Use of quantum computation in satellite data analysis

Artur Miroszewski

European Space Agency project:
'Quantum machine learning for analyzing multi- and hyperspectral satellite images'

Quantum Cosmos Lab, Jagiellonian University

JAGIELLONIAN UNIVERSITY
IN KRAKÓW

Quantum Cosmos Lab

European Space Agency

The ESA Φ-lab

KP LABS Mission Complete

Quantum machine learning for analyzing multi- and hyperspectral satellite images

Project's goal:
Investigate the possibility of using quantum computing and quantum machine learning for remote sensing and earth observation

Task: Cloud detection

- Data reduction - cloudy regions can be removed from further analysis
- Cloud cover important for meteorological and climate research.

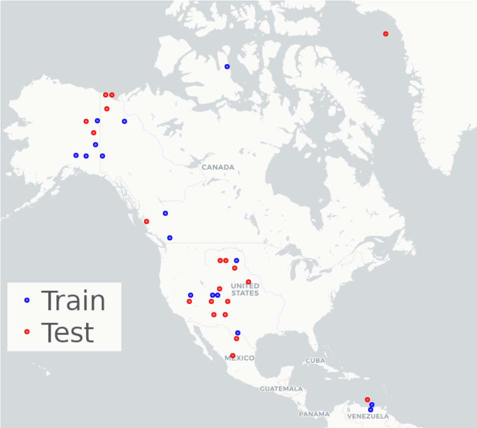
Methods:

- Hybrid SVM
- Quantum SVM
- Quantum neural networks

Recently we have started another ESA project: 'Quantum Advantage for Earth Observation'

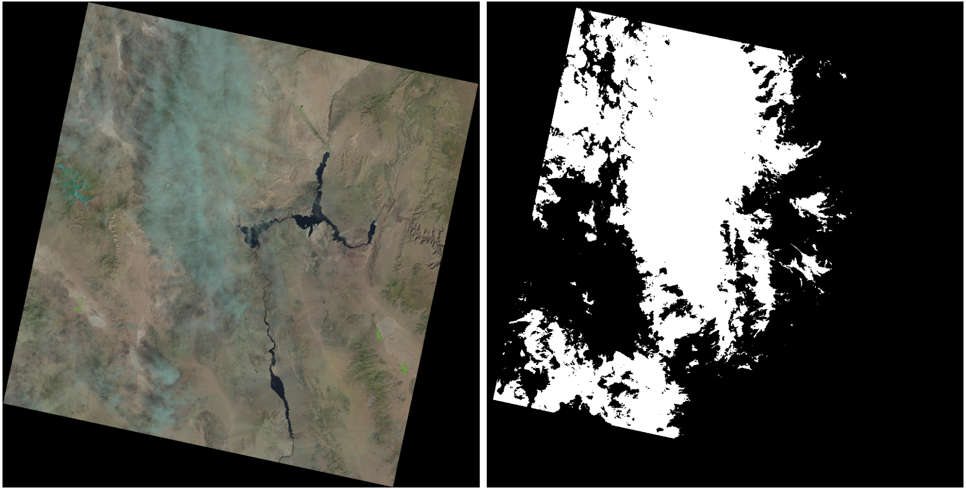
Data: 38-Cloud data set

- 38 Landsat-8 scene images
  - 18 training scenes
  - 20 test scenes
- Each scene image divided into ∼420 square patches of size $384px \times 384px$ patches
  - ∼ 1.2 bln pixels in training set
  - ∼ 1.4 bln pixels in test set
- Four spectral bands + ground-truth cloud mask:
  - red: 630-680 nm
  - green: 520-600 nm
  - blue: 450-515 nm
  - NIR: 845-885 nm



S. Mohajerani et al. "A Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks," 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), Vancouver, BC, 2018

38-Cloud data: scenes



S. Mohajerani et al. "A Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks," 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), Vancouver, BC, 2018

Support Vector Machines

- Classification supervised learning algorithm
- Maximization of margins

$$\min_{w,b} \frac{1}{2}|w|^2,$$

$$\textit{such that}: \; y^{(i)}(w \cdot w^{(i)} + b) \geq 1, i = 1, \ldots, m$$

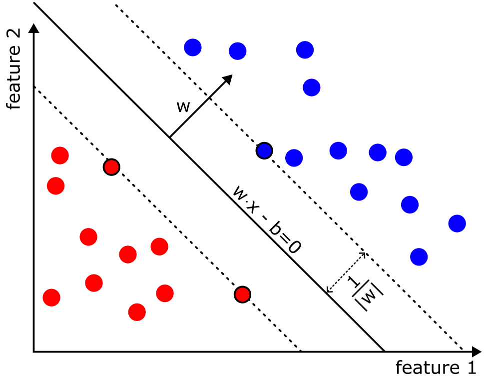
- Dual formulation

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\textit{such that}: \; \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

- Test phase

$$\textit{decision}: \; \text{sign}\left(\sum_{i=1}^{m} y^{(i)} \alpha_i \langle x, x^{(i)} \rangle + b \right)$$



Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

SVM kernels

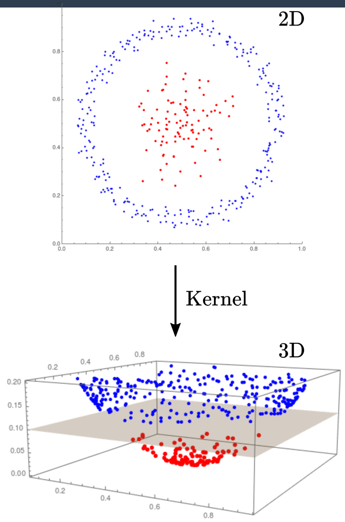
- Both training and test phase depends on $\langle x^{(i)}, x^{(j)} \rangle$
- Kernel trick: exchange inner product for some 'arbitrary' similarity measure

$$\langle x^{(i)}, x^{(j)} \rangle \mapsto K_{ij} = \langle \phi\left(x^{(i)}\right), \phi\left(x^{(j)}\right) \rangle$$

- The transformation ϕ:
  - leads to higher dimensional space - improved separability
  - allows for nonlinear class boundaries

Computational complexity for kernelized SVM:

- Training phase: $\sim \mathcal{O}(n \cdot m^2)$
- Test phase: $\sim \mathcal{O}(n \cdot m)$



Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

Why go quantum?

- Hilbert space grows exponentially (2^n) with number of qubits $\rightarrow$ **improved separability of data**
- The possibility of utilizing entanglement for finding complex correlations and patterns in the data
- Complicated kernel functions in Hilbert space (especially with entanglement) can be estimated **faster on quantum computer**
- The noise in current quantum computing devices is **tolerable in machine learning** applications

Challenges:

- Classical data encoding into quantum Hilbert space
- Untrainability
- Noise issues

“Introduction” to quantum computers

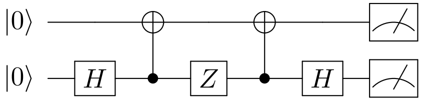
Noisy Intermediate-Scale Quantum (NISQ) devices era

Main types of QC

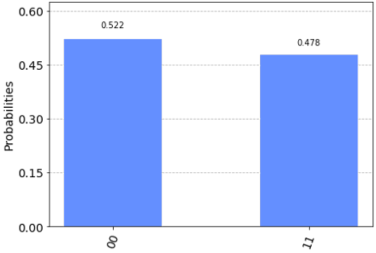
- **Gate based**
- Adiabatic

Gate based quantum computers

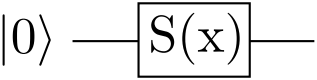
- **superconducting qubits**
- photonic
- ion traps
- topological
- quantum dots
- ...



- The measurement is customarily done in a Z-basis
- Probabilities are obtained by repeating the circuit runs

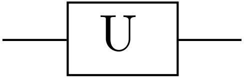


State preparation



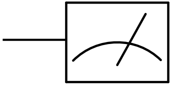
- QRAM
- Data encoding
  - basis
  - amplitude
  - angle
  - IQP

Quantum algorithm



- Most ‘first wave’ QML algorithms considered only this part
- Exponential speedups

Measurement



- Makes the algorithm non-deterministic
- Introduces additional polynomial scaling to the algorithm’s complexity

The idea:
Perform kernel trick on a quantum mechanical Hilbert space

1. Embed the classical data into the Hilbert space

$$\mathcal{U}_{\phi(x)} = \exp\left(i \sum_{S \subseteq [n]} \phi_S(x) \prod_{i \in S} Z_i \right) \cdot H^{\otimes n}$$

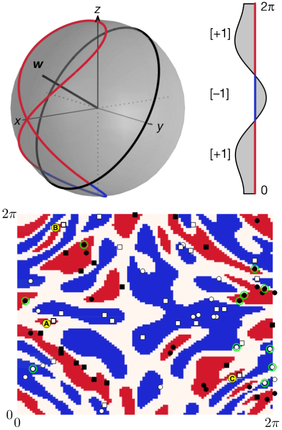
$$\left(\mathcal{U}_{\phi(x)}\right)^d |0\rangle = |\psi(x)\rangle$$

2. Estimate the kernel matrix with fidelity kernel function

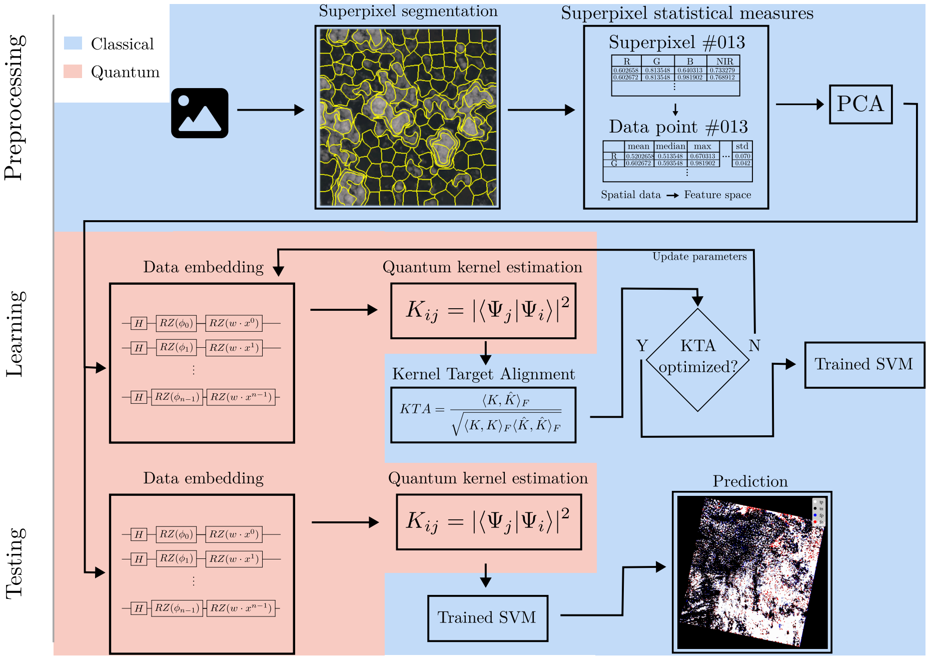
$$K_{ij} = |\langle \psi(x_j) | \psi(x_i) \rangle|^2 =$$

$$= |\langle 0| \left(\mathcal{U}^\dagger_{\phi(x_j)}\right)^d \left(\mathcal{U}_{\phi(x_i)}\right)^d |0\rangle|^2$$

3. Perform classical SVM on quantum kernel matrix



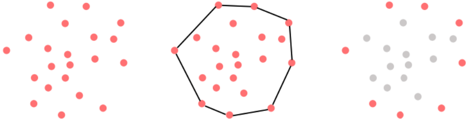
Havlíček, Vojtěch, et al. "Supervised learning with quantum-enhanced feature spaces." Nature 567.7747 (2019): 209-212.



Preprocessing
Learning
Testing
Classical
Quantum
Superpixel segmentation
Superpixel statistical measures
Superpixel #013
R G B NIR
0.602658 0.813548 0.640313 0.733279
0.602672 0.813548 0.981902 0.768912
Data point #013
mean median max std
R 0.5202658 0.513548 0.670313 0.070
G 0.602672 0.593548 0.981902 0.042
Spatial data → Feature space
PCA
Data embedding
H $RZ(\phi_0)$ $RZ(w \cdot x^0)$
H $RZ(\phi_1)$ $RZ(w \cdot x^1)$
H $RZ(\phi_{n-1})$ $RZ(w \cdot x^{n-1})$
Quantum kernel estimation
$K_{ij} = |\langle \Psi_j | \Psi_i \rangle|^2$
Kernel Target Alignment
$KTA = \frac{\langle K, \hat{K} \rangle_F}{\sqrt{\langle K, K \rangle_F \langle \hat{K}, \hat{K} \rangle_F}}$
Update parameters
KTA optimized?
Y
N
Trained SVM
Data embedding
Quantum kernel estimation
$K_{ij} = |\langle \Psi_j | \Psi_i \rangle|^2$
Trained SVM
Prediction

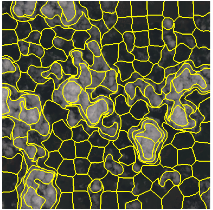
Training data reduction

Geometric reduction



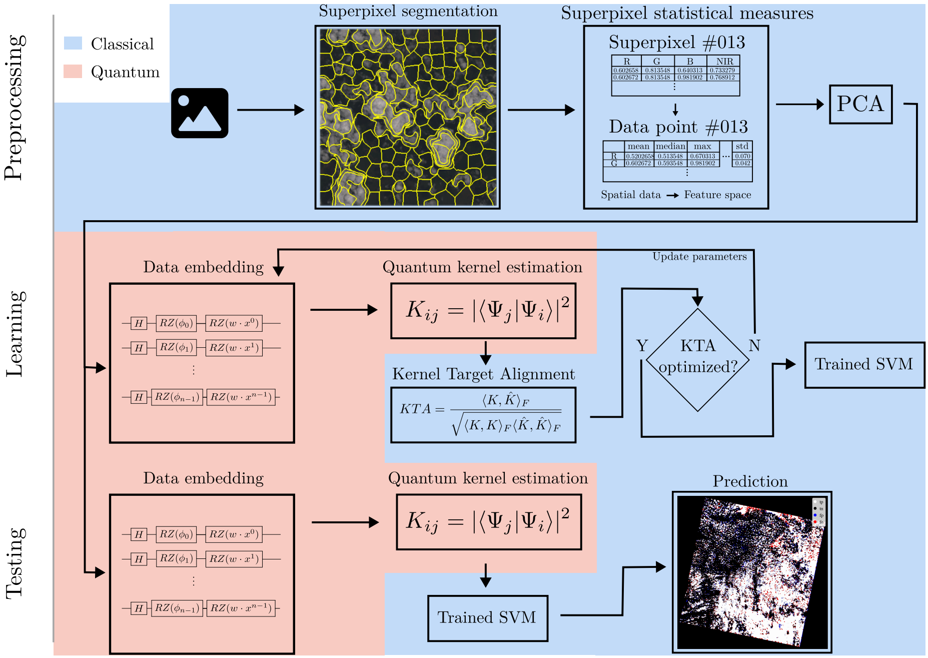
- Reduction performed in the feature space
- The border of data clusters important (especially for hm-SVM)
- Clusters obtained by K-means or X-means algorithms
- ‘Guaranteed’ classification performance
- $\sim 10\times$ reduction obtained

Superpixel segmentation

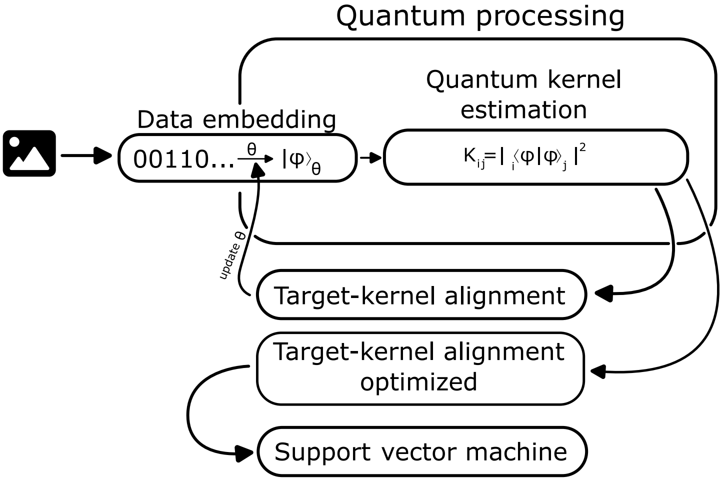


- Reduction performed on images
- Segments obtained with Simple Linear Iterative Clustering (SLIC)
- Statistical measures extracted from segments
- Testing also has to be done on superpixels
- $\sim 1300\times$ reduction obtained
- Bottlenecks: unsupervised ($\sim 96\%$ pixel homogeneity), RGB features

Nalepa, Jakub, and Michal Kawulok. "Selecting training sets for support vector machines: a review." Artificial Intelligence Review 52.2 (2019): 857-900.



Preprocessing
Learning
Testing
Classical
Quantum
Superpixel segmentation
Superpixel statistical measures
Superpixel #013
R G B NIR
0.602658 0.813548 0.640313 0.733279
0.602672 0.813548 0.981902 0.768912
Data point #013
mean median max std
R 0.5202658 0.513548 0.670313 0.070
G 0.602672 0.593548 0.981902 0.042
Spatial data → Feature space
PCA
Data embedding
H $RZ(\phi_0)$ $RZ(w \cdot x^0)$
H $RZ(\phi_1)$ $RZ(w \cdot x^1)$
H $RZ(\phi_{n-1})$ $RZ(w \cdot x^{n-1})$
Quantum kernel estimation
$K_{ij} = |\langle \Psi_j | \Psi_i \rangle|^2$
Kernel Target Alignment
$KTA = \frac{\langle K, \hat{K} \rangle_F}{\sqrt{\langle K, K \rangle_F \langle \hat{K}, \hat{K} \rangle_F}}$
Update parameters
KTA optimized?
Y
N
Trained SVM
Data embedding
H $RZ(\phi_0)$ $RZ(w \cdot x^0)$
H $RZ(\phi_1)$ $RZ(w \cdot x^1)$
H $RZ(\phi_{n-1})$ $RZ(w \cdot x^{n-1})$
Quantum kernel estimation
$K_{ij} = |\langle \Psi_j | \Psi_i \rangle|^2$
Trained SVM
Prediction


Quantum processing
Data embedding
00110... $\xrightarrow{\theta}$ $|\varphi\rangle_\theta$
Quantum kernel estimation
$K_{ij} = |{}_i\langle\varphi|\varphi\rangle_j|^2$
update θ
Target-kernel alignment
Target-kernel alignment optimized
Support vector machine

Target kernel alignment

Ideal kernel

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

For supervised learning problems

$$\bar{\mathcal{K}}_{ij} = y_i y_j$$

Target kernel alignment

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

$$\langle \mathcal{A}, \mathcal{B} \rangle_F = Tr\{\mathcal{A}^T \mathcal{B}\}$$

- Similarity measure between kernel $\mathcal{K}$ and the ideal kernel $\bar{\mathcal{K}}$
- Related to “angle” between matrix vectors

$$cos(\alpha) \text{``}=\text{''} \frac{\mathcal{K} \cdot \bar{\mathcal{K}}}{||\mathcal{K}|| \, ||\bar{\mathcal{K}}||}$$

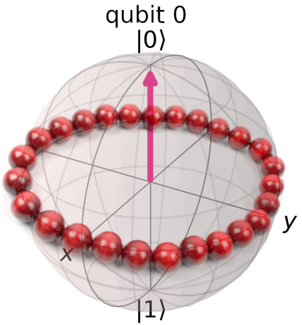
- For general kernels

$$\mathcal{T}(\mathcal{K}) \leq 1,$$

for QKE

$$\mathcal{T}(\mathcal{K}) \leq \frac{1}{\sqrt{2}}.$$

Toy model: Quantum beads



- Create N points on interval the $[0, 1]$ with alternating classes

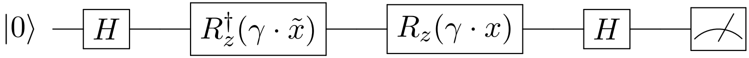
- Distribute them on one qubit with good separation between classes
- One-parameter embedding map

$$H \cdot R_Z(\gamma \cdot x) \cdot H|0\rangle$$

$$(H|0\rangle = |+\rangle, H|+\rangle = |0\rangle, H|-\rangle = |1\rangle)$$

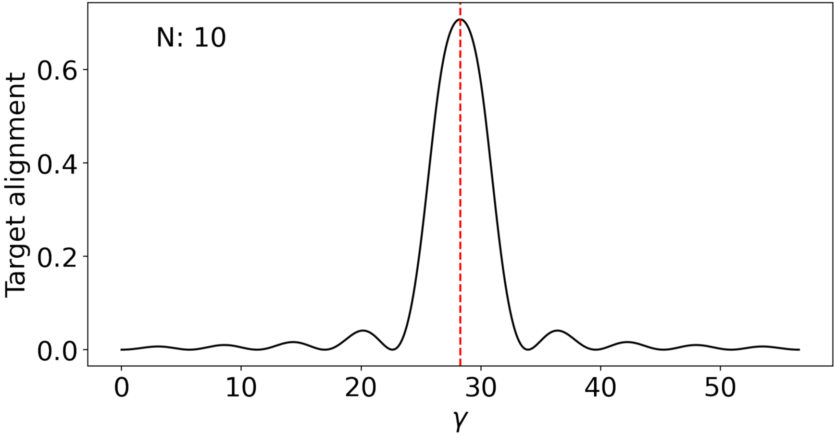
- For $\gamma = \pi \cdot (N-1)$ we have a perfect separation

$$-1 \to |0\rangle, 1 \to |1\rangle$$

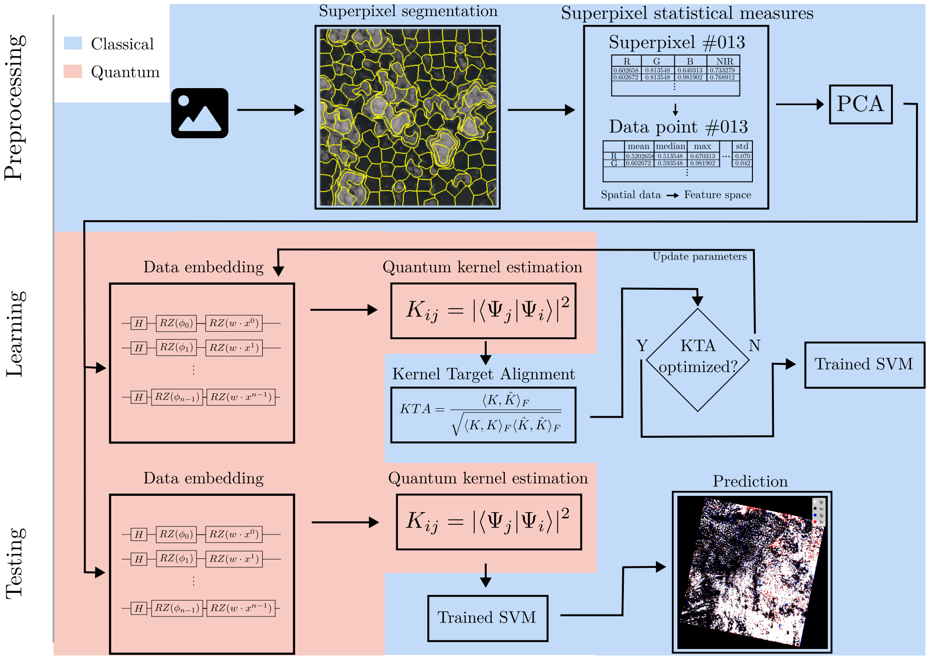


https://github.com/Quantum-Cosmos-Lab/quantum_maps_for_beads

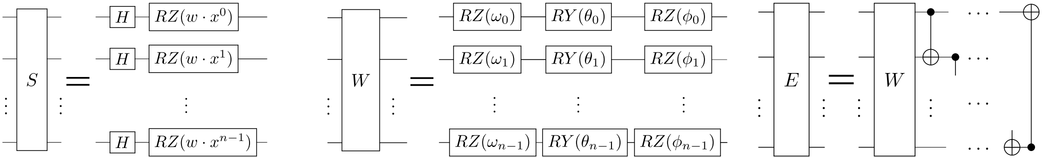
Target kernel alignment landscape



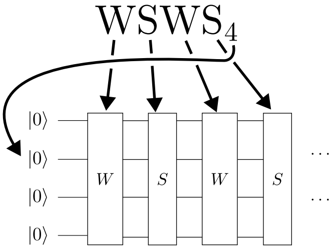
https://github.com/Quantum-Cosmos-Lab/quantum_maps_for_beads

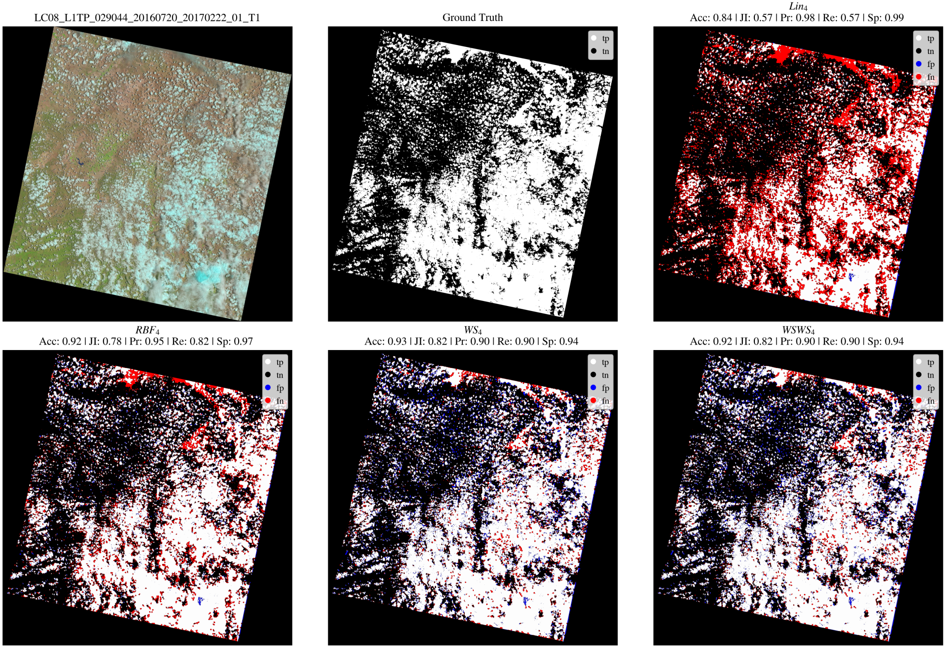

Preprocessing
Learning
Testing
Classical
Quantum
Superpixel segmentation
Superpixel statistical measures
Superpixel #013
Data point #013
Spatial data → Feature space
PCA
Data embedding
Quantum kernel estimation
$K_{ij} = |\langle \Psi_j | \Psi_i \rangle|^2$
Kernel Target Alignment
$KTA = \frac{\langle K, \hat{K} \rangle_F}{\sqrt{\langle K, K \rangle_F \langle \hat{K}, \hat{K} \rangle_F}}$
Update parameters
KTA optimized?
Y
N
Trained SVM
Prediction

Data embeddings



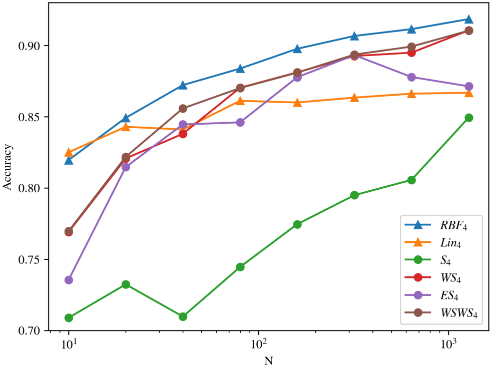
The studied quantum kernels are labeled with the data embedding and number of PCA components used.




LC08_L1TP_029044_20160720_20170222_01_T1
Ground Truth
tp
tn
Lin_4
Acc: 0.84 | JI: 0.57 | Pr: 0.98 | Re: 0.57 | Sp: 0.99
tp
tn
fp
fn
RBF_4
Acc: 0.92 | JI: 0.78 | Pr: 0.95 | Re: 0.82 | Sp: 0.97
tp
tn
fp
fn
WS_4
Acc: 0.93 | JI: 0.82 | Pr: 0.90 | Re: 0.90 | Sp: 0.94
tp
tn
fp
fn
$WSWS_4$
Acc: 0.92 | JI: 0.82 | Pr: 0.90 | Re: 0.90 | Sp: 0.94
tp
tn
fp
fn

Results

- First quantum-classical machine learning model for cloud detection in satellite imagery
- Proof of concept that quantum-classical algorithms are capable of performing useful tasks on real-life data
- There is no advantage in simple quantum kernels over classical kernels
- Comparison to the state of the art algorithms: U-Nets: 95 – 97% - after improving superpixel segmentation QSVM comparable
- Built a baseline model on an open multispectral data set



Accuracy

N	RBF_4	WS_4	$WSWS_4$
1280	0.919(0.010)	0.911(0.031)	0.910(0.031)

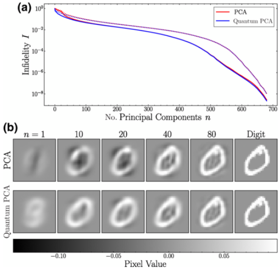
AM et al. "Detecting Clouds in Multispectral Satellite Images Using Quantum-Kernel Support Vector Machines" arXiv:2302.08270

Improvement of the current model

- superpixel algorithm - including the *NIR* spectral band in segmentation
- other data reduction methods
- dealing with underparameterization of the kernels
- Including ZZ feature map-like data embeddings

Feature extraction and selection for hyperspectral satellite data

- Recursive Feature Elimination (RFE)
- Feature selection based on quantum optimization
- Feature selection based on variational methods - VQE, VQC
- Quantum PCA
- Quantum autoencoders

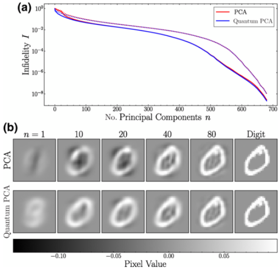


Improvement of the current model

- superpixel algorithm - including the *NIR* spectral band in segmentation
- other data reduction methods
- dealing with underparameterization of the kernels
- Including ZZ feature map-like data embeddings

Feature extraction and selection for hyperspectral satellite data

- Recursive Feature Elimination (RFE)
- Feature selection based on quantum optimization
- Feature selection based on variational methods - VQE, VQC
- Quantum PCA
- Quantum autoencoders



Thank you!

Quantum Kernel Estimation

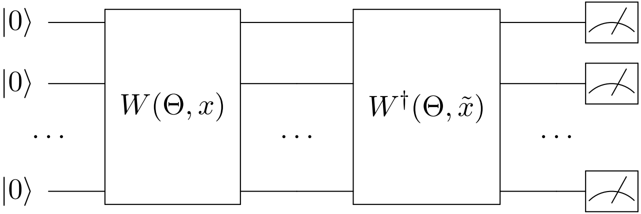
Kernel matrix elements given by an embedding state's overlap

$$\mathcal{K}_{ij} = |\langle \phi(x_j) | \phi(x_i) \rangle|^2, \quad \mathcal{O}(m^{4.67}/\epsilon^2)$$

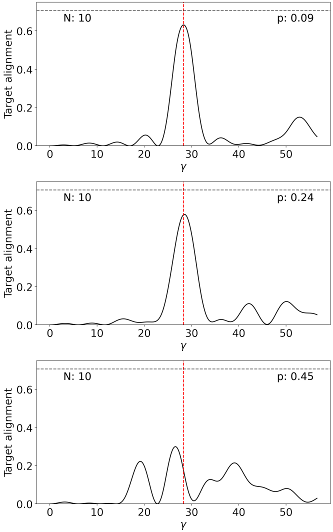
Calculated by:

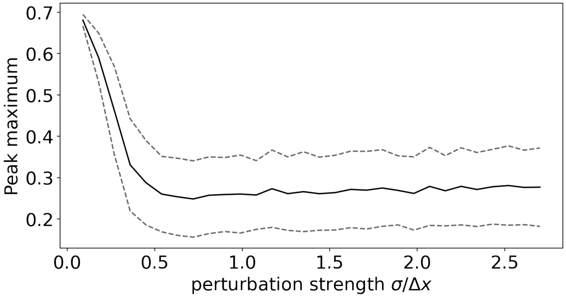
- Swap test
- Hadamard test
- Concatenating Hermitian conjugate

$$\mathcal{K}_{ij} = \langle 0|^{\otimes n} \, W^\dagger(\Theta, \tilde{x}) W(\Theta, x) |0\rangle^{\otimes n}$$



Target kernel alignment landscape: Barren Plateau?





https://github.com/Quantum-Cosmos-Lab/quantum_maps_for_beads