QUANTUM REINFORCEMENT LEARNING FOR COGNITIVE SAR

Jan Meyer[†*], *Kay Glatting*[†], *Sigurd Huber*[†], *Gerhard Krieger*[†]

[†]German Aerospace Center (DLR)
Oberpfaffenhofen Germany
[*]jan.meyer@dlr.de

ABSTRACT

Cognitive radar involves a closed loop between the radar receiver, radar transmitter, and the environment, similar to the perception-action cycle in human cognition. It adapts acquisition parameters based on the acquired information, which offers more efficient resource management of future SAR missions in varying scenes. We propose a spaceborne cognitive SAR concept for ship detection, which uses a two-stage process to improve ship detection capabilities. This radar concept can be implemented on a single platform using a hybrid mode or distributed on separate SAR satellites operated in a convoy configuration. We implement variational quantum circuits for the adaptation to the scene in a reinforcement learning approach to explore possible advantages of quantum computing combined with parameter optimisation on classical computers.

***Index Terms*—** SAR, cognitive radar, quantum computing, reinforcement learning, radar remote sensing

1. INTRODUCTION

At its core, the purpose of radar remote sensing is to gain up-to-date information about a scene. Respecting physical and technical constraints, radar systems use various modes and techniques to extract the desired scene parameters. Examples are a variety of data acquisition geometries, different waveforms and the synthesis of large synthetic apertures (SAR). While many of these concepts are employed more or less agnostically of the terrain and actual scene content, there are good reasons to adapt the acquisition process dynamically to the scene, an idea coined as cognitive radar by Simon Haykin [1]. He refers to the behaviour of some bats to adapt their echolocation transmission when they assume objects in front of them such as possible prey. This strategy, induces a richer

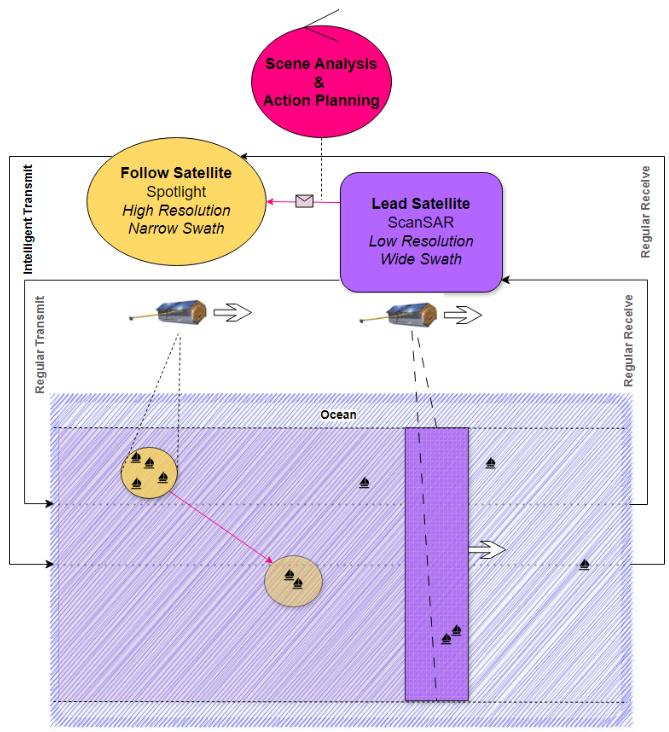

Fig. 1: The cognitive radar concept involves a feedback loop where a primary scanning satellite with a high false alarm rate directs another satellite to validate potential ship detections with higher resolution but smaller swath width. The second satellite relocates its spotlight to the most promising positions in the ocean based on the current state of accumulated SAR data. The spotlight repositioning and the decision process are denoted in pink, and multiple such beams could be active simultaneously.

neural response and perception that allows bats to gain a deeper understanding of their surroundings [2].
In radar systems, waveform diversity and digital beamforming (DBF) as well as active phased array antennas allow for a similar level of dynamic adaptation of the sensors. Challenging aspects are the real-time scene analysis and the corresponding action planning which are both crucial for improving the information retrieval and resource management beyond established 'static' concepts.
The application range of cognitive radar systems encompasses a wide spectrum. One promising example in the area of remote sensing is the detection of small ships or oil spills in large maritime areas or other forms of object detection in large-scale SAR imaging.
These scenarios suffer from a well-known trade-off between azimuth resolution and swath width on the ground [3]. DBF effectively enables to trade resolution for swath width and vice versa. This allows for a variety of different hybrid acquisition modes as discussed in [4]. An example of a cognitive radar concept for spaceborne maritime ship detection is illustrated in Fig. 1. A leading satellite in a 'static' ScanSAR mode scans the ocean with a low false negative rate for ships with a low resolution yet wide swath. It analyses the collected sensory data in real-time either onboard or on a ground station leading to strong demand for computational hardware dedicated to (quantum-) machine learning for a fast automated mode adaptation. Based on statistical similarities it then segments the scene into classes that likely represent either ship or ocean, accepting a high false alarm rate to ideally include all ships in the scene corresponding to a low false negative rate. Based on this prior analysis the following satellite is tasked to use one or more high-resolution spotlight beams centred on the most promising positions to clarify whether a ship is truly present.
Instead of first principle methods, this paper aims for efficient data-driven approaches to determine policies for the action planning step of the follow satellite, utilising reinforcement learning. These methods typically involve exploring large state-action spaces for finding good policies as well as rewarding, e.g. in our scenario, detecting ships and saving resources. Different reinforcement learning schemes are discussed in [5]. A promising quantum embedding of an actor-critic approach for flight path optimisation of UAVs is described in [6] which reports better trainability compared to non-quantum implementations. The aforementioned advantages of quantum computing seem to enable faster convergence for neural networks utilizing trainable quantum layers as reported in [6] [7] and [8]. In contrast to classical artificial neural networks, their quantum pendant is typically built of parameterized variational quantum circuits whose parameters function as trainable weights in the network. Inputs are encoded very similarly into the circuit. Exploring large state-action spaces remains a major challenge in reinforcement learning [5] which might profit from high dimensional quantum mechanical correlations.

2. PROBLEM DESCRIPTION

The goal of our concept is to achieve a better detection quality of ships in large maritime areas without prior knowledge of the scene using fewer resources than with a scene-agnostic acquisition. With the chosen satellite constellation two important tasks arise. The first is to find potential candidates in the wide swath SAR acquisition of the lead satellite and the second is the automatic steering of spotlight beams towards these candidates. The concept relies on a good choice of positions to steer towards instead of a second full coverage of the scene. Heuristic approaches could involve ordering possible targets by likelihood yet sparing sufficiently plausible positions to save resources. Since modern phased array antennas as well as

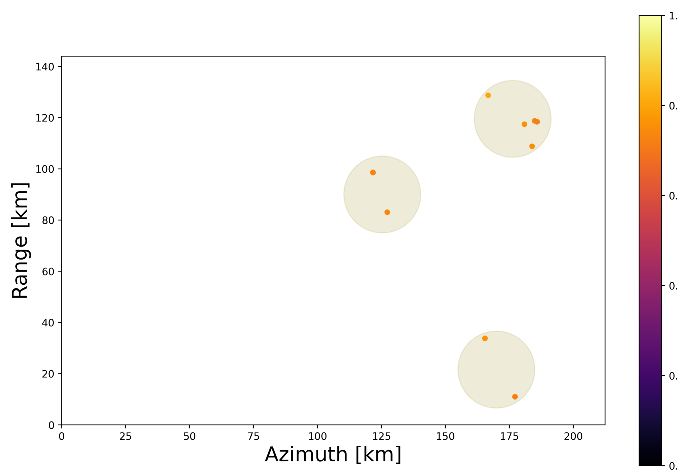

Fig. 2: Simplification of the beam steering problem: Only nine potential ship positions in a large scene are considered. Their probability is inferred from a prior low-resolution image of the whole scene and colour-coded. A limited number of additional acquisitions with a high-resolution mode (three circles in dark khaki) ideally covers all of the potential ships for validation with high gain.

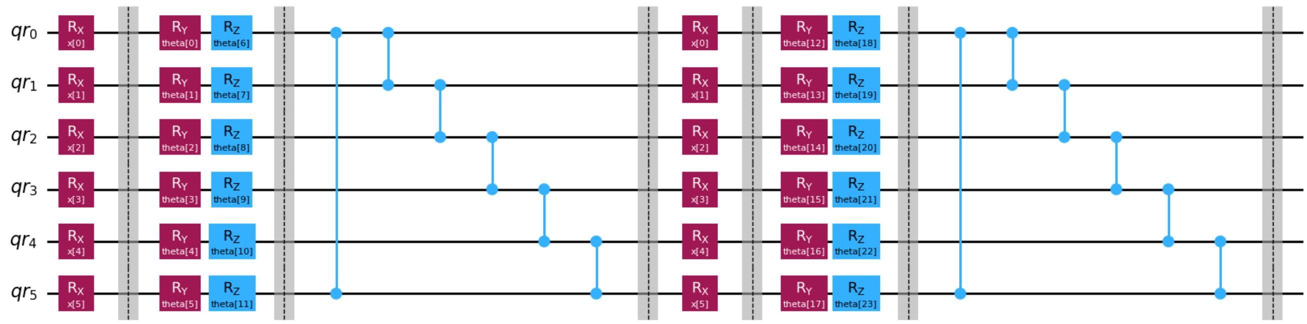

Fig. 3: Illustration of a variational quantum circuit: Horizontal lines correspond to the state of one qubit each. A vector $\boldsymbol{x}$ of six input variables is encoded into the collective quantum state of six qubits via single qubit R_x rotations. Trainable parameters $\boldsymbol{\vartheta}$ are encoded via single qubit R_y and R_z rotations of the state. Entanglement is implemented via two-qubit, controlled z gates (vertical blue lines) and the whole procedure is repeated another time with independent parameters.

DBF techniques allow a quick adaptation on the technical level, sophisticated adaptations of multiple high-resolution beams are possible in principle, allowing for a rich operational complexity. This makes first principle approaches challenging and motivates us to test data-driven approaches using quantum machine learning. Due to the current limitations of real and classically simulated quantum computers regarding the number of logical quantum bits (qubits), we simplify the problem of planning the beam steering.

We restrict the number of candidates to consider to ten or fewer positions in the scene and their estimated probability of finding ships among them, as shown in Fig. 2. We further restrict the beam steering to three high-resolution acquisitions which we approximate as circles of 15km radius on the ground. For the lead satellite, we use labelled focused SAR images acquired by a Sentinel-1 wide swath mode of 144 km $\times$ 212 km. We further ignore ship movement during the acquisitions. To quantify the success of the steering we suggest the true positive rate given by the number of true ships covered by the three acquisitions together divided by all true ships in the scene as we are interested in covering as many of the true ships with high-resolution acquisitions as possible within the above constraints.

3. METHODS

We use labelled data of large maritime scenes from [9] and train a UNET model [10] with an additional sigmoid function to predict a probability map for all pixels of the scene using focused SAR images as input. Due to the sparsity of the ships in the ocean, fast and stable convergence is enabled by using focal binary cross entropy [11] as a loss function during training of the UNET model.

Ideally one would like to compare this predicted probability map before and after the high-resolution acquisition of the following satellite to quantify the concept as a whole. Yet, current quantum computers and classical simulators allow not much more than 30 logical qubits to encode information. In principle, each qubit can encode a real number, for practical reasons a rational number. Therefore, the probability map needs to be compressed to such a size to be encoded into a quantum circuit which motivates the choice of our simplifications of the steering task. To encode k potential ships with their probability $q = 3k$ qubits are necessary.

The dataset contains at least an order of magnitude more ships per scene, such that we uniformly sample $k \in [3, 10]$ candidates with a probability above a threshold each. As an ansatz for a solution, we choose a variational quantum circuit similar to Fig. 3 which consists of a quantum circuit and a classical optimisation for training. The circuit encodes inputs as well as trainable parameters into rotations of entangled qubits in a high dimensional (2^q) Hilbert space. The network can be rewarded based on the true positive rate of its choice of beam locations.

We encode the ship candidates and trainable parameters in the style of the variational quantum circuit (VQC) in Fig. 3 with the appropriate number of qubits and subsequently steer our spotlights to the positions corresponding to the VQC's decoded outputs. Initialized in the ground state $|0\rangle$, a ship's position in azimuth is encoded into the angle α towards one qubit's new state $R_x^\alpha\big(\,|0\rangle\,\big)$ via rotation around the x axis of its Bloch sphere. The maximal position corresponds to an angle of 2π. The equivalent for the position in range. Its probability can be stored in the same way such that a value of 1 corresponds to an angle of 2π.

Consequently, k potential ships can be encoded into $3k$ qubits together with their probability or $2k$ without. The state of each qubit is then modified by different rotations around the y axis, R_y, and the z axis, R_z, with variable angles which function as trainable parameters of the circuit. To break their independence, neighbouring qubits are entangled via controlled z gates which rotate one qubit's state by π around its z axis if another control qubit is in the $|1\rangle$ state. The whole unitary operation is repeated multiple times with independent parameters to increase the circuit's complexity and establish a nonlinear relation between inputs and outputs. A higher complexity extends the space of functions that can be learned by the VQC allowing for better expressivity in theory. Measuring the state of the same $2l$ qubits at the end of the circuit many times (for statistical significance) allows decoding l pairs of positions in azimuth and range for the centre of l spotlight beams. The different angles can then be optimised for a desired output configuration of the beams while the remaining $3k - 2l$ qubits are not measured.

4. DISCUSSION & OUTLOOK

We are working on different reward-based optimisation strategies to reinforce useful steering behaviour. The variational quantum circuit can also be used inside machine learning frameworks that allow gradient descent-based optimisation via symbolic differentiation of an objective function. Therefore it is desirable to construct a differentiable reward function as a surrogate for the true positive rate. Alternative gradient-free methods for reinforcement could also turn out as useful. Moreover, we plan to explore the effect of weighting the actual knowledge gain from the spotlight acquisitions compared to the wide acquisition of our lead satellite alone for rewarding our model. Furthermore, we will study the influence of our model's hyperparameters, e.g., the number of repetitions in the circuit as well as different entangling schemes.

References

[1] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, Jan. 2006.

[2] Nobuo Suga, "Cortical computational maps for auditory imaging," *Neural Networks*, vol. 3, no. 1, pp. 3–21, Jan. 1990.

[3] Felipe Queiroz de Almeida, "Multichannel staggered sar for high-resolution wide-swath imaging," Tech. Rep., Karlsruher Institut für Technologie (KIT), Februar 2018.

[4] Gerhard Krieger, Nicolas Gebert, and Alberto Moreira, "Multidimensional Waveform Encoding: A New Digital Beamforming Technique for Synthetic Aperture Radar Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 31–46, Jan. 2008, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[5] Marc Vincent, *Reinforcement Learning for Multi-Function Radar Resource Management*, phdthesis, Sorbonne Université, Sept. 2023.

[6] Silvirianti, Bhaskara Narottama, and Soo Young Shin, "Layerwise Quantum Deep Reinforcement Learning for Joint Optimization of UAV Trajectory and Resource Allocation," *IEEE Internet of Things Journal*, pp. 1–1, 2023, Conference Name: IEEE Internet of Things Journal.

[7] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, June 2021, Number: 6 Publisher: Nature Publishing Group.

[8] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles, "Generalization in quantum machine learning from few training data," *Nature Communications*, vol. 13, no. 1, pp. 4919, Aug. 2022, Number: 1 Publisher: Nature Publishing Group.

[9] Tianwen Zhang, Xiaoling Zhang, Xiao Ke, Xu Zhan, Jun Shi, Shunjun Wei, Dece Pan, Jianwei Li, Hao Su, Yue Zhou, and Durga Kumar, "Ls-ssdd-v1.0: A deep learning dataset dedicated to small ship detection from large-scale sentinel-1 sar images," *Remote Sensing*, vol. 12, no. 18, 2020.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, arXiv:1505.04597 [cs].

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal Loss for Dense Object Detection," Feb. 2018, arXiv:1708.02002 [cs].