Quantum metric learning for image classification

Ahmed Hammad

Seoul national University of science and technology

Outline

1- QML classifier with fixed data embedding

2- QML with one qubit and variational data embedding (Toy example)

3- Why metric learning is important ?

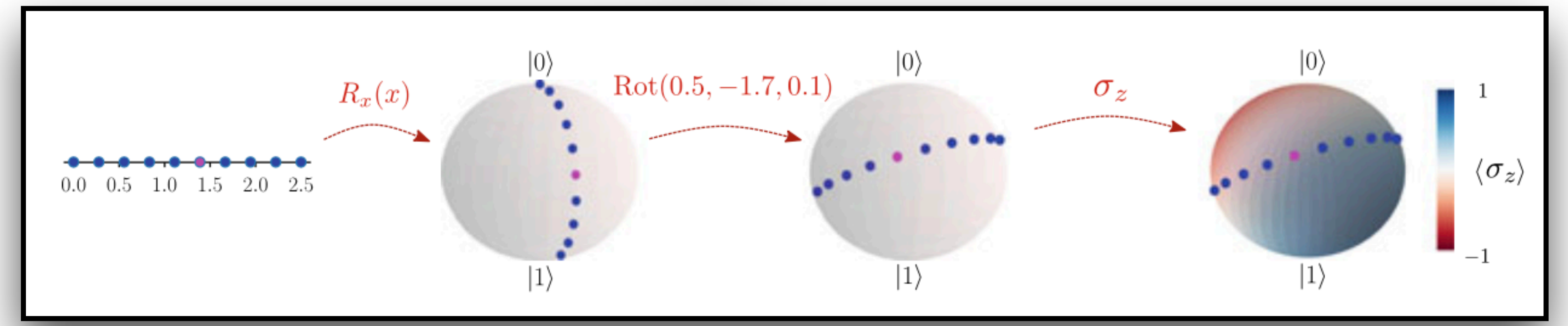
4- Metric learning in classical machine learning

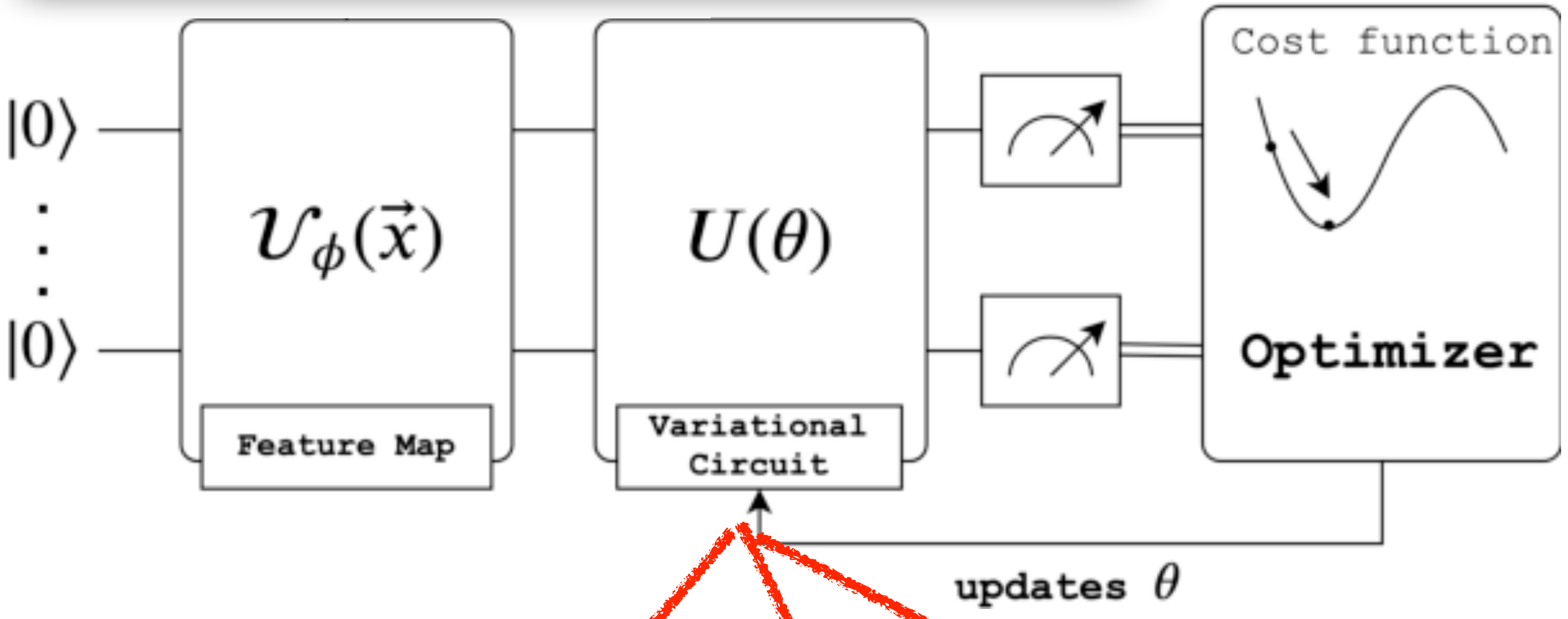
5- Hybrid classical-quantum model for image classification

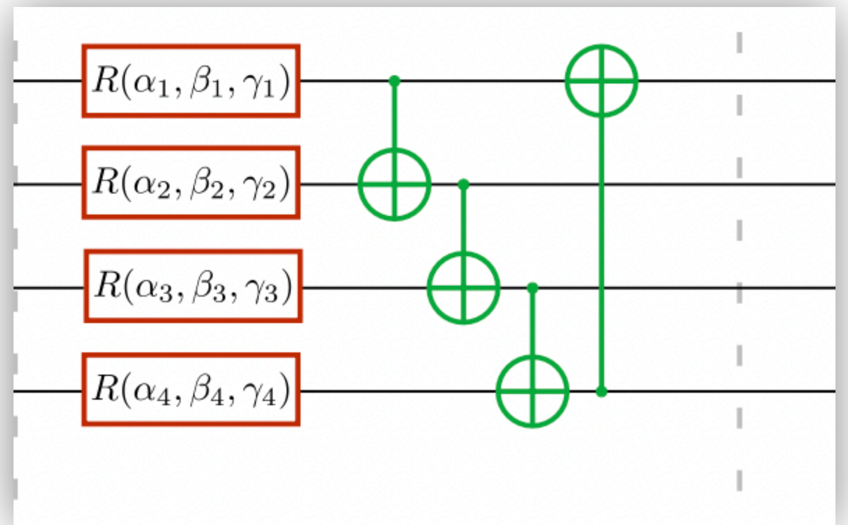
6- Preliminary results for jet images classification using QML with variational embedding

Background

How does quantum computer work as ML ?

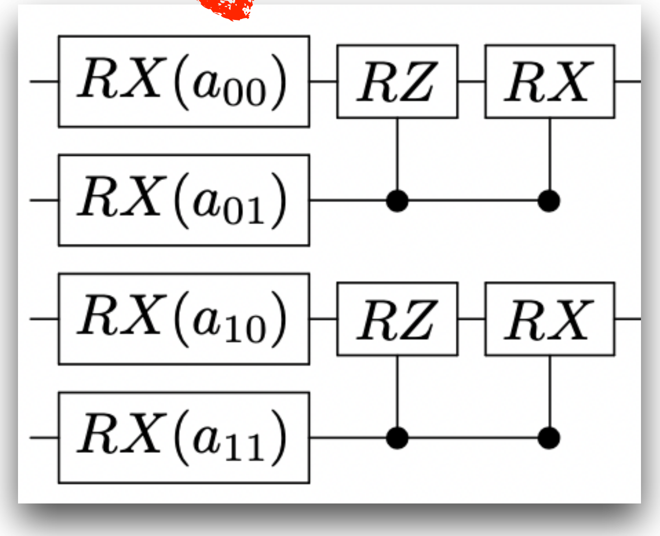






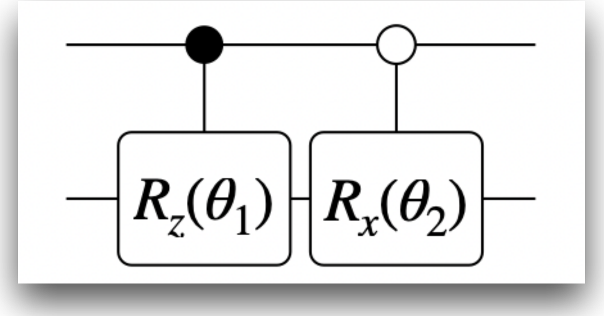
Fully connected

arXiv:2202.13943v1



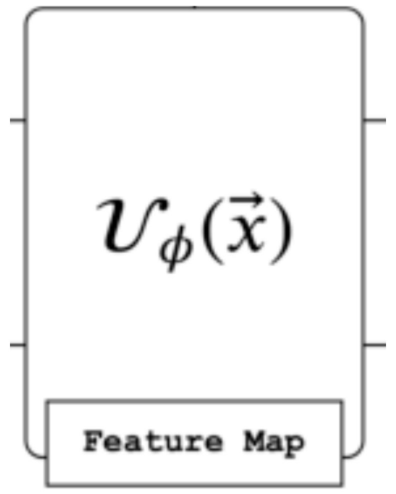
Conv circuit

arXiv:2108.00661



Pooling circuit

arXiv:2108.00661



Fixed

1- Basis embedding

2- Amplitude embedding

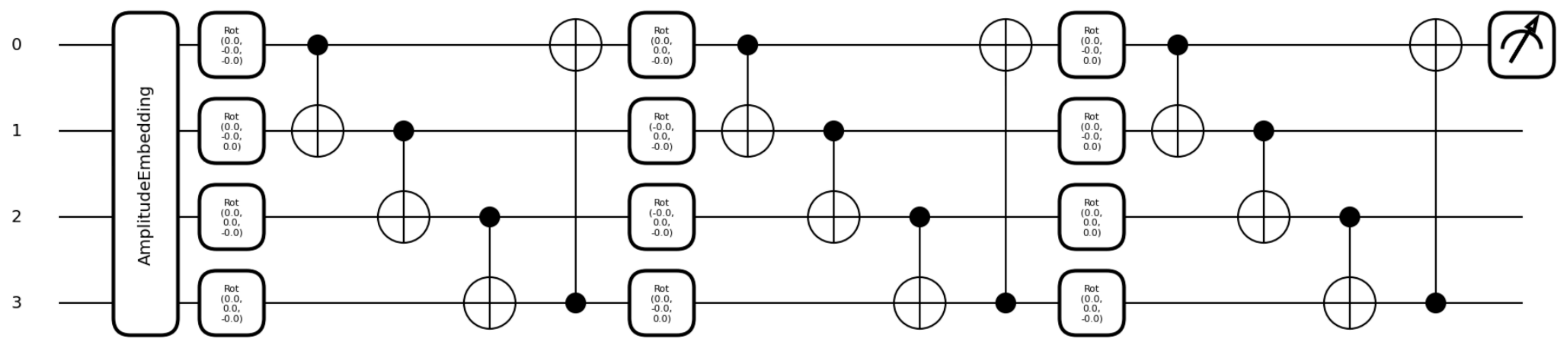
3- Angle embedding

4- Hamiltonian embedding

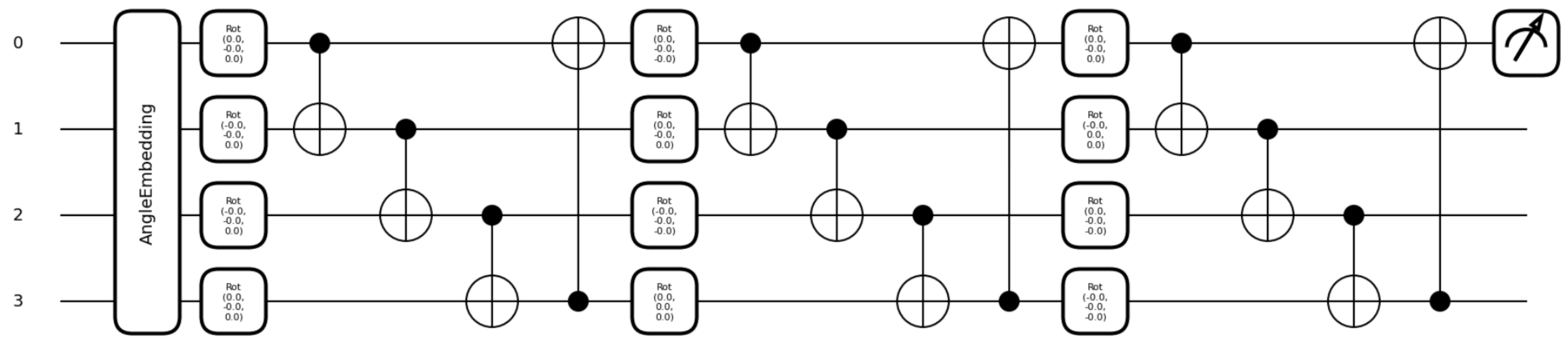
Variational embedding

Background:Fixed embedding

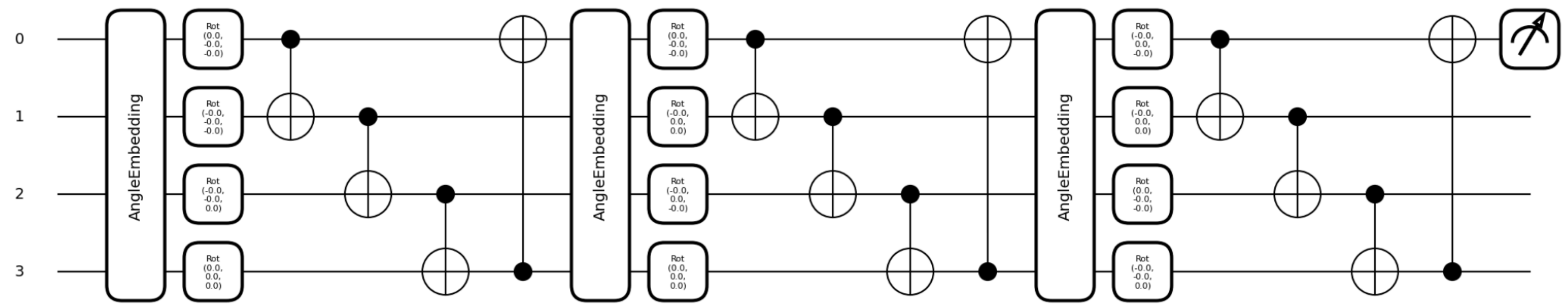
3 basic entangled layers + Amplitude embedding

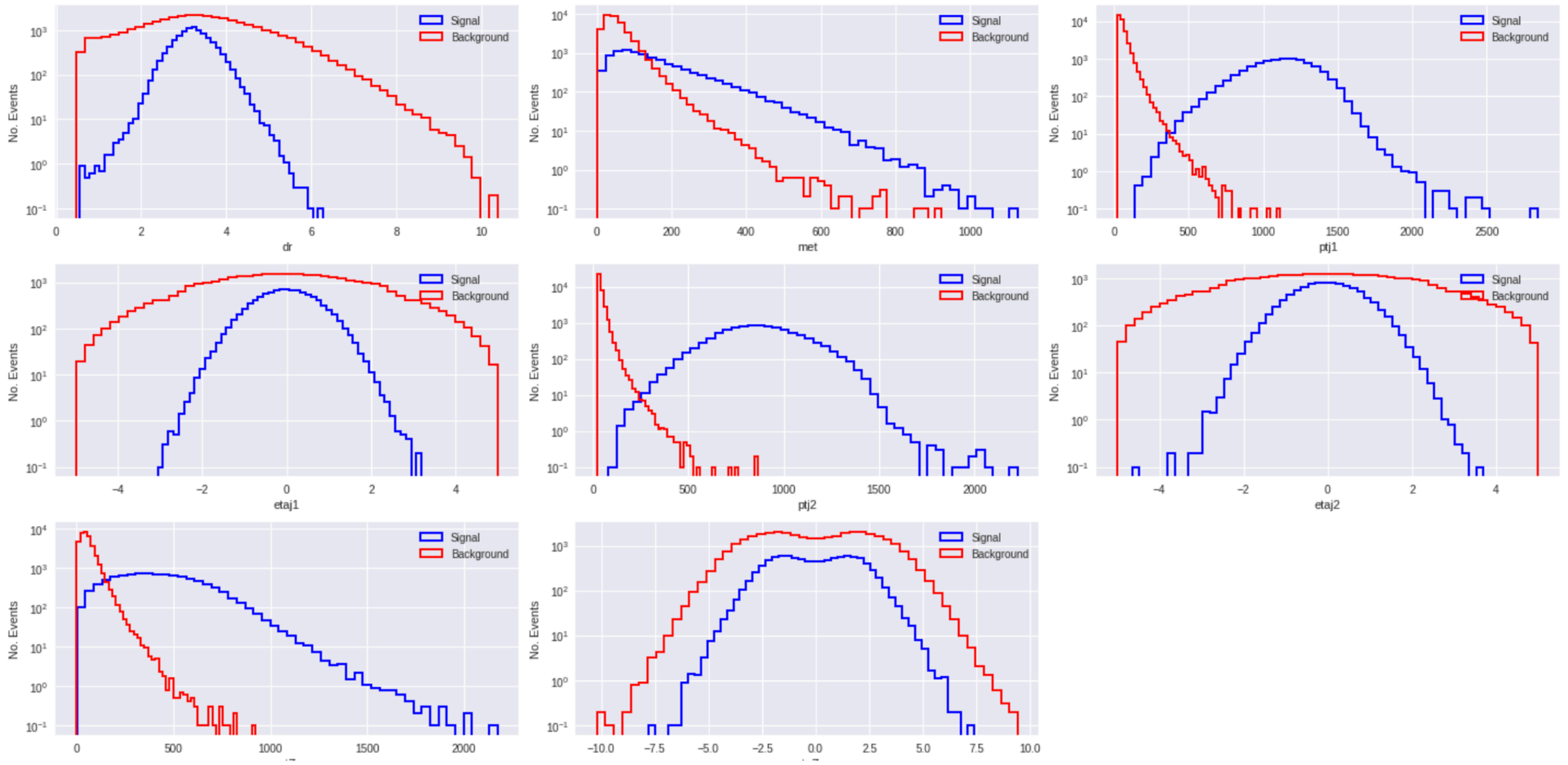


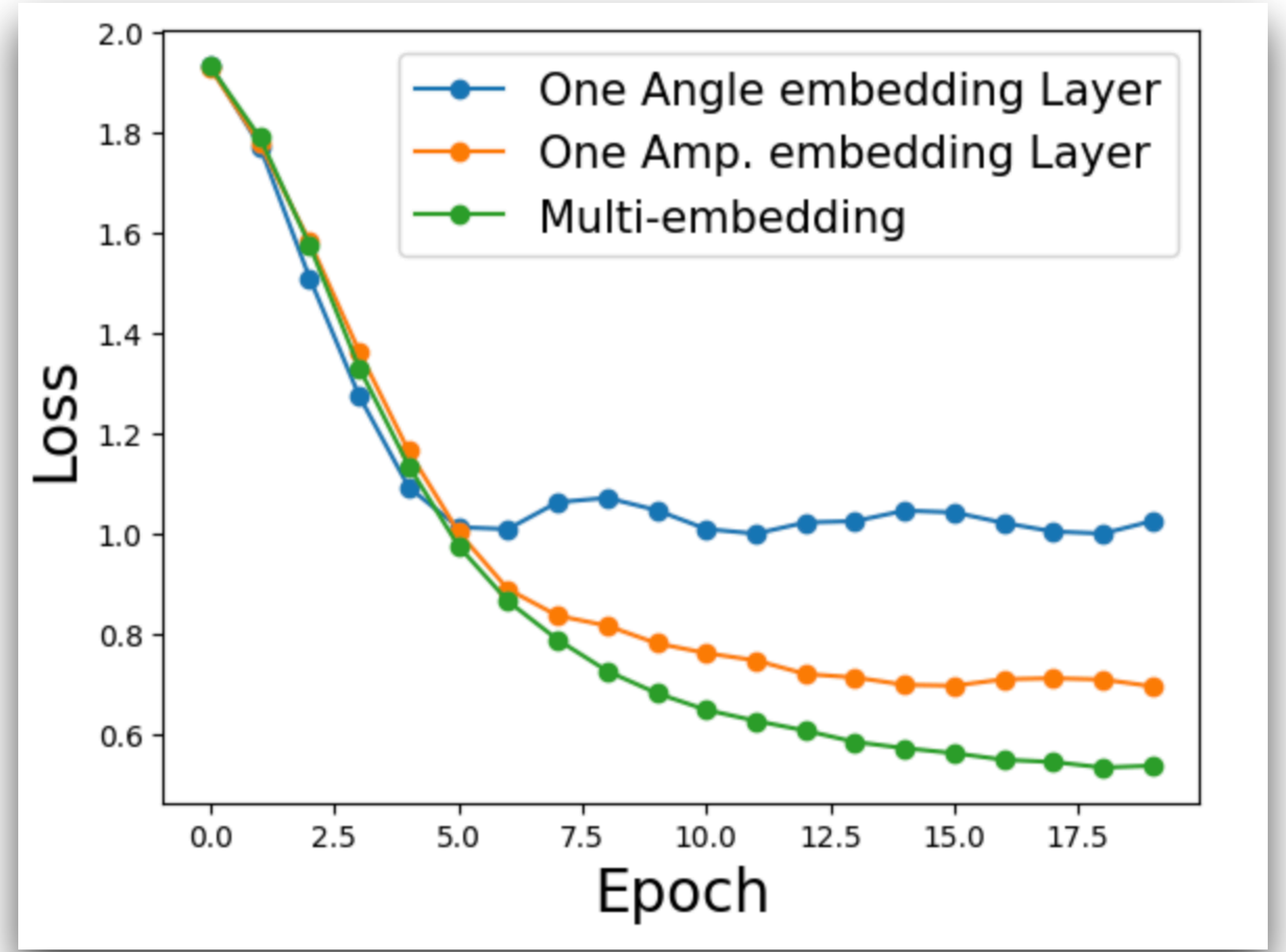
3 basic entangled layers + Angle embedding



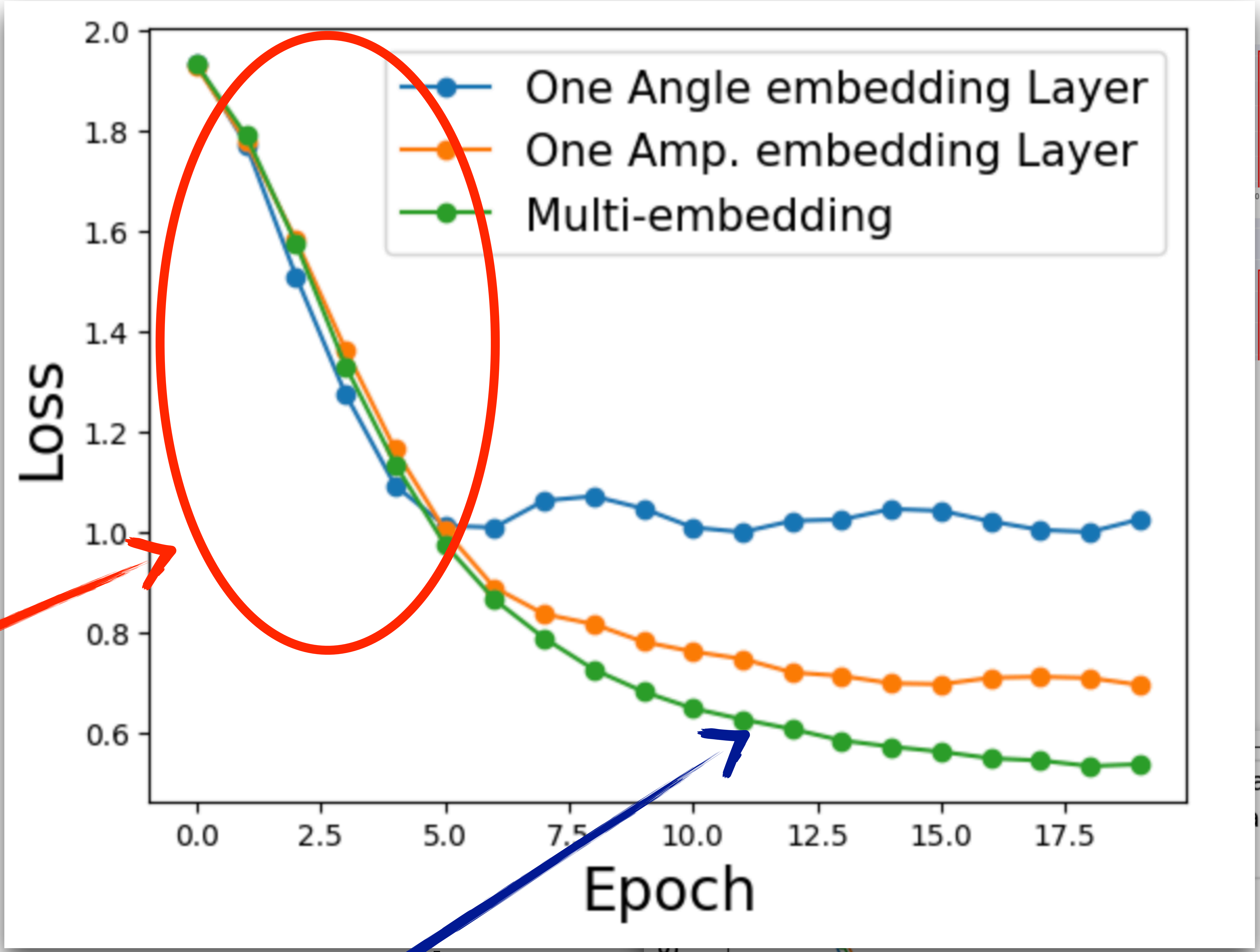
3 basic entangled layers + repeated embedding

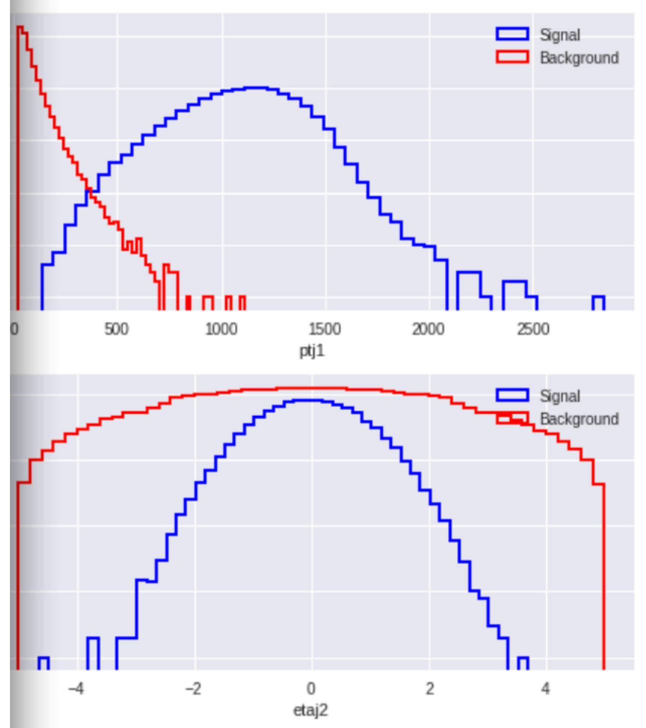


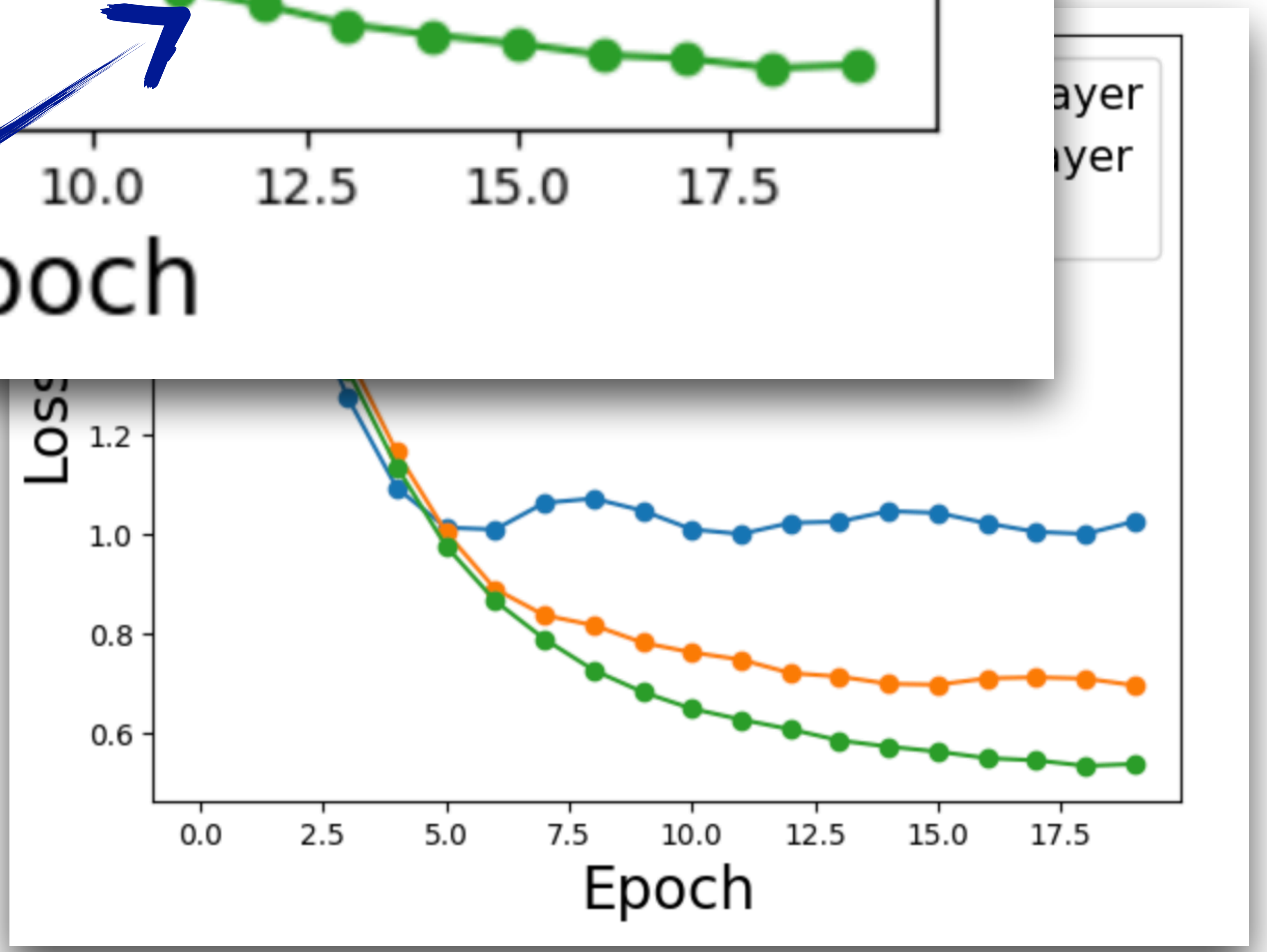




Fixed embedding







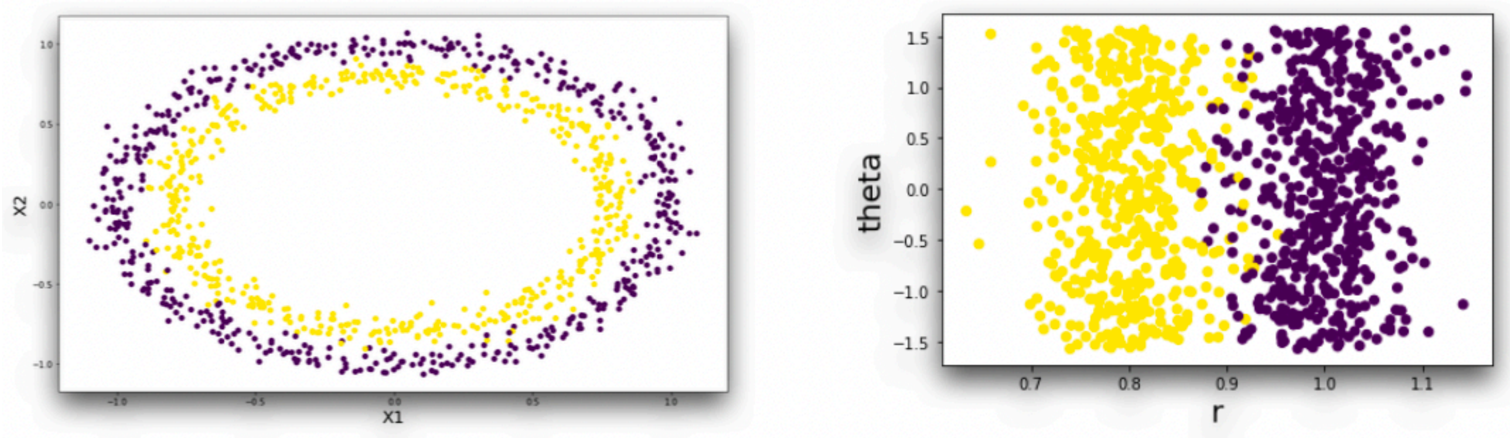
Large performance for
small training data/few steps
(Quantum kernel mapping)

Repeated embedding can
achieve higher performance

Kernel mapping

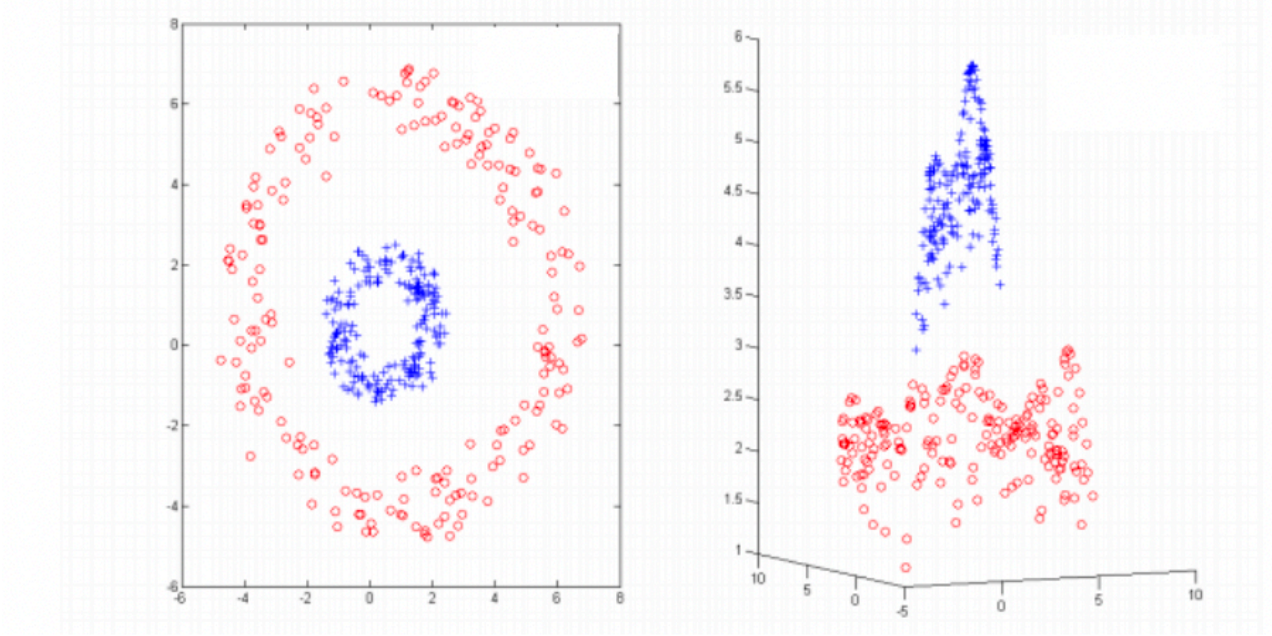
Mapping non-linear separable data from low dimensional space to other coordinates by using specific kernel, one can find a hyper-plane that can easily separate between the data

Classical mapping



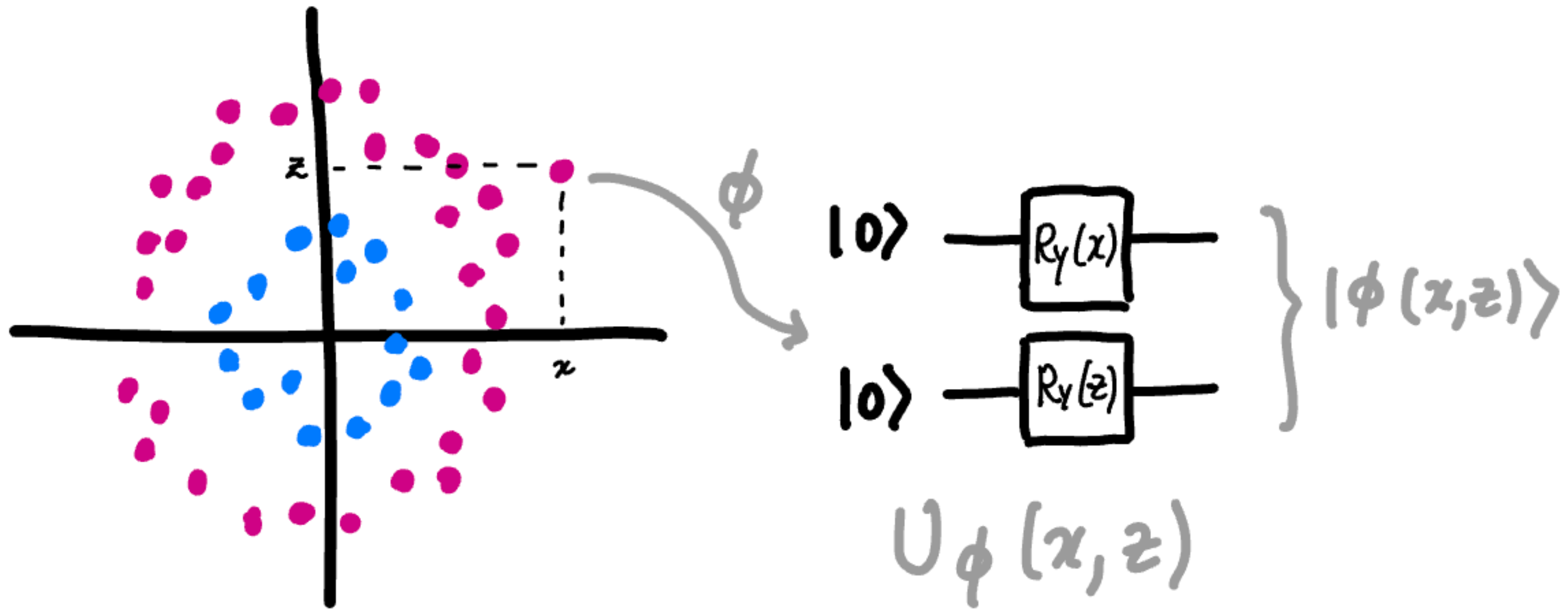
Cartesian coordinates -> polar coordinates

$$\theta = \tan^{-1}(\frac{x_1}{x_2}) \quad r = \sqrt{(x_1^2 + x_2^2)}$$



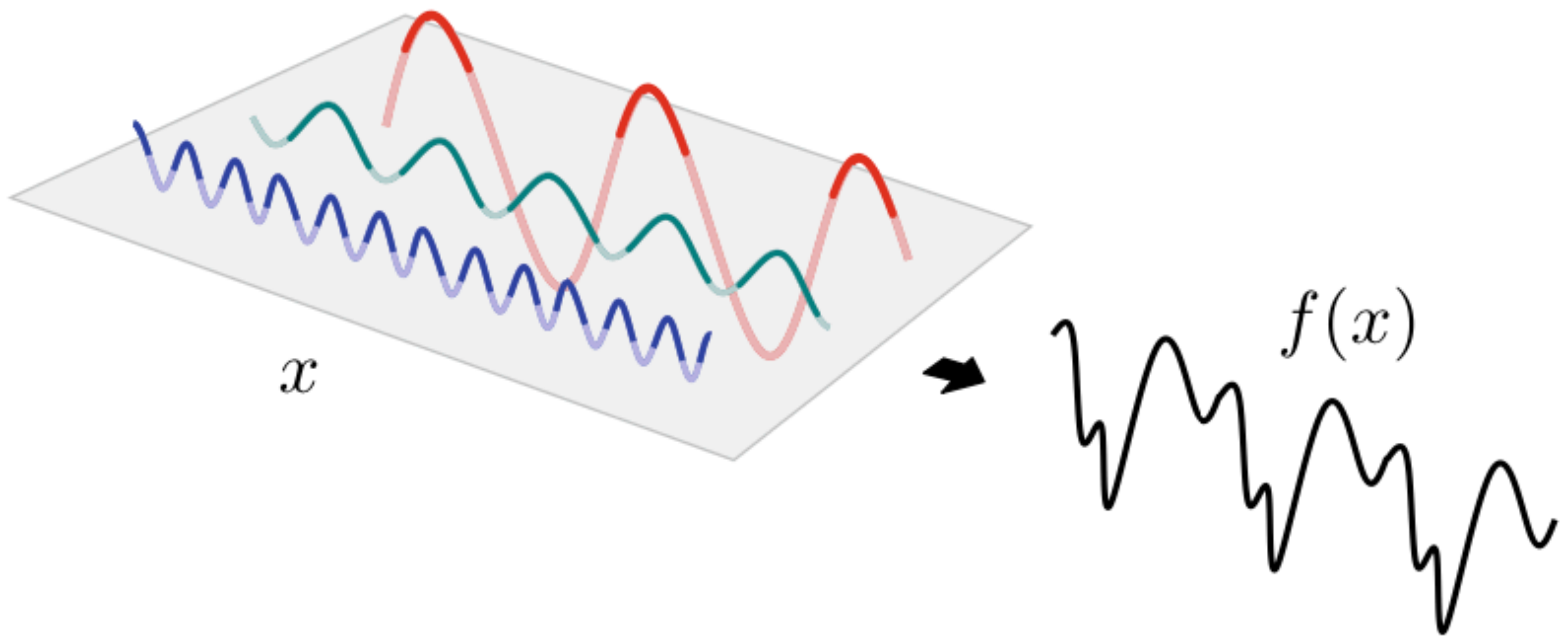
2d -> 3d using Gaussian RBF kernel

Quantum mapping



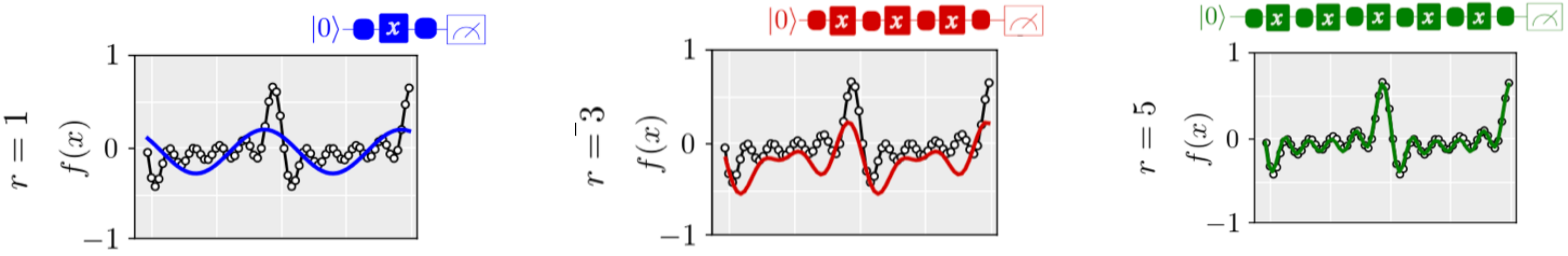
https://pennylane.ai/qml/glossary/quantum_feature_map.html

Repeated embeddings of the classical data



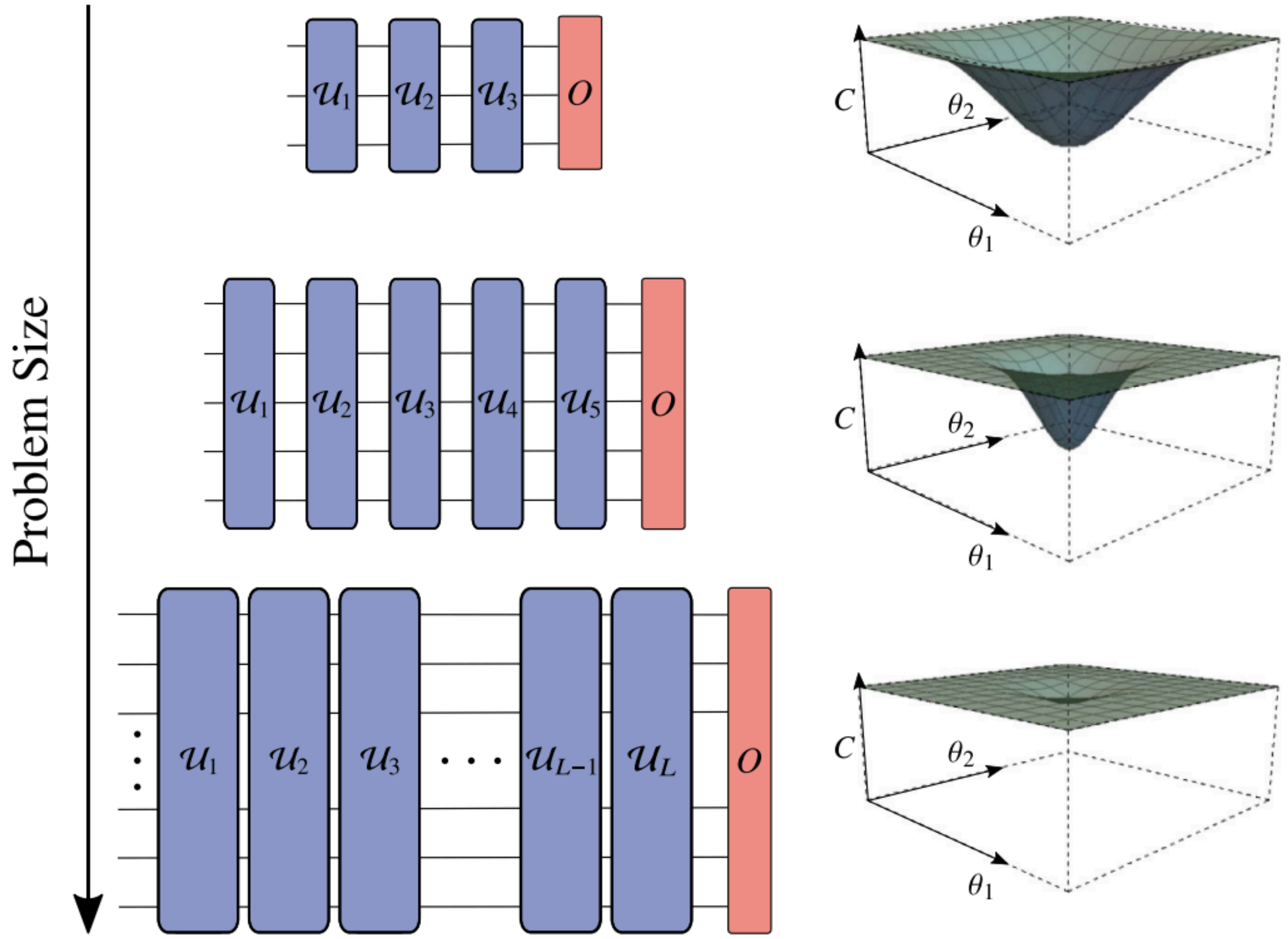
Quantum model with data re-uploading is a sum of trigonometric functions.

arXiv:2008.08605v2 [quant-ph]



Limitation of the repeated embedding

https://doi.org/10.1038/s41467-021-27045-6



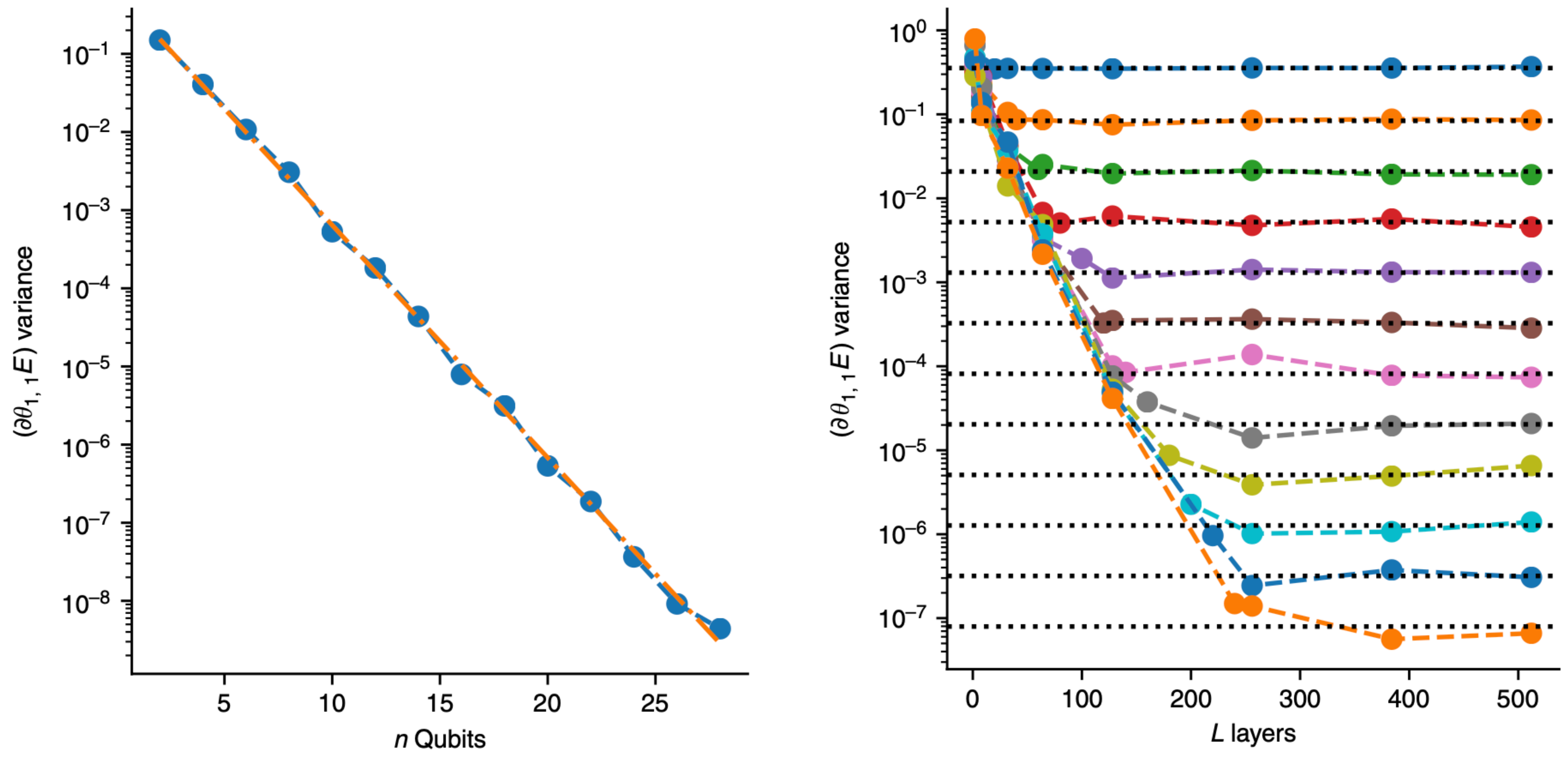
Barren plateaus affects the
variational quantum optimization algorithms, QAOA

ARTICLE
DOI: 10.1038/s41467-018-07090-4 OPEN

Barren plateaus in quantum neural network training landscapes

Jarrod R. McClean[1], Sergio Boixo[1], Vadim N. Smelyanskiy[1], Ryan Babbush[1] & Hartmut Neven[1]

Many experimental proposals for noisy intermediate scale quantum devices involve training a parameterized quantum circuit with a classical optimization loop. Such hybrid quantum-classical algorithms are popular for applications in quantum simulation, optimization, and machine learning. Due to its simplicity and hardware efficiency, random circuits are often proposed as initial guesses for exploring the space of quantum states. We show that the



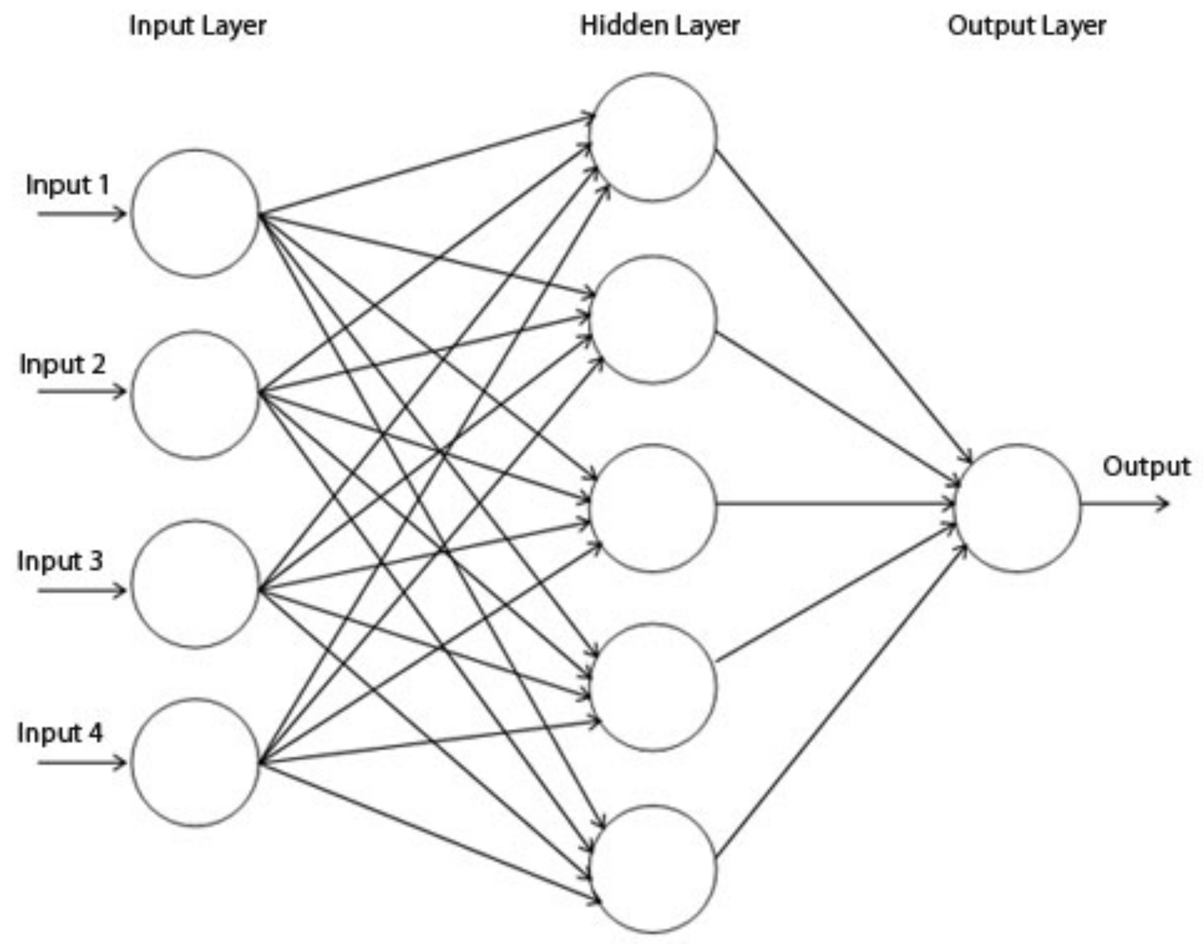
Evenly increasing qubits

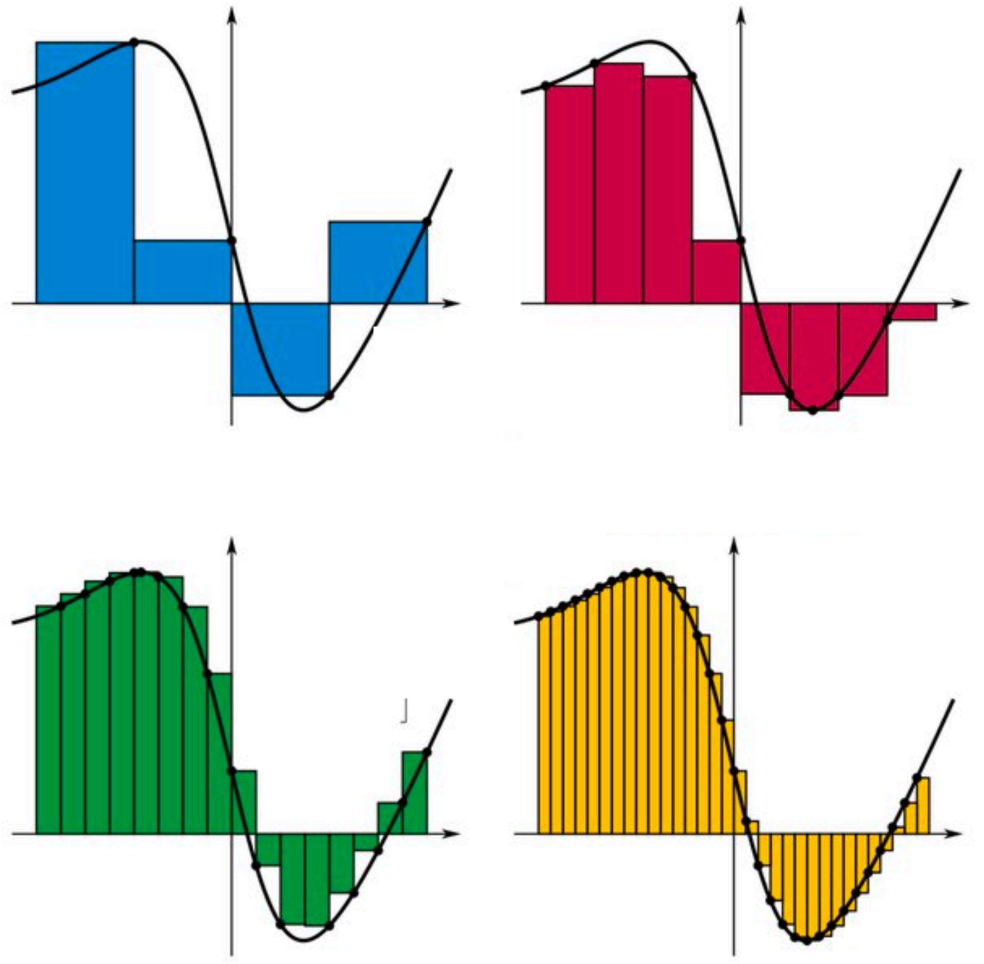
Question:

Can we construct a variational quantum classifier with only one qubit ?

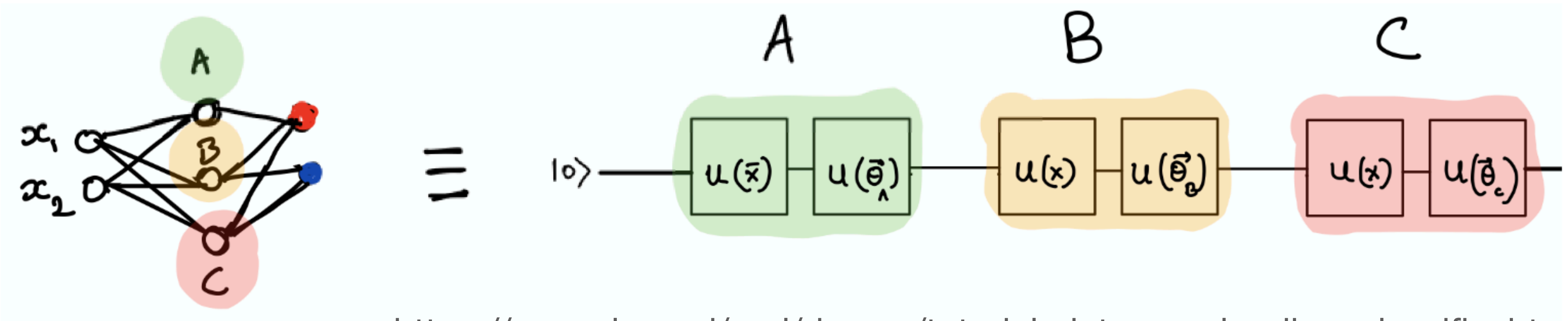
Universal approximation theorem

Given any continuous function, no matter how complicated it is, there is always exist a network that can approximate this function

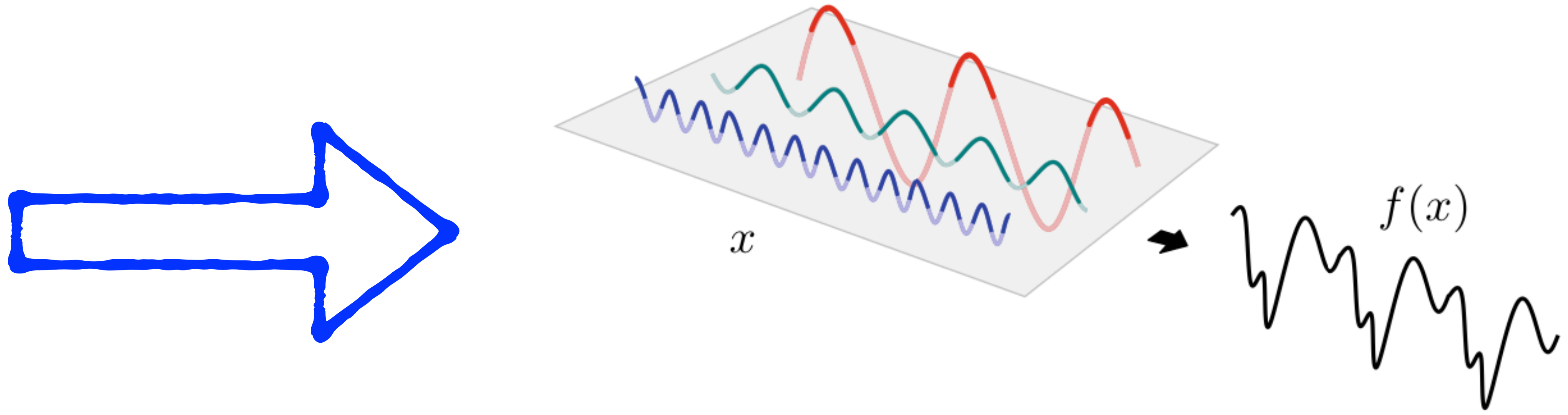




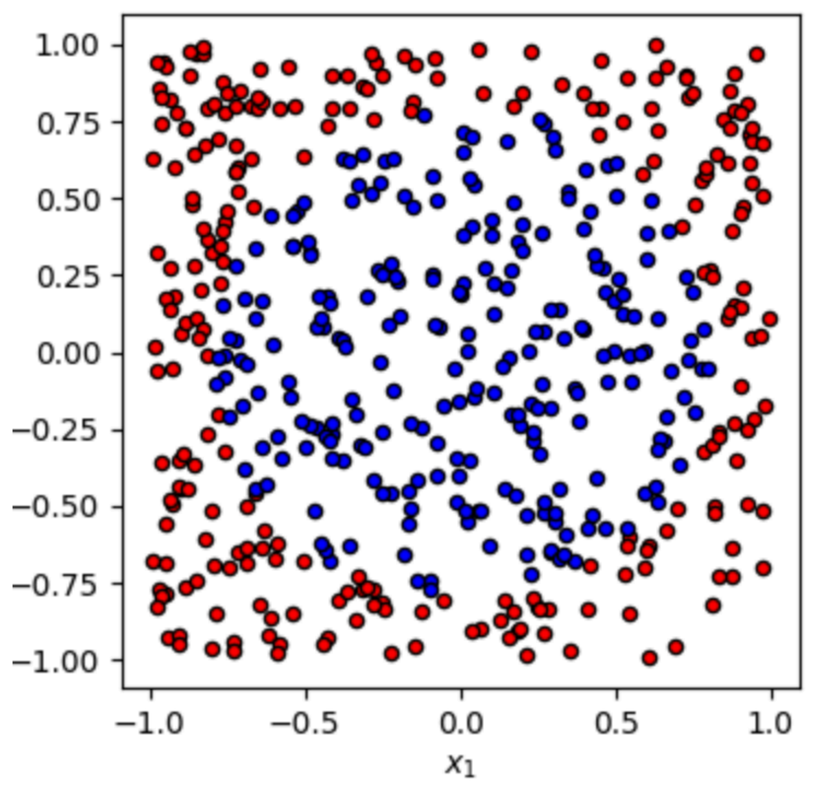
Single layer network can fit any function no matter how complicated it is



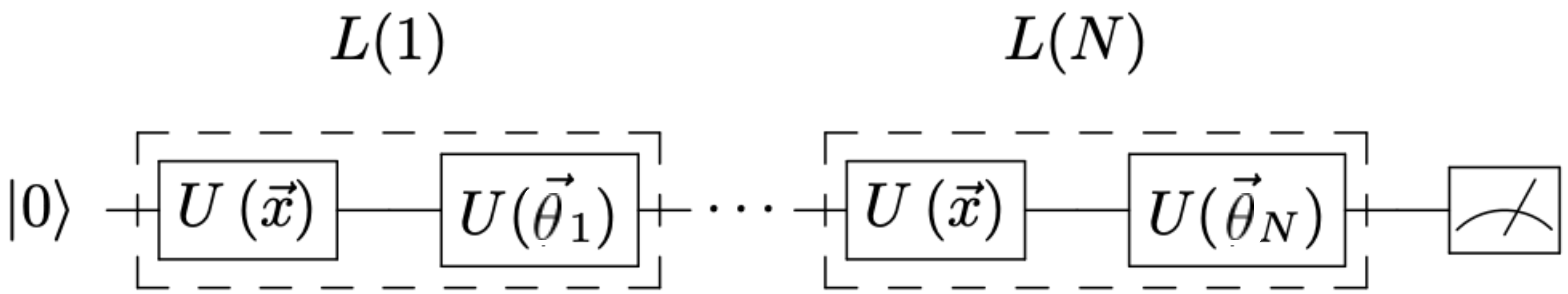
https://pennylane.ai/qml/demos/tutorial_data_reuploading_classifier.html



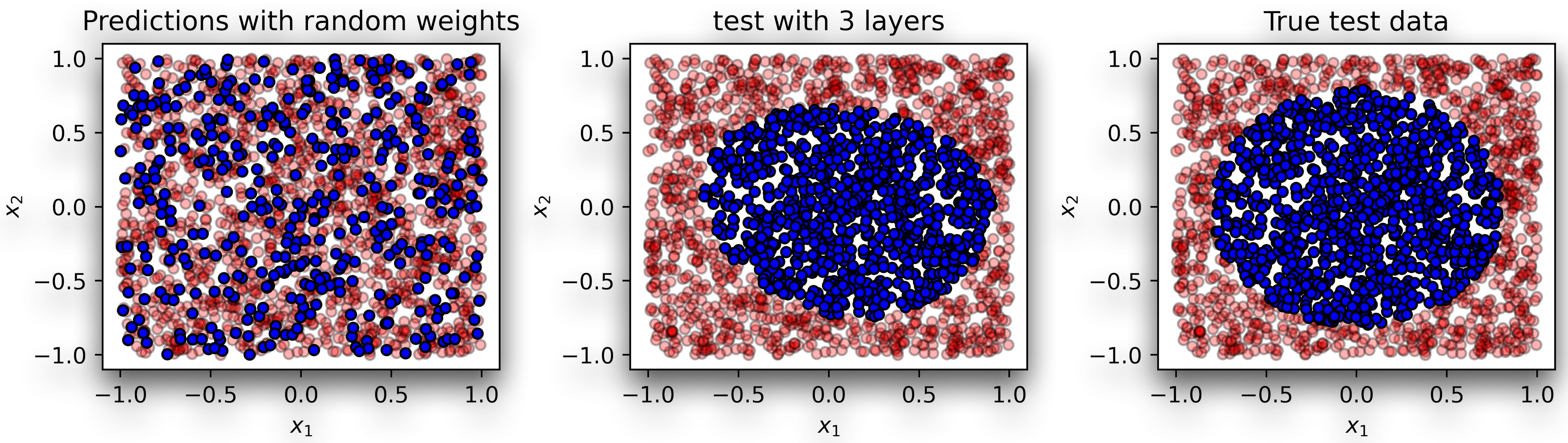
One qubit classifier: Toy example



Task: classify non-linear two dimensional data using one qubit with data re-uploading



$$\boxed{\text{Cost} = \sum_{\text{data points}} \left(1 - \text{fidelity}(\psi_{\text{output}}(\vec{x}, \vec{\theta}), \psi_{\text{label}})\right)}$$



Training 200 points with 10 epochs

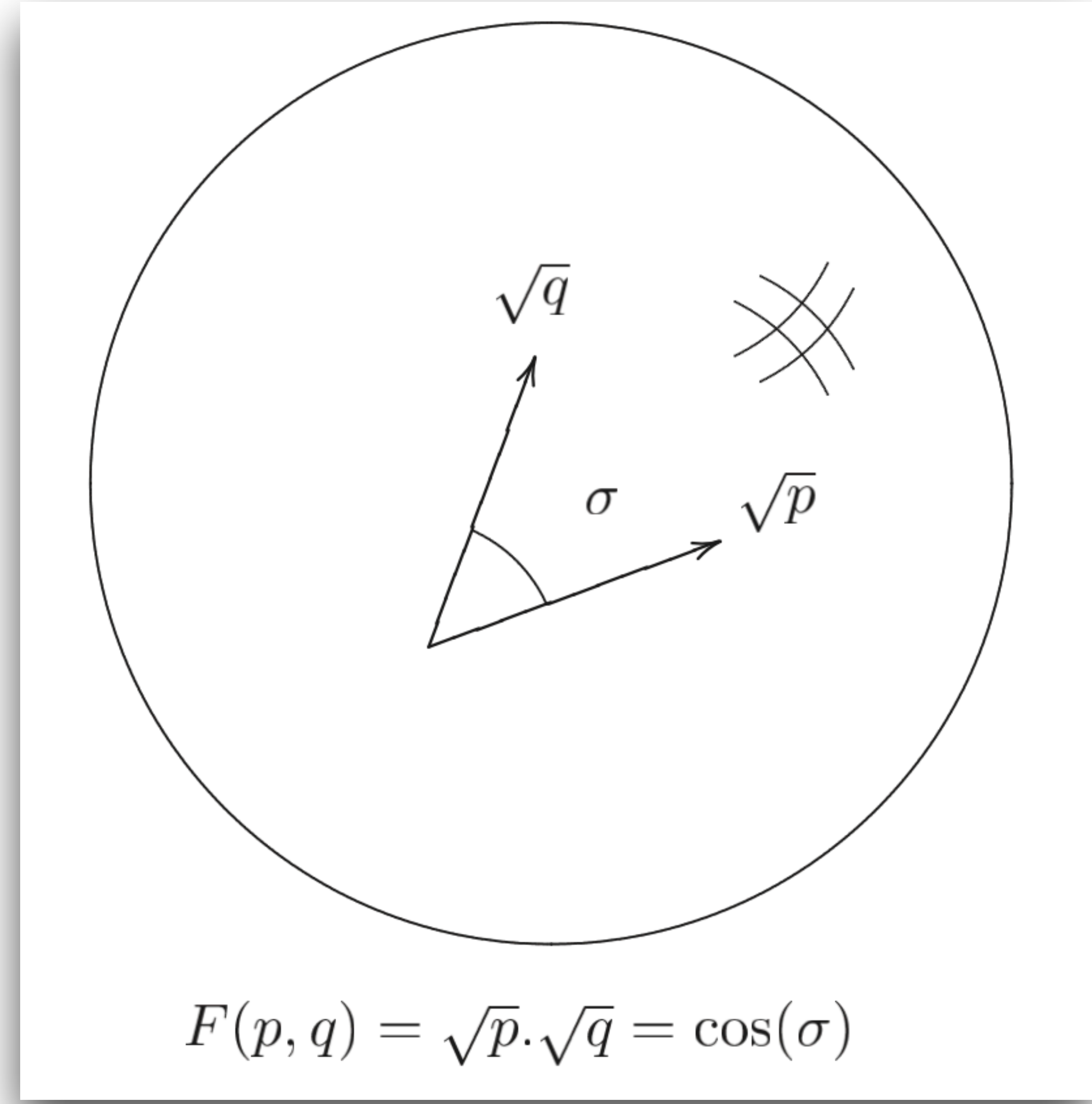
One qubit classifier: Toy example

$$\text{Cost} = \sum_{\text{data points}} (1 - \texttt{fidelity}(\psi_{\text{output}}(\vec{x},\vec{\theta}), \psi_{\text{label}}))$$

What does this cost function mean ?

Fidelity

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi, \phi\rangle|^{2}$$



Fidelity is the distance (similarity measure) between two quantum states

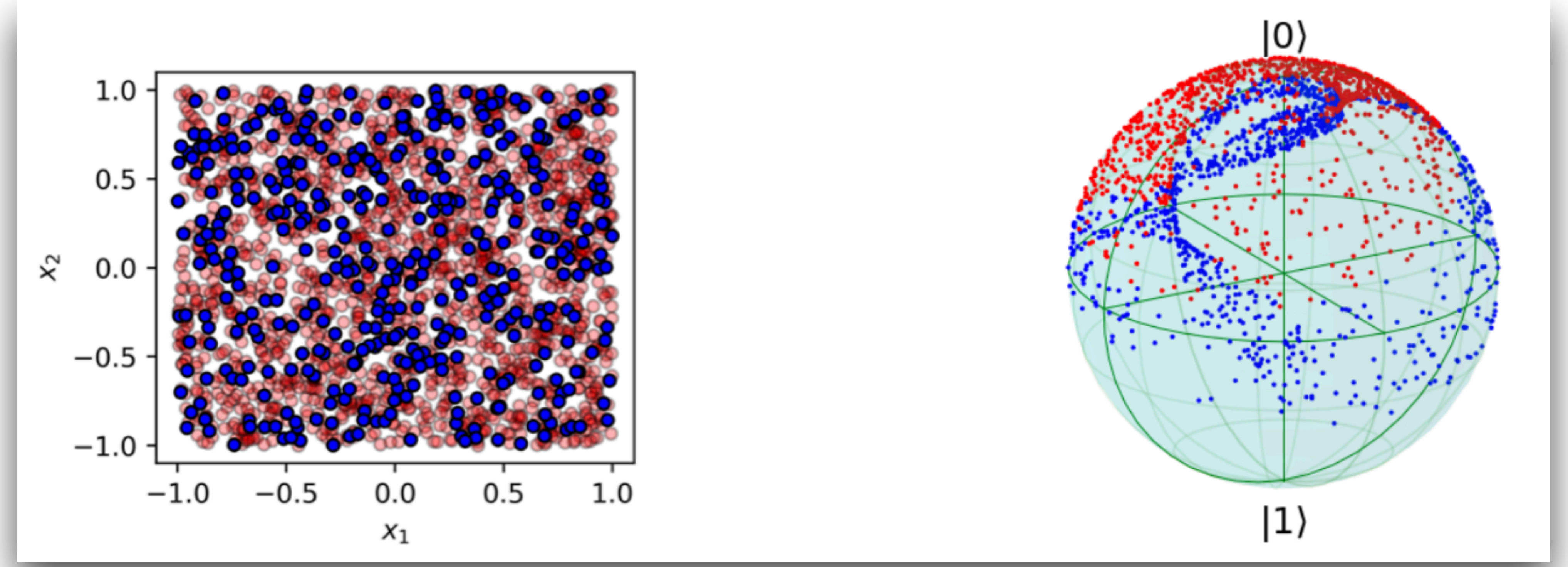
Minimizing the cost function means we increase the purity of quantum states

Quantum information and quantum computation Book

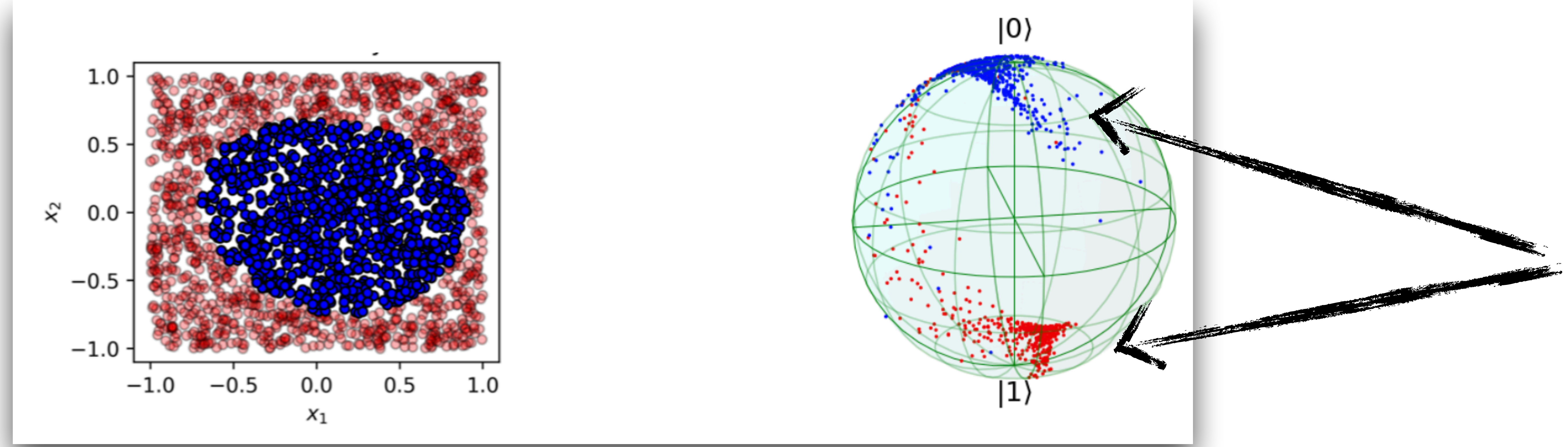
One qubit classifier: Toy example

Bloch sphere representation of the embedded data on the single qubit

Before training



After training

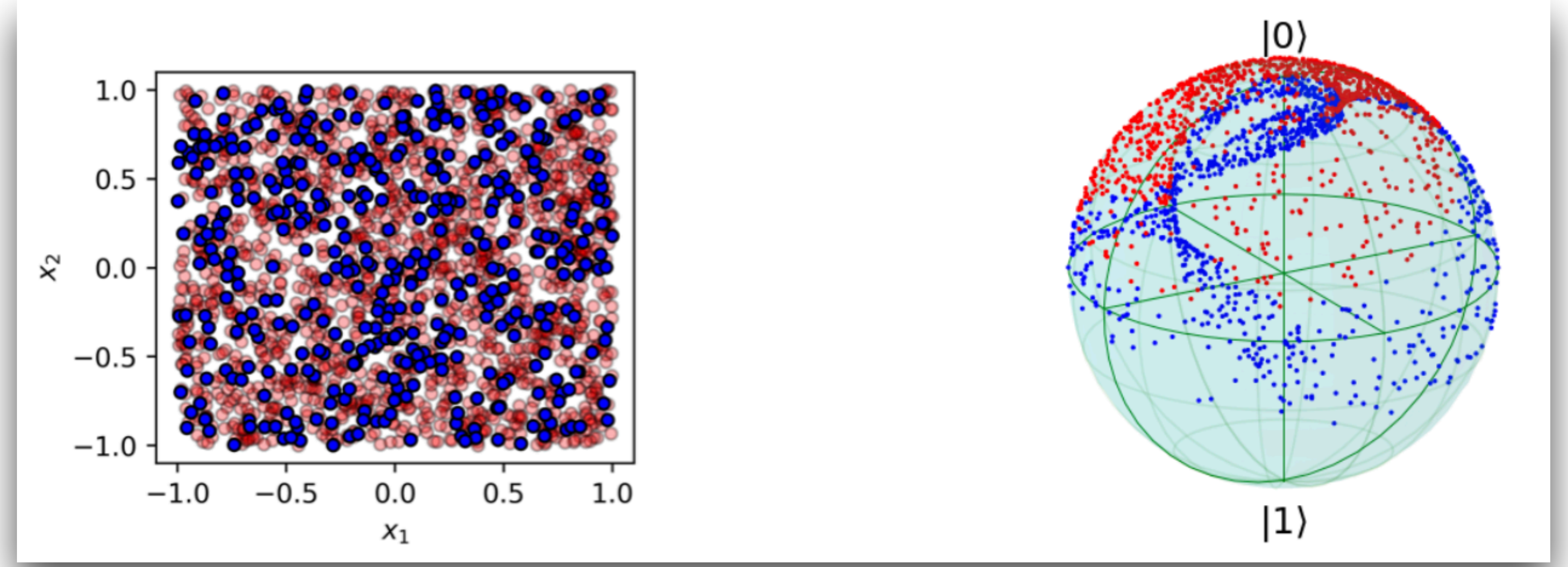


Maximize the quantum fidelity = maximize the distance between the measured quantum states on the qubit

One qubit classifier: Toy example

Bloch sphere representation of the embedded data on the single qubit

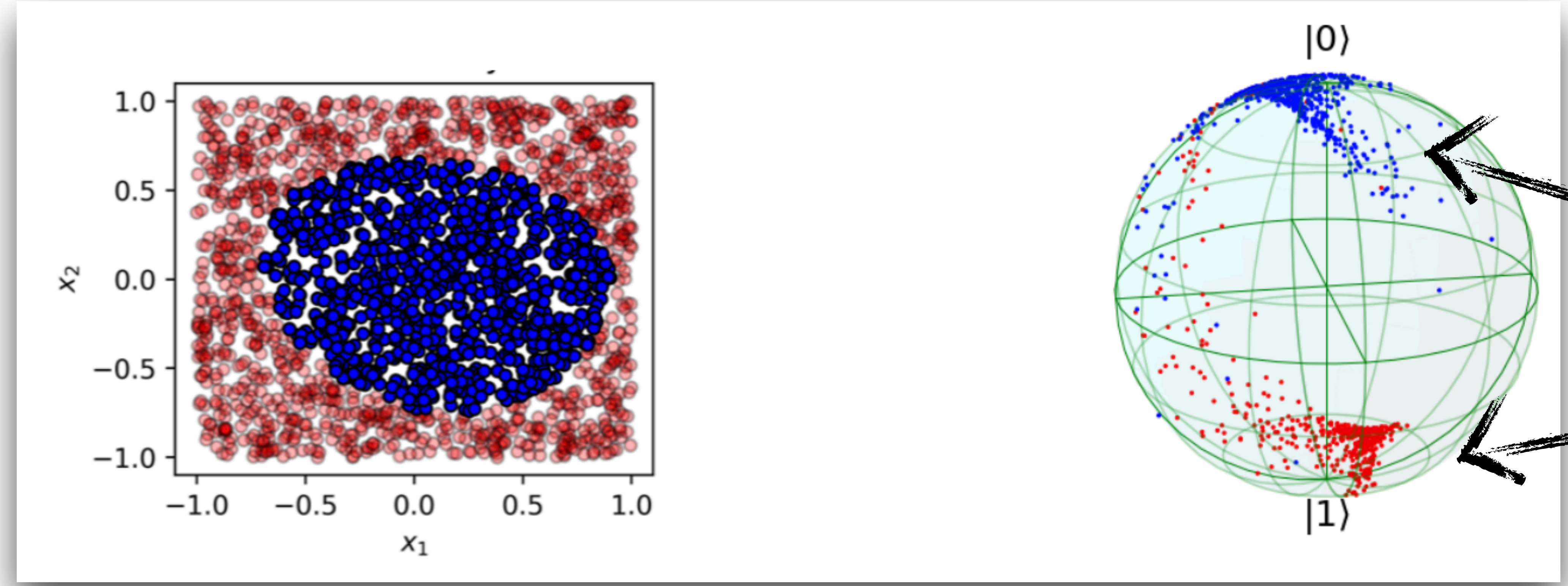
Before training



$$\text{Cost} = \sum_{\text{data points}} (1 - \mathtt{fidelity}(\psi_{\mathtt{output}}(\vec{x},\vec{\theta}),\psi_{\mathtt{label}}))$$

Is indeed a metric cost function

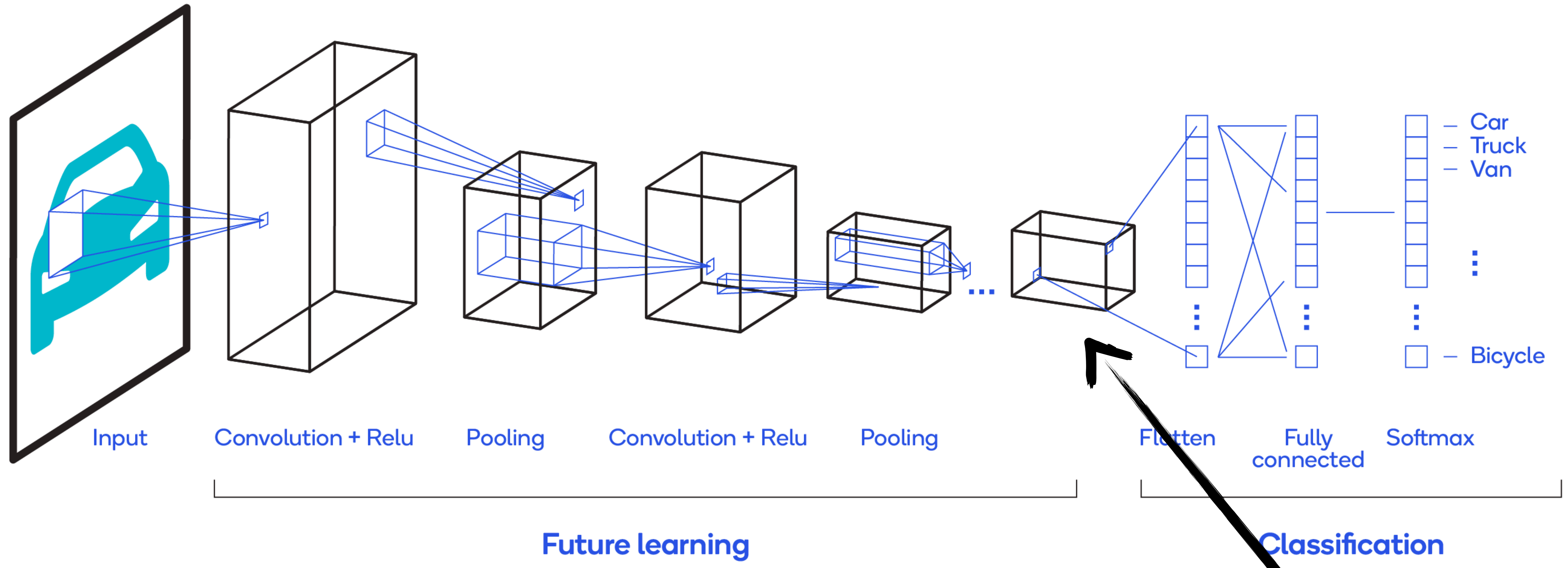
After training



Maximize the quantum fidelity = maximize the distance between the measured quantum states on the qubit

Metric Learning

Well, QVC of one qubit and repeated embedding
can work as metric learning but why is the metric learning ?



Information in the bottleneck is controlled by the structure of the network and thus we need an extra metric to control it.

Any convolution based model works as the following:

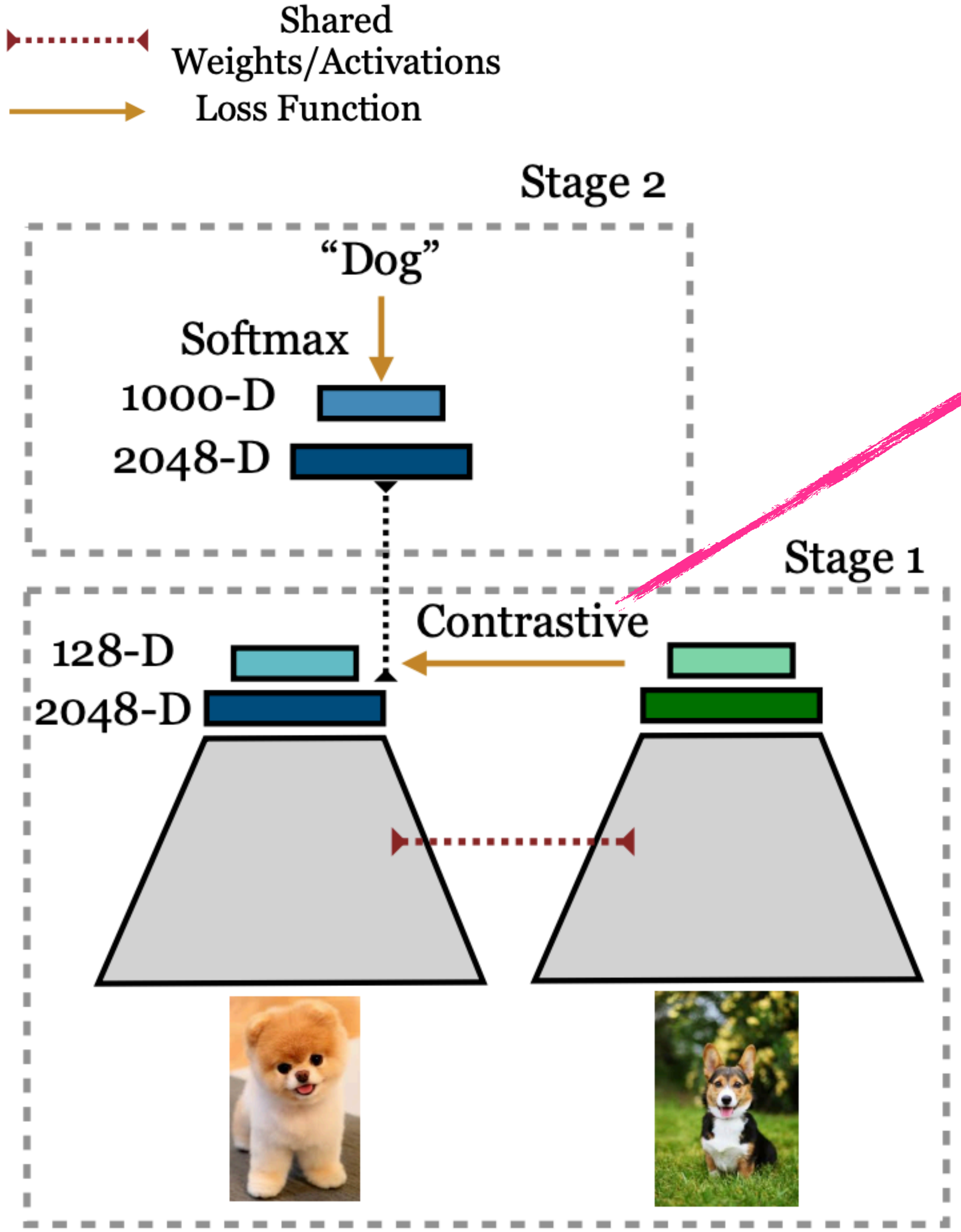
1- Feature extraction in high dimensions feature space

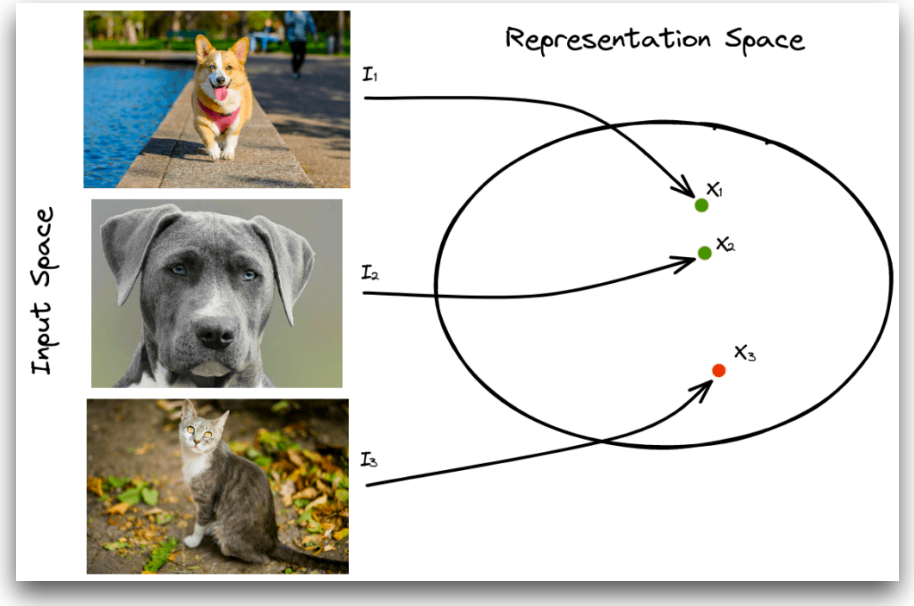
2- Decompose the extracted features into lower dimensions space (bottleneck)

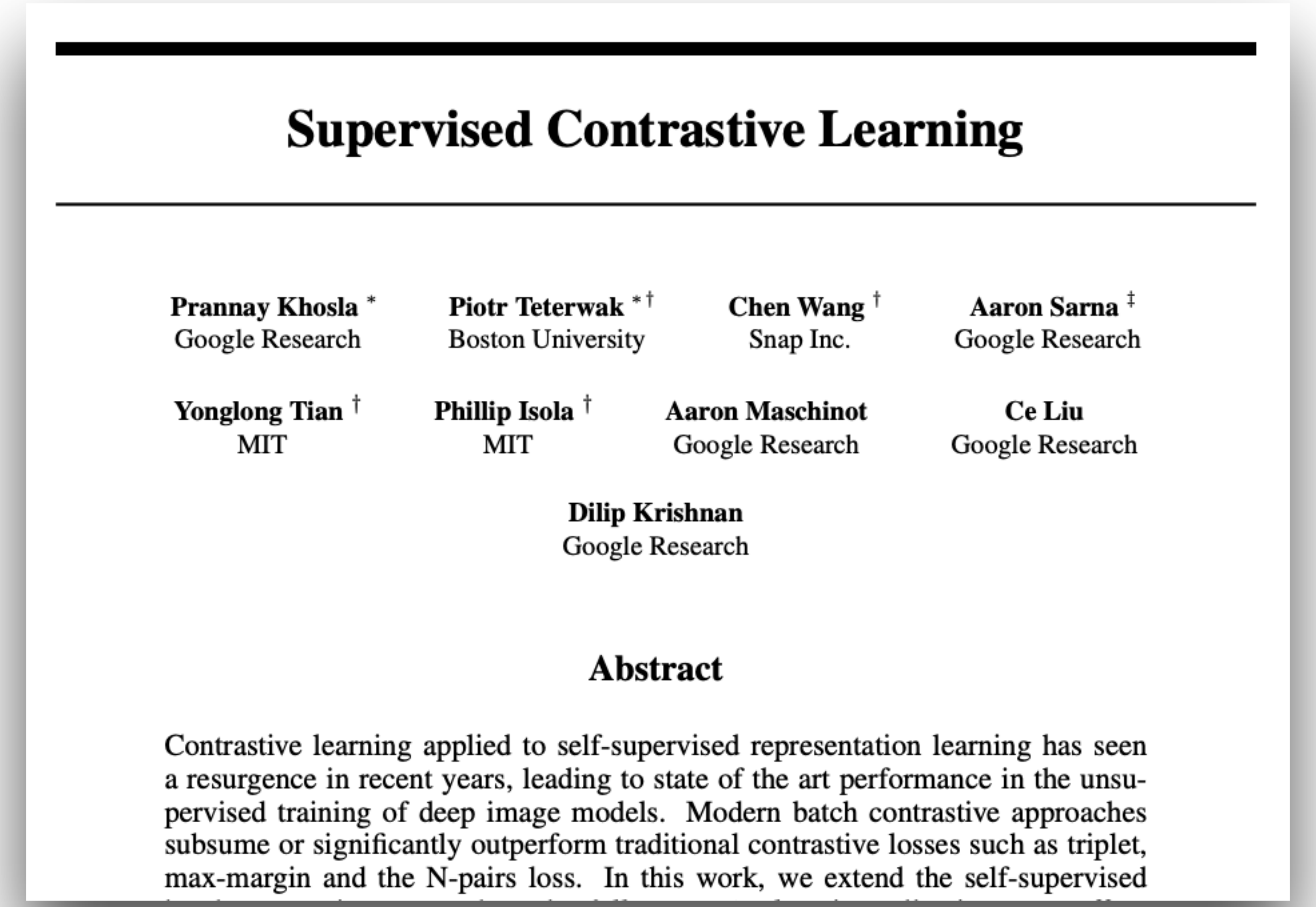
3- Use FC network to analyze the information in the bottleneck

Contrastive Learning with augmentation

Twins encoder with shared weights, two steps training with data augmentation





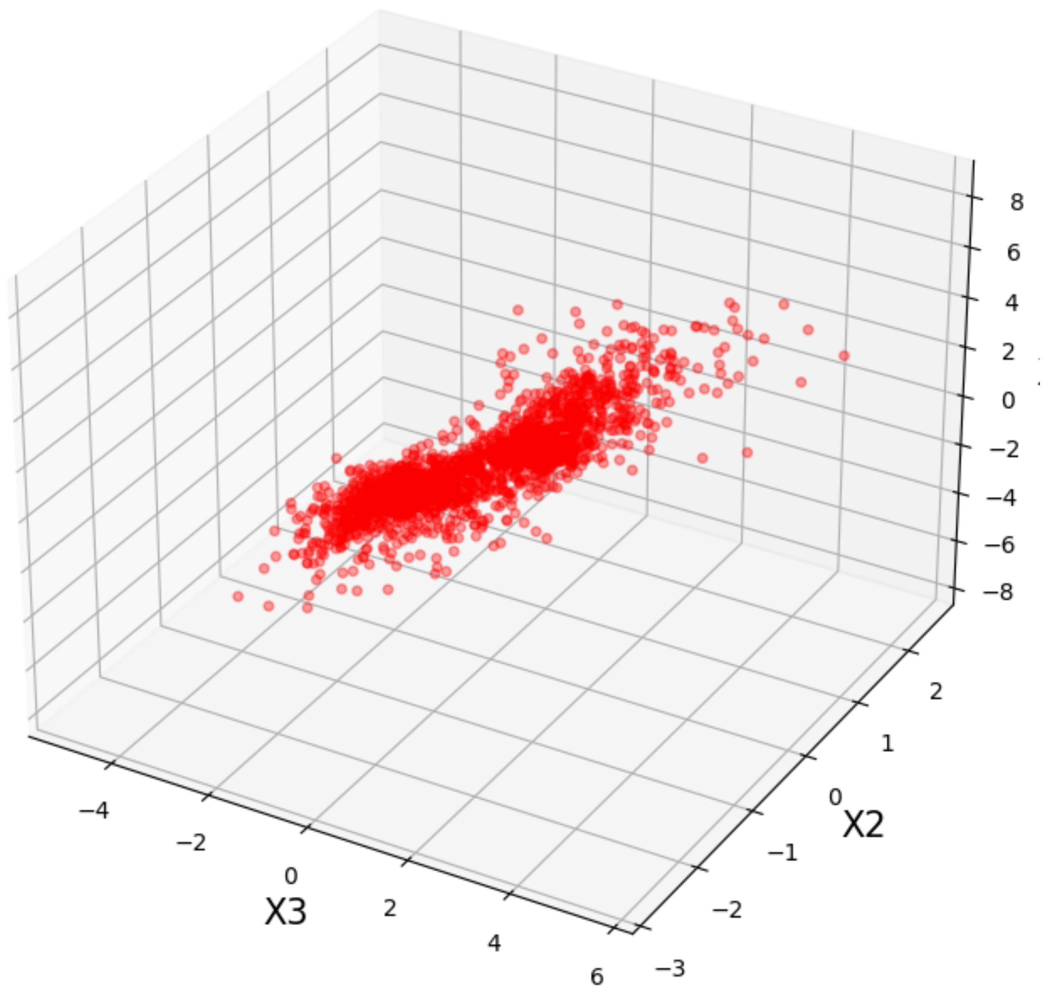

Supervised Contrastive Learning

Prannay Khosla * Google Research
Piotr Teterwak *† Boston University
Chen Wang † Snap Inc.
Aaron Sarna ‡ Google Research
Yonglong Tian † MIT
Phillip Isola † MIT
Aaron Maschinot Google Research
Ce Liu Google Research
Dilip Krishnan Google Research

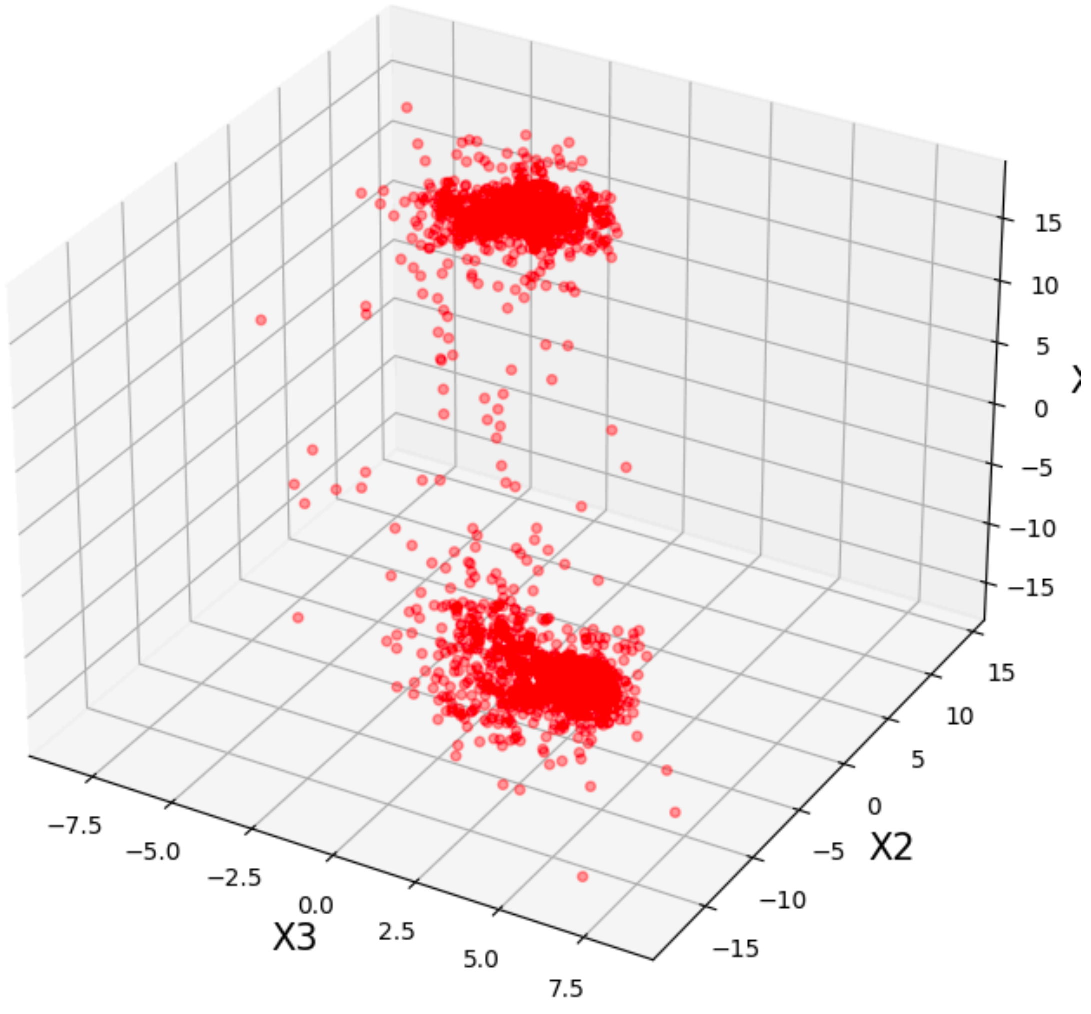
Abstract

Contrastive learning applied to self-supervised representation learning has seen a resurgence in recent years, leading to state of the art performance in the unsupervised training of deep image models. Modern batch contrastive approaches subsume or significantly outperform traditional contrastive losses such as triplet, max-margin and the N-pairs loss. In this work, we extend the self-supervised

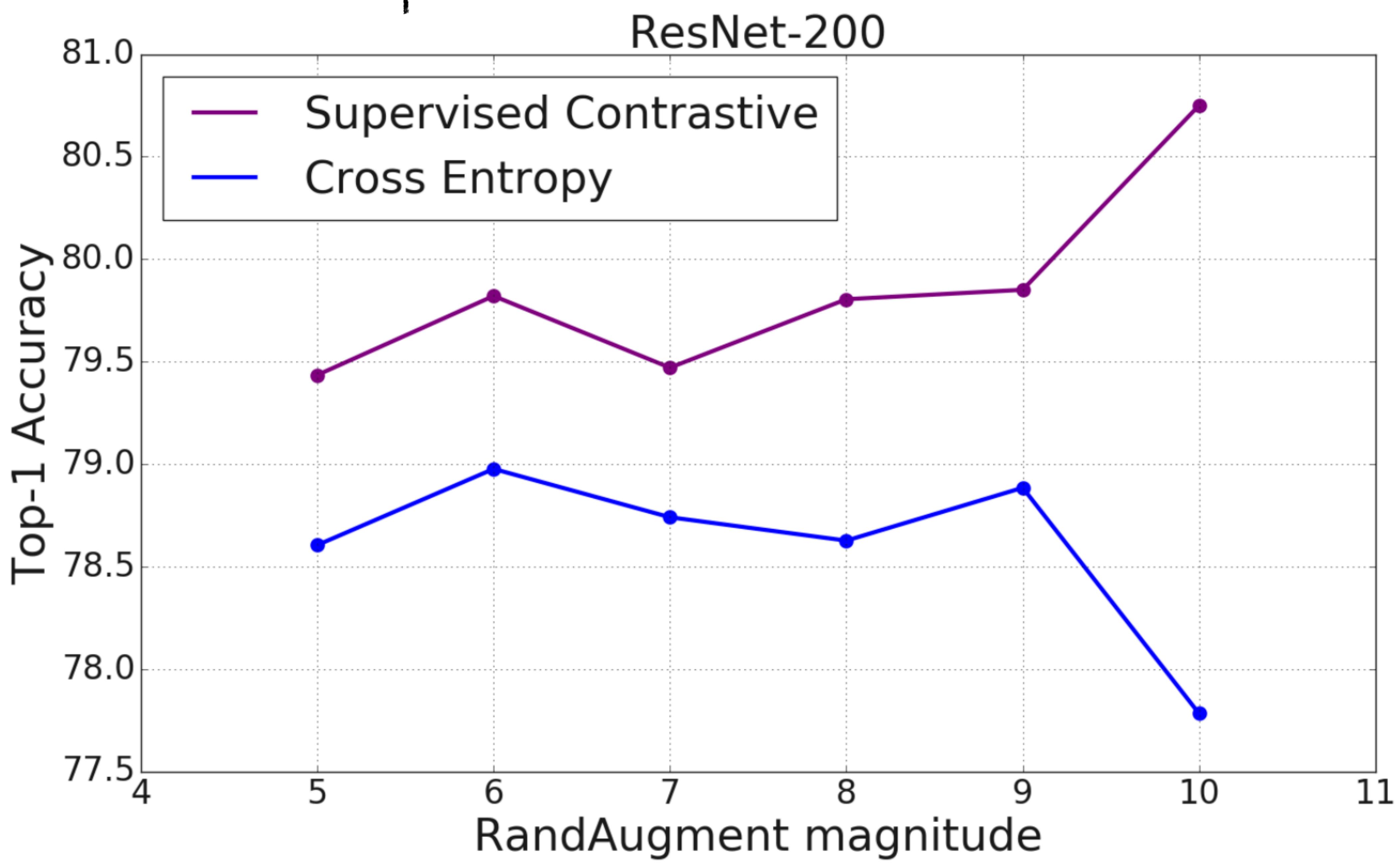
Before training



After training

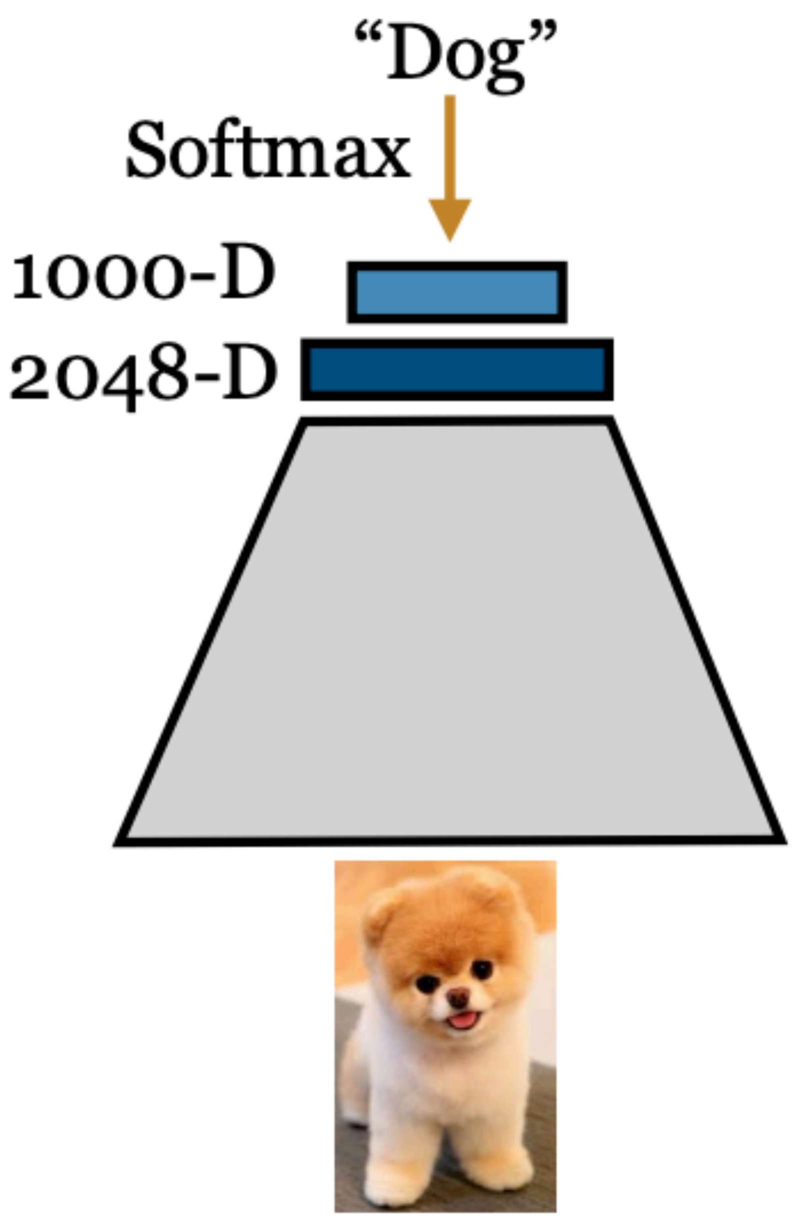


Contrastive Learning with augmentation

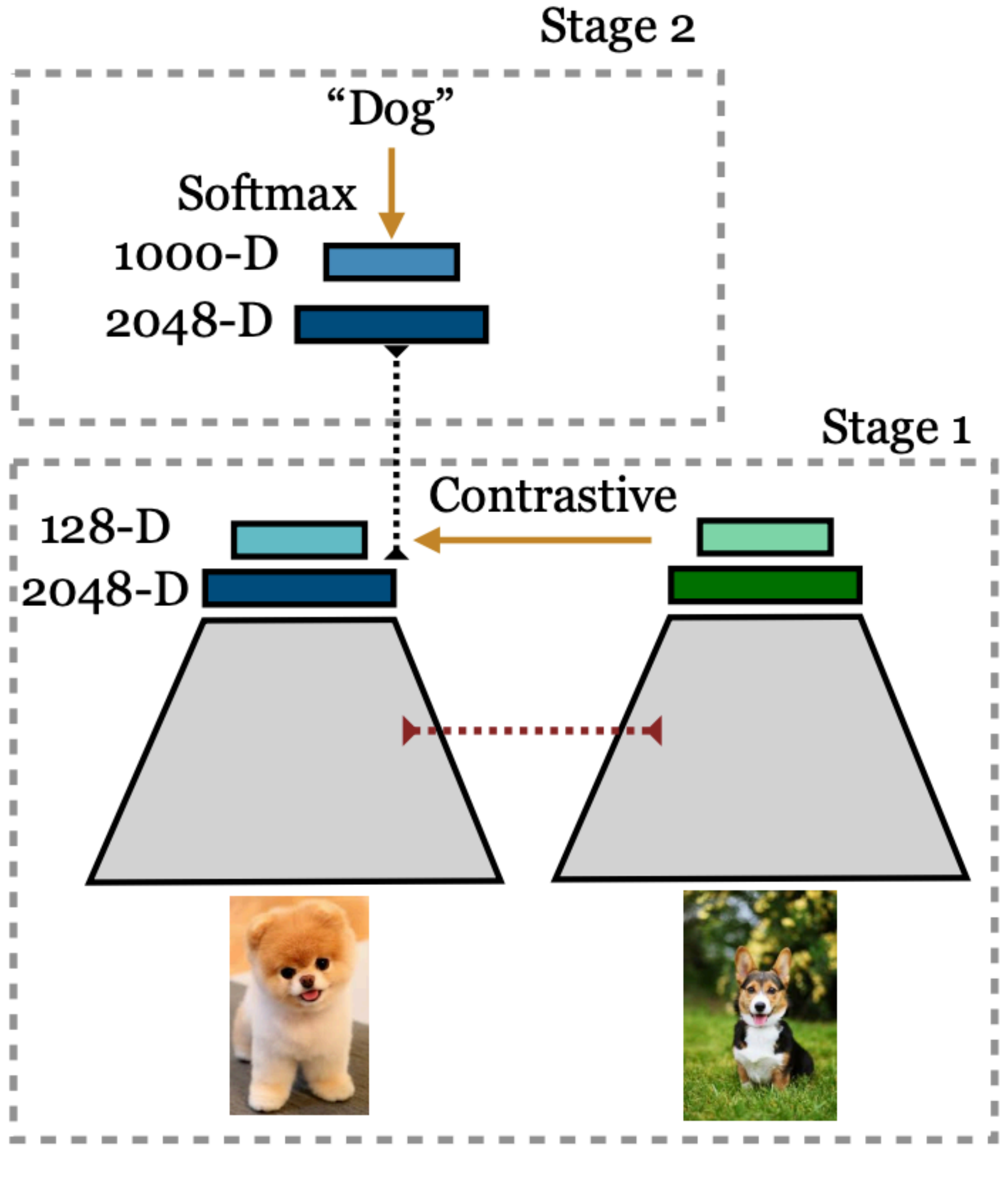


Classification performance of the classical contrastive models depends on the augmented data

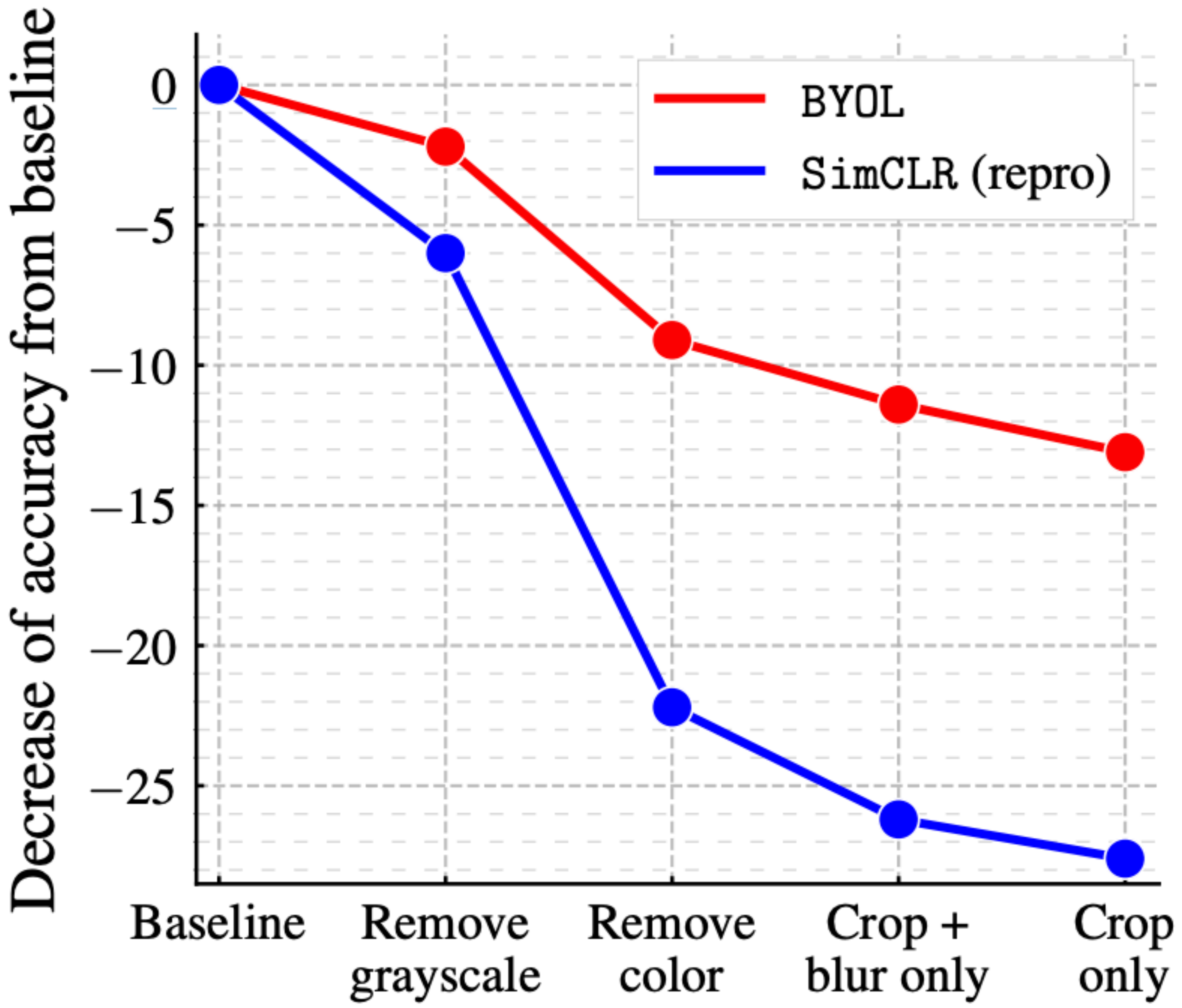
Cross Entropy Loss

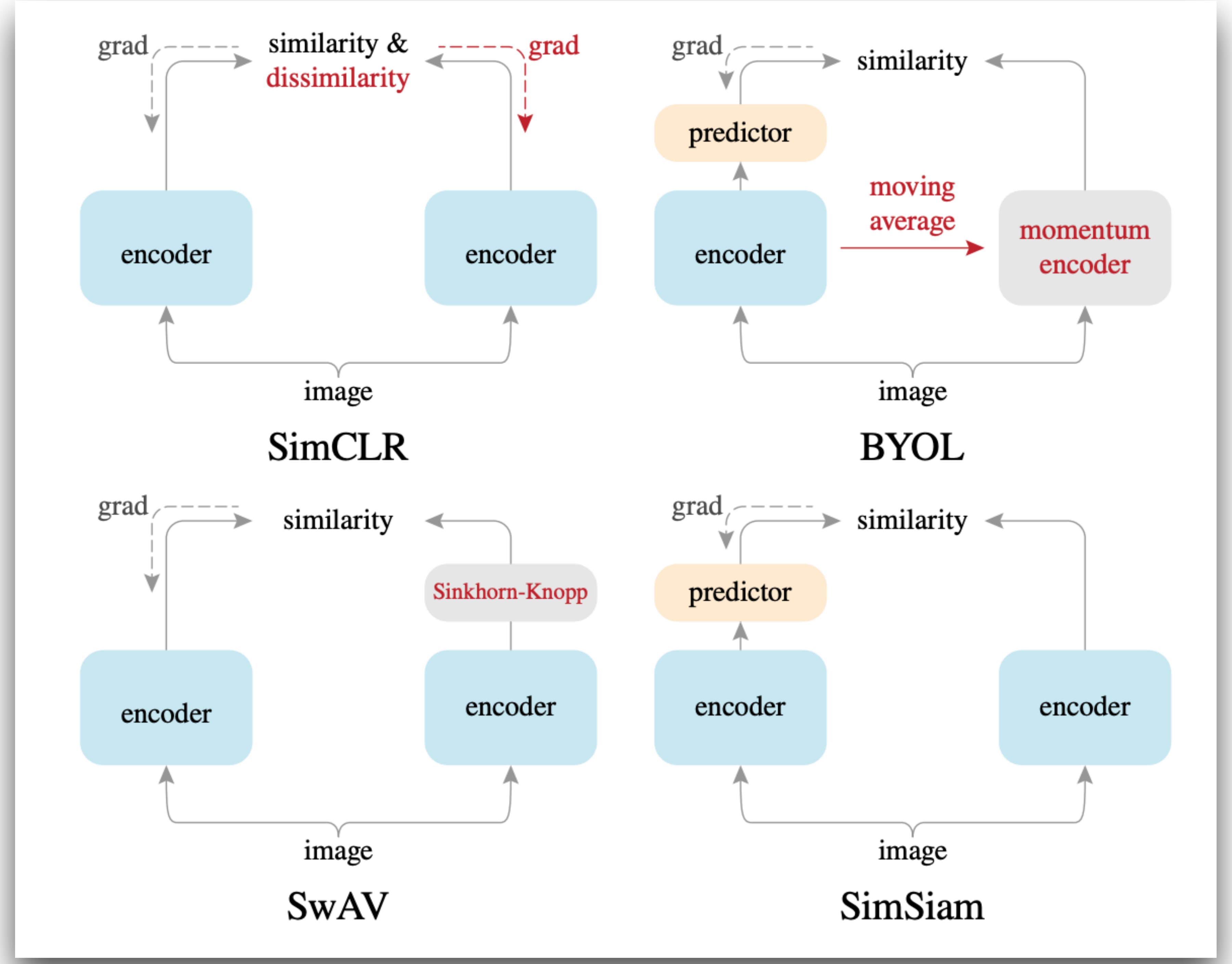


Supervised Contrastive Loss

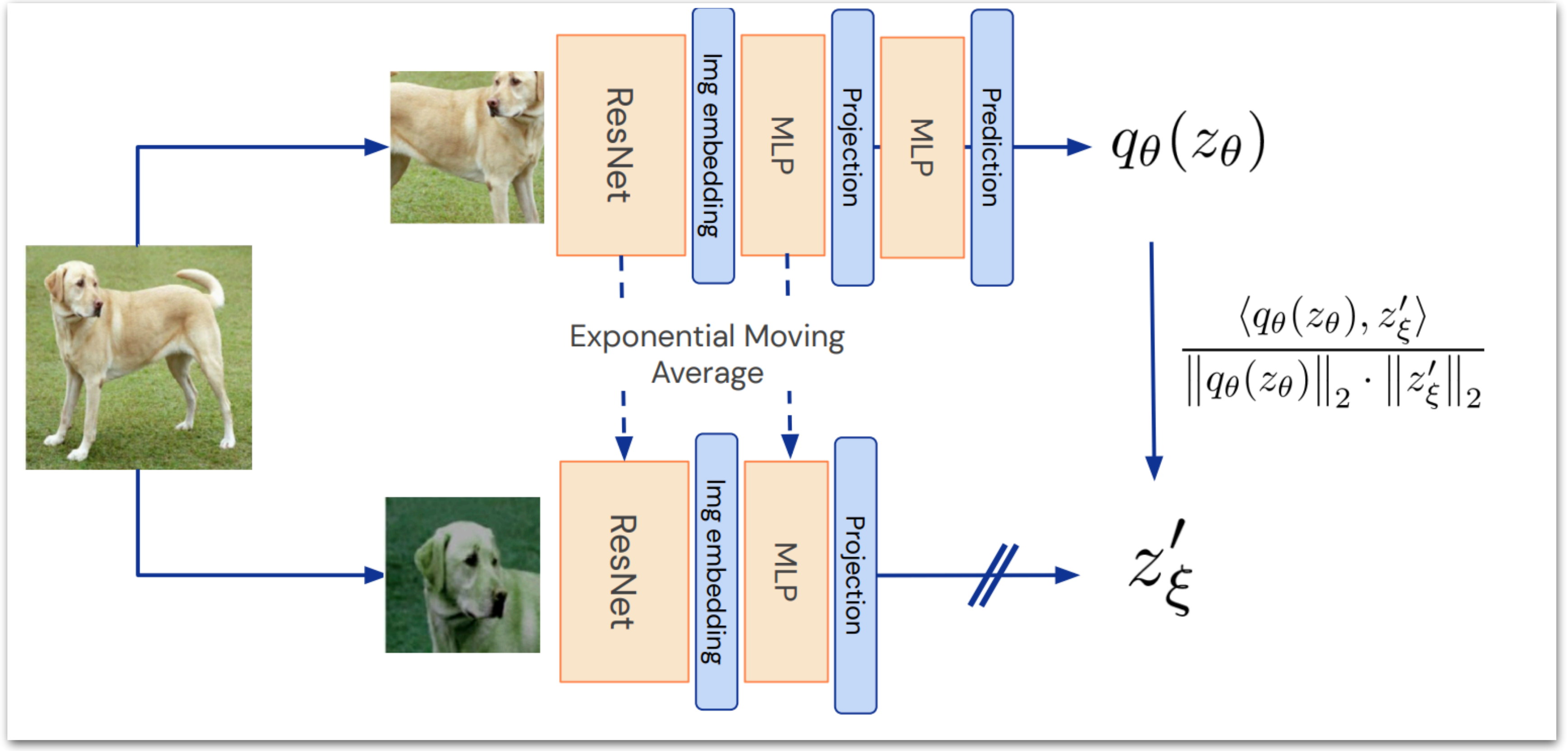


Side Remark





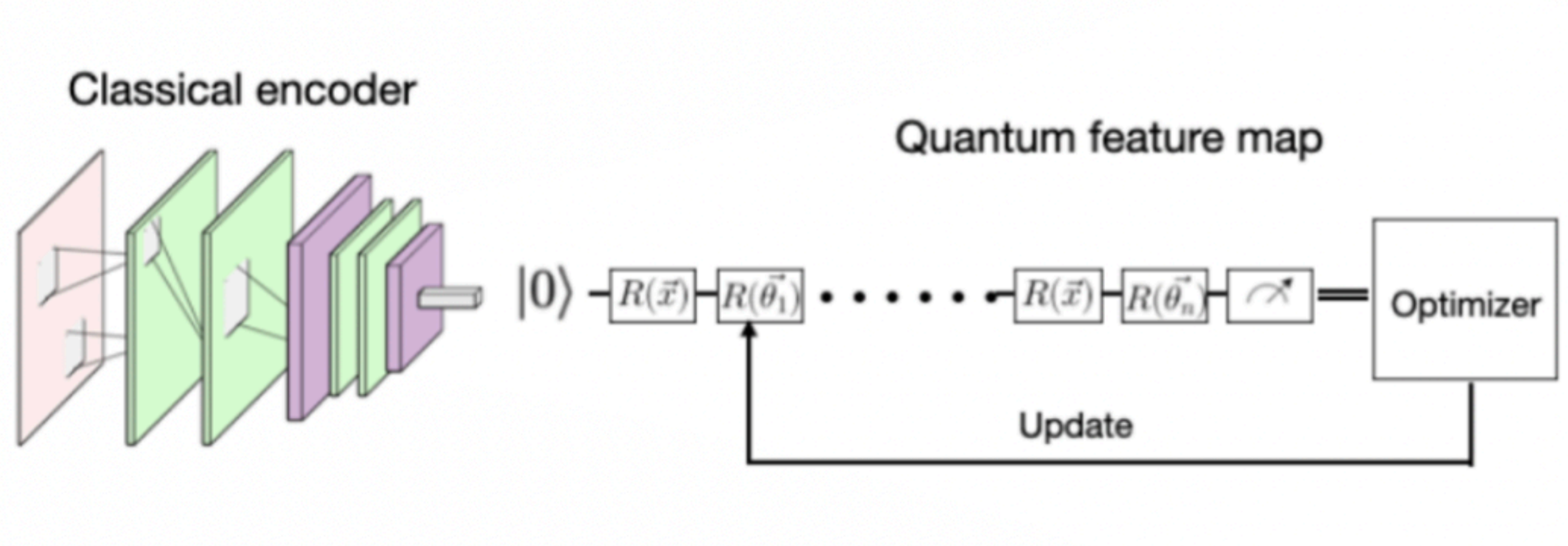
Bootstrap Your Own Latent (BYOL)

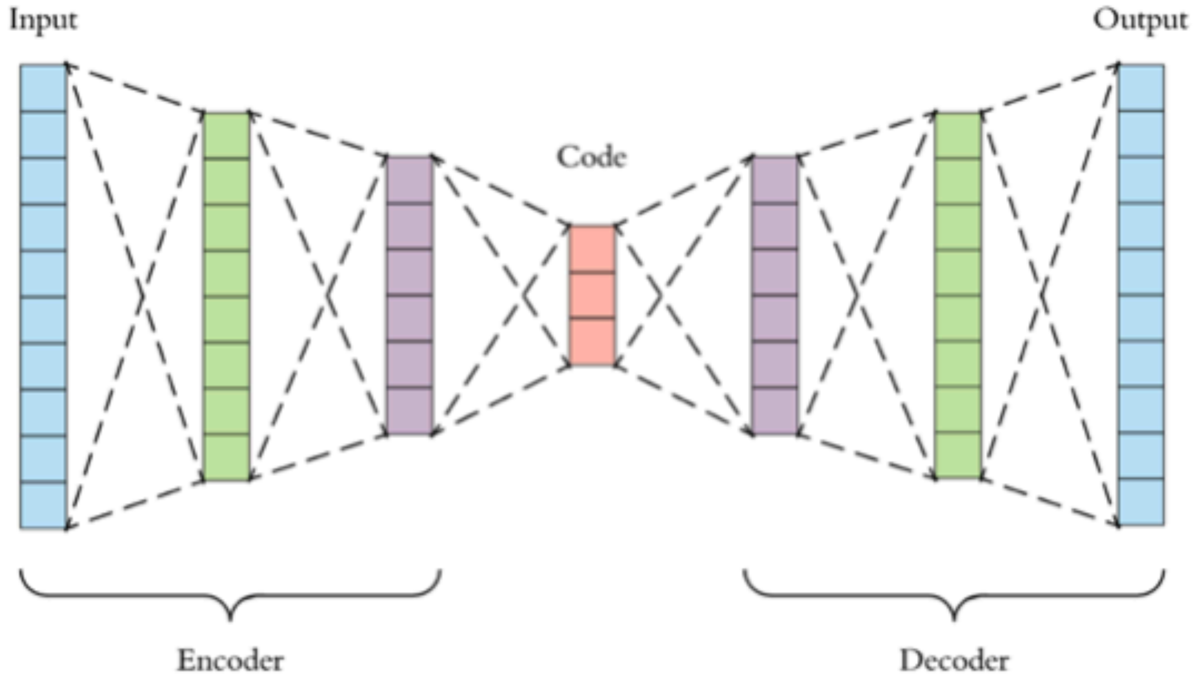


Shared weights with updating grads of both encoders causes the output from the projection head to collapse. Instead, bootstrap the weights with stopping grades of one encoder makes BEYOL less dependent on the augmentation

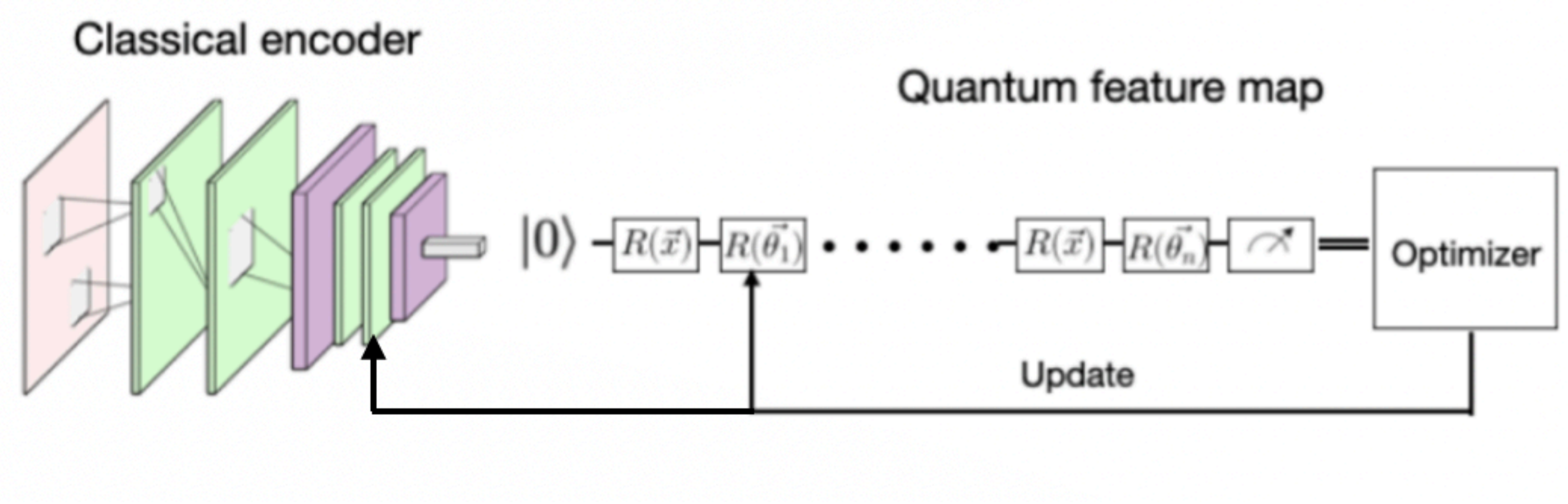
Quantum metric learning for image classification

Classical pre-trained Encoder
+
Trainable quantum layers





Classical trainable Encoder
+
Trainable quantum layers



Quantum metric learning for image classification

Training images: 200

Test images: 6000

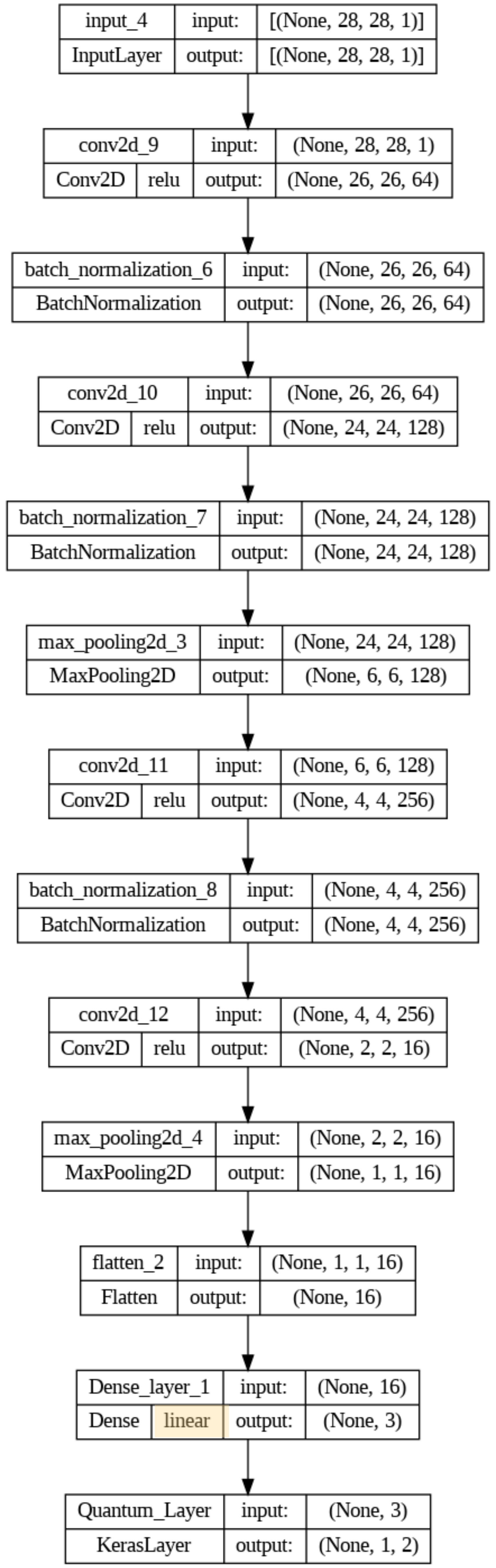
Validation images : 100

Optimizer: Adam (lr= 0.001)

Epochs: 10

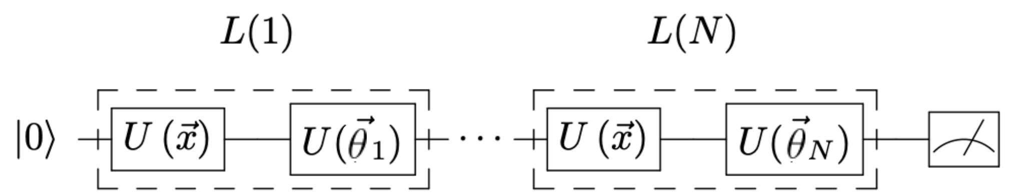
Batch size: 5

Quantum block: 10



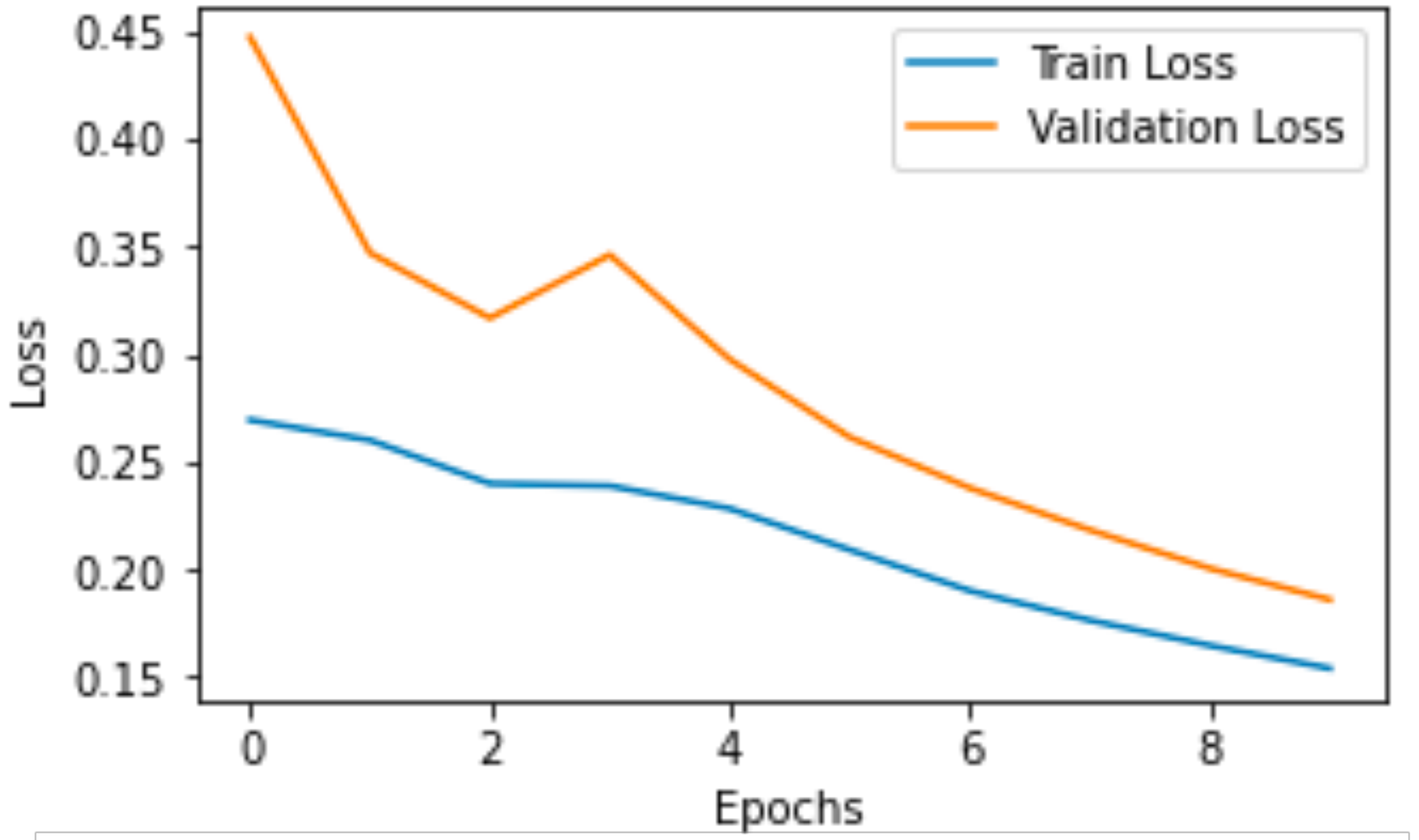
Classical Encoder

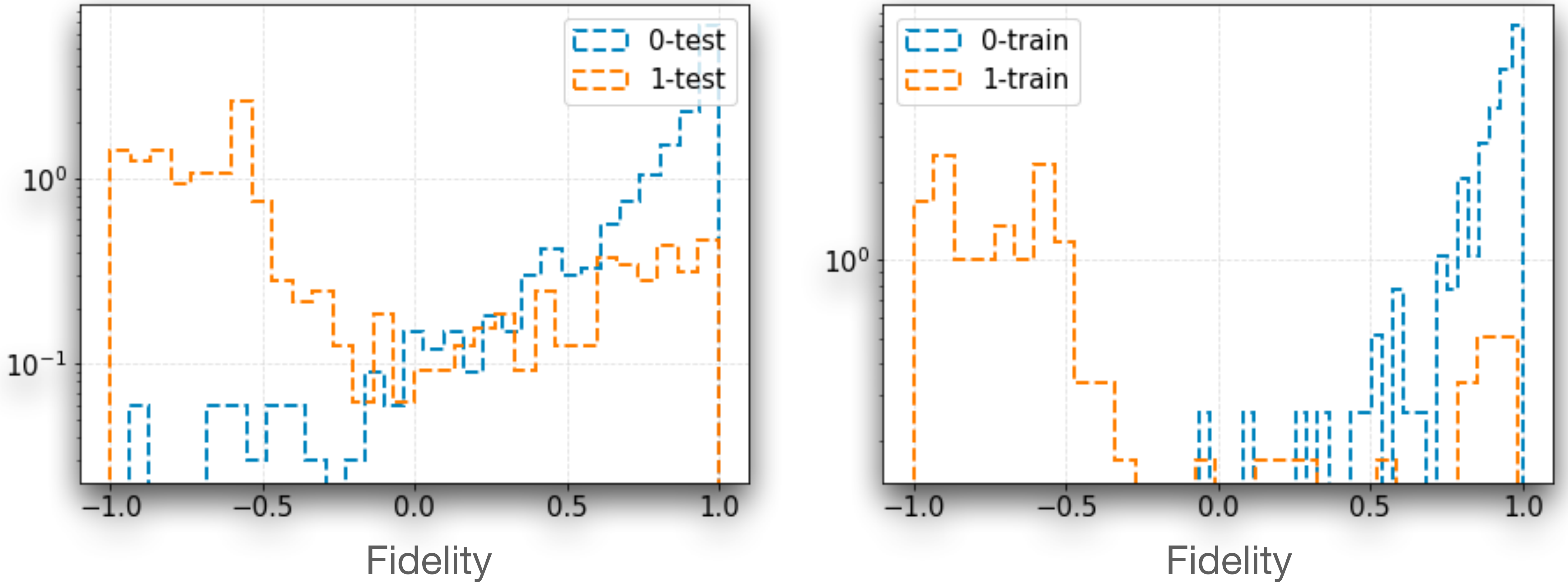
Quantum circuit

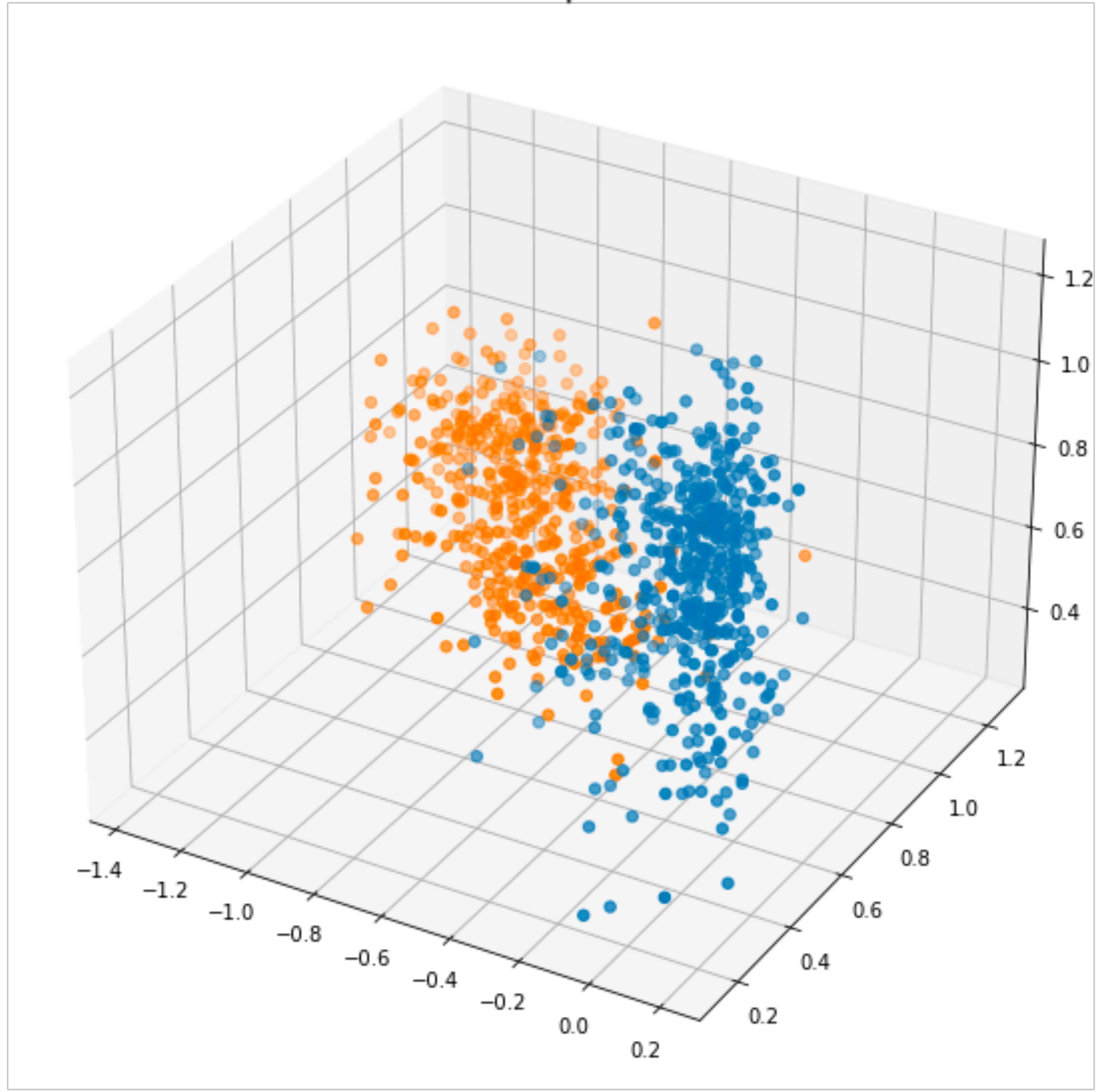


Quantum metric learning for image classification

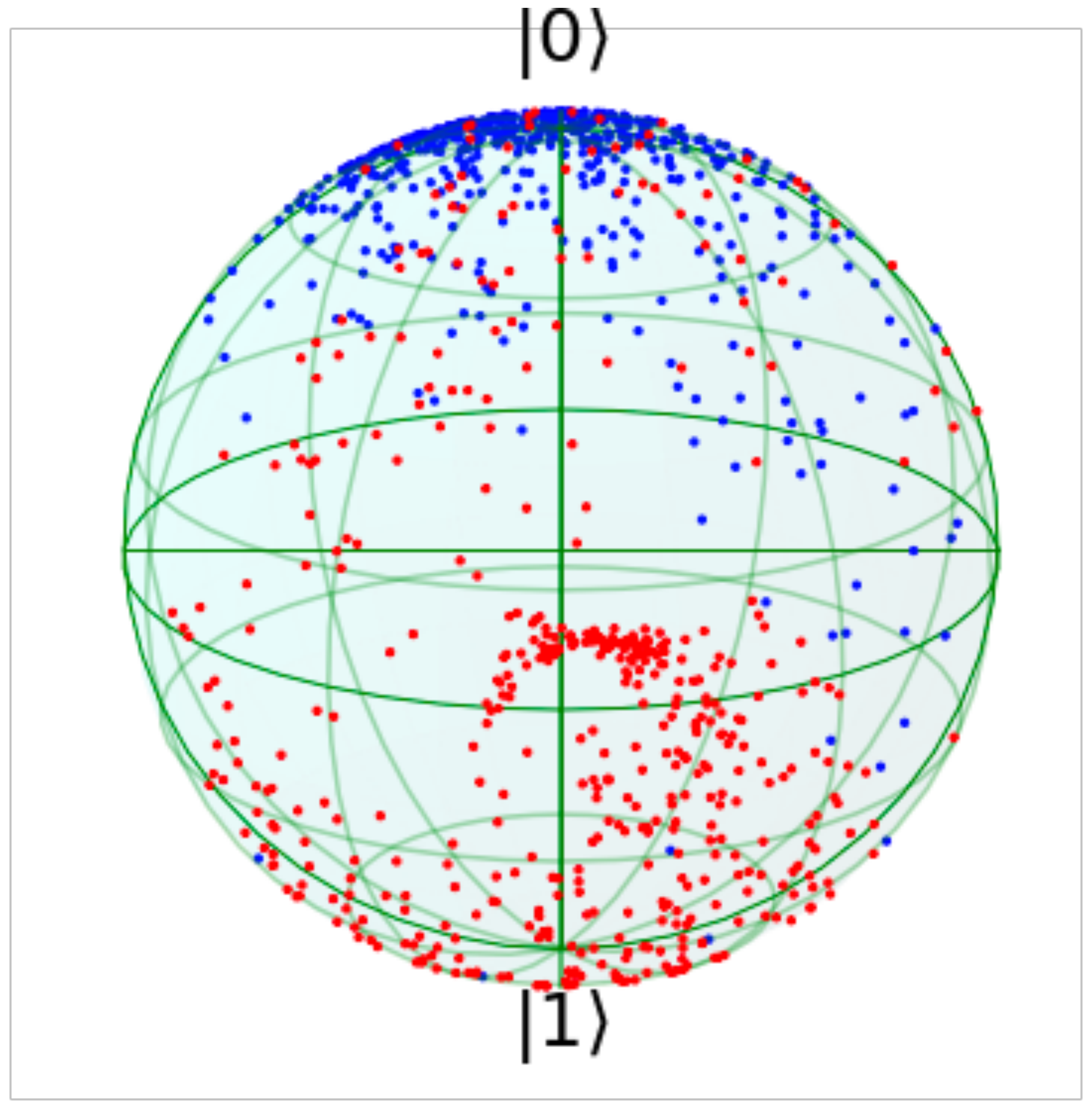
Test accuracy = 86.8% | Train accuracy= 93.5%







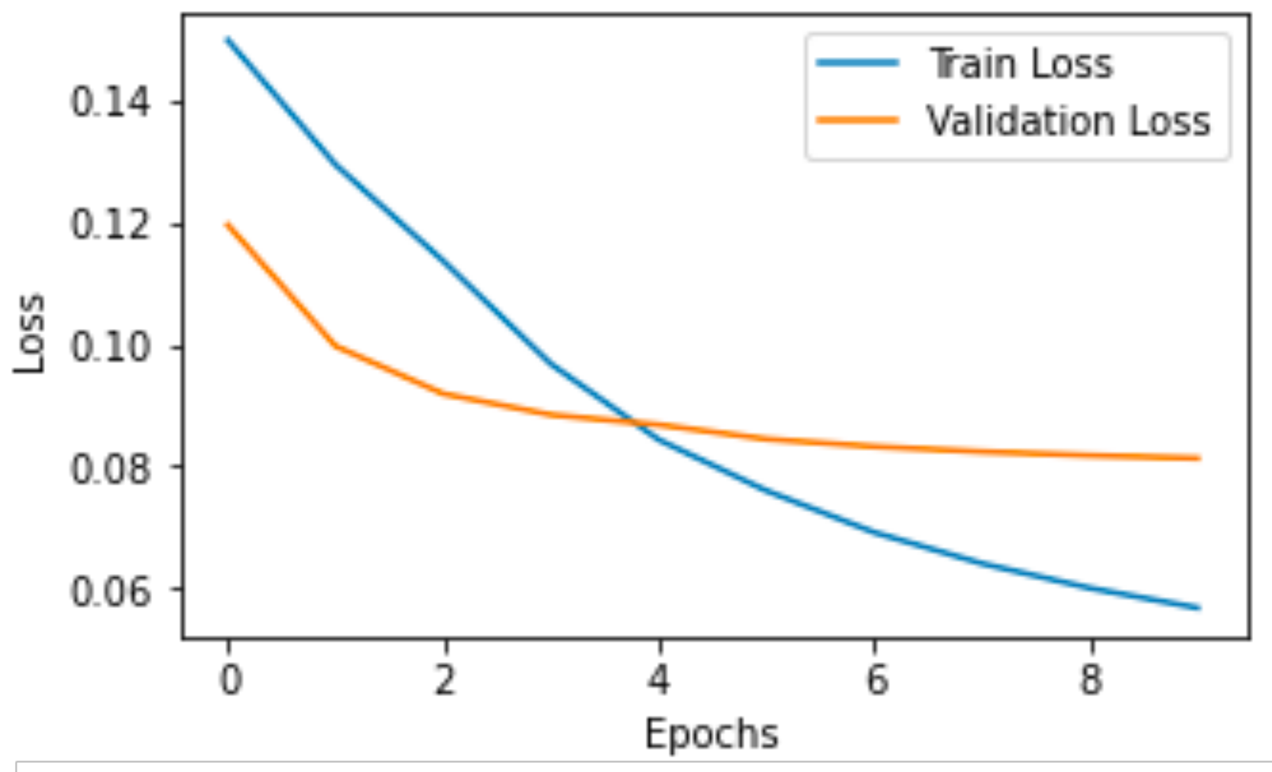
Latent space of
the classical encoder
Not-normalized

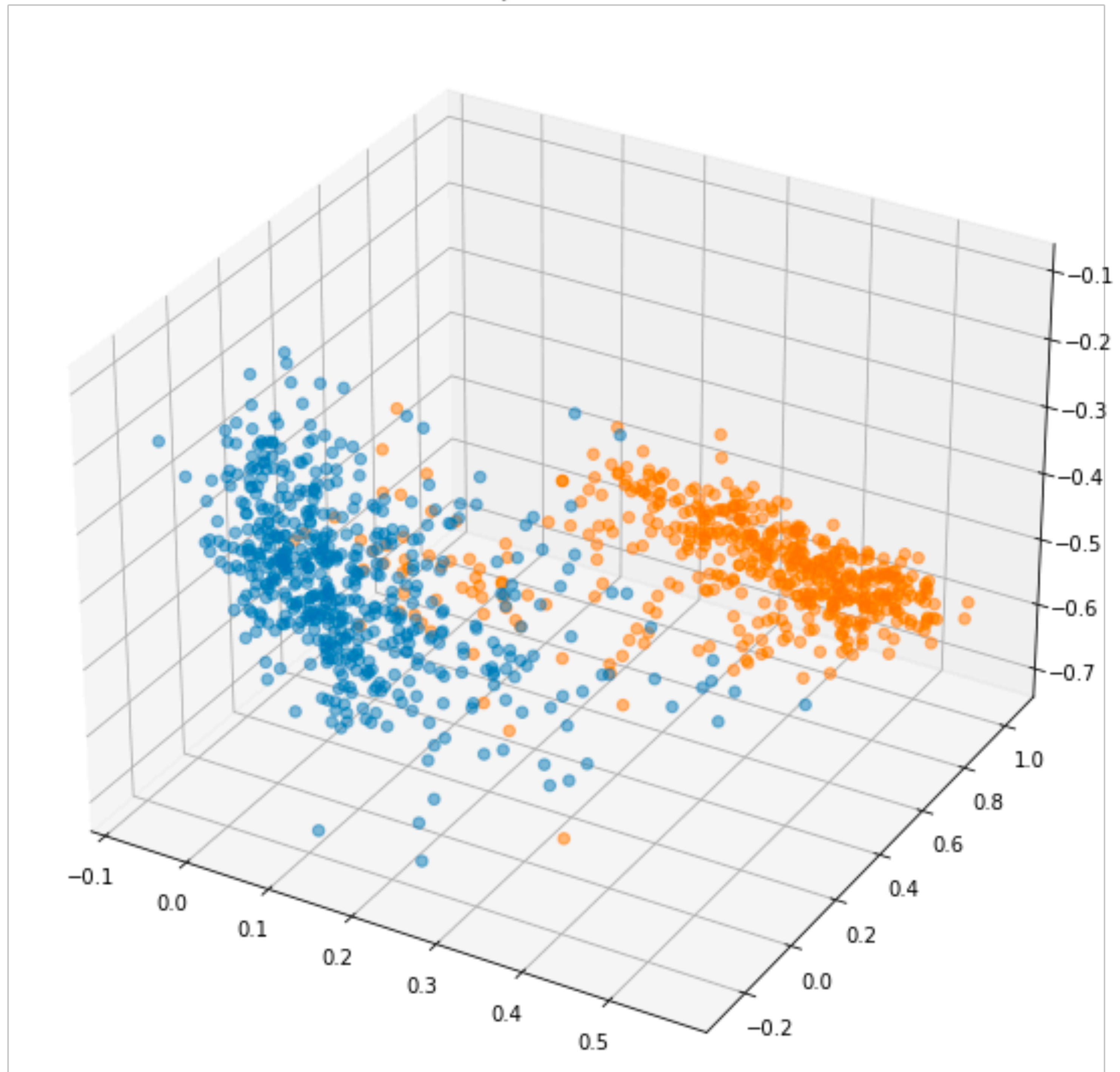


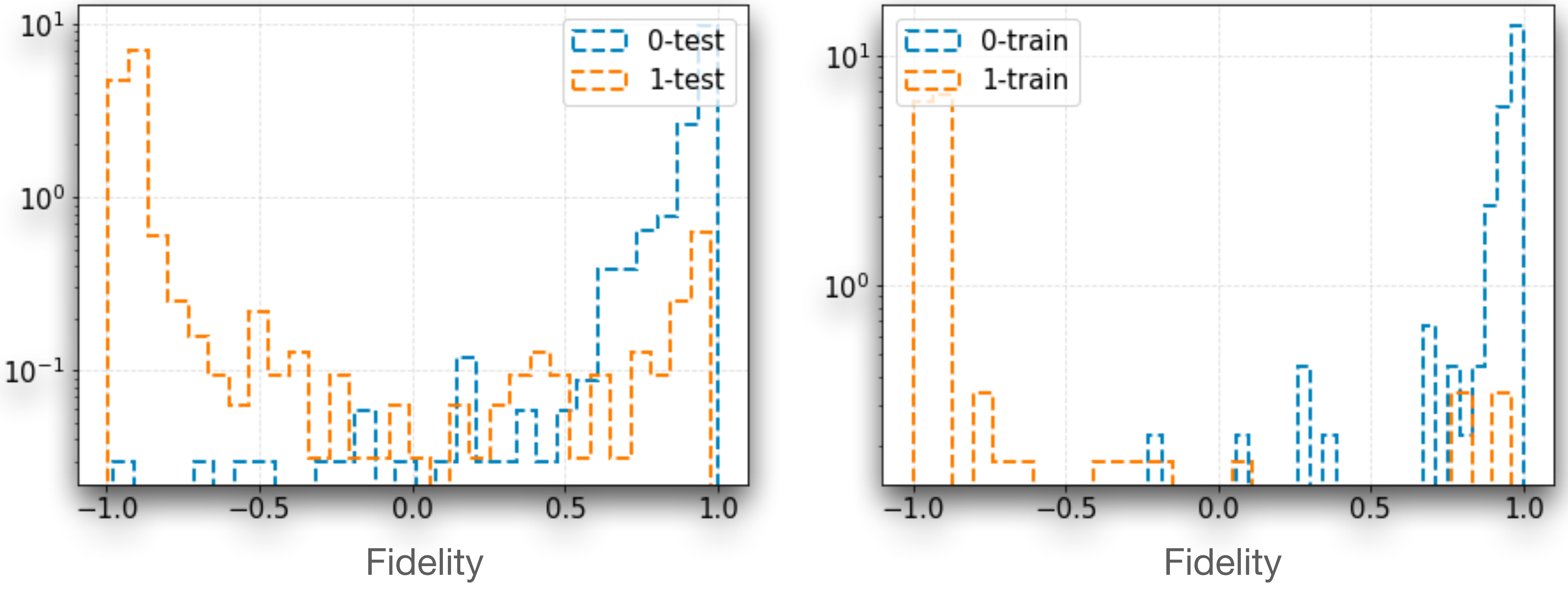
Bloch sphere
representation of the
embedded test images

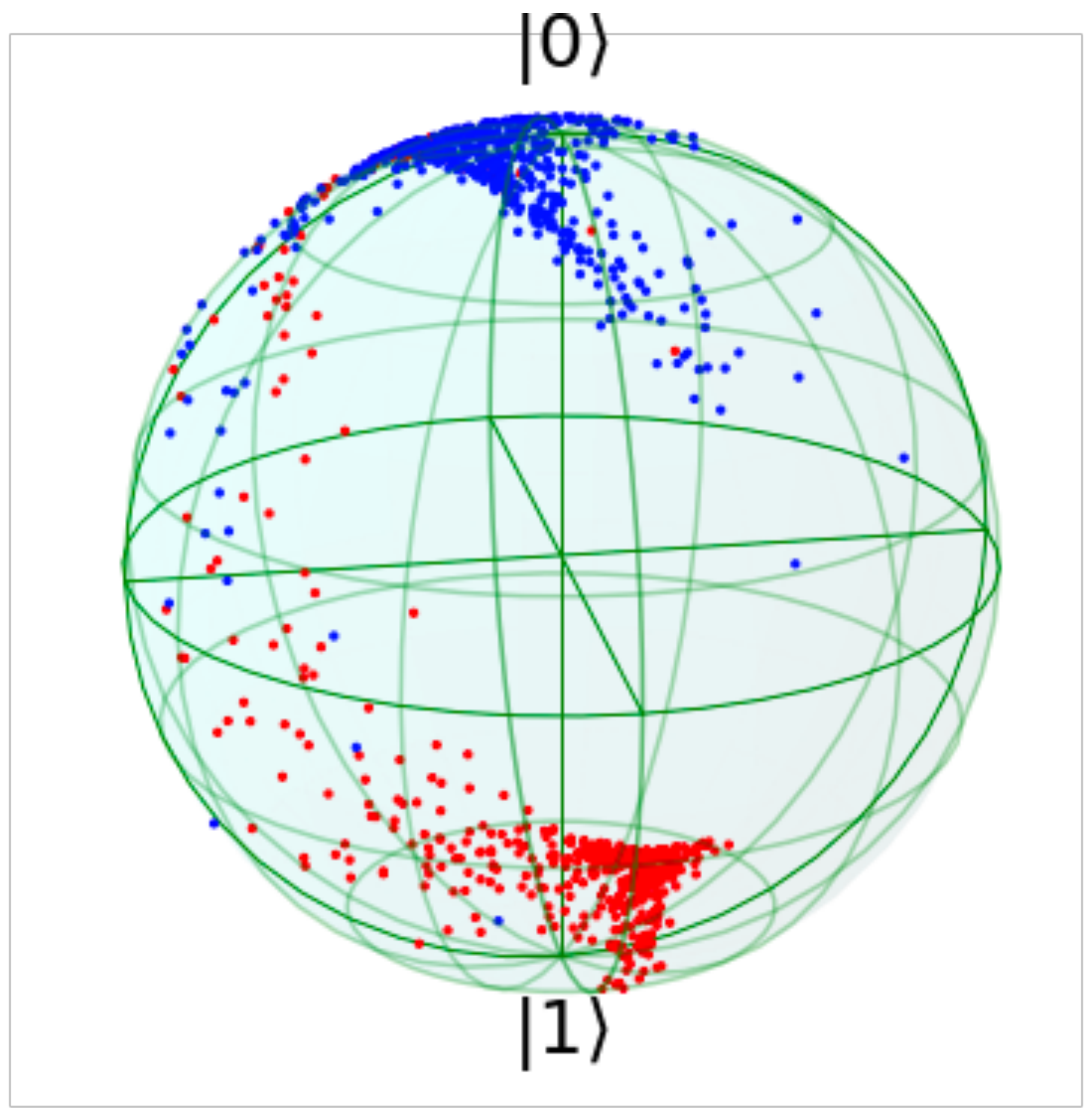
Normalizing the latent space (UnitNorm weights constrain)

Test accuracy = 93.5% | Train accuracy= 97.0%





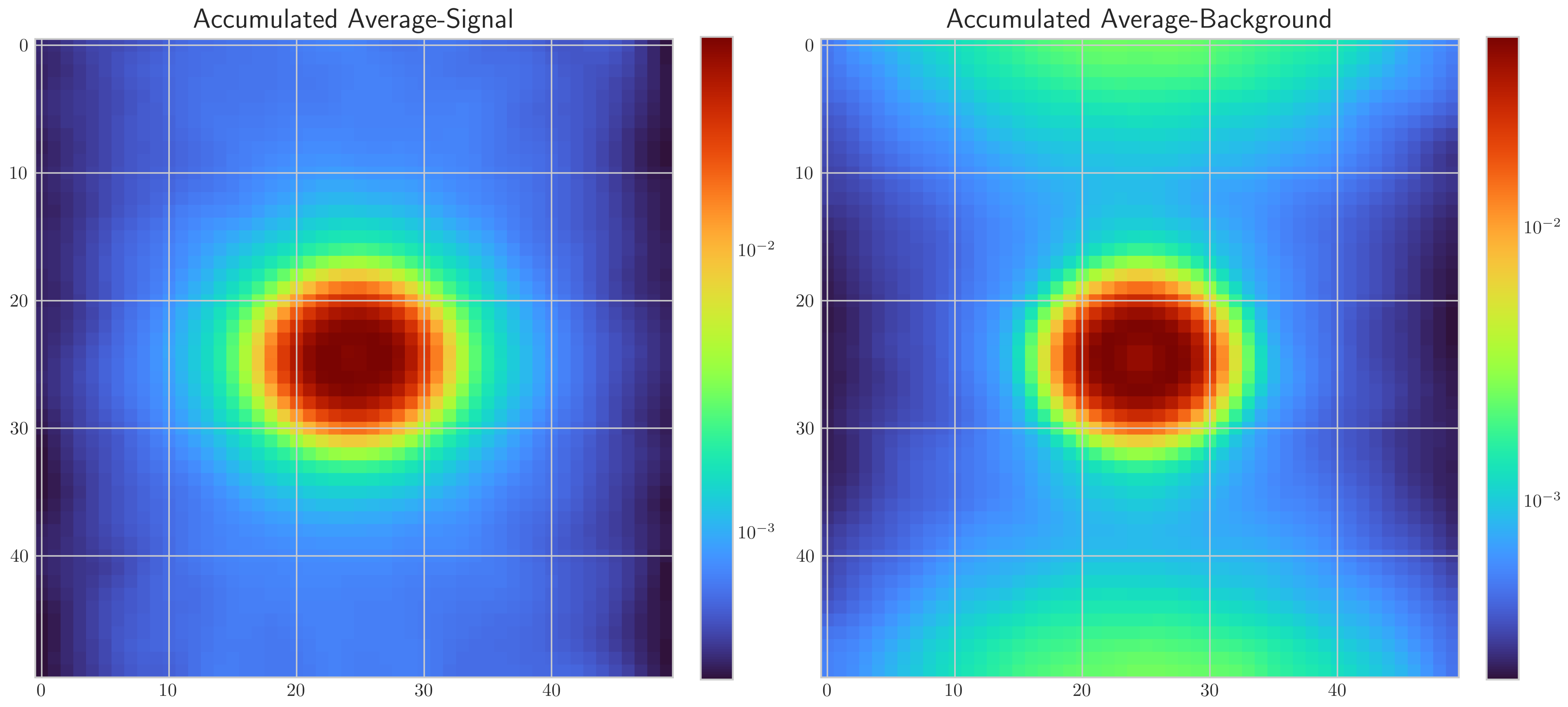




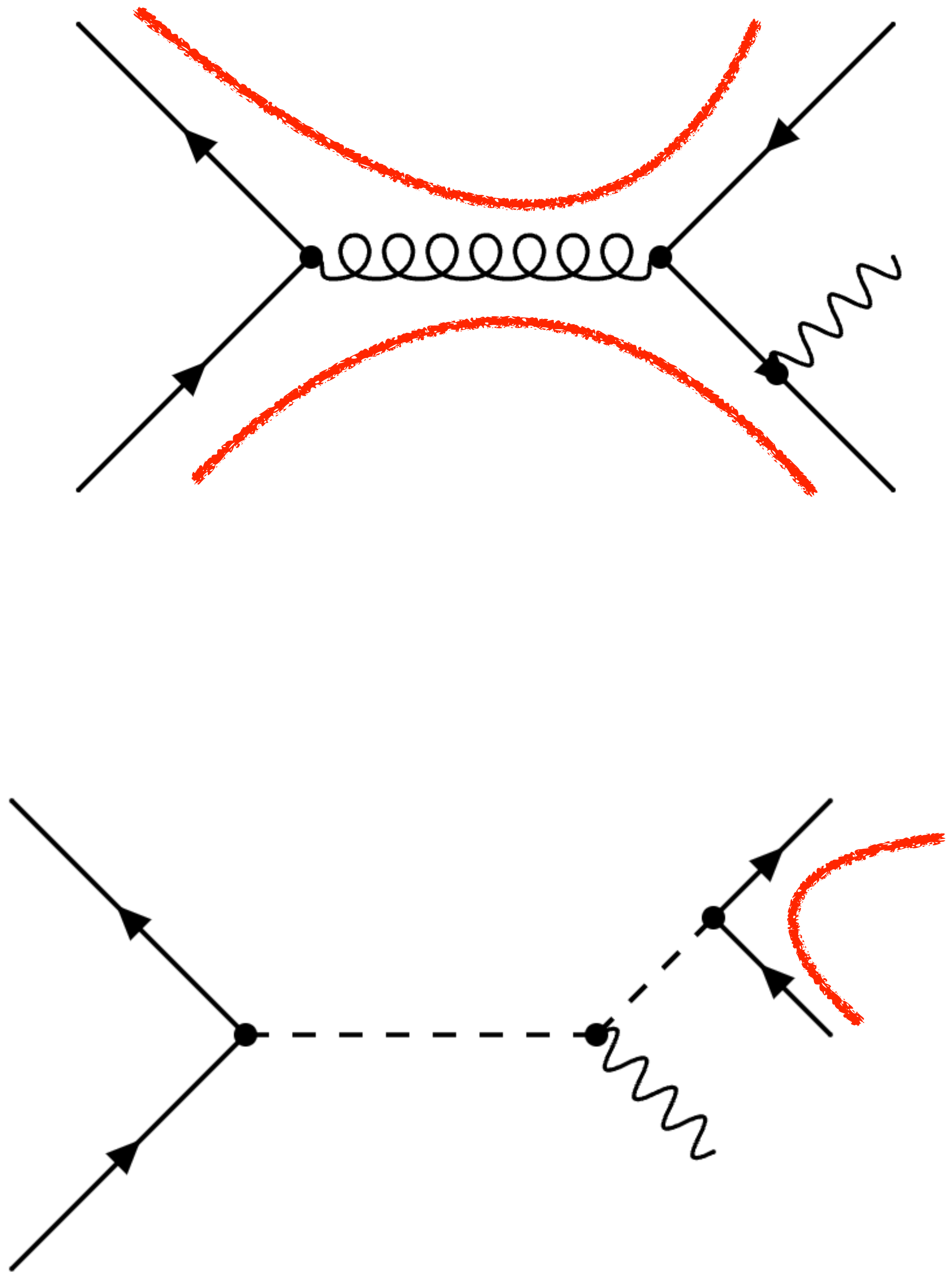
Quantum metric learning for jet image classification

Accumulated
50k images

$$pp \to Zh \qquad\qquad pp \to Zbb$$

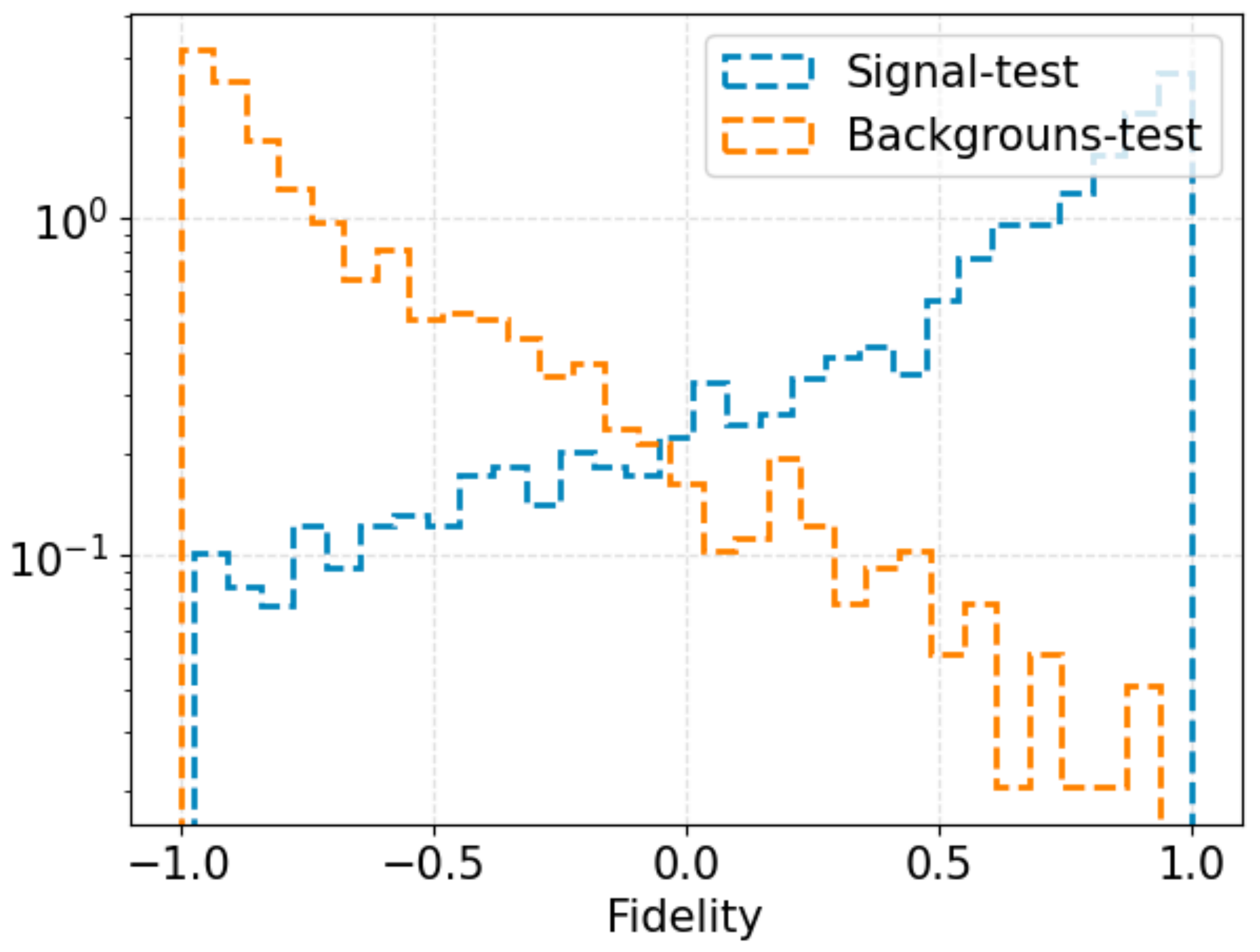


$$P_T(bb) > 250 \text{ GeV}$$

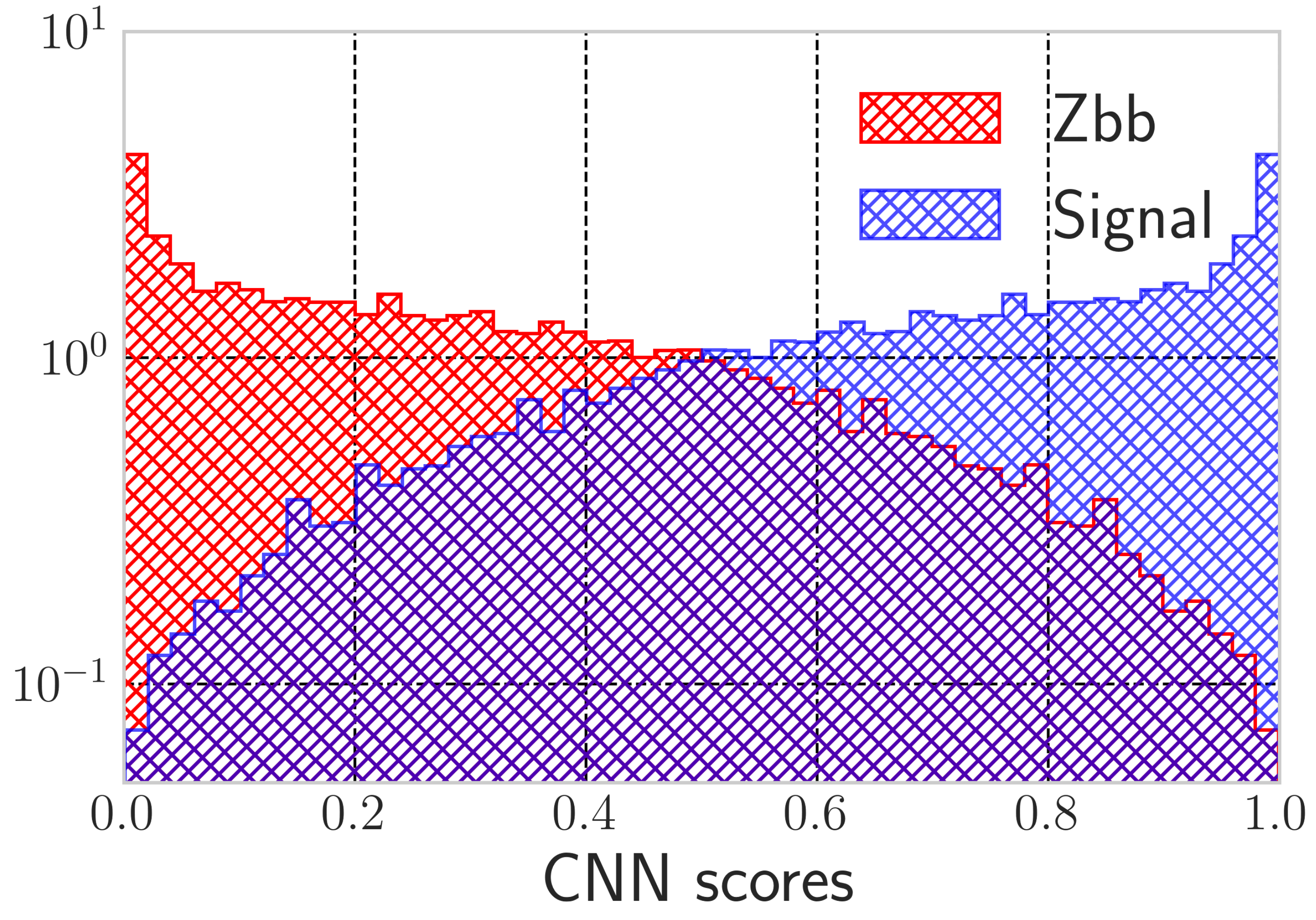


Quantum metric learning for jet image classification

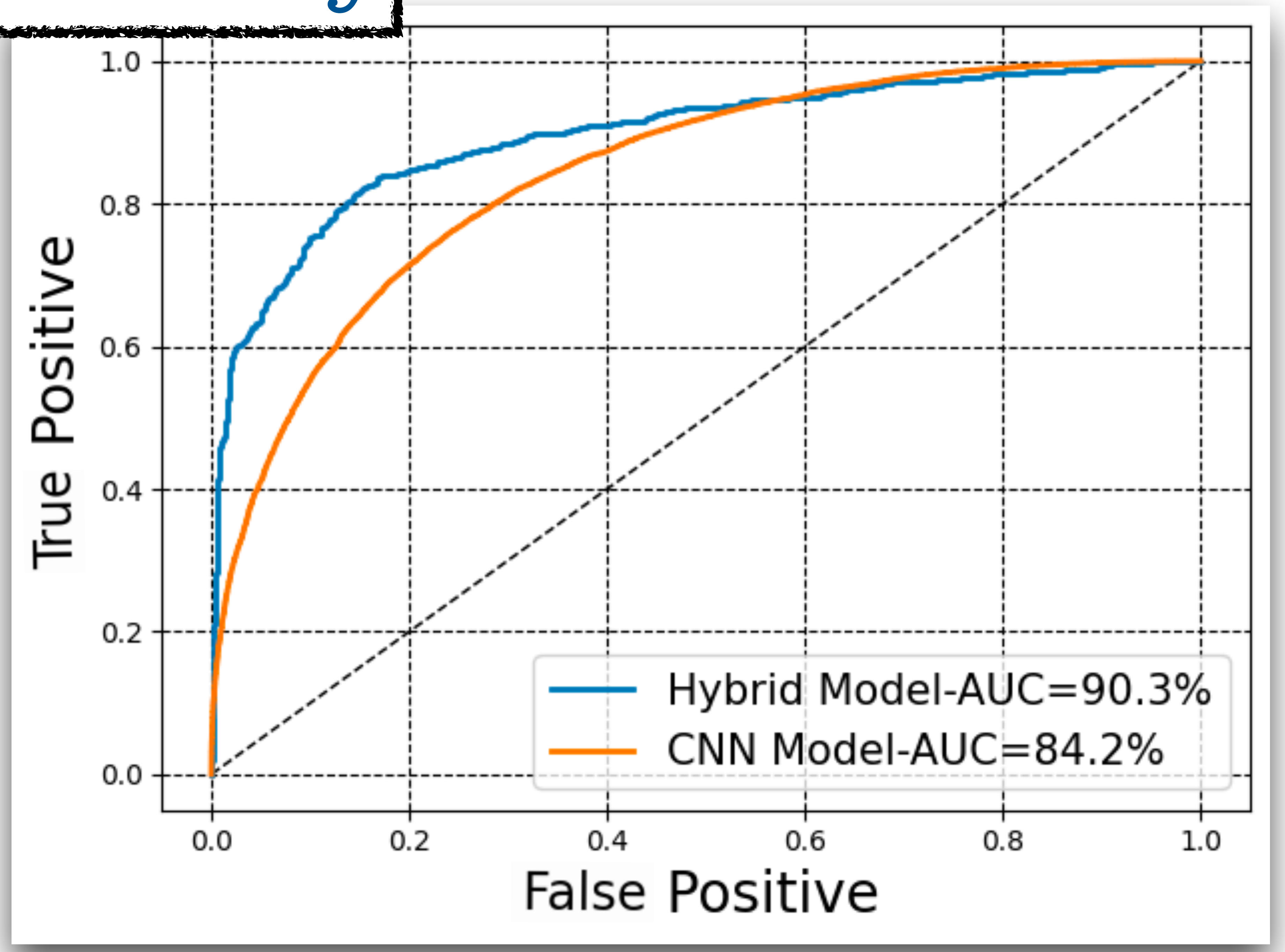
Accuracy = 82%



Accuracy = 75.9%

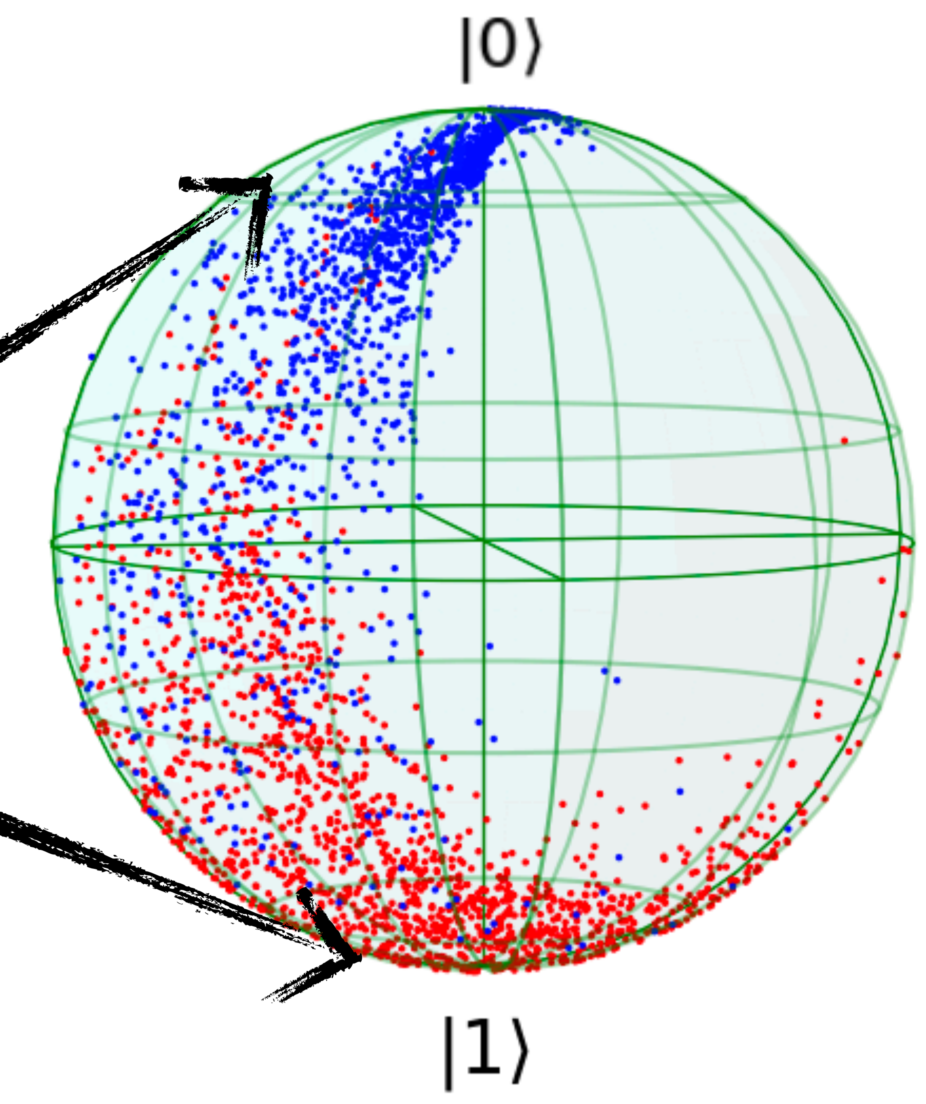


Preliminary



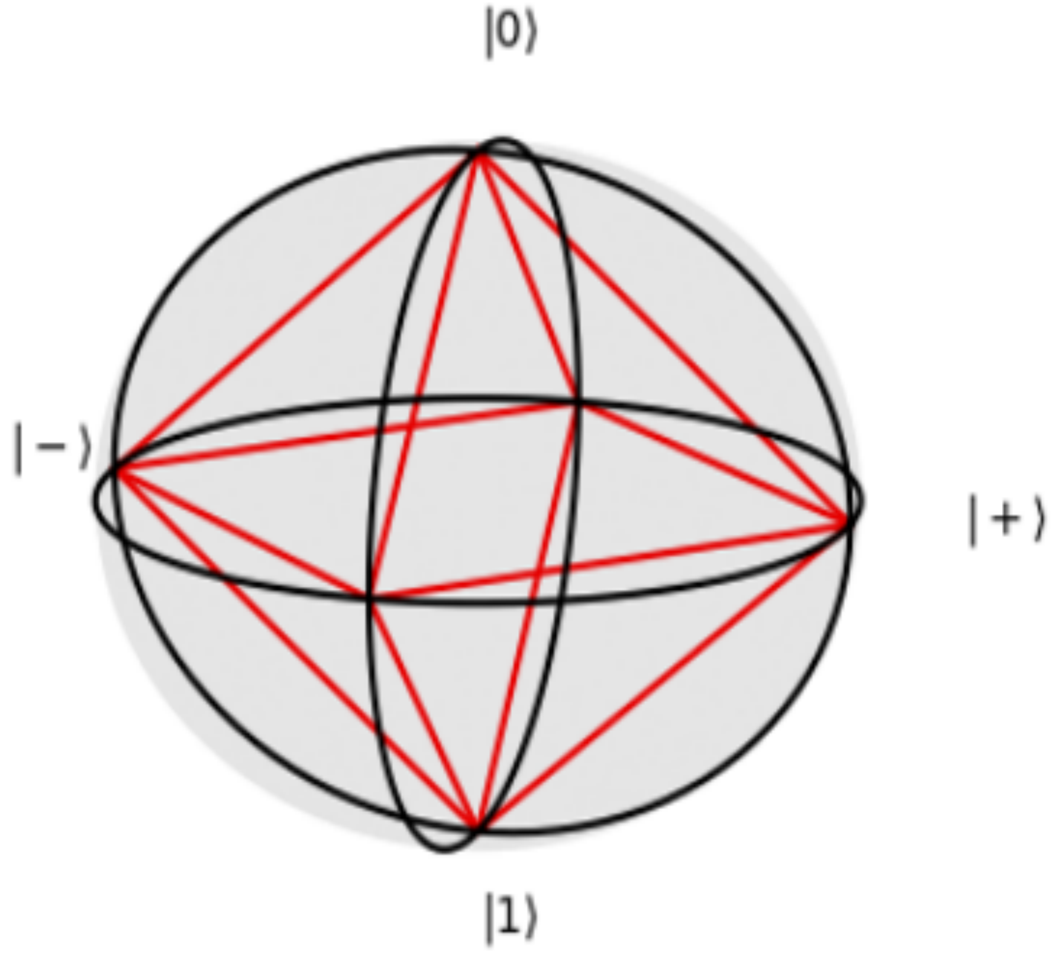
The classical encoder structure is the same in both models

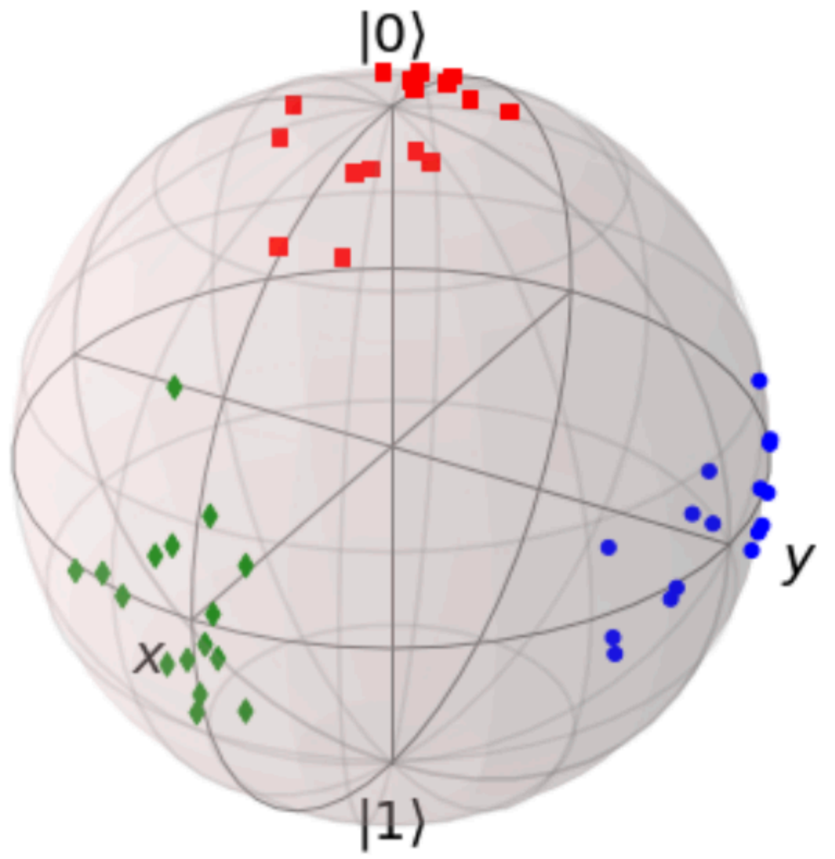
Signal and background images are embedded
to different states after training

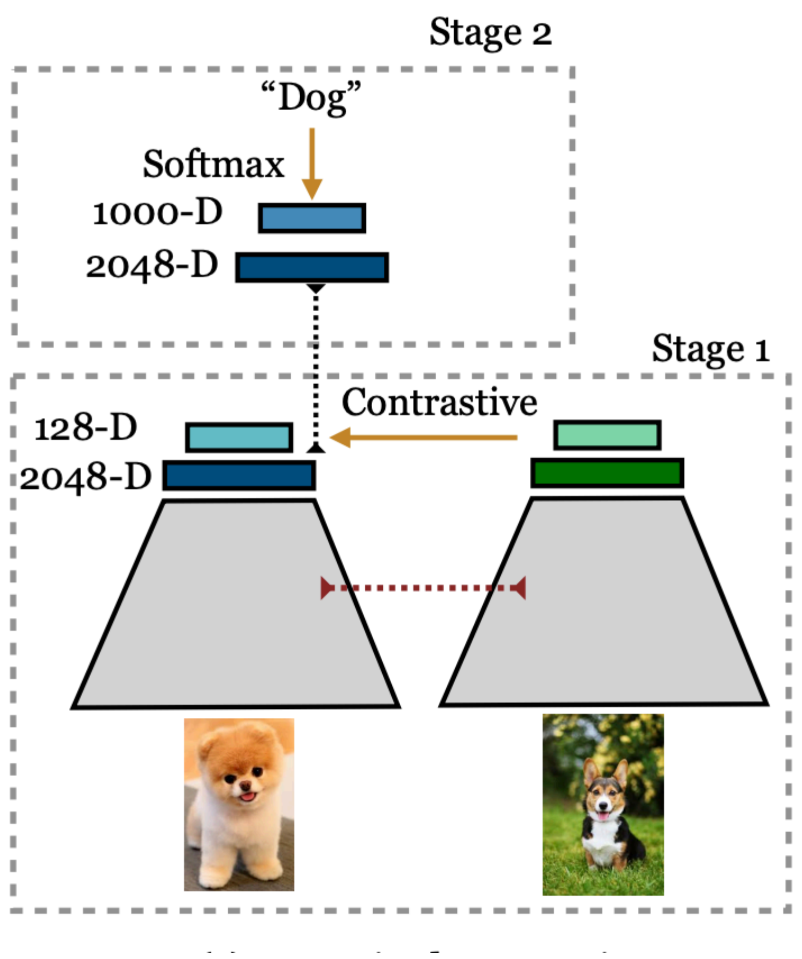


Side Remark

Multi-class classification problems







Contrast Learning Visual Attention for Multi Label Classification

Son D.Dao Ethan Zhao Dinh Phung Jianfei Cai
Faculty of Information Technology
Monash University

[cs.CV] 24 Jul 2021

Abstract

Recently, as an effective way of learning latent representations, contrastive learning has been increasingly popular and successful in various domains. The success of constrastive learning in single-label classifications motivates us to leverage this learning framework to enhance distinctiveness for better performance in multi-label image classification. In this paper, we show that a direct application of contrastive learning can hardly improve in multi-label cases. Accordingly, we propose a novel framework for multi-label classification with contrastive learning in a fully supervised setting, which learns multiple representations of an image under the context of different labels. This facil-

minibatch. More recently, supervised CL [20] has been proposed, where all the images with the same label as the anchor are considered as the positive samples and vice versa for the negative ones. Supervised CL has shown improvements in single-label image classifications than the self-supervised counterpart. With the above successful examples, CL has drawn significant research attention and has been applied in other tasks including image segmentation [40], adversarial training [21] or text to image learning [35].

Given the appealing properties and promising results of CL in single-label classification, it is natural to adapt it into multi-label cases to boost performance. However, this adaptation is non-trivial. In single-label cases, an image usually contains one salient object, thus, the label of the object can

To be done:

1- Increasing the quantum latent space by adding more qubits

2- Compare the Hybrid model against the classical contrastive models

3- The impact of the size of the training set

4- Test different structure and different entanglement for larger quantum circuit

5- Study the overfitting in the hybrid model

6- Can we use the Hybrid model for imbalanced data sets? Like anomalous data

arXiv:2301.10780 [quant-ph]

To be done:

1- Increasing the quantum latent space by adding more qubits

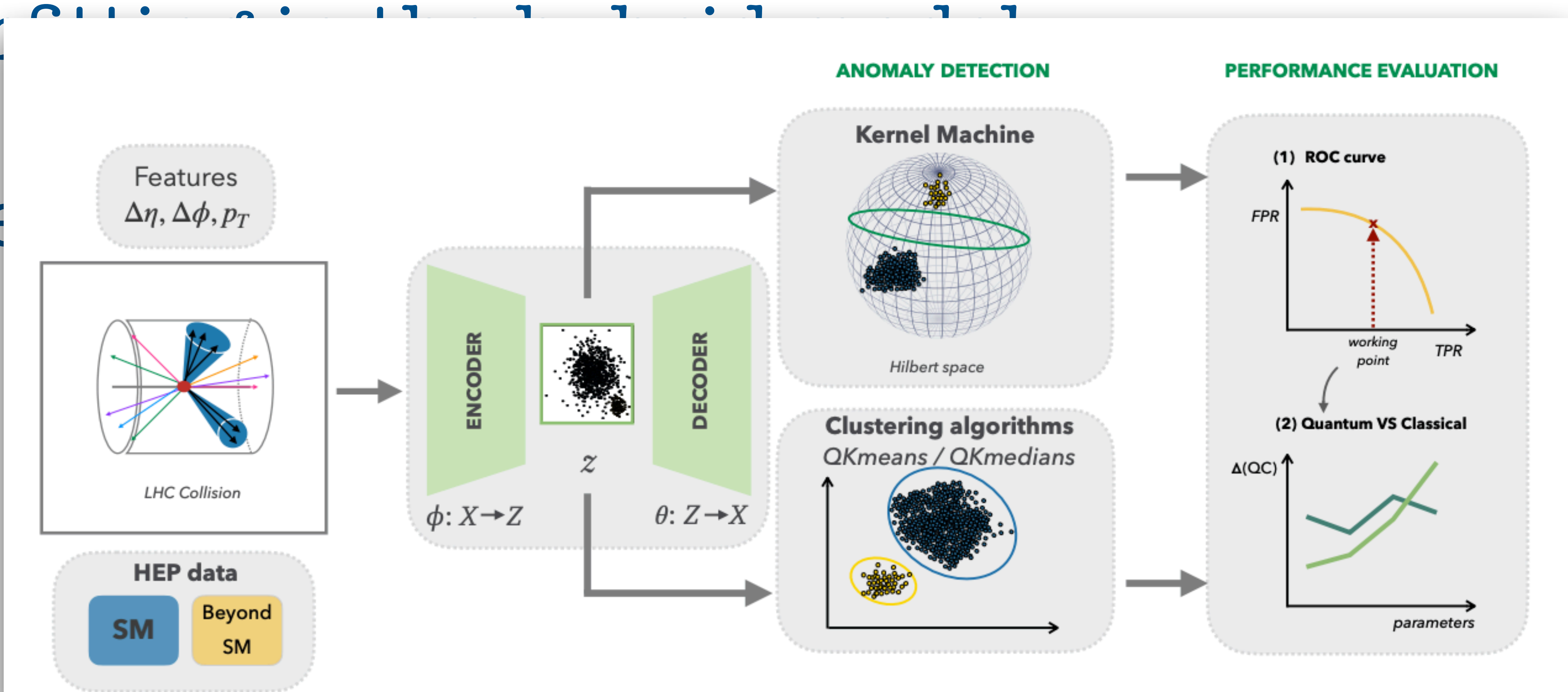
2- Compare the Hybrid model against the classical contrastive models

3- The impact of the size of the training set

4- Test different structure and different entanglement for larger quantum circuit

5- Study the over

6- Can we use the Like anomalous data



arXiv:2301.10780 [quant-ph]

Thank you!