A Data-Driven Waveform Adaptation Method for Mm-Wave Gait Classification at the Edge

Soheil Hor, *Student Member, IEEE*, Mert Pilanci, *Member, IEEE*, and Amin Arbabian, *Senior Member, IEEE*

***Abstract*—The ever-increasing need for real-time 3D perception for autonomy has resulted in development of a new generation of high-resolution perception sensors paired with powerful edge processors and advanced perception algorithms. However, processing of the large data volumes generated by such high-resolution sensors still imposes a great challenge for resource-limited edge devices and systems. In this work, we visit this problem in the context of pedestrian gait analysis using high-resolution mm-Wave sensors with application in autonomous driving. We propose reducing the total computational load of a resource-limited perception system by dynamically controlling the mm-Wave sensor's resolution on the fly. We provide results on three dynamic sensor adaptation approaches including an end-to-end Deep Q-Learning approach that can overcome the resource and performance limitations inherent to conventional static methods.**

***Index Terms*—Deep neural networks, mm-wave sensing, gait recognition, reinforcement learning.**

I. Introduction

RADAR is one of the most versatile and robust perception sensors used in robotics and autonomy applications today. Radar's unique ability to maintain range and Doppler resolution over long distances makes it an ideal sensor for perception tasks including pedestrian classification and traffic behaviour analysis. However, the power and computational limitations of perception systems limit the extent that the high dimensional raw radar data can be processed at the edge. As a result, even the state of the art automotive radars deployed on cars today are usually used only for collision avoidance or as a localization sensor. One approach for better utilisation of the high-resolution capabilities of radar is to adapt the radar sensor's settings (also known as "waveforms") on the fly to maximise the perception performance and computational efficiency simultaneously.

A. Mm-Wave Gait Analysis

One problem that has proven to be challenging using vision-based perception sensors is remote pedestrian classification and behaviour analysis over medium to far distances (see [1] for a review). High-resolution mm-Wave radars on the other hand are capable of providing rich feature-sets by analysis of small body movements of pedestrians 10 s to 100 s of meters away. These feature-sets are commonly known as gait micro-Doppler signatures.

Micro-Doppler features are widely used in literature for applications like human identification [2], classification of cars from pedestrians [3], classification of different activities like walking or crawling in indoor and outdoor environments [4]–[6], distinguishing walking vs. running patterns using micro-Doppler features [7], [8], identifying hand motions in gait micro-Doppler signatures [9] and estimating height of pedestrians from micro-Doppler signatures by directly estimating parameters of the Boulic-Thalmann gait model [10]–[12]. The perception model used in this letter uses the same principle to classify children from adults based on the effect of the pedestrian's height on the micro-Doppler gait patterns.

B. Cognitive Radar

Changing the radar's settings on the fly and in response to the dynamics of a scene, has been subject of study in the field of cognitive radar since the introduction of the concept by Haykin in [13].

Although Haykin has used the term cognition very broadly in his work, the body of literature following his work has mainly focused on classical channel estimation and adaptation methods with the goal of optimal spectrum allocation and interference suppression in complex multi-radar or dynamic environments.

Recent work in this field includes using Reinforcement Learning (RL) techniques for spectrum allocation [14]–[17], direct control of the radar beam or waveform for optimum target detection performance [18]–[21] or using Q-Learning for optimising tracking performance of a radar system in presence of interfering communication systems or other radars [22]–[24].

In contrast, we focus on dynamically adapting the radar settings to optimise the classification performance of a non-linear deep perception model with computational load as the main resource constraint (instead of spectrum or SINR). This is a new angle to cognition in radar because unlike the methods used for classical radar tasks like detection and tracking, classifiers used for analysis of complex targets found in modern urban environments are highly non-intuitive and non-linear.

Manuscript received August 16, 2021; revised October 5, 2021; accepted October 9, 2021. Date of publication October 26, 2021; date of current version January 20, 2022. This work was supported in part by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Beibei Wang. *(Corresponding author: Soheil Hor.)*

The authors are with the Electrical Engineering Department, Stanford university, Stanford, CA 94305 USA (e-mail: soheilh@stanford.edu; pilanci@stanford.edu; arbabian@stanford.edu).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LSP.2021.3122355, provided by the authors.

Digital Object Identifier 10.1109/LSP.2021.3122355

II. Waveform Adaptation Methods

In this section we introduce three different approaches to waveform adaptation for edge perception: A myopic open-loop waveform design approach, a myopic closed-loop waveform adaptation approach and an end-to-end Reinforcement Learning approach. In our experiments we assume a simplified adaptation space containing only two settings (waveforms): the low resolution waveform WF_{LR} (Doppler resolution of 75 cm/s), and the high resolution waveform WF_{HR} (Doppler resolution of 20 cm/s).

A. Myopic Approaches

1) Dynamic Open-Loop Approach: In conventional static waveform design approaches the sensor is set to work in either the high-resolution setting resulting in a high raw data volume (and higher power and memory consumption) or a low resolution setting resulting in lower accuracy.

The alternative approach used as a baseline in this letter is to randomly switch between the two settings based on a pre-calculated prior corresponding to the desired operating point for the perception system. This is a very simple strategy that can result in a better efficiency than the high-resolution static approach and better performance than statically selecting the low-resolution waveform. However, such open-loop methods can only yield in a linear trade-off between the average data-volume generated by the sensor and the average perception accuracy.

2) Dynamic Closed-Loop Approach: The proposed algorithm for the Myopic baseline is shown in Algorithm 1. The closed-loop algorithm uses the prediction results of the perception model at the previous m steps to estimate the most probable class. It then compares the model's confidence in the most probable class with a preset confidence threshold. If the criteria is met, then the algorithm reduces the resource consumption by choosing a waveform with lower resource consumption. Otherwise it switches to the waveform with the greatest average expected accuracy (i.e. the high-resolution waveform). Using this method one can trade accuracy and efficiency by choosing the minimum confidence threshold parameter.

Algorithm 1: Closed-loop Myopic Approach.

```
Calculate expected Resource Consumption (RC) for each
sensor setting in action space A
Calculate expected Average Accuracy (AA) for each
sensor setting in action space A
for A desired confidence threshold CT do
    Initialize label memory M with capacity m
    Initialize the sensor with a random setting
    for t = 1, T do
        Acquire S_t from the sensor assuming setting a_t
        y_t = argmax_y[P(Y = y|S = S_t)]
        Push y_t in label memory M
        y_m = Mode(M)
        if P(Y = y_m|S = S_t) >= CT then
            a_t = argmin_{a∈A}(RC(a))
        else
            a_t = argmax_{a∈A}(AA(a))
        end
    end
end
```

B. Reinforcement Learning Approach

The first step in our proposed Reinforcement Learning approach is to formulate the sensor adaptation task a Markov Decision Process (MDP). Markov Decision Processes model interaction of an agent with a dynamic environment through a set of states S, a set of actions A, a transition probability P defined over the state-space, and an immediate reward function r corresponding to each state. Our goal is to find the optimum policy $\pi : S \rightarrow A$ that can maximise the (discounted) sum of rewards by selecting the optimum action at each state. In our scenario we define the Markov Decision Process as follows:

a) *State:* We choose a 600 ms snapshot of the target's Doppler spectrogram as the state for the Markov Decision Process. In principle, a radar system can adapt its waveform much faster than the 600 ms window needed for the gait analysis task. However, for the sake of simplicity we have assumed that every snapshot contains data only from a single waveform and therefore the maximum waveform adaptation rate is the same as the perception rate.

b) *Action:* We define set of actions as the set of available waveforms which control the Doppler resolution of the radar, WF_{LR} and WF_{HR}, as explained above.

c) *Reward:* Since the value of each state should be a function of both the performance of the perception model and the resource consumed for achieving that performance, we define the reward function for each state is as follows:

$$r_t = 1[y_p(s_t) = y_{true}(t)] - \alpha RC_{norm}(a_t) \quad (1)$$

In which $1[\cdot]$ is the indicator function, $y_p(s_t)$ is the label predicted by the perception model at given state s_t, y_{true} is the target ground truth label, and RC_{norm} is the normalised Resource Consumption (RC) cost defined as:

$$RC_{norm}(a_t) = \frac{RC(a_t)}{\max_{a\in\mathcal{A}}(RC(a))} \quad (2)$$

In this equation $RC(a)$ is the resource consumption of the system assuming that the sensor is operating with the waveform corresponding to action a. The α coefficient in Equation (1) can be used to to balance the perception model's expected accuracy and expected resource consumption over the entire training set without explicitly requiring a knowledge of the target's true class or the perception confidence level.

d) *Transient Probability:* The transient probability variable in MDPs signifies the stochastic nature of the state variables representing the dynamics of the observed environment. In our experiments we include target trajectories that constantly change direction in order to simulate the dynamics

Algorithm 2: Deep Q-Learning Approach.

Calculate expected Resource Consumption (RC) for each sensor setting in A
Choose a desired operation point by setting the α coefficient
Initialize a replay memory M with capacity T
Initialize the Deep Q-Network with random weights
Initialize the sensor with a random setting
for $t = 1, T$ *in each episode* **do**
  Select action a_t = $argmax_a[Q(s_t, a)]$
  Acquire S_t from the sensor assuming setting selected by a_t
  $y_t = argmax_y P(Y = y|S = S_t)$
  Calculate r_t for given s_t and a_t according to equation 1
  Store (s_t,r_t,a_t) in replay memory M
  Sample a random batch of transition tuples (s_m,r_m,a_m)
  if s_m *is a terminal state* **then**
    $\hat{Q}_m = r_m$
  else
    $\hat{Q}_m = r_m + \gamma max_a[Q(s_m, a)]$
  end
  Perform a gradient descent step on $(\hat{Q}_m - Q(s_m, a_m))^2$
end

of the Doppler signatures in real-world scenarios. However, instead of explicitly modeling the transient probability of the MDP, we use a model-free approach called Deep Q-learning to directly find an optimum policy that yields in the maximum reward.

e) *Deep Q-learning:* Based on the definitions of the MDP above, the state value function can be defined as:

$$V_\pi(s) = \mathbf{E}_\pi \left[\sum_t \gamma^t r_t | S_t = s \right]. \quad (3)$$

In which γ is the discount factor controlling relative importance of immediate and delayed rewards ($\gamma = 0.99$ in our experiments). The Q-value can be defined as the expected value of the state value function (V_π) conditioned on action a. In other words:

$$Q_\pi(s, a) = \mathbf{E}_\pi[V_\pi(s)|S_t = s, A_t = a]. \quad (4)$$

The Deep Q-Network (DQN) approach [25] proposes using a deep neural network to find the mapping between the observed (state, action) pair and the corresponding Q-function. The learned Q-function can in turn be used to find an optimum policy $\pi_{opt} : S \rightarrow A$ simply by choosing the action that maximises the Q-function at each state. A step by step pseudo-code of the proposed DQN approach is shown in Algorithm 2.

III. Experiments and Results

The first step in comparison of the different approaches explained above is to set up a perception pipeline that can handle data from both waveforms with minimal performance degradation. For this purpose we trained and validated a ResNet20 [26] over a static dataset consisting of 600 ms snapshots of 3 classes of pedestrians (Child, Adult, and Child accompanied with Adult), walking on randomized straight lines, and observed using each of the two waveforms. We then formed an independent dynamic dataset by simulating 10 s episodes of pedestrians walking in constant velocity circles with different speeds and locations.

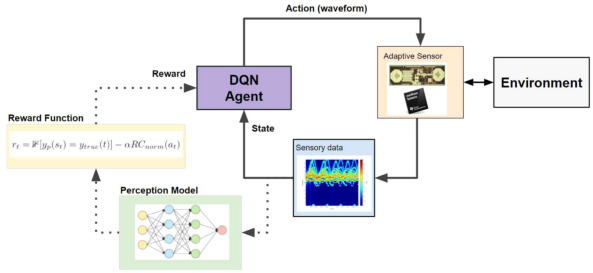

Fig. 1. Block diagram of the proposed Q-learning approach.

Fig. 2 shows three examples of such dynamic episodes.The red curve in Fig. 2 shows the performance of the proposed DQN method with $\alpha = 0.45$ in a scenario with both intra-class dynamics (change of Doppler signature caused by change in direction and location) and inter-class dynamics (change of the target class). Comparing the blue and green curves it can be seen the perception model's confidence is highly variable as a function of both the waveform and the dynamics of the episode. However, the DQN model has learned to select the higher resolution waveform in cases that the low-resolution waveform would result in a low confidence prediction (for instance at t=7 to 9 seconds) and has decided to reduce the resolution to reduce the data volume in cases that increasing the resolution would not necessarily help with the perception model's performance (for instance at t=10 to 15 seconds).

Summarizing the performance of the three baselines over this dynamic example, it can be seen that at the lowest resolution, the perception model correctly classifies the target in only 16 out of 44 time steps. This results in confusion between the Child and Adult classes even when the predictions are aggregated across time. On the other hand, the DQN method has managed to double the number of correct classifications (30 out of 44) resulting in correct aggregated classification for all the three classes. A static high-resolution waveform design approach would require around 2.5X more computational resources to achieve the same aggregated classification results.

A. Ablation Analysis

Each of the approaches introduced in previous section provide access to a tunable parameter that controls the performance versus efficiency trade-off in each of the waveform adaptation schemes. In order to have a fair comparison for all the methods over all the possible choices of the tuning parameters, we re-trained the DQN method for α parameters between 0 and 1, and swept the tunable parameters of the other baselines (Confidence Threshold CT for the closed-loop approach and the waveform prior probability P for the open-loop approach) over all the possible values.

Fig. 3 shows the resulting average accuracy vs. (normalized) data volume plot for over 1700 dynamic episodes sampled from

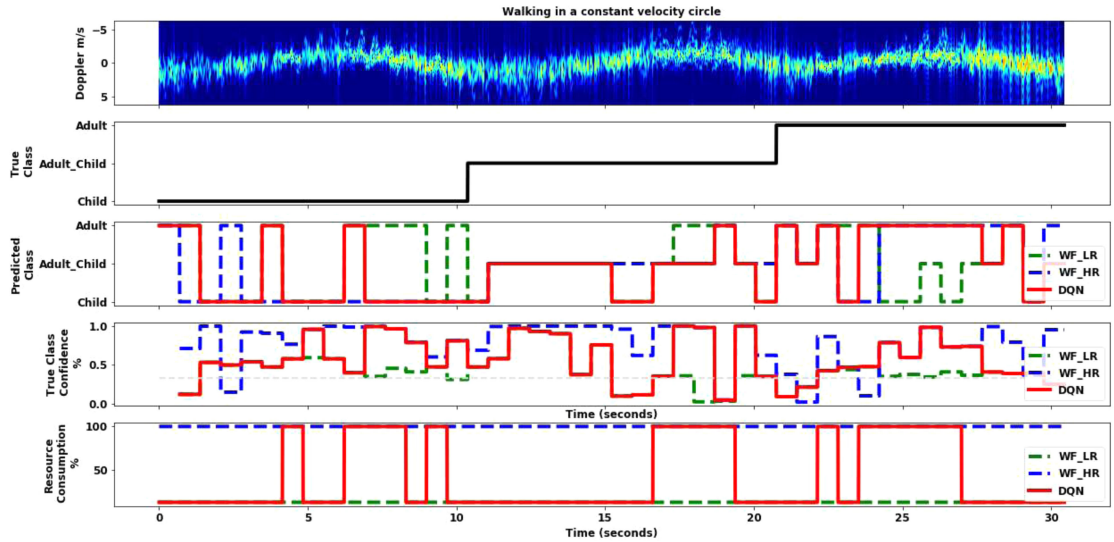

Fig. 2. The proposed DQN approach in action. The DQN agent adapts the waveform in response to the dynamics of the scene to maximise the perception performance while requiring minimal computational resources ($\alpha = 0.45$).

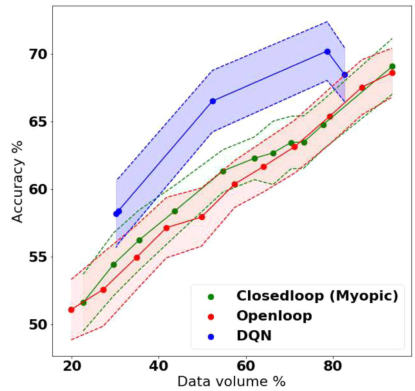

Fig. 3. Accuracy versus resource consumption for different sensor adaptation schemes.

TABLE I
LOWEST RESOURCE CONSUMPTION ACHIEVED BY EACH METHOD ASSUMING MAXIMUM PERFORMANCE DROP OF 5%

	Static High-res	Open-loop	Closed-loop	DQN	
Tunable Parameter	-	P	CT	α	
	-	0.81	0.91	0.45	0.25
Average Accuracy (%)	68.59	63.36	64.76	66.49	**70.19**
Resource Consumption (%)	100	79.18	77.73	**52.30**	78.63

57 different walking trajectories (30 episodes per walking trajectory, 19 trajectories per class). As it was expected, the open-loop approach only provides a linear improvement of accuracy for an increase in the resolution (and data volume). But the closed-loop approach can break this linear trade-off by dynamic adaptation of resolution based on the expected perception performance. The error bars show standard deviation of the average accuracy over the entire test-set. It should be noted that the accuracy numbers reported here (and in Table I) are single-snapshot performance figures of the perception model. One could achieve much higher accuracy figures by utilizing label aggregation and filtering methods in post-processing (assuming that the system is allowed to make a prediction after a longer time period). For instance, a single-snapshot accuracy of 70% reported in this work can translate to an expected aggregated accuracy of up to 99.97% for a 6 s (10-step) label memory. This is comparable with observation windows commonly used for gait analysis tasks. As an example see [12] and [8] using observation window sizes of 12 and 16 seconds respectively.

Table I provides a more quantitative view of the same results by showing the lowest resource consumption achievable by each method assuming that a single-snapshot accuracy drop of 5% is acceptable to the perception system. As it can be seen, unlike the Myopic approach that has marginal improvement over the open-loop method, the proposed DQN approach can meet the performance criteria with 21% to 48% less raw data volume. We expect this gap to improve if the DQN method is provided access to more degrees of freedom in waveforms or is trained on longer dynamic episodes.

Since in our experiments the effect of the waveform on accuracy can be much smaller than the effect of co-factors (evident by the standard deviation of curves shown in Fig. 3), the comparison between the performance of the models for a fixed data volume budget is challenging. We will investigate this angle further in future work.

IV. CONCLUSION

In this work we demonstrated effect of trajectory-related dynamic co-factors on performance of a mm-Wave gait analysis system and proposed a data-driven method to adapt the waveform to mitigate the effect. The proposed DQN approach enables better utilization of the state of the art mm-Wave sensors and reduces the computational load of resource-limited perception systems by reducing the raw data volumes generated by mm-Wave sensors.

REFERENCES

[1] M. A. Khan, M. Mittal, L. M. Goyal, and S. Roy, "A deep survey on supervised learning based human detection and activity classification methods," *Multimedia Tools Appl.*, pp. 1–57, 2021.

[2] S. Dong, W. Xia, Y. Li, and K. Chen, "Human identification under multiple gait patterns based on FMCW radar and deep neural networks," in *Proc. Int. Conf. Commun., Signal Process., Syst.*, Springer, 2020, pp. 176–185.

[3] L. Du, Y. Ma, B. Wang, and H. Liu, "Noise-robust classification of ground moving targets based on time-frequency feature from micro-Doppler signature," *IEEE Sensors J.*, vol. 14, no. 8, pp. 2672–2682, Aug. 2014.

[4] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.

[5] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.

[6] S. Björklund, H. Petersson, A. Nezirovic, M. B. Guldogan, and F. Gustafsson, "Millimeter-wave radar micro-Doppler signatures of human motion," in *Proc. 12th Int. Radar Symp.*, 2011, pp. 167–174.

[7] L. Vignaud, A. Ghaleb, J. Le Kernec, and J.-M. Nicolas, "Radar high resolution range & micro-Doppler analysis of human motions," in *Proc. Int. Radar Conf.*, 2009, pp. 1–6.

[8] L. Senigagliesi, G. Ciattaglia, A. De Santis, and E. Gambi, "People walking classification using automotive radar," *Electronics*, vol. 9, no. 4, 2020, Art. no. 588.

[9] F. H. C. Tivive, A. Bouzerdoum, and M. G. Amin, "A human gait classification method based on radar Doppler spectrograms," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–12, 2010.

[10] R. Boulic, N. M. Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification," *Vis. Comput.*, vol. 6, no. 6, pp. 344–358, 1990.

[11] A. Sona, G. Giorgi, and R. Ricci, "A measurement approach based on micro-Doppler maps for human motion analysis and detection," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2012, pp. 354–359.

[12] M. B. Guldogan, F. Gustafsson, U. Orguner, S. Björklund, H. Petersson, and A. Nezirovic, "Human gait parameter estimation based on micro-Doppler signatures using particle filters," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 5940–5943.

[13] S. Haykin, "Cognitive radar: A way of the future," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 30–40, Jan. 2006.

[14] P. Liu, Y. Liu, T. Huang, Y. Lu, and X. Wang, "Decentralized automotive radar spectrum allocation to avoid mutual interference using reinforcement learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 190–205, Feb. 2021.

[15] L. O. Wabeke and W. A. Nel, "Utilizing Q-learning to allow a radar to choose its transmit frequency, adapting to its environment," in *Proc. 2nd Int. Workshop Cogn. Inf. Process.*, 2010, pp. 263–268.

[16] L. Kang, J. Bo, L. Hongwei, and L. Siyuan, "Reinforcement learning based anti-jamming frequency hopping strategies design for cognitive radar," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2018, pp. 1–5.

[17] P. Liu, Y. Liu, T. Huang, Y. Lu, and X. Wang, "Cognitive radar using reinforcement learning in automotive applications," vol. 57, no. 1, pp. 190–205, 2019, doi: 10.1109/TAES.2020.3011869.

[18] A. M. Ahmed, A. A. Ahmad, S. Fortunati, A. Sezgin, M. Greco, and F. Gini, "A reinforcement learning based approach for multi-target detection in massive MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 5, pp. 2622–2636, Oct. 2021, doi: 10.1109/TAES.2021.3061809.

[19] E. Selvi, R. M. Buehrer, A. Martone, and K. Sherbondy, "Reinforcement learning for adaptable bandwidth tracking radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 3904–3921, Oct. 2020.

[20] B. Wang, J. Wang, X. Song, and F. Liu, "Q-learning-based adaptive waveform selection in cognitive radar," *Int. J. Commun. Netw. Syst. Sci.*, vol. 2, no. 7, pp. 669–674, 2009.

[21] Q. Yang, Z. Han, H. Wang, J. Dong, and Y. Zhao, "Radar waveform design based on multi-agent reinforcement learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 35, no. 10, 2021, Art. no. 2159035, doi: 10.1142/S0218001421590357.

[22] M. Kozy, J. Yu, R. M. Buehrer, A. Martone, and K. Sherbondy, "Applying deep-Q networks to target tracking to improve cognitive radar," in *Proc. IEEE Radar Conf.*, 2019, pp. 1–6.

[23] C. E. Thornton, M. A. Kozy, R. M. Buehrer, A. F. Martone and K. D. Sherbondy, "Deep reinforcement learning control for radar detection and tracking in congested spectral environments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 4, pp. 1335–1349, Dec. 2020.

[24] Q. Cheng, L. Chen, K. Zhang, and Z. Wang, "Reinforcement learning for robust radar tracking," in *Proc. IET Int. Radar Conf.*, 2021, pp. 1540–1544.

[25] V. Mnih *et al.*, "Playing Atari with deep reinforcement learning," vol. 518, no. 7540, pp. 529–533, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.