Deep Reinforcement Learning for UAV Navigation Through Massive MIMO Technique

Hongji Huang*, *Member*, *IEEE*, Yuchun Yang‡, Yue Hao*, *Student Member*, *IEEE*,
Guan Gui*, *Senior Member*, *IEEE*, Hikmet Sari†, *Fellow*, *IEEE*, Fumiyuki Adachi§, *Life Fellow*, *IEEE*
*FocusLab, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.
‡Information Management and Information System, Jilin University, Changchun 130012, China.
§Research Organization of Electrical Communication (ROEC), Tohoku University, Sendai 980-8577, Japan.
E-mails: {guiguan, hikmet}@njupt.edu.cn

***Abstract*—Unmanned aerial vehicles (UAVs) technique has been witnessed as a promising solution in future wireless connectivity from the sky. UAV navigation is one of the most significant open research problems and it has attracted wide interest in the research community. However, current UAV navigation schemes are unable to capture the UAV motion and select the best UAV-ground links, and these weaknesses overwhelm the performance of the UAV navigation. To tackle these fundamental limitations, in this paper, we merge the state-of-the-art deep reinforcement learning with the UAV navigation issue through massive multiple input multiple output (MIMO) technique. To be specific, we carefully design a deep Q-network (DQN) for optimizing the UAV navigation by selecting the optimal policy, and then we propose an efficient learning mechanism for processing the DQN. Then, we train the DQN so that the agent is capable of making decisions based on the received signal strengths for navigating the UAVs with the aid of the powerful Q-learning, aiming at being close to the signal sources. Furthermore, simulation results are provided to corroborate the effectiveness and superiority of the proposed schemes in terms of the UAV navigation through the massive MIMO compared with that of other schemes.**

I. Introduction

Future communication networks will evolve not only for incremental throughput and explosive traffic, but also for low energy consumption, ultra-reliability, and supporting highly diversified applications with heterogeneous quality-of service (QoS) requirements [1]. Therefore, wireless connectivity techniques have drawn universal attention in academy and industry communities, and several emerging techniques have been proposed, such as helikites [2], balloons [3], and unmanned aerial vehicles (UAVs) [4]. Particularly, thanks to their wide applications, high mobility, and superior line-of-sight (LoS) propagation, UAVs have a great potential as airborne nodes such as relays and terminals, and therefore they are considered as an essential part of future communication systems [5]–[8].

In recent years, a large quantity of works have been devoted to enhancing the performance of UAV-enabled communication. In [9], the authors proposed a UAV-enabled data collection system for optimizing the energy consumption issue, where a UAV is assigned to collect data from a ground terminal at fixed location. For the sake of dealing with endurance problem, a novel scheme which leverages proactive caching at the users was provided and numerical results have demonstrated that proactive caching is a good candidate to resolve the endurance issue in UAV-based systems [11]. Then, by deploying UAV as a quasi-stationary base station (BS), the UAV placement issue in two-dimensional or three-dimensional space has been well investigated for utilizing UAVs-BS channel statistics [10], [12]. Additionally, taking the advantages of the splendid massive multiple input multiple output (MIMO) technique [13] which is capable of boosting the system capacity, a UAV cellular-based system through massive MIMO was explored to improve the channel reliability [14].

Thanks to the high mobility of the UAVs, UAV navigation is a promising technique and it has been applied in public safety, emergency rescue, and search and rescue operation. In [15], an evolutionary-based scheme which integrates classic genetic algorithm into modified breeder genetic algorithm was proposed, but this method cannot realize high-accuracy UAV navigation because of the randomness of the genetic algorithm and furthermore, its implementation is too complicated. With the aid of sensors, the authors of [15] presented an autonomous UAV navigation approach (TF-UAV) to address the simultaneous localization and mapping issues but it requires robots which hinder its flexibility. Other strategies are based on received signal strength indicator (RSSI), but strong deep fades degrade the UAV navigation performance [16]. The appealing deep learning-based wireless communication method which adopts deep learning into wireless communication [17] provides an alternative mean for optimizing the UAV navigation problem, whose performance has been corroborated in non-orthogonal multiple access (NOMA) [18], massive MIMO [19]–[23], beamforming [24], [25], resource allocation [26], [27] , traffic control [28], and millimeter-wave (mmWave) communication [29], etc.. In 2015, Google proposed an evolutionary framework called deep reinforcement learning [30], which integrates the deep learning into the reinforcement and is a selected tool to address dynamic optimization problems such as the UAV navigation.

Inspired by the above considerations, in this paper, we incorporate the deep reinforcement learning technique into the UAV navigation through the massive MIMO. The main contributions of this paper are listed below

1) First, we attempt to employ the deep reinforcement learning technique to achieve UAV navigation based on the massive MIMO system. By constructing a deep Q-network (DQN) [30], we can obtain the optimal location selection policy based on the received signal strengths and the optimal link selection policy through the massive MIMO.
2) Second, based on the developed DQN, we propose two efficient deep learning-based algorithms for optimizing UAV navigation performance. After training the DQN based on the Rayleigh fading channel, we run an environment simulator and employ the well-trained DQN for realizing super-resolution UAV navigation. Furthermore, extensive numerical results have verified the superior performance of the proposed algorithms.

II. System Model

Consider a special massive MIMO system, which is comprised of one BS with a uniform linear array (ULA) of N_t antennas and K UAVs with single antenna. According to the well-known ray-tracing-based wireless channel model [31], the channel model of the k-th UAV is modeled as

$$\mathbf{h}_k = \int_{\varphi_{\min}}^{\varphi_{\max}} \mathbf{a}(\varphi_k) g_k(\varphi_k) d\varphi_k, \tag{1}$$

where $\mathbf{a}(\varphi_k)$ and $g_k(\varphi_k)$ represent the array response and the complex gains function of the k-th UAV, respectively. To be specific, φ_k denotes the incidence angle of the k-th UAV, and $\mathbf{a}(\varphi_k)$ can be written as

$$\mathbf{a}(\varphi_k) = \frac{1}{\sqrt{N_t}}[1, e^{-j2\pi\frac{d}{\lambda}\sin\varphi_k}, \cdots, e^{-j2\pi\frac{d}{\lambda}(N_t-1)\sin\varphi_k}]^T, \tag{2}$$

where d is the antenna spacing, while λ represents the carrier wavelength . The gain function $g_k(\varphi_k)$ of the k-th UAV is assumed to have the following property

$$\mathbb{E}\{g_k(\varphi_k) g_k^*(\varphi_k^{'})\} = \gamma_k \upsilon_k(\varphi_k)\delta(\varphi_k - \varphi_k^{'}), \tag{3}$$

where γ_k represents a random variable at UAV k for tracking Rayleigh fading information, $\upsilon_k(\varphi_k)$ represents the power azimuth spectrum that describes the power distribution in the angle domain. It must be noted that γ_k may induce deep fades, leading to high bias in UAV navigation, as shown in Fig. 1. This phenomenon is inevitable even if the UAV is extraordinarily close to

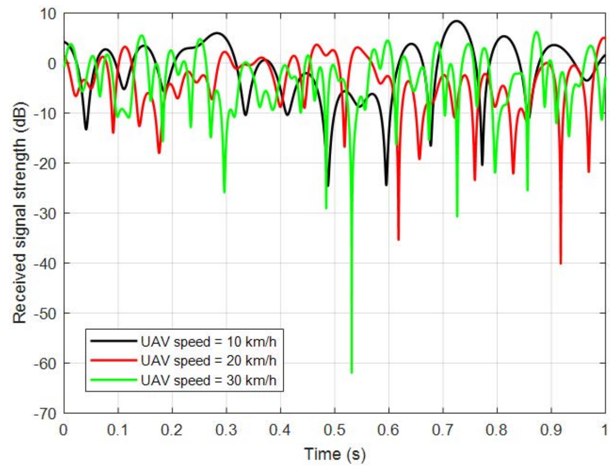

Fig. 1. Rayleigh fading when the UAV speed is 10 km/h, 20 km/h, and 30 km/h.

the destination. Therefore, we provide a deep reinforcement learning-based scheme to deal with this issue for boosting the UAV navigation performance through the massive MIMO. $\delta(\cdot)$ is denoted as the delta function and it has the property that $\int_{-\infty}^{+\infty}\delta(\varphi)d\varphi = 1$. Furthermore, the signal-to-interference-plus-noise ratio (SINR) η_k at the k-th UAV is expressed as

$$\eta_k = \frac{P_{\text{tr}}|\mathbf{h}_k^H \mathbf{w}_k|^2}{P_{\text{tr}}\sum_{j\neq k, j\in K}|\mathbf{h}_j^H \mathbf{w}_j|^2 + \sigma_k^2}, \tag{4}$$

where P_{tr} is the transmitted power at the BS, and $\mathbf{w}_k$ is the unit-norm vector of the k-th UAV. Also, we assume that the channel is corrupted by the additive white Gaussian noise (AWGN) with zero mean and variance σ_k^2 for the k-th UAV. Since the UAV navigation is based on the RSSI, the received signal strengths need to be obtained based on the given channel model. The SINR η_k is introduced as the RSSI for the k-th UAV.

III. DQN-based UAV Navigation Framework

In this section, we provide a deep reinforcement learning-based framework for UAV navigation through the massive MIMO. To be specific, we first develop a deep reinforcement learning-based framework, and then formulate a novel learning policy to train the developed network. Furthermore, we propose an efficient deep reinforcement learning-based strategy for UAV navigation.

A. Deep Q-network

As an appealing embranchment of machine learning, reinforcement learning has been attracting a great attention among aca-demia and industry. For attaining the best situation, agents interact with environment and they search for the optimal stages with maximal rewards. Generally speaking, reinforcement learning can be regarded as a specific description of Markov decision processes (MDPs). It is comprised of four elements: a policy, a reward signal, an environment, and a value function, which is a good candidate for resolving high-complexity situations and capturing realistic scenarios.

However, conventional reinforcement learning requires the agents to adopt appropriate representations of the environment based on the high-dimensional input and generate past know-ledge to the new state. Meanwhile, its applicability only covers the area where the features can be fully exploited or the low-dimensional domain. To break up these gaps, DQN which integrates the deep neural networks into the reinforcement learning has been provided, and deep reinforcement learning has become a remarkable tool to handle complex problems. Therefore, we introduce the DQN to optimize the UAV navigation issue through the massive MIMO.

In the proposed DQN framework, the input layer is a $32 \times 32 \times 4$ space and the first hidden layer is a convolutional (conv.) layer with 8 4×4 filters with stride 4. Followed by a rectifier nonlinear operation, the second hidden layer is designed as a conv. layer with 16 filters of 2×2 filters with stride 2. Then, the next layer is also a conv. layer with 16 filters, and the dimension of these filters is 3×3 with stride 1. The remaining hidden layer is a fully-connected (FC) layer with 256 neurons. Additionally, the output layer is a FC layer which provides valid actions in the UAV navigation optimization.

B. Learning Policy

To enable the UAV navigation, a novel learning policy is proposed based on the developed DQN. At first, the state space S is supposed to represent the received signal strengths, and this set is formulated as $S = \{P_R^k < -120dBm, -120dBm \leq P_R^k \leq -40dBm, P_R^k \geq -40dBm|\forall k\}$. Following the state space S, we assume R, P, and V as the mean value of the immediate reward, the transition probability, and the value function, respectively, and the Q-function is expressed by

$$Q^{\pi}(s,a) = R(s,a) + \tau \sum_{s^{'} \in S} P_{ss^{'}} V^{\pi}(s^{'}), \tag{5}$$

where π is denoted as the policy, and our goal is to obtain the best policy π^{*}. Also, s and a represent state and action, respectively. Concretely, the action a is performed through to the environment simulator and updates its state and its reward based on the information from the BS. Furthermore, τ defines the discount factor in the domain $0 < \tau < 1$, while S represents the state space.

Then, we obtain the maximum Q-function as

$$\begin{aligned} Q^{\pi^{*}}(s,a) &= R(s,a) + \tau \sum_{s^{'} \in S} P_{ss^{'}} V^{\pi^{*}}(s^{'}) \\ &= \mathbb{E}[r(t) + \tau \max_{a^{'}} Q^{\pi^{*}}(s^{'}, a^{'})|s,a], \end{aligned} \tag{6}$$

Here, $r(t)$ is assumed as the immediate reward function, while t is defined as the time index. To be specific, we formulate $r(t)$ as

$$\begin{aligned} r(t) &= a(t)\eta_k \\ &= a(t)\frac{P_{\text{tr}}|\mathbf{h}_k^H \mathbf{w}_k|^2}{P_{\text{tr}} \sum_{j \neq k, j \in K} |\mathbf{h}_j^H \mathbf{w}_j|^2 + \sigma_k^2}, \end{aligned} \tag{7}$$

Algorithm 1 DQN Based Training Method for UAV Navigation.

Input: Environment simulator, DQN.
Output: Well-trained DQN.

Initialize the $Q(s,a;\omega)$ table with weights ω randomly.
Initialize the target DQN parameters with $\omega^{-} = \omega$.
Initialize the replay memory M.
Construct the DQN structure.
Start the environment simulator according to the Rayleigh fading channel.
For episode = $1, 2, \cdots, num$ **do:**
Initialize all the beginning stages s as zero.
For $t = 1, 2, \cdots, T$ **do:**
Obtain the received signal strengths at the UAVs.
Obtain the instant reward $r(t)$ based on the received signal strengths.
Choose an action based on the given random probability ε.
Observe the instant reward r(t) and the next state $s(t+1)$.
Save the knowledges $(s(t), s(t+1), r(t), r(t+1))$ in M.
Sample mini-batch of examples from the M randomly.
Adopt the stochastic gradient descent (SGD) mean to train the DQN according to Eq. (9).
Update network parameters ω of the DQN.
Update the $Q(s,a;\omega)$ table.
End for
End for

Afterwards, noting that A as the action space, the discounted cumulative state function presented in Eq. (7) is formulated as

$$V^{\pi^{*}}(s) = \max_{a \in A}[Q^{\pi^{*}}(s,a)]. \tag{8}$$

After obtaining the maximum Q-function, we need to derive the optimal policy. Using the recursive mechanism, the Q-function can be updated as

$$\begin{aligned} Q_{t+1}(s,a) =& Q_t(s,a) \\ &+ \beta(r + \tau[\max_{a^{'}} Q_t(s^{'}, a^{'})] - Q_t(s,a)), \end{aligned} \tag{9}$$

where β is denoted as the learning rate of the DQN. Since the received signal strength is fluctuating as the changing of the UAVs' position, β needs to vary from different position. For example, in order to collect the received signal strengths when the UAV is close to the destination, the learning rate should be increased.

Thereafter, supposing ω_j as the weight at the j-th iteration of the DQN, the target values of the DQN can be given as

$$y = r + \tau \max_{a'} Q_t(s', a'; \omega_j), \tag{10}$$

Afterwards, to find the optimum solution, the loss function of the DQN can be designed as

$$\text{loss}(\omega) = \mathbb{E}[(y - r + Q(s, a; \omega))^2]. \tag{11}$$

After deriving the learning policy, it is noted that action selection and execution for the agents should be processed and we propose an efficient ε-greedy-based policy for selecting behavior distribution. To be specific, ε is denoted as the exploration probability. We can select the behavior distribution which follows the greedy strategy with probability $1 - \varepsilon$ and choose an action randomly.

Algorithm 2 DQN Based Testing Method for UAV Navigation.

Input: Environment simulator, well-trained DQN.
Output: Network output.

1: Load the DQN framework.
2: Start the environment simulator according to the Rayleigh fading channel.
3: **Loop**
4: Select a received signal strengths of the massive MIMO system.
5: Select an action with the largest $Q(s, a; \omega)$ value in the $Q(s, a; \omega)$ table.
6: Update the environment simulator
7: Update the network output, i.e., the optimal location at the UAVs, the links between the UAVs and the BS.
8: **End Loop**
9: **return:** Network output.

C. Proposed DQN-based UAV Navigation Scheme

In the context of the proposed DQN-based UAV navigation scheme, we separate this DQN-based method into training algorithm and testing algorithm. First of all, we conduct training procedure to train the proposed DQN, and the output of each stage (e.g., experience replay) of the DQN is stored in the replay memory M. Based on the mini-batch mean, the DQN is trained and its network parameters can be updated. After training the DQN, we employ the well-trained DQN to learn the massive MIMO system for UAV navigation. Concretely, the proposed DQN-based UAV navigation strategy is provided in Algorithm 1 and Algorithm 2.

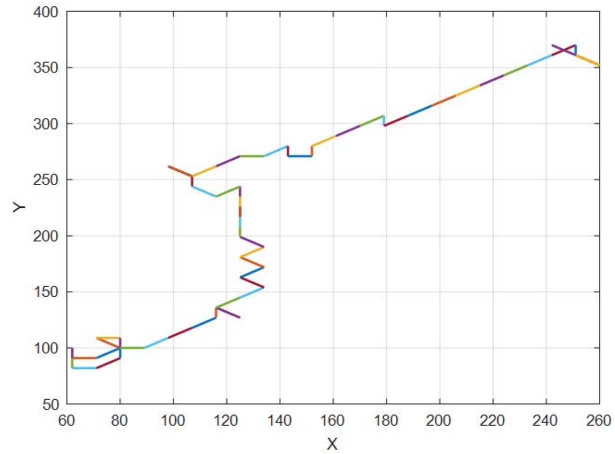

Fig. 2. UAV navigation performance of the proposed DQN-based scheme.

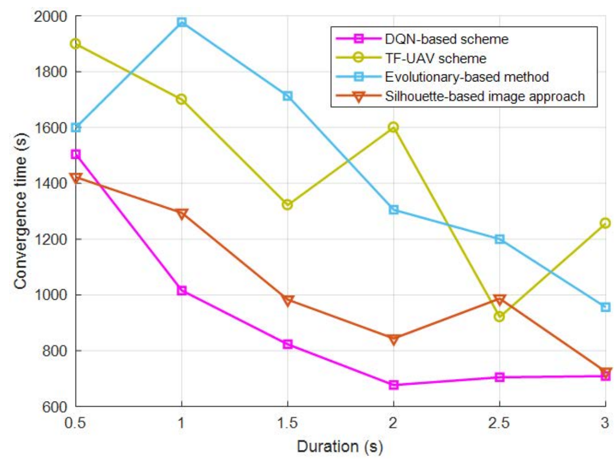

Fig. 3. Convergence performance of the UAV navigation of the DQN-based scheme, the TF-UAV scheme [16], the evolutionary-based method [15], and the Silhouette-based image approach [32].

IV. Simulation Results and Analysis

In this section, we present numerical results of the proposed DQN-based UAV navigation scheme through massive MIMO. Here, the *Tensorflow 1.11* framework is introduced to design the DQN, and our model is deployed on a GPU-based server with 2 Nvidia Quado P6000 GPUs and 2 Intel Xeon Gold 6136 CPUs. In our experiment, we consider a massive MIMO system with $N_t = 128$ transmit antennas and $K = 32$ single-antenna UAVs. Also, we set the total transmitted power as 20 W and the sampling period is initialized as 0.02 ms, while $d = \frac{\lambda}{2}$ is set in our simulation. Furthermore, the batch size is 100 and the number of training examples is 250000, while the amount of testing examples is assumed as 50000.

Fig. 2 shows the navigation performance of the DQN-based scheme, which exhibits the trajectory of the UAVs in a $2-D$ space. Here, the initial learning rate is set as 0.001 and the UAV is simulated in a 500m $\times$ 500m indoor space. It can be seen from Fig. 2 that the UAV is moving away from the origin of the cartesian coordinate system as time goes by, which indicates that the UAV can fly under accurate UAV navigation. Also, we observe that some parts of the navigation curve change sharply, which are induced by the fact that the UAV encounters obstacles such as walls during flying, implying that the proposed DQN-based UAV navigation method is capable of extracting the environment information and making the best decisions.

The performance comparison of the convergence time of the UAV navigation against sampling duration is presented in Fig. 3, in which the DQN-based scheme, the TF-UAV scheme [16], the evolutionary-based method [15], and the Silhouette-based image approach [32] are considered, respectively. Here, the speed of the UAV is set as 10 km/h. It can be observed from Fig. 3 that the convergence time of all the UAV navigation algorithms is decreasing as sampling duration increases, for the reason that longer distances are traveled by the UAV but this behavior may increase the probability of making wrong decisions. Meanwhile, it can be seen from Fig. 3 that when the sampling duration is approaching a threshold time, the speed of the convergence of all the algorithms is reducing and the speed is required to reduce to 0 in theory when the algorithms converge. In particular, the proposed DQN-based scheme can converge when the sampling duration is 2 s, while other methods still shake sharply. Also, the proposed scheme requires less convergence time compared with that of other schemes in most cases, although the Silhouette-based image approach requires less convergence time when the sampling duration is increased from 0.5 s to 0.6 s. Hence, the proposed DQN-based scheme outperforms other state-of-the-art schemes in terms of UAV navigation which is due to the powerful mapping capacity of the DQN.

Fig. 4 exhibits the convergence performance of the UAV navigation with different learning rate of the DQN-based method, in which the initial learning rate is set as 0.1, 0.01, 0.005, and 0.001. Initially, the speed of the UAV is set as 10 km/h. We observe from Fig. 4 that the DQN-based scheme can converge quickly when adopting larger initial learning rate, which is due to the fact that larger initial learning rate can facilitate the convergence behavior. However, it should be noted that larger learning rate leads to strong vibration and the algorithm may fail to converge. As shown in Fig. 4, the curve is more stable and it becomes totally smooth finally when introducing the learning rate as 0.001, indicating that we should use smaller learning rate for enhancing the UAV navigation performance.

Fig. 5 shows the convergence performance of the UAV navigation when the UAV flies at different velocity, in the case of 10 km/h, 15 km/h, 20 km/h, 25 km/h, and 30 km/h. Here, the initial learning rate is set as 0.001. It is observed from Fig. 5 that larger velocity requires less convergence time compared to that of smaller velocity, since larger velocity reduces responsiveness of the UAV system. However, it can also be seen from Fig. 5 that the tendency of the curves are not decreasing in general and they shake frequently when adopting larger UAV speed, illustrating that larger UAV speed degrades the UAV navigation performance.

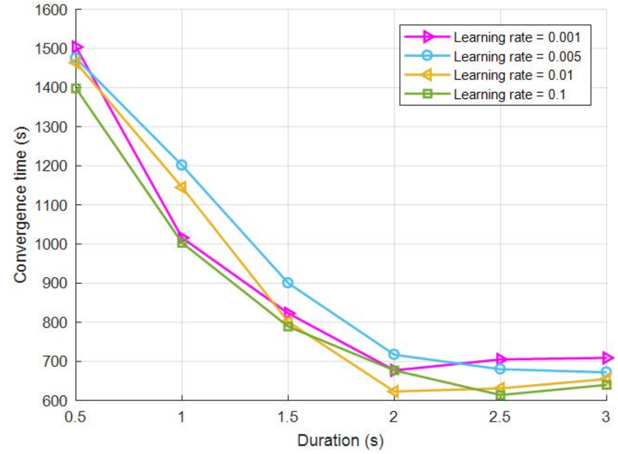

Fig. 4. Comparison of the convergence time via duration performance of the UAV navigation with different initial learning rates, in the case of 0.1, 0.01, 0.005, and 0.001.

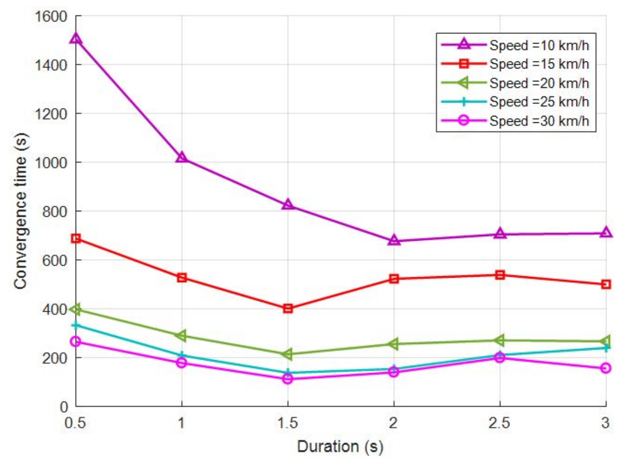

Fig. 5. Convergence performance of the proposed scheme when the UAV speed is set as 10 km/h, 15 km/h, 20 km/h, 25 km/h, and 30 km/h.

V. Conclusions

In this paper, we have provided a deep reinforcement learning-based scheme for UAV navigation through massive MIMO. Specifically, we first design an efficient DQN which comprises conv. layers and FC layers to extract useful features of the massive MIMO, followed by proposing a Q-learning-based learning policy to realize UAV navigation. Here, we regard each UAV-ground link as an agent, and the optimal location at the UAVs is obtained based on the received signal strengths without requiring global information. Numerical results also show the superior UAV navigation performance of the DQN-based strategy compared with several typical strategies, and it is revealed that the proposed scheme can eliminate the bias induced by the Rayleigh fading. Our research has demonstrated that the deep reinforcement learning is a magnificent tool to handle dynamic tracking and resource management, and future research can be done towards in this direction to optimize dynamic wireless communication systems.

References

[1] A. Osseiran et al., "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26-35, May 2014.

[2] *Project Loon*. Accessed: Jul. 15, 2017. [Online]. Available: https://x.company/loon/.

[3] S. Chandrasekharan et al., "Designing and implementing future aerial communication networks," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 26-34, May 2016.

[4] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36-42, May 2016.

[5] Q. Wu, and R. Zhang, "Common Throughput Maximization in UAV-Enabled OFDMA Systems With Delay Consideration," *IEEE Trans. Commun.*, no. 12, pp. 6614-6627, Dec. 2018.

[6] F. Cheng, et al., "UAV Relaying Assisted Secure Transmission With Caching," *IEEE Trans. Commun.*, to be published, doi: 10.1109/TCOMM.2019.2895088

[7] X. Liu, et al., "Transceiver Design and Multi-hop D2D for UAV IoT Coverage in Disasters," *IEEE Internet of Things*, to be published, doi: 10.1109/JIOT.2018.2877504

[8] N. Zhao, et al., "Caching UAV Assisted Secure Transmission in Hyper-Dense Networks Based on Interference Alignment," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2281-2294, May 2018.

[9] D. Yang, Q. Wu, Y. Zeng, and R. Zhang, "Energy Tradeoff in Ground-to-UAV Communication via Trajectory Design," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6721-6726, Jul. 2018.

[10] R. I. Bor-Yaliniz, A. Ei-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1-5.

[11] X. Xu, Y. Zeng, Y. L. Guan, and R. Zhang, "Overcoming Endurance Issue: UAV-Enabled Communications With Proactive Caching," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1231-1244, Jun. 2018.

[12] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-mounted mobile base stations," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 604-607, Mar. 2017.

[13] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[14] G. Geraci, A. Garcia-Rodriguez, L. G. Giordano, D. Lopez-Perez, and E. Bjoernson, "Supporting UAV Cellular Communications through Massive MIMO," in *ICC Workshops*, Kansas City, MO, 2018, pp. 1-6.

[15] I. K. Nikolos, K. P. Valavanis, N. C. Tsourveloudis and A. N. Kostaras, "Evolutionary algorithm based offline/online path planner for UAV navigation," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 33, no. 6, pp. 898-912, Dec. 2003.

[16] T. Tomic et al., "Toward a Fully Autonomous UAV: Research Platform for Indoor and Outdoor Urban Search and Rescue," *IEEE Robotics & Automation Mag.*, vol. 19, no. 3, pp. 46-56, Sept. 2012.

[17] Hinton, Geoffrey E., S. Osindero, and Y. W. The, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol.18, 2006.

[18] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep Learning for An Effective Non-Orthogonal Multiple Access Scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440-8450, Sept. 2018.

[19] H. He, C. Wen, S. Jin, G. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852-855, 2018.

[20] H. Huang, J. Yang, Y. Song, H. Huang, and G. Gui, "Deep Learning for Super-Resolution Channel Estimation and DOA Estimation based Massive MIMO System," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549-8560, Sept. 2018.

[21] C. Wen, W, Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748-751, 2018.

[22] H. Huang, G. Gui, H. Sari, and F. Adachi, "Deep Learning for Super-Resolution DOA Estimation in Massive MIMO Systems," in *VTC-Fall*, Chicago, USA, Aug. 27-30, 2018, pp. 1-6.

[23] H. Huang, Song Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," submitted to *IEEE Wireless Commun.*, Jan. 2019.

[24] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, X. Zhu, "Unsupervised Learning Based Beamforming Design for Downlink MIMO," *IEEE Access*, vol. 7, pp. 7599-7605, 2019.

[25] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, D. Tujkovic, "Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems," *IEEE Access*, vol. 6, pp. 37328-37348, 2018.

[26] M. Liu, T. Song, G. Gui, "Deep Cognitive Perspective: Resource Allocation for NOMA based Heterogeneous IoT with Imperfect SIC," *IEEE Internet of Things*, to be published, doi: 10.1109/JIOT.2018.2876152

[27] M. Liu, J. Yang, T. Song, J. Hu, G. Gui, "Deep Learning-Inspired Message Passing Algorithm for Efficient Resource Allocation in Cognitive Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641-653, Jan. 2019

[28] N. Kato et al., "The Deep Learning Vision for Heterogeneous Network Traffic Control: Proposal, Challenges, and Future Perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146-153, Jun. 2017.

[29] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-Learning based Millimeter-Wave Massive MIMO for Hybrid Precoding," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2019.2893928, Jan. 2019.

[30] Mnih V, Kavukcuoglu K, Silver D, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015.

[31] D. Tse, and P. Viswanath, *Fundamentals of Wireless Commun.*. New York, NY: Cambridge University Press, 2005.

[32] S. Luan, X. Kong, B. Wang, Y. Guo and X. You, "Silhouette coefficient based approach on cell-phone classification for unknown source images," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, 2012, pp. 6744-6747.