].

Cellular-connected UAV communication possesses substantially different characteristics that pose new technical challenges as opposed to cellular communication with terrestrial mobile devices. Such challenges include dominance of line-of-sight (LoS) interference and reduced ground base stations (GBSs) antenna gain [2]. UAVs flying at high altitudes are served from the side-lobes of the base station (BS) antennas [3]. Several studies suggest that cellular networks with communication techniques toward 5G such as higher antenna gains due to beamforming can potentially be utilized to provide improved connectivity for UAVs [2] [4]. As such, several recent efforts and open work items are directed towards the integration of the UAVs with 5G cellular networks. Our objective is the autonomous localization of the UAV without the need to actually interact with the cellular network. Accordingly, we focus on the broadcast signals transmitted by the 5G cellular networks which can be detected by any cellular-enabled device. Specifically, the 5G cellular network periodically broadcasts the Synchronization Signal Block (SSB), which is composed of 4 subblocks including the Primary Synchronization Sequence (PSS), the Secondary Synchronization Sequence (SSS), the Physical Broadcast Channel (PBCH) and the Demodulation Reference Signal (DMRS), mainly for synchronization, cell search and initial beamforming [13]. The UAV can detect and measure the Secondary Synchronization Reference Signal Received Power (SS-RSRP) from the surrounding base stations for self-localization purposes.

The UAV localization problem formulation based on cellular measurements is non-trivial and difficult to solve. While exhaustive optimization techniques can be applied to accurately solve for such a problem, a real-time calculation is needed in such dynamic environments. Machine Learning (ML) based techniques could potentially provide an attractive alternative to provide a near-optimal localization solution that provides a practical real-time calculation that is needed in such dynamic applications. Deep Learning (DL) is a branch of supervised machine learning involving neural

networks with several layers capable of capturing complex non-linear relationships between the inputs and the outputs. Deep learning-based approaches typically fall into one of two categories, namely, regression-based models and the fingerprint-based techniques which require extensive data collection to build a fingerprint database for the training. Reinforcement Learning (RL) is another branch of machine learning where the agent can learn through direct interaction with the environment. The Reinforcement Learning model is based on the Markov decision process (MDP). Specifically, Reinforcement Learning agents make decisions, observe the results, and then automatically adjust their policies to achieve their objective to select an action that maximizes a reward.

#### A. RELATED WORK

There are several recent studies that have investigated the use of cellular signals and/or Machine Learning based approaches in the localization of mobile devices or UAVs.

In [5], the authors propose utilizing higher order Voronoi tessellations at the base station to localize mobile devices in an outdoor environment. In the proposed base station ordering localization technique (BoLT) algorithm, the mobile device sends back a list containing the order of neighboring BSs based on the received signal power achieving a localization accuracy of a few meters. In [6], CellinDeep, a deep learning-based localization system for mobile devices in an indoor environment is proposed by creating a footprint map of RSSI measurements at different locations during the training phase. The area is divided into grid cells of  $1\text{m}^2$  and RSSI measurements from 17 cell towers are recorded prior to the training phase at each cell. The localization problem is then presented as a classification-problem to find the grid cell with the maximum likelihood followed by a fine localizer module achieving a fine grained accuracy of 0.78m. Alternatively in [7], the authors introduce DeepCReg, a convolutional neural network based regressor that leverages cellular data to estimate the location of a mobile device in an outdoor environment. The system achieves median localization accuracy of 2.82m in the 2-D localization problem for mobile devices. In [8], an optimization-based approach utilizing carrier phase measurements is proposed to localize and navigate a UAV in 3-D assuming limited GPS presence. According to the authors, this technique realizes a Root Mean Square Error (RMSE) location error of 0.8m using 7 CDMA BSs and 0.36m using measurements from 9 CDMA BSs. In both cases, the UAV had access to the GPS for 10 seconds then the GPS was cut off. During the time the GPS was available, the cellular signals were used to cluster and characterize the clock deviations.

However, to the best of our knowledge, none of the presented works provide an accurate and effective real-time solution to the 3-D localization problem for the UAV in an outdoor environment. The presented techniques either mostly focus on the localization of the agent in the 2-D space assuming the height of the mobile device is known yielding low accuracy when extended to 3-D or require extensive

knowledge of the environment during the training phase to build a fine-grained accurate fingerprint map. Such proposed approaches and data collection processes would be infeasible and unscalable in large 3-D outdoor geographic areas and unknown environments.

**B. PAPER CONTRIBUTIONS AND STRUCTURE**

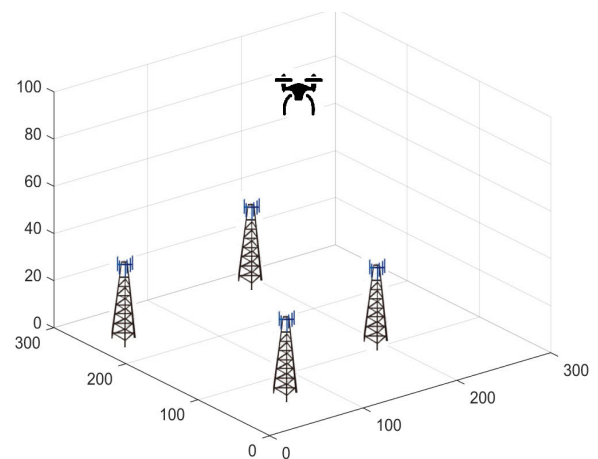
The objective of this paper is to provide an accurate real-time solution for the 3-D autonomous localization of UAVs in an outdoor environment using existing 5G networks independent of GPS or other detectable mobile signals. This is the first work, to the best of our knowledge, to analyze the effectiveness of various machine learning based approaches to provide a real-time solution to the 3-D UAV self-localization problem using 5G cellular networks in an outdoor environment with near optimal accuracy. The major contributions of this paper can be summarized as follows

- We formulate the UAV localization problem as an optimization problem in which the drone needs to rely on the RSSI measurements of the surrounding 5G base stations without having to actually interact with these base stations. The objective function minimizes the overall mean least square error to determine the UAV location.
- We propose a 3D UAV localization algorithm through multi-lateration that is based on 5G RSSI measurements from 4 base stations. We develop an optimization-based approach to determine the optimal bound of the solution to the formulated localization problem. We present Nelder-Mead as a heuristic approach as well as Exhaustive Search based solutions.
- We propose a deep supervised learning approach to provide a near-optimal localization solution that provides a real-time calculation that is needed in such dynamic environments. We adopt a black-box mapping approach between the inputs, which are the noisy measured distances based on 5G RSSI readings and the 4 gNBs' coordinates, and the output, which is the estimated UAV coordinates.
- We propose a reinforcement learning based approach to provide a practical real-time calculation for the localization problem.
- We provide an in-depth comparative analysis between the proposed deep neural network and reinforcement learning based approaches to assess the efficiency and specific use case for each approach to solve the localization problem.
- We provide a comparative analysis with a benchmark solution to the UAV localization problem using cellular signals proposed in the literature based on cellular carrier phase measurements and weighted non-linear least squares estimation.
- We conduct the complexity analysis for our proposed deep and reinforcement learning approaches to examine the fulfilment of the objective of meeting the real time calculation requirement.

The rest of this paper is organized as follows. In Section II, we introduce our environmental assumptions and present our system model. In Section III, we formulate the 5G based 3-D UAV localization problem. In Section IV, we solve the formulated UAV localization problem through applying the proper optimization techniques to determine the optimal bound of the solution. In Sections V and VI, we propose two machine learning based techniques to provide real-time near-optimal localization. Specifically, we propose a deep supervised learning approach in Section V as well as a reinforcement learning approach in Section VI. In Section VII, we present and analyze our analytical results as well as present an alternative approach proposed in the literature based on Carrier phase measurements for comparison purposes. Finally, Section VIII concludes this paper.

**II. SYSTEM MODEL**

The RSSI measurements from at least four cellular base stations are needed to localize the UAV in 3D given that the intersection of 4 spheres (one sphere per gNB) is a point. Accordingly, we consider a suburban outdoor environment with four 5G gNBs located around 100–300m apart as shown in Fig. 1.



**FIGURE 1. Environmental setup and system model.**

Let the index  $1 \leq i \leq 4$  specify a given gNB<sub>*i*</sub> and let  $d_i$  represent the Euclidian distance between each gNB<sub>*i*</sub> and the UAV. The estimated coordinates of the UAV are denoted by the vector  $\mathbf{x}$  while the coordinates of the antennas of every gNB<sub>*i*</sub> are denoted by the vector  $\mathbf{g}_i$  where  $\mathbf{x}, \mathbf{g}_i \in \mathbb{R}^n$  and  $n = 3$ . Let  $h_i$  represent the height of the UAV reference to the antenna of gNB<sub>*i*</sub> and  $r_i$  represent the distance between the vertical projection of the UAV in the plane  $z = h_{gNB_i}$  and the gNB<sub>*i*</sub>.

We adopt the free-space path loss model in accordance with [9] and assume standard practical antenna configurations in accordance with [3], [4]. The measured received power  $Pd_i$  from each base station, given a transmit power of  $Pt_i$ , is given by

$$Pd_i = Pt_i - L(h_i, r_i), \tag{1}$$

where  $L(h_i, r_i)$  represents the probabilistic mean model path loss in both LoS and NLoS links and is given by

$$L(h_i, r_i) = L_{LOS} \times P_{LOS} + L_{NLOS} \times P_{NLOS}, \quad (2)$$

where  $L_{LOS}$  and  $L_{NLOS}$  represent the mean path losses in case of LoS and NLoS scenarios, respectively.  $P_{LOS}$  is the probability of LoS conditions between the UAV and the gNB<sub>i</sub> while  $P_{NLOS}$  is the probability of NLoS conditions.  $L_{LOS}$  and  $L_{NLOS}$  are given by

$$L_{LOS} = 20 \log \left( \frac{4\pi f_c d_i}{c} \right) + \eta_{LOS}, \quad (3)$$

$$L_{NLOS} = 20 \log \left( \frac{4\pi f_c d_i}{c} \right) + \eta_{NLOS}, \quad (4)$$

where  $f_c$  is the carrier frequency,  $c$  is the speed of light and  $L_{LoS}$  and  $\eta_{NLoS}$  represent the mean additional losses for LoS and NLoS, respectively.  $P_{LOS}$  and  $P_{NLOS}$  are functions of environment-dependent parameters,  $a$  and  $b$  and are given by

$$P_{LOS} = \frac{1}{1 + a \exp(-b(\frac{180}{\pi} \tan^{-1}(\frac{h_i}{r_i}) - a))}, \quad (5)$$

$$P_{NLOS} = 1 - P_{LOS} \quad (6)$$

Consequently, the path loss model can be expressed as

$$L(h_i, r_i) = \frac{A}{1 + a \exp(-b(\frac{180}{\pi} \theta_i - a))} + 20 \log \left( \frac{r_i}{\cos(\theta_i)} \right) + B, \quad (7)$$

where,

$$A = \eta_{LOS} - \eta_{NLOS}, \quad (8)$$

$$B = 20 \log \left( \frac{4\pi f_c d_i}{c} \right) + \eta_{NLOS} \quad (9)$$

$$\theta_i = \tan^{-1} \left( \frac{h_i}{r_i} \right) \quad (10)$$

Then the estimated distance,  $d_i$ , between gNB<sub>i</sub> and the UAV is given by

$$d_i = 10^{((L(h_i, r_i) - \frac{A}{1 + a \exp(-b(\frac{180}{\pi} \theta_i - a))} - B)/20)} \quad (11)$$

For every gNB<sub>i</sub>, measuring the received SNR and determining the estimated distance,  $d_i$ , between the UAV and gNB<sub>i</sub> will translate into an estimated location of the UAV denoted by the vector  $\mathbf{x}_i$  which is potentially any point on the surface of the sphere with center  $g_i$  and radius  $d_i$  given by:

$$(x_i - x_{gi})^2 + (y_i - y_{gi})^2 + (z_i - z_{gi})^2 = d_i^2 \quad (12)$$

The multi-lateration algorithm, to be described shortly, uses the RSSI measurements,  $P$ , and the measured path losses,  $L$ , between the UAV and each of the 4 gNBs given by

$$P = [Pd_1 Pd_2 Pd_3 Pd_4] \quad (13)$$

$$L = [L(h_1, r_1) L(h_2, r_2) L(h_3, r_3) L(h_4, r_4)] \quad (14)$$

The estimated distances,  $D$ , based on the noisy RSSI reading between the UAV and every gNB<sub>i</sub> are given by

$$D = [d_1 d_2 d_3 d_4] \quad (15)$$

We assume that the probability density function of the distance measurement noise,  $\varepsilon_i$ , follows a zero mean gaussian distribution in accordance with [1]. Accordingly, we model the distance  $d_i$  between the UAV and gNB<sub>i</sub> as

$$d_i = a_i + \varepsilon_i, \quad (16)$$

where  $a_i$  corresponds to the actual distance between the UAV and gNB<sub>i</sub>.

According to [16], 3 spheres in the 3-D space with radii  $d_i$ ,  $d_j$ , and  $d_k$  and centers given by the coordinates of gNBs  $g_i$ ,  $g_j$  and  $g_k$  are the vertices of the triangle defining a 2-D plane,  $T_{ijk}$ . The 3 spheres intersect at straight line orthogonal to the plane defined by  $T_{ijk}$ .

The intersection of this straight line with the plane  $T_{ijk}$  is given by point  $Z_l$  with barycentric coordinates given by

$$\begin{aligned} 16T_{ijk}^2 Z_l = & -2d_i^2 \|g_j - g_k\|^2 \\ & + d_j^2 (\|g_j - g_k\|^2 + \|g_k - g_i\|^2 - \|g_i - g_j\|^2) \\ & + d_k^2 (\|g_j - g_k\|^2 - \|g_k - g_i\|^2 + \|g_i - g_j\|^2) \\ & + \|g_j - g_k\|^2 (-\|g_j - g_k\|^2 + \|g_k - g_i\|^2 \\ & + \|g_i - g_j\|^2) \end{aligned} \quad (17)$$

Accordingly, we approximate the UAV location by  $\mathbf{x}_L = (x_L, y_L, z_L)$  as

$$x_l = Z_l + \frac{n_l \cdot |h_l|}{|n_l|}, \quad (18)$$

where the index  $L = [1, 4]$ ,  $n_l$  is the normal vector to the 2-D plane  $T_{ijk}$  and the vector  $h_l$  corresponds to the height of the UAV relative to plane  $T_{ijk}$ . Next, our objective is to find the estimated location  $\mathbf{x}_e = (x, y, z)$  that minimizes the error of measurements from the 4 gNBs.

### III. PROBLEM FORMULATION

We formulate the UAV localization problem as an optimization problem to minimize the sum of the least square estimate for the errors of the measured distances between the UAV and the  $m = 4$  gNBs. This problem can be written as

$$\begin{aligned} \text{Min}_{\mathbf{x}_e} \sum_{i=1}^m \text{Err}_i^2 \\ = \sum_{i=1}^m (\|\mathbf{x}_e - g_i\| - d_i)^2 \\ = \sum_{i=1}^m \left( \sqrt{(x - g_{xi})^2 + (y - g_{yi})^2 + (z - g_{zi})^2} - d_i \right)^2 \end{aligned}$$

subject to  $L(d_i, \theta_i) < L_{th_i}$

$$\text{Err}_i < \text{Err}_{th_i}$$

$$h_{uav} > h_{min}, \quad (19)$$



where  $L_{th_i}$  is the maximum tolerable path loss and is given by

$$L_{th_i} = \frac{A}{1 + a \exp(-b(\frac{180}{\pi}\theta_{opt} - a))} + 20 \log \left( \frac{r_{max}}{\cos(\theta_{opt})} \right) + B \quad (20)$$

The objective function in (19) results in a constrained non-linear programming problem (NLP) classified as NP-hard problem which is difficult to solve. Accordingly, we utilize the penalty method to reformulate the objective function in the form of an unconstrained NLP problem given by

$$\begin{aligned} & \text{Min}_{x_e} \psi_i C_i \\ & + \sum_{i=1}^m \left( \sqrt{(x - g_{xi})^2 + (y - g_{yi})^2 + (z - g_{zi})^2} - d_i \right)^2 \end{aligned} \quad (21)$$

where the penalty coefficient  $\psi_i > 0$ , and the constraints are given by

$$C_i = \max(0, L(h, r_i) - L_{th}, Err_i - Err_{thi}, h_{min} - h_{uav}) \quad (22)$$

For each gNB<sub>i</sub>, we assume each distance measurement noise,  $\varepsilon_i$ , is an independent random variable with standard deviation  $\sigma_i$  and probability density function given by

$$f(\varepsilon_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\varepsilon_i}{\sigma_i} \right)^2} \quad (23)$$

The joint probability density function of the distance measurement errors is given by the product of the four probability densities [31] given by

$$f(\varepsilon) = \prod_{i=1}^m \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\varepsilon_i}{\sigma_i} \right)^2}, \quad (24)$$

where

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m]^T \quad (25)$$

Next, we derive theoretical lower and upper bounds for the localization error vector,  $\varepsilon$ , to analyze the impact of the accuracy of the RSSI measurements from the four gNBs on the UAV localization error. The Cramér-Rao Lower Bound (CRLB) is one of the commonly utilized metrics for characterizing the lower bound of the accuracy limitations in RF-based localization applications [5]. The CRLB is calculated through the inverse of the covariance matrix of the RSSI measurements errors known as the Fisher information matrix [32]. Let the Fisher information matrix,  $F$ , be defined as an  $m \times m$  matrix with element  $F_{i,j}$  given by

$$F_{i,j} = -\mathbb{E} \left[ \frac{\partial^2}{\partial \varepsilon_i \partial \varepsilon_j} \log f(\varepsilon) \right] \quad (26)$$

The CRLB,  $\varepsilon_{LB}$ , for the distance measurement error,  $\varepsilon_i$ , is given by

$$\varepsilon_{LB} = F_{i,i}^{-1} \quad (27)$$

We derive an upper bound to guarantee a minimum confidence level,  $C_{min}$ , for the UAV localization error. Specifically, the upper bound,  $\varepsilon_{UB}$ , for the distance measurement noise,  $\varepsilon_i$ , satisfies

$$\int_{-\varepsilon_{UB}}^{\varepsilon_{UB}} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\varepsilon_i}{\sigma_i} \right)^2} d\varepsilon_i \geq C_{min} \quad (28)$$

As such, the theoretical lower and upper bounds for the UAV localization error can be established based on the channel propagation conditions of the RSSI measurements from the four gNBs.

#### IV. THE OPTIMIZATION BASED APPROACH

Our objective in this section is to accurately solve the localization problem in (21). Several techniques can be utilized to provide optimal and near optimal solutions with high accuracies. The exact solution can be found by utilizing the Exhaustive Search method bounded by the coverage area of the 4 gNBs. Despite its accuracy, this technique involves a significant complexity given the large area in which the UAV is allowed to fly. To illustrate its complexity, we perform a complexity analysis to be presented in the form of the big-O notation. We let  $n$  to denote the algorithm's variable input size including the RSSI readings from  $m$  gNBs as measured by a UAV to be localized in 3-D at a given time instant  $t$  and their corresponding coordinates. In Table 1, we illustrate the computational function of each complexity notation. Initially, we solve the multilateration algorithm which contributes a linear complexity of  $O(n)$ . To this end, we state the worst-case complexity of the computational steps for each optimization algorithm. The complexity is proportional to the flying space  $A$  of the UAV bounded by the coverage area of the 5G gNBs which is typically very large. The complexity can thus be given by

$$\text{Complexity}_{ES} = O(nA) \quad (29)$$

To solve the given localization problem at hand without the substantial number of iterations and complexity, we propose to utilize the Nelder-Mead (NM) algorithm as a heuristic optimization approach. The Nelder-Mead algorithm is convenient and efficient in solving unconstrained multidimensional objective functions and does not require any differentiation to search for the optimal point [17]. The Nelder-Mead algorithm is composed of four operations, namely, reflection, expansion, contraction, and shrink [18]. The reflection, expansion, contraction and shrink coefficients are given by  $\alpha$ ,  $\gamma$ ,  $\rho$  and  $\sigma$ , respectively. As shown in Algorithm 1, we use the estimates obtained by equation (15) as the initial test points.

Next, we analyze the computational complexity of the steps described in Algorithm 1. This method includes an outer loop of  $n$  iterations and an inner loop of  $t$  iterations. Accordingly, the computational complexity of the Nelder-Mead based localization technique can be given in the form of big-O notation by

$$\text{Complexity}_{NM} = O(nt) + O(nt \log(m)) \quad (30)$$

**Algorithm 1** Nelder Mead Based Optimization

- Input:**  $x_1, \dots, x_{n+1}, \alpha, \gamma, \rho, \sigma$
- 1: **while** (termination criteria not met)
  - 2: Order the vertices of the simplex such that:  
 $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$ .
  - 3: Calculate the centroid  $x_o$  of all points except  $x_{n+1}$ .
  - 4: Calculate the reflection of  $x_{n+1}$  given by  $x_r = x_o + \alpha(x_o - x_{n+1})$  where  $\alpha > 0$ .
  - 5: Expansion. If  $f(x_r) < f(x_1)$ , then calculate  $x_e = x_o + \gamma(x_r - x_o)$ , where  $\gamma > 1$ . If  $f(x_e) < f(x_1)$ , replace  $x_{n+1}$  with  $x_e$ ; otherwise, replace  $x_{n+1}$  with  $x_r$ .
  - 6: Reflection. If  $f(x_1) \leq f(x_r) < f(x_n)$ , then replace  $x_{n+1}$  with  $x_r$ .
  - 7: Contraction. If  $f(x_r) > f(x_n)$ , then compute  $x_c = x_o + \rho(x_{n+1} - x_o)$  where  $0 < \rho \leq 0.5$ . If  $f(x_c) < f(x_{n+1})$ , then replace  $x_{n+1}$  by  $x_c$ .
  - 8: Shrink. Replace all points except  $x_1$  by  $x_i = x_1 + \sigma(x_i - x_1)$  where  $\sigma > 0$ .
  - 9: **End**

The computational complexity for both approaches is proportional to the number of iterations needed to reach the optimal solution. As such, we conclude that Exhaustive Search and Nelder-Mead algorithms do not guarantee to meet the real-time requirement but are to be used to determine the true bound of the solution as benchmarks for comparison to our state-of-the-art machine learning based approaches to be proposed in the next sections.

**TABLE 1.** Computational complexity illustration for algorithm steps.

Time Complexity Description	Big-O notation	Computational Function
Constant time	$O(1)$	Basic calculation
Logarithmic time	$O(\log n)$	Binary search to find the best in $n$ combinations
Linear time	$O(n)/O(t)$	Basic calculations for $n$ combinations or $t$ times
Linearithmic time	$O(n \log A)$	Finding the best in $A$ combinations for $n$ iterations

**V. THE PROPOSED DEEP LEARNING BASED APPROACH**

Our objective in this section is to develop a machine learning based model capable of capturing the non-linear co-related relationships between the UAV location and the measured 5G RSSI readings from the 4 gNBs. In this section, we deal with the optimization problem as a black-box mapping process between the inputs and the outputs. We propose to leverage deep supervised learning to solve the formulated error minimization problem in a computationally efficient manner such that it can be practically used for real-time UAV localization. We propose utilizing an Artificial Neural Network (ANN) as a framework to learn the mapping between the inputs and outputs. There are various types of ANNs. We propose to leverage a Multi-Layer Perceptron (MLP) composed of a deep feedforward neural network architecture as no feedback loop is required between the inputs.

The model is to be trained offline to resolve the 3-D UAV localization problem as a regression problem as opposed to a classification problem to enable scalability with the 3-D UAV localization area. The training data needed for the proposed deep supervised learning approach are to be synthetically generated through applying the Nelder-Mead and Exhaustive Search optimization-based techniques that we discussed in section IV.

As summarized in Algorithm 2, we deal with the optimization problem as a black-box mapping process between the input, which is the noisy measured distances based on the 5G RSSI readings and the gNB coordinates and the output, which is the estimated UAV coordinates. Therefore, we train a feed-forward neural network to learn this mapping. As shown in Fig. 2, the input vector in a  $16 \times 1$  vector composed of the 4 gNBs 3-D coordinates and their corresponding RSSI readings while the output is a  $3 \times 1$  vector corresponding to the estimate of the 3-D UAV location. We study the effect of various neural network architectures, backpropagation training functions and hyper parameter tuning on the performance evaluation of the proposed deep learning algorithm. The neural network weights are trained in a supervised manner with forward and back-propagation algorithms to optimize a given performance function [19].

There are various implementations for the backpropagation building on the standard backpropagation algorithm. Specifically, we use Levenberg-Marquardt backpropagation and Bayesian Regularization which are two backpropagation training functions used for neural network approximations. The Levenberg-Marquardt (LM) algorithm is a robust backpropagation algorithm for performance function approximation [21], [22]. It is a pseudo-second order method which estimates the Hessian matrix, which is a square matrix used to describe the local curvature of the function, using the sum of outer products of the gradients. The Levenberg-Marquardt algorithm does not consider the outliers in the data, which may lead to overfitting noise. To overcome this problem, Bayesian Regularization can be applied to the neural network learning problem. Bayesian Regularization expands the cost function used by the model to minimize the error as well as the effective number of parameters using the minimal weights [23]. Bayesian Regularization adds little overhead to the Levenberg-Marquardt process given that a Hessian approximation is already computed.

Next, we perform the complexity analysis of the proposed deep learning approach to be presented in the form of the big-O notation. The complexity of the proposed deep learning algorithm operating in the online phase has a linear value that is calculated as

$$Complexity_{DL} = O(Cn), \quad (31)$$

where  $C > 0$  is a constant corresponding to the neural network dimensions given by

$$C = n_{inputs} \times n_{Layer1} \times n_{Layer...} \times n_{LayerH_d} \times n_{outputs}, \quad (32)$$

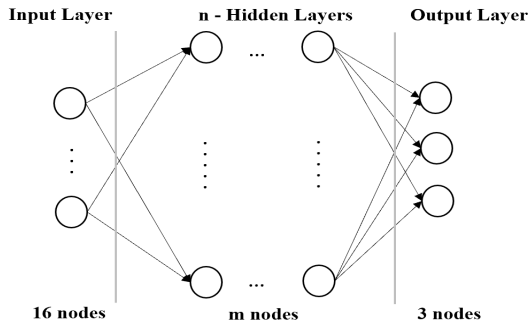


FIGURE 2. n-layer feed-forward neural network architecture.

**Algorithm 2** Deep Supervised Learning Approach

**Input:** The set of gNB coordinates, the corresponding RSSI readings and estimated UAV coordinates.

- 1: Algorithm hyperparameters: Initial  $\mu \in [0,1]$ ,  $\mu$  increase factor,  $\mu$  decay factor, max  $\mu$ , no. of training epochs  $K$ , Performance goal  $\epsilon$ , Batch size.  $\mathcal{B}$ , back propagation training function FCN.
- 2: Random initialization of network weights and biases
- 3: **for** training epochs  $\leftarrow 1$  to  $K$
- 4: **while** ( $MSE > \epsilon$ ) **do**
- 5: Sample random batch size  $\mathcal{B}$  from the training data.
- 6: Compute performance evaluation for training and validation data sets.
- 7: Perform forward and back propagation gradient descent training function FCN.
- 8: Update neural network hidden layer weights and biases.
- 9: **if** ( $gradient < min\_grad$ ) **then**
- 10: break // end Epoch
- 11: **if** ( $trainFCN=trainLM$ ) **and** ( $validation > max\_val$ ) **then**
- 12: break // end Epoch
- 13: **if** ( $MU > mu\_max$ ) **then**
- 14: break // end Epoch
- 15: **end**
- 16: **end**

where  $H_d$  represents the number of hidden layers of the neural network. The calculated computational complexity in case of the deep learning based technique is proportional only to the neural network size as opposed to the number of iterations required to compute the optimization based techniques. As such, the proposed deep learning-based technique can be used in real-time to solve the formulated UAV localization problem.

**VI. THE PROPOSED REINFORCEMENT LEARNING BASED APPROACH**

As mentioned earlier, our objective is to come up with a real-time efficient solution for the UAV 3-D localization problem. Therefore, we also devised a RL based algorithm in order to be able to determine the most suitable ML based technique for this problem. RL is a branch of ML that learns directly by interacting with the environment. It addresses problems where there is no explicit training data available. Q-Learning (QL) is a type of RL learning involving agents which make decisions, observe the results, and then automatically adjust

their policies to achieve their objective to select an action that maximizes a reward. It is worth noting that a QL agent can be trained in an offline or online implementation. The traditional table-based QL, proposed by Watkins [10], maximizes the expected value of the total reward over any and all successive steps by taking action in the current state and follows an optimal policy afterwards. The tabular QL does not scale with the increase in the size of state space. In most real applications, there are too many states to visit and keep track of. For scalability, function approximation is used to approximate a value function of each state-action pair through a number of iterations. An artificial neural network representation, also known as deep Q-network (DQN), can be used for non-linear function approximations [24]–[27]. The goal is to select the action which has the maximum Q-value using the following update rule

$$Q(s, a) \leftarrow (1 - \alpha) Q(s, a) + \alpha \left[ r(s) + \gamma \max_{a'} Q(s', a') \right] \tag{33}$$

where  $s$  is the state,  $a$  is the action,  $\alpha$  is the learning rate and  $r$  is the reward attained for the current state-action pair. The discount factor  $\gamma$  determines the importance of future rewards, i.e. a high discount factor sets the priority towards distant rewards, whereas a lower value will force the agent to consider only immediate rewards. In the next subsections, we first present the Q-learning environmental setup. We then detail the proposed QL architecture and the overall operation for the formulated UAV localization problem.

**A. Q-LEARNING ENVIRONMENT**

In this subsection, we define the state, action and reward for the Q-learning agent. We define the state,  $s$ , as a  $16 \times 1$  vector composed of the 3-D coordinates and the corresponding RSSI readings from the 4 gNBs as follows

$$s = [g_{x1} \ g_{y1} \ g_{z1} \ g_{x2} \ g_{y2} \ g_{z2} \ g_{x3} \ g_{y3} \ g_{z3} \ g_{x4} \ g_{y4} \ g_{z4} \ d_1 \ d_2 \ d_3 \ d_4] \tag{34}$$

Let the action space  $A$  correspond to the allowable flight area of the UAV bounded by the coverage area of the 4 gNBs to maintain the SINR threshold along with the minimum height requirement of the UAV as defined by the constraints in equation (19). We discretize the 3-D action space,  $A$ , to equally spaced increments. For example, considering a  $200 \times 200 \times (h_{max} - h_{min}) m^3$  area, it can be divided into 1m increments along each axis. Assuming  $h_{min} = 30m$  and  $h_{max} = 100m$ , then  $A$  is a  $3 \times 1$  vector with  $n = 1 \times 2868471$  discrete possible elements. We also investigate the effect of discretizing the action space into smaller sized increments e.g. 0.1m in our evaluation to enable localizing the UAV to a decimeter accuracy.

We define the action  $a \in A$  as a  $3 \times 1$  vector corresponding to the estimated UAV location. It can be written as

$$a = [x_{uav} \ y_{uav} \ z_{uav}] \tag{35}$$

We define the reward as the negative of mean square error as follows

$$r = - \sum_{i=1}^m Err_i^2 = - \sum_{i=1}^m (||\mathbf{x}_e - \mathbf{g}_i|| - d_i)^2 \quad (36)$$

We approximate Q-value function by setting the discount factor  $\gamma$  to zero to consider only the immediate reward.

### B. DEEP Q-NETWORK

In this subsection, we present the proposed architecture to estimate the Q-value function. We utilize a DQN representation, termed as the critic network, to estimate the Q-value for a given state-action pair. The critic network architecture is shown in Fig. 3. The input to the neural network is a  $19 \times 1$  input vector composed of the state and action pair and the output is a scalar estimate of the immediate reward. The network is composed of an addition layer followed by H-hidden layers with tanh activation functions. The Q-learning agent interacts directly with the environment and the critic network is trained according to set hyper parameters. We train the critic network by applying experience replay and batch training where the union of the state, action, reward, and transition of each step is stored as an item in the experience pool, and a particular number of samples from the experience pool are selected to feed into the neural network in each step to do the weighting parameters' update [30]. We apply Bayesian regularization back propagation to determine the critic network learnable parameter,  $\theta$ , that minimizes the MSE for the Q-value approximation.

### C. THE OVERALL Q-LEARNING APPROACH

In this subsection, we detail the overall algorithm and computational complexity for the proposed reinforcement learning based approach. As shown in Fig. 4, the current state  $16 \times 1$  vector,  $s$ , is input to the agent. The agent multiplies the state vector,  $s$ , by a  $1 \times n$  unit vector  $J$ . The resulting  $16 \times n$  vector is then applied to the addition layer of critic network to evaluate the approximate Q-value for each state-action pair. The action with the maximum Q-value is then selected as the most likely estimate for the UAV location. The proposed reinforcement learning based approach is summarized in Algorithm 3.

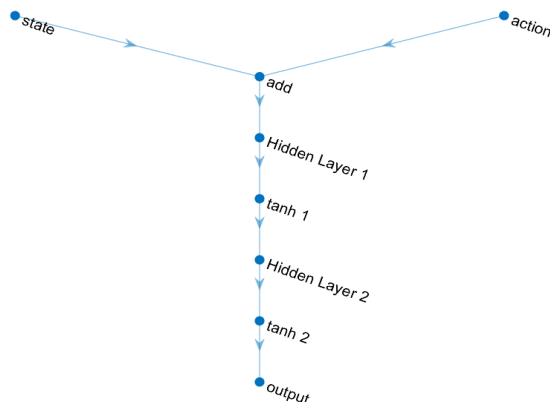


FIGURE 3. Critic network architecture.

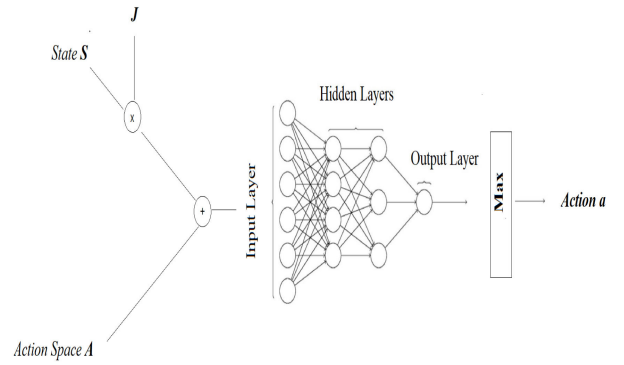


FIGURE 4. Q-learning algorithm representation for the online stage.

We now perform the complexity analysis of the proposed RL based approach to be presented in the form of the big-O notation. In this case, the complexity is proportional to the discretization step size of the flying space of the UAV bounded by the coverage area of the 5G gNBs. To estimate a UAV location at a single point, we first resize the state vector by performing a copy operation with constant time complexity. A critic network is used to approximate the function mapping between the inputs corresponding to the gNB coordinates, the RSSI readings and an estimate of the UAV location and the output which is a scalar value corresponding to the maximum reward. The complexity of the critic network can thus be given by

$$Complexity_{critic} = O(Cn), \quad (37)$$

where  $C > 0$  is a constant corresponding to the neural network dimensions given by

$$C = n_{inputs} \times n_{Layer1} \times n_{Layer2} \times n_{LayerH_r} \times n_{outputs}, \quad (38)$$

where  $H_r$  represents the number of the hidden layers of the critic network. Finally, the action with the maximum reward is selected as the estimated UAV location. Consequently, the complexity of the reinforcement learning algorithm is given by

$$Complexity_{RL} = O(Cn) + O(n \log A) \quad (39)$$

The calculated  $Complexity_{RL}$  is proportional to the critic network size and the discrete action space independent of any iterations needed to converge to an optimal solution as is the case of optimization-based techniques. As such, the proposed RL based technique can be used solve the formulated UAV localization problem to meet the real time requirement.

## VII. EVALUATION RESULTS

In this section we conduct simulations to evaluate the performance of each of proposed approaches to solve the UAV localization problem formulated in section III. We consider a  $500 \times 500 \times 100m^3$  outdoor flight area with the system and environmental parameters settings as shown in Table 2. We assume random noise with zero mean gaussian distribution and standard deviation,  $\sigma_1$ . As shown in Table 3,



**Algorithm 3** Reinforcement Learning Approach

**Input:** Action space  $A$ , Batch size  $\mathcal{B}$ , Experience replay buffer length  $\mathbb{E}$ .

- 1: Random initialization of network weights and biases
- 2: **for** training episodes  $\leftarrow 1$  to  $K$
- 3: **while**  $(MSE > \epsilon)$  **do**
- 4: Sample random batch size  $\mathcal{B}$  from the training data.
- 5: Compute performance evaluation for training and validation data sets.
- 6: Perform forward and back propagation gradient descent training function FCN.
- 7: Update neural network hidden layer weights and biases.
- 8: **if**  $(gradient < min\_grad)$  **then**
- 9: break // end Epoch
- 10: **if**  $(trainFCN=trainLM)$  and  $(validation > max\_val)$  **then**
- 11: break // end Epoch
- 12: **if**  $(MU > mu\_max)$  **then**
- 13: break // end Epoch
- 14: **end**
- 15: **end**

we consider various gNB-UAV placement scenarios. The deep learning hyper parameter settings used during the training phase are shown in Table 4. The reinforcement learning agent hyper parameter settings used to train the critic network are summarized in Table 5. We consider the mean-square error (MSE) merit function to approximate the mapping of the weights of the neural networks during the offline phase. We evaluate the optimal results solved by utilizing the Exhaustive search, Nelder-Mead, as well as a Carrier phase optimization-based approach proposed in [8] to be described in detail in section VII.A for benchmarking purposes. Then, we perform an in-depth comparative analysis for performance of each of our proposed deep neural network and reinforcement learning based approaches as compared to the optimal benchmark techniques. Specifically, we assess the efficiency and specific use case for each of the deep neural network and reinforcement learning based approaches to solve the localization problem. In our proposed deep and reinforcement learning approaches, the neural networks are trained during the offline phase with the objective of providing a real-time near-optimal localization solution. All the results are based on a data size of 100 runs and 98% confidence analysis.

**TABLE 2.** Environmental and system parameters.

<b>5G Parameters</b>	Range	FR1
	Band	n1
	F	2.1 GHz
<b>Environment Parameters</b>	A	5.0188
	B	0.3511
	$\eta_{LoS}$	0.1
	$\eta_{nLoS}$	21
<b>Penalty Coefficient</b>	$\psi_i$	1000
<b>Max. Error</b>	$Err_{th}$	3m

**A. THE CARRIER PHASE APPROACH**

In order to compare our proposed techniques to a representative technique from the literature, we present an optimization-based approach utilizing carrier phase measurements which

**TABLE 3.** Base station and UAV placement scenarios.

Scenario #	Placement
1	gNBs are uniformly placed ~150 m apart and the UAV is conveniently located in the middle above the gNBs.
2	gNBs are uniformly placed ~150 m apart and the location of the UAV relative to gNBs is random.
3	gNBs are randomly placed and the UAV is located above in the middle relative to the gNBs.
4	gNBs are randomly placed and the location of the UAV is random.

**TABLE 4.** Deep learning neural network parameters.

<b>Data</b>	Training Data Size	$10^5 - 1.5 \cdot 10^6$
	Testing Data Size	$10^4$
	Standard Deviation	Random [0.01-0.1]
<b>Neural Network Architecture</b>	Layers	1,2,3,4
	Nodes per Layer	40, 70, 100, 120
	Batch Size	$10^3 - 10^5$
	Training Epochs	40-240
<b>Levenberg-Marquardt (LM)</b>	Initial MU	0.001
	MU Inc. Factor	10
	MU Decay Factor	0.1
	Performance	MSE
<b>Bayesian Regularization (BR)</b>	Initial MU	0.005
	MU Inc. Factor	10
	MU Decay Factor	0.1
	Performance	MSE

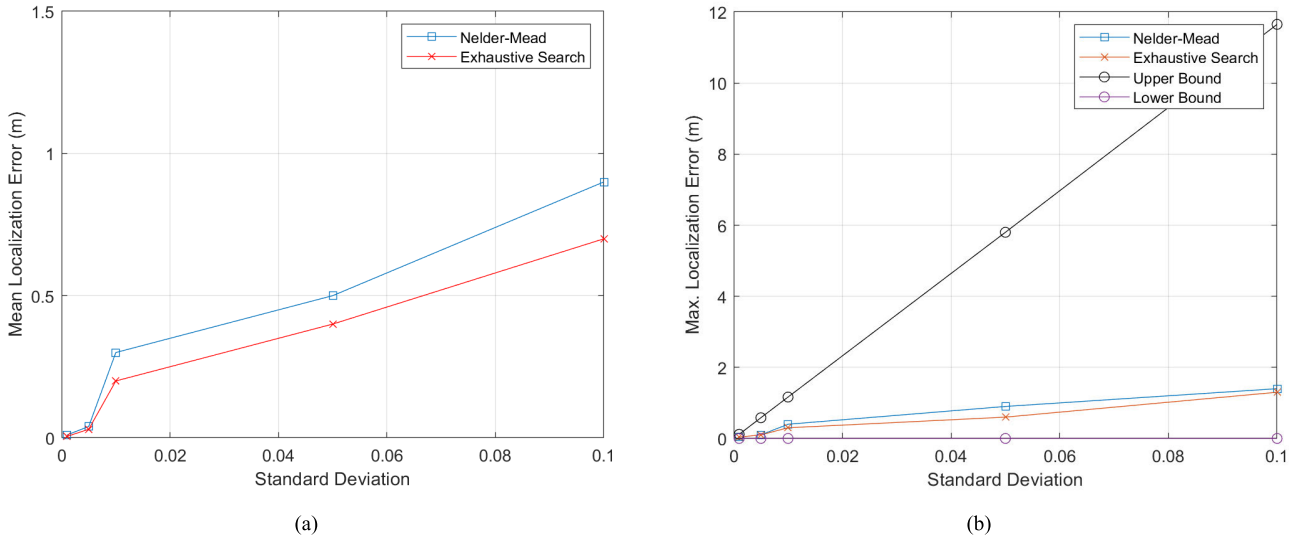
**TABLE 5.** Reinforcement learning agent hyper parameters.

<b>QL Agent</b>	Experience Replay Buffer Length $\mathbb{E}$	$10^5$
	Batch size $\mathcal{B}$	$10^3 - 10^5$
	Episodes	$10^3 - 10^6$
	Number of Steps	1
	Action space discretization step	1m - 0.1m
<b>Critic Network</b>	Number of nodes per hidden layer	10 - 30
	Number of layers	1 - 3
	Activation function	Tanh
	Backpropagation Train FCN	Bayesian Regularization
	Initial MU	0.005
	MU Inc. Factor	10
	MU Decay Factor	0.1
Training Epochs	MSE	

is proposed in [8] to solve the UAV localization problem. The authors utilize cellular signals to localize and navigate a UAV assuming limited GPS presence. The authors assume the initial presence of GPS to accurately determine the carrier phase ambiguity for each of the N base stations then leverage the relative stability of BSs clocks to estimate the UAV location at later stage when GPS is cut off. The carrier phase observable in meters  $z_i$  from  $BS_i$  at time  $t = k$ , corresponding to equation (11), can be expressed as

$$z_i(k) = \lambda \phi_i + \lambda T \sum_{l=0}^{k-1} f_{D_i}(l) + v_i(k), \quad (40)$$

where  $\lambda$  is the wavelength of the carrier signal,  $\phi_i$  is the observed carrier phase reading for  $BS_i$ , T is the carrier period



**FIGURE 5. Optimization-based localization results (a) Mean localization error for exhaustive search vs. Nelder-Mead; (b) Maximum localization error as compared to theoretical bounds for exhaustive search vs. Nelder-Mead.**

and  $v_i(k)$  is the zero mean gaussian measurement noise. The carrier phase observable can be re-parametrized in terms of UAV and BS positions as follows

$$z_i(k) = \left\| r_{uav}^T(k) - r_i^T \right\| + c [\delta t_{uav}(k) - \delta t_i(k)] + \lambda N_i + v_i(k), \quad (41)$$

where  $r_{uav}$  and  $r_i$  are the UAV and BS<sub>*i*</sub> 2-D position vectors, respectively, whereas  $\delta t_{uav}$  and  $\delta t_i$  are the receiver and cellular BS clock biases, respectively,  $c$  is the speed of light and  $N_i$  is the carrier phase ambiguity for BS<sub>*i*</sub>. The terms  $c [\delta t_{uav} - \delta t_i] + \lambda N_i$  can be combined into one term that can be defined as

$$c\delta t_i(k) = c \left[ \delta t_{uav}(k) - \delta t_i(k) + \frac{\lambda}{c} N_i \right] \quad (42)$$

According to the study, cellular BSs possess tight carrier phase synchronization, which results in very similar clock biases up to an initial bias  $c\delta t_{0i}$ , and proceed to leverage this relative frequency stability to reduce the parameters to be estimated. Accordingly, the authors reparametrize the clock biases in terms of a time varying common bias term,  $c\delta t$ , as well as BS<sub>*i*</sub> specific clock bias deviation term,  $\epsilon_i(k)$ , as follows

$$c\delta t_i(k) = c\delta t(k) + \epsilon_i(k) \quad (43)$$

Accordingly, the carrier phase observable is expressed as

$$z_i(k) = \left\| r_{uav}^T(k) - r_i^T \right\| + c\delta t(k) + c\delta t_{0i} + \eta_i(k), \quad (44)$$

where  $\eta_i(k) = \epsilon_i(k) + v_i(k)$  is the overall measurement noise and  $c\delta t_{0i}$  is the initial carrier phase bias term that is estimated at the beginning of the setup during the GPS availability duration by knowing the initial UAV position and measurement  $z_i(0)$  and is given by

$$c\delta t_{0i} = \left\| r_{uav}^T(0) - r_i^T \right\| - z_i(0) \quad (45)$$

The above estimation ignores the initial measurement noise. However, the authors propose to average several measurements for the total duration in which the GPS is initially available. To estimate the common bias term  $c\delta t(k)$ , the authors propose to lump the  $N$  BS<sub>*i*</sub> clocks into clusters, each of size  $N_l$  where  $N = \sum_{l=1}^L N_l$  and  $L$  is the total number of clusters. Extending this approach to our application scenario to localize the UAV through measurements from four base stations, we let  $N = 4$  and  $L \leq 2$ . Note that, given the 2-D position of the UAV is being estimated with  $L$  clock clusters, the number of clusters,  $L$ , cannot exceed  $N-2$ . Next, the authors utilize a weighted non-linear least squares (WNLS) estimator to estimate the UAV location given by

$$x^{j+1}(k) = x^j(k) + \left( H^T R_\eta^{-1} H \right)^{-1} H^T R_\eta^{-1} \delta z(k), \quad (46)$$

where  $j$  is the WNLS iteration index,  $R_\eta$  is the measurement noise covariance and  $H$  is the measurement Jacobian and  $\delta z(k)$  is given by

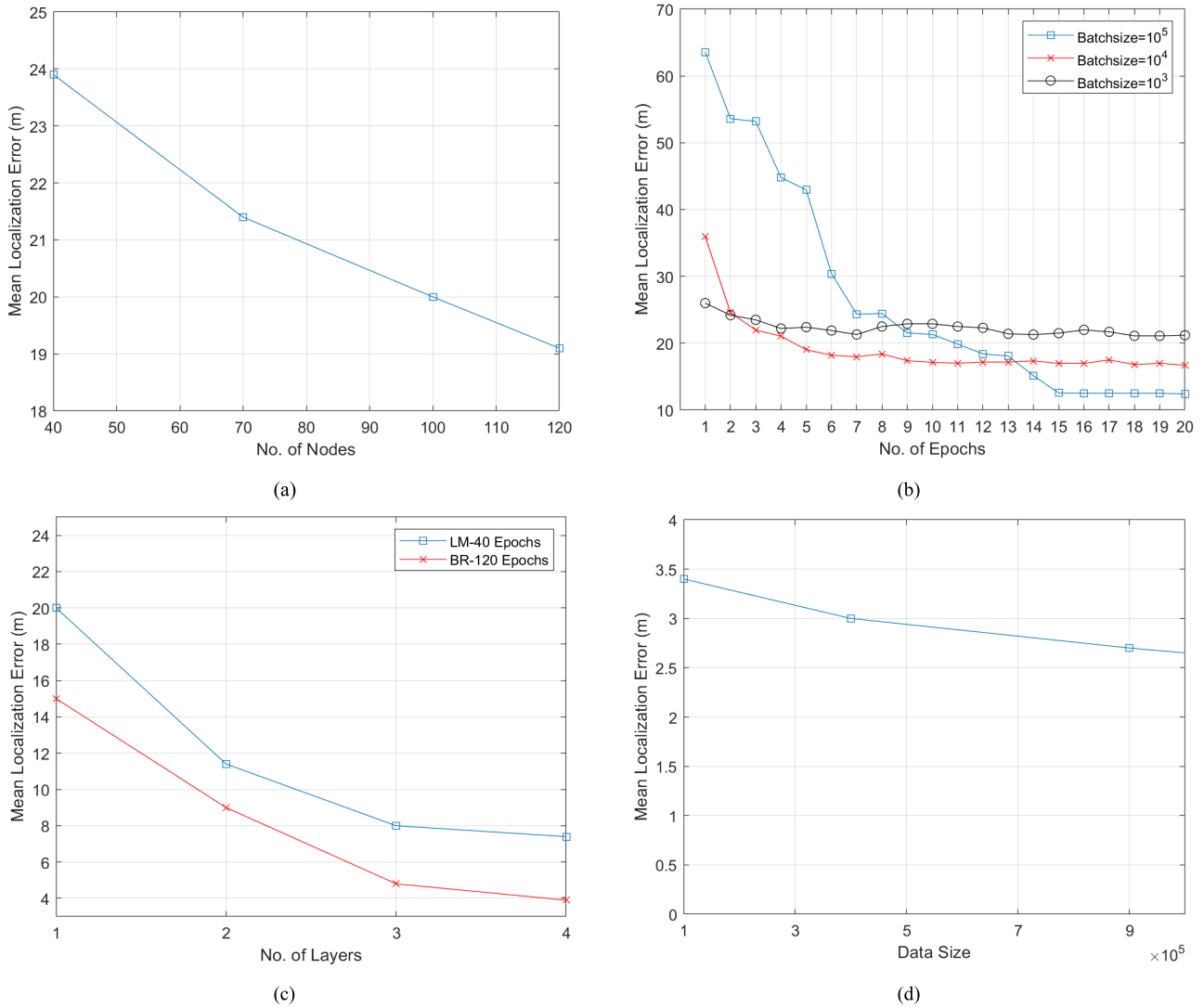
$$\delta z(k) = [\delta z(k)_1, \dots, \delta z(k)_N]^T \quad (47)$$

This method contains an outer loop of iterations  $n$  and an inner loop of iterations  $j$ . Accordingly, the computational complexity of this localization technique is given by

$$Complexity_{CP} = O(njN) + O(nj \log(N)) \quad (48)$$

The calculated  $Complexity_{CP}$  is proportional to the needed number of iterations of the computational steps to converge to an optimal solution and as such is not guaranteed to meet the real time requirement.

It is worth noting that this approach has major drawbacks that render it impractical where GPS presence in urban areas has the limitations that we discussed earlier. Moreover, the study's results are based on assuming an altimeter is available to accurately estimate the UAV height and it then proceeds to



**FIGURE 6.** Deep learning-based localization results (a) Mean localization error vs. no. of nodes for a 1-Layer NN; (b) Mean localization error vs. no. of Epochs for a 2-Layer NN; (c) Mean localization error vs. no. of layers; (d) Mean localization error vs. data size.

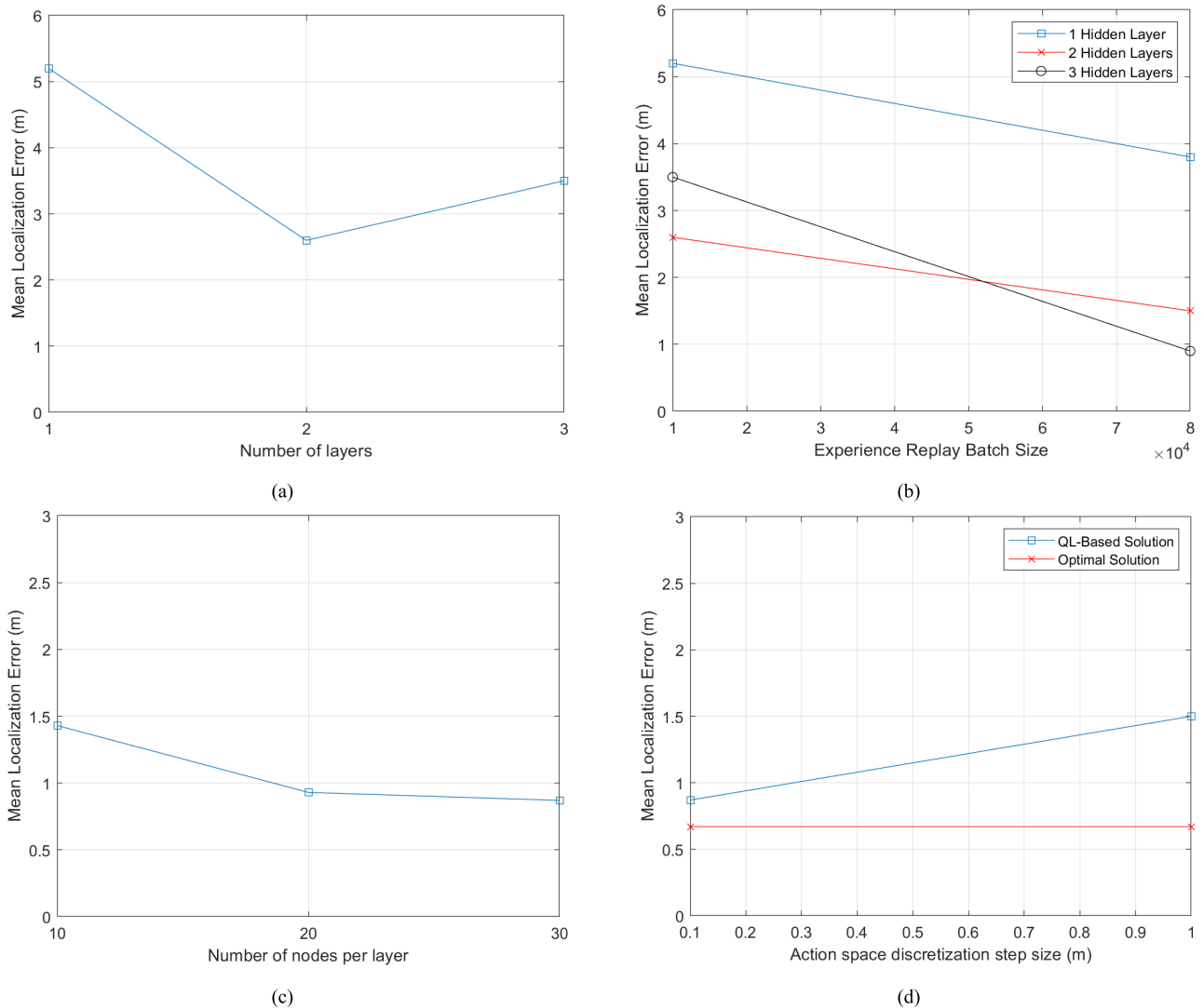
estimate the 2-D position vector for the UAV due to the poor vertical diversity of cellular phase measurements.

**B. THE OPTIMIZATION BASED RESULTS**

We first perform simulations to determine the optimal bound of the solution by evaluating the performance of Exhaustive Search technique which can usually be done on smaller system scales due to its complexity. We then assess the performance of the Nelder-Mead technique in comparison to the exhaustive search to examine its relative performance to the global optimal solution obtained by the Exhaustive Search. The objective is to enable us to examine the possibility of using the Nelder-Mead technique at larger scales, instead of resorting to the use of exact optimization techniques, due to its lower complexity. This is to help researchers decide on the use of such an alternative in similar problems to save

considerable time and resources while obtaining reasonably close optimal bounds.

As shown in Fig. 5(a), we demonstrate the mean UAV localization error assuming random placement for each of the four gNBs and UAV versus different standard deviation values of the noisy RSSI readings. The results show that Nelder-Mead as a heuristic optimal approach performs closely to Exhaustive Search. By examining the results, we find that the Nelder-Mead and Exhaustive Search methods result in a mean localization error of 0.7 and 0.5 for a random UAV-gNB placement scenario and random standard deviations  $0.01 < \sigma_i < 0.1$ , respectively. In Fig. 5(b), we show the maximum localization error for both algorithms as compared to the bounds derived in section III. We evaluate  $\epsilon_{UB}$  based on a 98% confidence analysis and evaluate  $\epsilon_{LB}$  assuming optimal propagation conditions for each of the four RSSI measurements. As expected, the maximum UAV



**FIGURE 7. Reinforcement learning-based localization results (a) Mean localization error vs. number of layers; (b) Mean localization error vs. experience replay batch size; (c) Mean localization error vs. number of nodes per hidden layer; (d) Mean localization error vs. action space discretization step size.**

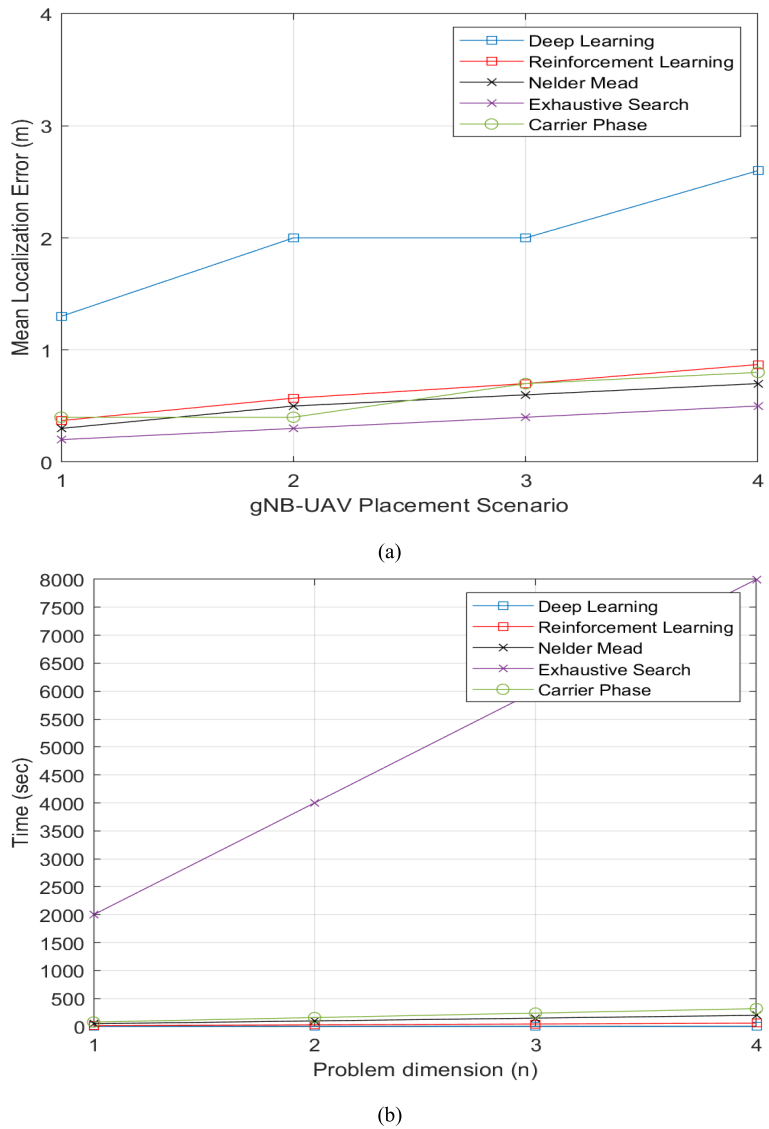
localization error for both is within the theoretical bounds for both optimization based algorithms.

### C. THE DEEP LEARNING BASED RESULTS

In this section we evaluate the performance of the deep supervised learning-based approach in solving our formulated UAV localization problem. Fig. 6 demonstrates the effect of various neural network architectures, backpropagation training functions and hyper parameter tuning on the performance evaluation of the proposed deep learning algorithm. As shown, the localization accuracy improves as the number of nodes increases since the higher number of nodes enables the neural network to learn the non-linear correlations. We also investigate the effect of varying the batch size on the neural network performance. As shown in Fig. 6, the smaller batch sizes accelerate the training converging

to suboptimal accuracy as opposed to using larger batch sizes. The performance also improves with increasing the number of network layers. However, the rate of improvement slows down as the number of layers increases as the network becomes susceptible to overfitting due to the increase in the number of parameters. We also demonstrate the effect of training the neural network using different functions. The mean square error is lower in case of applying Bayesian Regularization as opposed Levenberg-Marquardt given that the validation criteria prevents overfitting while allowing the training for a larger number of epochs. We also show that the localization accuracy increases with increasing the number of training epochs. However, the performance improvement decreases significantly when the number of training epochs exceeds 220 epochs as the network weights converge to the local suboptimal solution.





**FIGURE 8.** Overall analysis (a) Mean localization error vs. gNB-UAV placement scenario; (b) Time complexity vs. problem dimension.

Our simulation results show that a neural network architecture that consists of 4 hidden layers with at least 100 nodes per layer is needed to capture the non-linear relationships between the inputs and the outputs. We also studied the effect of training the neural network by utilizing training data obtained from applying the different optimization algorithms. Our simulation results show a neural network with a 4 hidden layers architecture trained for at least 260 epochs with a minimum training data size of  $10^6$  and utilizing Bayesian regularization yields the best results. Our simulation results also show that the mean localization error of a 4-layer neural network trained for 280 epochs utilizing Bayesian regularization with a training data size of  $10^6$  obtained by applying the Exhaustive Search and Nelder-Mead algorithms is 2.6m and 2.7m, respectively. This confirms our earlier observation regarding the closeness of the optimal bound results of both

the Exhaustive Search and the Nelder-Mead optimization techniques.

**D. THE REINFORCEMENT LEARNING BASED RESULTS**

In this section, we evaluate the performance of the reinforcement learning-based approach that we proposed as compared to the optimal solution obtained by applying the Exhaustive Search and Nelder-Mead techniques and the near optimal solution obtained by applying the deep learning approach. As shown in Fig. 7, we see the effect on performance with increasing number of layers of the critic network. As expected, the performance improves with increasing the number of network layers from 1 to 2 given the quadratic relationship between the inputs and the outputs. However, the performance degrades as the number of layers increases

to 3. This can be attributed to the need for additional training episodes and/or experience buffer size to accurately train the network with a higher number of weights and biases. We also show the effect of the experience replay batch size in training the critic network. The performance improves when the experience replay batch size used to train the critic network is increased. Finally, we demonstrate the effect of action space discretization step size by which we are able to reach near optimal localization results when the action space is discretized to a decimeter accuracy. Our simulation results show that a 2-layer critic network with 30 nodes per layer is capable to capture the non-linear relationship between the inputs consisting of the RSSI readings, gNB coordinates and the UAV action space, and the output which is the expected reward or localization error to be minimized. We show that a QL agent with experience replay batch size of  $8 \times 10^4$  utilizing Bayesian regularization for at least 11,000 episodes and action space discretization step size of 0.1 m yields best error result of 0.87m, which is comparable to the optimal solution. Accordingly, we conclude that the proposed reinforcement learning based approach can be utilized to effectively solve the formulated localization problem with comparable accuracy to the optimization based techniques.

#### E. OVERALL RESULTS AND ANALYSIS

In this section, we provide a comparative analysis for each of the techniques including Exhaustive search, Nelder-Mead, Deep Learning, Reinforcement learning and the Carrier phase technique proposed in [8] in terms of mean localization error and time complexity. First, we show the mean localization error for each approach under different gNB-UAV placement scenarios. As shown in Fig. 8(a), the results demonstrate that our proposed deep and reinforcement learning approaches can perform closely to optimal results realized by iterative Exhaustive Search, Nelder Mead and Carrier Phase approaches.

Next, we show in Fig. 8(b) the time complexity derived in the form of big-O notation that we have derived for each of the presented localization techniques as a function of the problem dimension  $n$ . The results demonstrate that the computational complexity of our proposed deep and reinforcement learning based approaches is lower with increasing the problem dimension as compared to the other optimization based approaches and accordingly provide the advantage of meeting real-time constraints needed in such dynamic environments. Our simulation results show the effectiveness of our proposed deep and reinforcement learning based approaches to solve the UAV localization problem. We demonstrate that our proposed reinforcement learning algorithm achieves a lower localization error as compared to deep supervised learning-based algorithm. However, the reinforcement learning algorithm requires higher computational complexity that is proportional to the discrete action space size as compared to the deep supervised learning-based approach.

As such, we conclude the following

- The reinforcement learning based approach is better suited in case when the flight space of the UAV is pre-constrained to a relatively small area and a high localization accuracy is needed.
- In case the flight space of the UAV is very large, the deep supervised learning based approach is recommended to solve the UAV localization with lower computational complexity in real time.
- The use of the heuristic approach provides optimal bound results that are quite close to those of the exact optimal solution at a much lower computational cost than that of the exact optimal solution. Hence the use of heuristic solutions to obtain the optimal bounds is recommended to save a lot of time and effort while providing the needed guidance.

#### VIII. CONCLUSION

In this study, we proposed autonomous UAV 3-D localization solutions that work independently of the GPS or other detectable mobile signals transmitted by the UAV. This is done through utilizing the existing cellular infrastructure to enable the UAV to determine its location using the locations of four surrounding base stations of the cellular network. We formulate the UAV localization problem as an optimization problem to minimize the error of the RSSI measurements from the surrounding cellular base stations. We solved this problem using exact and heuristic optimization techniques to determine the optimal bound of the solution. We then proposed a deep supervised learning-based approach as well as a reinforcement learning based approach to solve the problem in real time, which is critically needed in such applications. Simulations results show that the proposed machine-learning based solutions perform closely to the optimal bound while proving the required real-time results. While both the proposed deep learning and reinforcement learning approaches provide a near optimal solution, the results show that the proposed reinforcement learning algorithm achieves a lower localization error as compared to deep supervised learning-based algorithm at the expense of the added complexity that it sustains.

#### REFERENCES

- [1] S. Ragothaman, M. Maaref, and Z. M. Kassas, "Multipath-optimal UAV trajectory planning for urban UAV navigation with cellular signals," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–6, doi: [10.1109/VTCFall.2019.8891218](https://doi.org/10.1109/VTCFall.2019.8891218).
- [2] R. Amer, W. Saad, and N. Marchetti, "Toward a connected sky: Performance of beamforming with down-tilted antennas for ground and UAV user co-existence," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1840–1844, Oct. 2019, doi: [10.1109/LCOMM.2019.2927452](https://doi.org/10.1109/LCOMM.2019.2927452).
- [3] R. Amer, W. Saad, B. Galkin, and N. Marchetti, "Performance analysis of mobile cellular-connected drones under practical antenna configurations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–7, doi: [10.1109/ICC40277.2020.9148841](https://doi.org/10.1109/ICC40277.2020.9148841).
- [4] A. E. García, M. Ozger, A. Baltaci, S. Hofmann, D. Gera, M. Nilson, C. Cavdar, and D. Schupke, "Direct air to ground communications for flying vehicles: Measurement and scaling study for 5G," *Proc. IEEE 2nd 5G World Forum (5GWF)*, Sep. 2019, pp. 310–315, doi: [10.1109/5GWF.2019.8911712](https://doi.org/10.1109/5GWF.2019.8911712).

- [5] H. Elsawy, W. Dai, M. Alouini, and M. Z. Win, "Base station ordering for emergency call localization in ultra-dense cellular networks," *IEEE Access*, vol. 6, pp. 301–315, 2018, doi: [10.1109/ACCESS.2017.2759260](https://doi.org/10.1109/ACCESS.2017.2759260).
- [6] H. Rizk, M. Torki, and M. Youssef, "CellinDeep: Robust and accurate cellular-based indoor localization via deep learning," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2305–2312, Mar. 2018, doi: [10.1109/JSEN.2018.2885958](https://doi.org/10.1109/JSEN.2018.2885958).
- [7] K. Elawaad, M. Ezzeldin, and M. Torki, "DeepCReg: Improving cellular-based outdoor localization using CNN-based regressors," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6, doi: [10.1109/WCNC45663.2020.9120714](https://doi.org/10.1109/WCNC45663.2020.9120714).
- [8] J. Khalife and Z. M. Kassas, "Precise UAV navigation with cellular carrier phase measurements," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2018, pp. 978–989, doi: [10.1109/PLANS.2018.8373476](https://doi.org/10.1109/PLANS.2018.8373476).
- [9] A. Fahim and Y. Gadallah, "Optimized 3D drone placement and resource allocation for LTE-based M2M communications," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5, doi: [10.1109/VTC2020-Spring48590.2020.9129408](https://doi.org/10.1109/VTC2020-Spring48590.2020.9129408).
- [10] M. M. U. Chowdhury, F. Erden, and I. Guvenc, "RSS-based Q-Learning for indoor UAV navigation," in *Proc. MILCOM IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2019, pp. 121–126.
- [11] W. Jingjing, G. De, and L. Fei, "Research on autonomous positioning method of UAV based on binocular vision," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2019, pp. 3588–3593, doi: [10.1109/CAC48633.2019.8996413](https://doi.org/10.1109/CAC48633.2019.8996413).
- [12] J. A. del Peral-Rosalado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1124–1148, 2nd Quart., 2017.
- [13] *3GPP TS 38.215 Version 15.2.0 Release 15*, 3rd Gener. Partnership Project, Sophia Antipolis Cedex, France, 2018.
- [14] D. Wang, G. Hosangadi, P. Monogioudis, and A. Rao, "Mobile device localization in 5G wireless networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2019, pp. 185–190, doi: [10.1109/ICNC.2019.8685597](https://doi.org/10.1109/ICNC.2019.8685597).
- [15] K. W. Cheung, W. K. Ma, and H. C. So, "Accurate approximation algorithm for TOA-based maximum likelihood mobile location using semidefinite programming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, p. 145, doi: [10.1109/ICASSP.2004.1326215](https://doi.org/10.1109/ICASSP.2004.1326215).
- [16] M. Petitjean, "Spheres unions and intersections and some of their applications in molecular modeling," in *Distance Geometry*, A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, Eds. New York, NY, USA: Springer, 2013, pp. 61–83, doi: [10.1007/978-1-4614-5128-0\\_4](https://doi.org/10.1007/978-1-4614-5128-0_4).
- [17] K. Lee, H. Lee, and K. You, "Optimised solution for hybrid TDOA/AOA-based geolocation using Nelder–Mead simplex method," *IET Radar, Sonar Navigat.*, vol. 13, no. 6, pp. 992–997, Jun. 2019.
- [18] A. Corominas, A. Garcia-Villoria, and R. Pastor, "Technical note: Relating to the parameter values given by Nelder and Mead in their algorithm," *Comput. J.*, vol. 58, no. 1, pp. 157–159, Jan. 2015, doi: [10.1093/comjnl/bxt150](https://doi.org/10.1093/comjnl/bxt150).
- [19] A. Payal, C. S. Rai, and B. V. R. Reddy, "Comparative analysis of Bayesian regularization and Levenberg–Marquardt training algorithm for localization in wireless sensor network," in *Proc. 15th Int. Conf. Adv. Commun. Technol. (ICACT)*, *PyeongChang*, Jan. 2013, pp. 191–194.
- [20] I. Khan, M. A. Z. Raja, M. Shoaib, P. Kumam, H. Alrabaiah, Z. Shah, and S. Islam, "Design of neural network with Levenberg–Marquardt and Bayesian regularization backpropagation for solving pantograph delay differential equations," *IEEE Access*, vol. 8, pp. 137918–137933, 2020, doi: [10.1109/ACCESS.2020.3011820](https://doi.org/10.1109/ACCESS.2020.3011820).
- [21] M. Kayri, "Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: A comparative empirical study on social data," *Math. Comput. Appl.*, vol. 21, no. 2, p. 20, May 2016.
- [22] M. T. Hagan and M. Menhaj, "Training feed-forward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, 1999, pp. 989–993, 1994.
- [23] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proc. Int. Conf. Neural Netw. (ICNN)*, Jun. 1997, pp. 1930–1935.
- [24] D. Ebrahimi, S. Sharafeddine, P.-H. Ho, and C. Assi, "Autonomous UAV trajectory for localizing ground objects: A reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 20, no. 4, pp. 1312–1324, Apr. 2021.
- [25] Y.-H. Hsu and R.-H. Gau, "Reinforcement learning-based collision avoidance and optimal trajectory planning in UAV communication networks," *IEEE Trans. Mobile Comput.*, early access, Jun. 19, 2020, doi: [10.1109/TMC.2020.3003639](https://doi.org/10.1109/TMC.2020.3003639).
- [26] Q. Liu, L. Shi, L. Sun, J. Li, M. Ding, and F. S. Shu, "Path planning for UAV-mounted mobile edge computing with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5723–5728, May 2020, doi: [10.1109/TVT.2020.2982508](https://doi.org/10.1109/TVT.2020.2982508).
- [27] K. Li, W. Ni, E. Tovar, and M. Guizani, "Deep reinforcement learning for real-time trajectory planning in UAV networks," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 958–963, doi: [10.1109/IWCMC48107.2020.9148316](https://doi.org/10.1109/IWCMC48107.2020.9148316).
- [28] *3GPP 5G RSI TS*, 3GPP TS 38.133 Version 15.3.0 Release 15, Sophia Antipolis Cedex, France, 2018.
- [29] A. S. Abdalla and V. Marojevic, "Communications standards for unmanned aircraft systems: The 3GPP perspective and research drivers," *IEEE Commun. Standards Mag.*, vol. 5, no. 1, pp. 70–77, Mar. 2021, doi: [10.1109/MCOMSTD.001.2000032](https://doi.org/10.1109/MCOMSTD.001.2000032).
- [30] S. Zhu, L. Gui, N. Cheng, F. Sun, and Q. Zhang, "Joint design of access point selection and path planning for UAV-assisted cellular networks," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 220–233, Jan. 2020, doi: [10.1109/JIOT.2019.2947718](https://doi.org/10.1109/JIOT.2019.2947718).
- [31] Y. Fateev, D. Dmitriev, V. Tyapkin, N. Kremez, and V. Shaidurov, "The probability distribution function for the sum of squares of independent random variables," in *Proc. Amer. Inst. Phys. Conf. Ser.*, Aug. 2016, Art. no. 020015, doi: [10.1063/1.4959629](https://doi.org/10.1063/1.4959629).
- [32] B. Liu, X. Zhu, Y. Jiang, Z. Wei, and Y. Huang, "UAV and piecewise convex approximation assisted localization with unknown path loss exponents," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12396–12400, Dec. 2019, doi: [10.1109/TVT.2019.2946142](https://doi.org/10.1109/TVT.2019.2946142).



**GHADA AFIFI** received the B.Sc. and M.Sc. degrees in electronics and communications engineering from The American University in Cairo (AUC), New Cairo, Egypt, in 2012 and 2016, respectively, where she is currently pursuing the Ph.D. degree in electronics and communications engineering. She worked as a Process Engineering Manager with Procter and Gamble, Egypt, for several years. Her current research interests include the Internet of Things issues, machine-to-machine communications, cellular communications, autonomous UAV localization and trajectory tracking solutions, machine learning, deep learning, and reinforcement learning.



**YASSER GADALLAH** (Senior Member, IEEE) is currently a Professor at the Department of Electronics and Communications Engineering, The American University in Cairo (AUC), New Cairo, Egypt. Before he joined AUC, he was an Associate Professor with the Electronics and Communications Department, Misr International University, Cairo. Earlier, he was an Assistant Professor of networking system with the College of Information Technology, UAE University. Prior to joining UAE University, he was a Research Scientist in the area of wireless sensor networks with the Communications Research Centre, Ottawa, ON, Canada. He has also worked for many years in the networking and telecommunications industry. He was a Software Development Manager with Cisco Systems for several years, where he was involved in numerous research and development projects, which resulted in the introduction of several high-profile Cisco products. He was also with Nortel Networks as the Software Project Leader and prior to that as a Software Designer. He has authored/coauthored several technical and scientific publications in the networking, communications, and IT fields. His current research interests include the Internet of Things issues, machine-to-machine communications, wireless sensor networks, mobile *ad-hoc* networks, and broadband wireless access networks. He was an organizer, a technical program committee member, and a reviewer for several international conferences and journals.