CLOUD DETECTION IN MULTISPECTRAL SATELLITE IMAGES USING SUPPORT VECTOR MACHINES WITH QUANTUM KERNELS

Artur Miroszewski¹, Jakub Mielczarek¹, Filip Szczepanek¹, Grzegorz Czelusta¹, Bartosz Grabowski², Bertrand Le Saux³, and Jakub Nalepa^{2,4}

¹Jagiellonian University, prof. S. Łojasiewicza 11, 30-348 Cracow, Poland
 ²KP Labs, Konarskiego 18C, 44-100 Gliwice, Poland
 ³European Space Agency, Largo Galileo Galilei 1, 00044 Frascati, Italy
 ⁴Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
 artur.miroszewski@uj.edu.pl, jnalepa@ieee.org

ABSTRACT

Support vector machines (SVMs) are a well-established classifier effectively deployed in an array of pattern recognition and classification tasks. In this work, we consider extending classic SVMs with quantum kernels and applying them to satellite data analysis. The design and implementation of SVMs with quantum kernels (hybrid SVMs) is presented. It consists of the Quantum Kernel Estimation (QKE) procedure combined with a classic SVM training routine. The pixel data are mapped to the Hilbert space using ZZ-feature maps acting on the parameterized ansatz state. The parameters are optimized to maximize the kernel target alignment. We approach the problem of cloud detection in satellite image data, which is one of the pivotal steps in both on-the-ground and on-board satellite image analysis processing chains. The experiments performed over the benchmark Landsat-8 multispectral dataset revealed that the simulated hybrid SVM successfully classifies satellite images with accuracy on par with classic SVMs.

1. INTRODUCTION

Satellite imaging plays an increasingly important role in various aspects of human activity. The spectrum of applications ranges from cartographic purposes [1, 2] through meteorology [3], ecology, and agronomy [4] to security and urban monitoring [5]. Consequently, dozens of terabytes of raw imaging data are generated daily from satellite constellations, such as the one built within the European Copernicus Programme. The large volume of multi- or hyperspectral images, which capture the detailed characteristics of the scanned materials, makes them difficult to transfer, store, and ultimately analyze. Therefore, their reduction through the extraction of useful information is a critical issue in real-world applications. An important step in the data analysis chain of optical satellite data is the identification of clouds. The interest is two-fold: on the one hand, such cloudy regions can be removed from further processing, as the objects of interest are likely to be obscured. On the other hand, efficient detection of cloud cover on the Earth surface is important in meteorological and climate research [6].

Since the reduction is performed on a huge amount of raw data, the efficiency of this process is a key factor in practice. Therefore, it is reasonable to search for new methods for analyzing such huge datasets, improving image data classification into cloudy and clear areas. In this paper, we investigate the classification performance of classic SVMs exploiting the radial basis function kernels, and those which benefit from the quantum kernels (introduced in Section 2). There are theoretical arguments [7–9] that the proposed quantum kernels are #P-hard to evaluate on a classical computer. Therefore, if they provide an advantage in classification accuracy, this would advocate a strong use case for quantum computers. Additionally, to get a deeper understanding of quantum machine learning mechanisms and show its usefulness in practice, it is pivotal to focus on widely adopted image data corresponding to real use cases. Thus, we tackle the cloud detection task in satellite image data, which is one of the most important processing steps for such imagery. Our experimental study performed over the benchmark multispectral image data acquired by the Landsat-8 satellite revealed that SVMs with quantum kernels offer a classification accuracy at least comparable to classic RBF kernel SVMs (Section 3).

2. MATERIALS AND METHODS

2.1. Data

We utilize satellite image data contained in the 38-Cloud dataset [10]. The data consist of Landsat 8 scene images cropped into 384×384 pixel patches. Each pixel has five values associated with it: intensity values in four spectral bands (blue: 450-515 nm, green: 520-600 nm, red: 630-680 nm, NIR: 845–885 nm) and a label corresponding to the fact whether a pixel contains a cloud or not. Therefore, the dimension of the data is m = 4. SVMs suffer from their high time and memory training complexity, which depend on the size of the training set. Since only a subset of all training vectors is annotated as support vectors during SVM training, we can effectively exploit only a subset of the most important examples [11]. To find the best training data, two metrics for patches were introduced: cloudiness C (the ratio of cloud pixels to all pixels in the patch), fill F (the ratio of physical pixels in the patch, as patches contain scene margins). Balanced training sets from patches with properties F = 100%, $40\% \leq C \leq 60\%$ are sampled by randomly selecting a fixed number of pixels.

2.2. Methods

In Fig. 1, we present a high-level flowchart of the proposed hybrid SVM procedure. First, we encode the data with the parameterized feature map consisting of ZZ-feature map acting on ansatz (see Fig. 2). Then we perform the Quantum Kernel Estimation (QKE) and obtain a quantum kernel, which target alignment is maximized with a classic optimization method. When optimization is finished, the final quantum kernel is passed to a classic SVM.



Fig. 1: Overview of the hybrid SVM design. The gray box indicates the part of the algorithm which was calculated on the quantum computer simulator.

2.2.1. Selected Feature Map

When considering the employment of quantum computation methods, a principal question that quickly arises pertains to the way in which classic input data will be loaded into the quantum circuit. In general, our aim will be to construct a unitary operator for each input datum x, such that applying it to the initial quantum zero state will leave us with a specified representation of x, $|\phi(x)\rangle$. This process is called *quantum embedding*, while any such map $x \mapsto |\phi(x)\rangle$ is known as a *quantum feature map*. Consider the unitary transformation

$$U_{\phi(x)} = \exp\left(i\sum_{S\subseteq[n]}\phi_S(x)\prod_{i\in S}P_i\right),\tag{1}$$

being a general quantum circuit Pauli expansion of an n-qubit unitary transformation. The index S describes the connectivities between different qubits: $S \in \{\binom{n}{k} \text{ combinations}, k \in$ $\{1, \ldots, n\}\}$, P_i are the basic Pauli gates that act on the i^{th} quantum register and the data mapping function is $\phi_{\{i\}}(x) =$ $\pi x_i, \phi_{\{i,j\}}(x) = \pi (1 - x_i)(1 - x_j)$. The number of qubits used can be identified with the dimension of the data n = m. Following [7], we restrict the above unitary to k = 2 connectivities with single-qubit gates $P_{\{a\}} = Z_a$, two-qubit gates $P_{\{b,c\}} = Z_b Z_c, a, b, c \in \{0, \dots, n-1\}$. This transformation is called a ZZ feature map with one repetition. It is already #P-hard to calculate classically [8], but shows no computational advantage over the classical kernel estimation, performed by random sampling [9]. To increase the complexity of the classical simulation of the ZZ feature map, additional bases-changing layers are included by repeating the $U_{\phi(x)}$ transformation

$$\mathcal{U}_{\phi(x)}^{d} = \left(\tilde{U}_{\phi(x)}\right)^{d} = \left(U_{\phi(x)}H^{\otimes n}\right)^{d}, d \in \mathbb{N}.$$
 (2)

The transformation $\mathcal{U}_{\phi(x)}^d$ is called the ZZ feature map with d repetitions.

2.2.2. Circuit Parameterization

Having defined the ZZ feature map, we turn our attention to the possibility of introducing additional parameters into the circuit to maximize the kernel target alignment of the data in the feature space. We follow the approach of modifying the initial state of the circuit [12] on which the feature map acts. The initial state will be parameterized with continuous variables θ (Fig. 2), with respect to which we will perform kernel target alignment optimization.

2.2.3. Kernel Target Alignment

Considering a collection of quantum states obtained by means of applying a quantum embedding $|\phi(x)\rangle = \mathcal{U}_{\phi(x_i)} V_{\theta} |0\rangle^{\otimes 2}$ to different classic input data, it is straightforward to reason about them in terms of kernel methods. The kernel \mathcal{K} with respect to any two embedded data x_i, x_j can be defined as the fidelity between the resulting quantum states,

$$\mathcal{K}_{ij} = |\langle \phi(x_i) | \phi(x_j) \rangle|^2 = |\langle 0|^{\otimes n} V_{\theta}^{\dagger} \mathcal{U}_{\phi(x_j)}^{d\dagger} \mathcal{U}_{\phi(x_i)}^{d} V_{\theta} | 0 \rangle^{\otimes n} |^2.$$
(3)

This kernel \mathcal{K} is known as the *quantum kernel*. Consider a kernel function given by

$$\bar{\mathcal{K}}_{ij} = \begin{cases} +1 & \text{if } x_i \text{ and } x_j \text{ are in the same class} \\ -1 & \text{if } x_i \text{ and } x_j \text{ are in different classes.} \end{cases}$$

$$|0\rangle \qquad R_{Y}(\theta_{0}) \qquad R_{Z}(\theta_{2}) \qquad R_{Y}(\theta_{4}) \qquad R_{Z}(\theta_{6}) \qquad R_{Y}(\theta_{8}) \qquad R_{Z}(\theta_{10}) \qquad \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Z}(\theta_{3}) \qquad R_{Y}(\theta_{5}) \qquad R_{Z}(\theta_{7}) \qquad R_{Y}(\theta_{9}) \qquad R_{Z}(\theta_{11}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Z}(\theta_{11}) \qquad R_{Z}(\theta_{11}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Z}(\theta_{11}) \qquad R_{Z}(\theta_{11}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Z}(\theta_{11}) \qquad R_{Z}(\theta_{11}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Z}(\theta_{11}) \qquad R_{Z}(\theta_{11}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Y}(\theta_{1}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad \\ \\ |0\rangle \qquad R_{Y}(\theta_{1}) \qquad R_{Y}(\theta_{1}) \qquad \\ \\ |0\rangle \qquad \\ \\ |0\rangle$$

Fig. 2: The ansatz $V_{\theta}|0\rangle^{\otimes 2}$ circuit, which prepares the initial state for the kernel estimation for classical data of dimension m = 2.

It shows a clear distinction between classes of data points and is called an *ideal* kernel. In general, in almost every situation, one will not be able to find the exact feature map, which gives rise to the ideal kernel. Therefore, parametrized families of feature maps are used to optimize the resulting kernel matrix in such a way that it resembles the ideal kernel as closely as possible. The function that indicates the similarity between a specific and ideal kernel matrices is called kernel target alignment [13]

$$\mathcal{T}(\mathcal{K}) = \frac{\langle \mathcal{K}, \bar{\mathcal{K}} \rangle_F}{\sqrt{\langle \mathcal{K}, \mathcal{K} \rangle_F \langle \bar{\mathcal{K}}, \bar{\mathcal{K}} \rangle_F}},\tag{4}$$

where $\langle A, B \rangle_F = Tr\{A^TB\}$ is a Frobenius inner product.

3. EXPERIMENTAL RESULTS

The objective of our study is to compare hybrid SVMs with their classic counterparts. The results, presented in Table 1, are obtained by using the Qiskit Aer simulator, whereas the optimization algorithm is the standard simultaneous perturbation stochastic approximation. The SVM score is obtained from the sklearn support vector classification, with the radial basis function (RBF) kernel with default $\gamma = \frac{1}{m\sigma^2}$ (where m = 2 is the number of features and σ^2 is the variance of the data) and C = 1, the latter being the regularization parameter. For each simulation run, we randomly sample 800 pixels for the training set, and 200 pixels for the test set (the training and test sets are non-overlapping). To keep the size of the quantum circuit minimal, we decrease the number of data features by running a principal component analysis before feeding it into the algorithm. The number of four spectral bands is reduced to two features. The ZZ-feature map is chosen to consist of d = 2 repetitions. The results of the Wilcoxon matched-pairs signed rank test show that there is no statistically significant difference between hybrid SVMs with a quantum kernel and classic SVMs with an RBF kernel (at p < 0.05). Therefore, it shows that the classification methods with quantum kernels based on ZZ-feature map are, at least, competitive with classical SVM models. With the further development of quantum kernels, we expect hybrid methods to be advantageous over classical classification methods.

4. CONCLUSIONS

We introduced the design and implementation of an SVM with quantum kernels. The proposed algorithm was experimentally verified on the cloud detection benchmark dataset. The main takeaway from the work is that-at this stage-SVMs with the quantum kernel have a classification accuracy on par with classic SVMs with RBF kernel. The experiment was performed with a quantum computer simulator. In [14] the use of underparametrized quantum circuits for similar task was performed on the whole 38-Clouds data set. Current results are consistent with this work. To estimate the effect of noise and better understand the computational time, more work should focus on running the algorithm on quantum computers. We anticipate that further investigation of the quantum feature maps-including using full dimension of the dataset (m = 4), new data mapping functions or different generators-will result in an additional improvement of the classification performance. This would indicate a strong use case for quantum computers in SVM models.

5. REFERENCES

- [1] "Copernicus land monitoring service," https://land.copernicus.eu/, Accessed: 2022-07-12.
- [2] Javiera Castillo-Navarro, Bertrand Le Saux, Alexandre Boulch, Nicolas Audebert, and Sébastien Lefèvre, "Semi-Supervised Semantic Segmentation in Earth Observation: The MiniFrance suite, dataset analysis and multi-task network study," *Machine Learning*, pp. 1–36, 2021.
- [3] Derek Hanson, James Peronto, and Douglas Hilderbrand, "NOAA's Eyes in the Sky - After five decades of weather forecasting with environmental satellites, what do future satellites promise for meteorologists and society?," *World Meteorological Organization (WMO) Bulletin*, vol. 62, no. 1, pp. 17–21, 2013.
- [4] Jakub Nalepa, Bertrand Le Saux, Nicolas Longépé, Lukasz Tulczyjew, Michal Myller, Michal Kawulok, Krzysztof Smykala, and Michal Gumiela, "The Hyperview Challenge: Estimating Soil Parameters from Hyperspectral Images," in 2022 IEEE International Conference on Image Processing (ICIP), 2022, pp. 4268– 4272.

Table 1: The results of the circuit simulations on 20 different training-test 38-Cloud splits: hSVM and SVM indicate classification accuracy for hybrid SVM and classic SVM methods. T_i and T_f show kernel target alignment before and after optimization.

	\mathcal{T}_i	\mathcal{T}_{f}	hSVM	SVM	
Average	0.049	0.081	0.778	0.788	
Standard deviation	0.018	0.024	0.038	0.029	

- [5] Nicolas Audebert, Alexandre Boulch, Hicham Randrianarivo, Bertrand Le Saux, Marin Ferecatu, Sébastien Lefevre, and Renaud Marlet, "Deep learning for urban remote sensing," in 2017 Joint Urban Remote Sensing Event (JURSE). IEEE, 2017, pp. 1–4.
- [6] Bartosz Grabowski, Maciej Ziaja, Michal Kawulok, and Jakub Nalepa, "Towards Robust Cloud Detection in Satellite Images Using U-Nets," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 4099–4102.
- [7] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta, "Supervised learning with quantumenhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.
- [8] Leslie Ann Goldberg and Heng Guo, "The complexity of approximating complex-valued ising and tutte partition functions," *computational complexity*, vol. 26, no. 4, pp. 765–833, 2017.
- [9] Tommaso F Demarie, Yingkai Ouyang, and Joseph F Fitzsimons, "Classical verification of quantum circuits containing few basis changes," *Physical Review A*, vol. 97, no. 4, pp. 042319, 2018.
- [10] S. Mohajerani, T. A. Krammer, and P. Saeedi, "A Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks," in 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), 2018.
- [11] Jakub Nalepa and Michal Kawulok, "Selecting training sets for support vector machines: a review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 857–900, Aug 2019.
- [12] Jennifer R Glick, Tanvi P Gujarati, Antonio D Corcoles, Youngseok Kim, Abhinav Kandala, Jay M Gambetta, and Kristan Temme, "Covariant quantum kernels for data with group structure," *arXiv preprint arXiv:2105.03406*, 2021.
- [13] Nello Cristianini, John Shawe-Taylor, Andre Elisseeff, and Jaz Kandola, "On kernel-target alignment," *Advances in neural information processing systems*, vol. 14, 2001.

[14] Artur Miroszewski, Jakub Mielczarek, Grzegorz Czelusta, Filip Szczepanek, Bartosz Grabowski, Bertrand Le Saux, and Jakub Nalepa, "Detecting clouds in multispectral satellite images using quantum-kernel support vector machines," *arXiv preprint arXiv:2302.08270*, 2023.

Acknowledgements

This work was funded by the European Space Agency, and supported by the ESA Φ -lab (https://philab.phi.esa.int/) AI-enhanced Quantum Computing for Earth Observation (QC4EO) initiative, under ESA contract No. 4000137725/22/NL/GLC/my. AM, JM, GC, and FS were supported by the Priority Research Areas Anthropocene and Digiworld under the program Excellence Initiative – Research University at the Jagiellonian University in Kraków. JN was supported by the Silesian University of Technology grant for maintaining and developing research potential.