arXiv:2305.04504v1 [quant-ph] 8 May 2023

数据编码、拟设表达能力与纠缠对混合量子神经网络可训练性的综合影响

Muhammad Kashif 和 Saif Al-Kuwari

哈马德·本·哈利法大学信息计算系、科学与工程学院、卡塔尔基金会，卡塔尔多哈

摘要

本文提出了一个框架，用于研究导致量子神经网络（QNN）中贫瘠高原问题的若干因素的综合影响。贫瘠高原问题是量子机器学习（QML）面临的一项关键挑战。这些因素包括数据编码、量子比特纠缠和拟设表达能力。为在真实场景中探究这些因素的共同影响，我们聚焦于用于多分类任务的混合量子神经网络（HQNN）。我们提出的框架旨在分析这些因素对混合量子神经网络训练景观的影响。研究结果表明，混合量子神经网络中的贫瘠高原问题取决于底层拟设的表达能力和数据编码类型。此外，我们发现纠缠也会影响贫瘠高原问题。通过采用多种分类任务评估指标评估混合量子神经网络的性能，我们针对不同的约束情形提出建议，凸显了该框架对于量子神经网络取得实际应用成功的重要意义。

关键词

量子机器学习；纠缠；数据编码；量子神经网络；可训练性。

1. 引言

构建实用量子计算机的探索在过去几年中不断加速。若干量子设备已经研发完成，其量子比特数量约为一百。这些设备被称为含噪中等规模量子（NISQ）设备[1]。尽管这些设备存在局限且容易出错，但与现有最先进的经典计算机相比，它们在特定应用中展现出了明显优势 [2–5]。由于含噪中等规模量子设备需要具有较浅深度且具备抗噪能力的量子程序，整合经典处理与量子处理的混合设计空间已成为释放量子计算在广泛应用中潜力的主流途径 [6]。

在混合设计空间中，变分量子算法（VQAs）是最受欢迎的一类算法。这些算法利用含噪中等规模量子设备，通过参数化量子电路（PQCs）评估目标函数，并利用经典设备针对目标应用进行函数优化。VQAs 已在广泛的应用领域中得到研究，包括量子化学 [7],态对角化 [8], 因子分解 [9], 量子优化 [10], 以及量子场论模拟 [11,12]。此外，这些算法也在噪声鲁棒性 [13], 可训练性 [14–16]和计算复杂性方面得到了研究

联系邮箱：mkashif@hbku.edu.qa

[17,18]。换句话说，VQAs 与机器学习（ML）算法十分相似，因为它们同样训练计算机学习模式 [19]。因此，VQAs 被提出作为各种 ML 算法的量子类比 [10,20–24]。由此，量子计算与 ML 融合，催生了量子机器学习（QML）这一新兴领域。

近年来，人们为量子机器学习（QML）应用提出了许多参数化量子电路（PQC）[25]。其中包括量子神经网络（QNN），它们作为经典深度神经网络（DNN）的量子扩展，已得到广泛研究 [26–33] 。QNN 的基本构建模块是量子感知机，其形式已有多种不同的提出方式 [34–40]。为此，参数化量子电路成为人工神经元一种很有前景的量子类比 [41–43]。鉴于含噪中等规模量子时代的局限性，人们通常使用混合量子神经网络（HQNN）来分析 QNN 中潜在的量子优势。HQNN 通过在典型 QNN 的前端加入经典输入预处理、末端加入经典输出后处理，复现 QNN 的架构。输入预处理通常旨在缩减输入规模，以应对含噪中等规模量子设备的限制（主要是量子比特数量），并通过神经元数量较少的经典神经元层或某种降维算法来完成。输出后处理则用于以有意义的方式解读封装的 QNN 输出。后处理通常由末端的经典神经元层完成，这也使得可以应用熟悉的非线性激活函数来获得最终输出或预测。尽管 QNN 已针对各种应用得到广泛研究，但现有文献仍缺乏关于其量子优势的可靠且明确的论述 [44–46]。

1.1. 贫瘠高原

混合量子神经网络中最具挑战性的问题之一是贫瘠高原（BP）现象 [14–16,47]。在贫瘠高原中，随着系统规模增大，混合量子神经网络训练过程中的代价函数景观会呈指数级变平。换句话说，待优化参数的梯度会随着量子比特数量的增加而呈指数级消失。这意味着，贫瘠高原出现在混合量子神经网络的训练景观中会影响其可训练性[1],，导致性能显著下降，进而限制混合量子神经网络在实践中的适用性。

已有研究表明 [14] ，如果一个足够随机的拟设（一种由一系列作用于特定量子线路的门组成的量子子程序）所对应的均匀酉分布在二阶矩上与之匹配，即构成一个酉 2-设计，那么它就会遭遇贫瘠高原。因此，拟设的选择是混合量子-经典算法成功的关键。对此，经常研究的一些拟设包括哈密顿量变分拟设 [48]，耦合簇拟设 [49–51],量子交替算子拟设 [10,52] 和硬件高效拟设 [53]。理想情况下，拟设既应具备可训练性，也应具备表达能力，从而达到最优解。特别需要拟设的表达能力，以便它能够准确逼近解。同时，训练景观也需要具备足够易于探索的特征，以找到解。量子拟设的表达能力 表明给定拟设能够以多均匀的方式探索酉空间 [54]。 [54] 中的作者将报告贫瘠高原的原始工作 [14] 扩展到拟设表达能力。他们提出了问题启发式拟设的概念，而非硬件高效拟设；这在含噪中等规模量子时代具有重要意义。其表达能力

[1]可训练性本质上确保目标函数在每次训练迭代后都得到优化，直至模型收敛。

拟设在一定程度上对克服贫瘠高原（BP）起着重要作用。表达能力越强的拟设越容易受到贫瘠高原（BP）的影响，反之亦然。同样，深层拟设被认为具有更强的表达能力 [14]。此外，拟设表达能力可以直接根据目标问题的性质推导得出（复杂问题采用表达能力更强的拟设，反之亦然）。

除了拟设表达能力之外，混合量子神经网络中的贫瘠高原问题也可能源于参数化量子电路所采用的纠缠类型 [55], 、高深度量子电路中的噪声水平 [1,47,56–58] ，以及数据在参数化量子电路中的编码方式 [43]。纠缠是量子力学的一项基本性质，也是构建具有表达能力的量子电路的关键。然而，如 [55] 所述，纠缠也可能成为贫瘠高原的潜在来源。数据编码也是混合量子神经网络中的关键步骤，通常被认为是性能瓶颈，而且也会影响混合量子神经网络的可训练性和表达能力 [59]。

1.2. 研究空白

如第 1.1 节所述，混合量子神经网络中同时采用了数据编码、拟设表达能力和纠缠。尽管已有不同研究分别探讨了贫瘠高原对这些概念的依赖关系 [43,54,55,59],但据我们所知，尚未有人从贫瘠高原的角度研究它们的联合（整体）效应（以及它们彼此之间的关系）对量子神经网络可训练性的影响。此外，现有研究侧重于量子神经网络独立的数学实现（公式化或建模），而混合量子神经网络则更贴近含噪中等规模量子设备上的实际应用，使我们能够使用真实世界数据集进行实验。因此，我们需要一种混合量子神经网络框架，以便同时分析数据编码、拟设表达能力以及量子比特之间的纠缠对混合量子神经网络可训练性的影响。

1.3. 提出的框架

本文提出了一个框架，用于实证分析（基于实验所得数据）数据编码、拟设表达能力和纠缠在混合量子神经网络中的联合影响及其相互关系。我们通常针对真实世界应用（多分类），研究上述概念对采用硬件高效周期拟设结构的前馈混合量子神经网络的影响。图 1 展示了我们分析的抽象示意图。

经典到量子的特征映射通过两种常用的数据编码策略实现，分别称为振幅编码和角度编码。对于参数化量子电路，我们考虑两种相似的拟设结构（用作混合量子神经网络中的隐藏量子层），两者仅在是否包含纠缠方面有所不同。我们将其称为纠缠拟设（图 1 中显示为最近邻纠缠）和无纠缠拟设（图 1 中显示为无纠缠）。针对每种编码方案，我们分别对这两种拟设结构进行实验。我们考虑底层量子层的不同宽度（n），以分析混合量子神经网络中的贫瘠高原。此外，对所有宽度均考虑量子层的不同深度（n），以分析对于两种拟设而言，拟设表达能力如何影响混合量子神经网络的可训练性。

我们将准确率和损失收敛作为本文所用模型的评估指标。对于具有特定 n 和 m的混合量子神经网络，能够以更快、更平稳的收敛速度达到最优解且准确率更高的模型被认为表现更好。

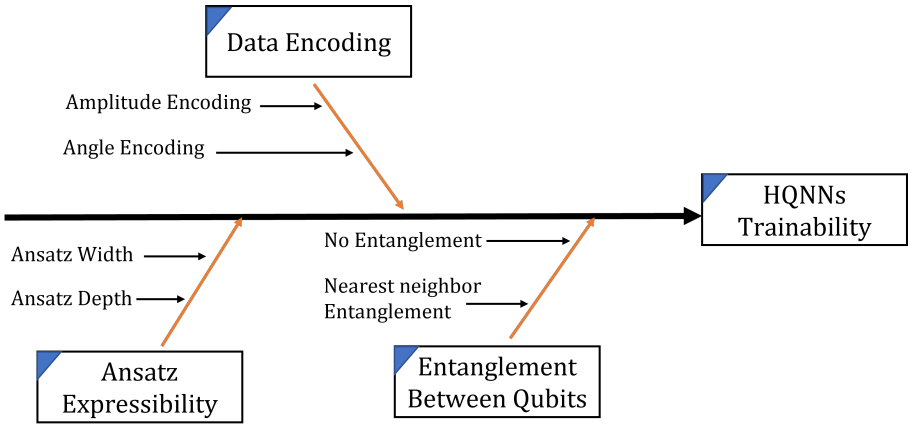


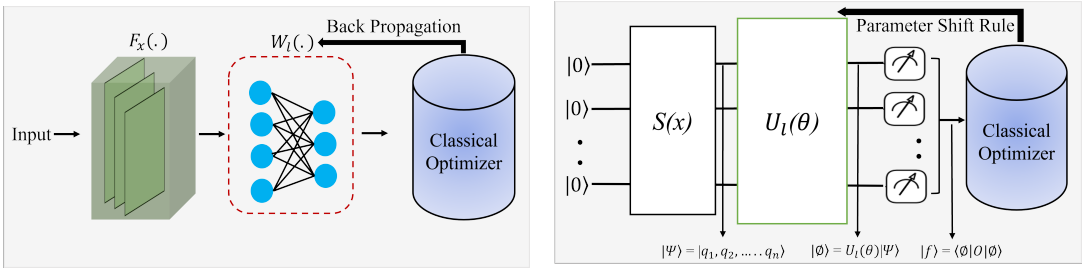
图1：所提方法概述

优于其他模型。这通常意味着这些模型尚未完全受到贫瘠高原（BP）的影响，因此能够获得更好的性能。除了准确率和损失收敛情况外，我们还评估了混合量子神经网络的一些其他性能参数。这些附加性能参数包括精确率、召回率和 F1分数。

1.4. 贡献

我们开展了一系列广泛的实验，研究数据编码、拟设表达能力以及量子比特之间的纠缠对混合量子神经网络在实际数据集上的可训练性的综合影响。我们通常使用 *sklearn* [60]中的手写数字数据集。之所以选择这个数据集，而不是 MNIST 等其他著名的类似数据集， [61], 是因为它的图像尺寸较小，更适合含噪中等规模量子设备。根据实验结果，我们观察到混合量子神经网络中也会出现贫瘠高原问题；随着量子比特数量增加，模型性能会下降。此外，无论采用何种数据编码策略，贫瘠高原的出现都取决于拟设深度（表达能力）。因此，我们首先针对纠缠拟设和无纠缠拟设，分别对两种数据编码进行可训练性和表达能力分析。这些分析有助于在实际应用场景中，确定给定宽度下量子层的合适深度。此外，所得结果也有助于判断哪种编码策略在贫瘠高原和可训练性方面略占优势。我们还观察到，量子层中量子比特之间的纠缠也会影响混合量子神经网络的可训练性。不过，纠缠对底层模型整体性能的影响（正面或负面）取决于数据的编码方式。此外，我们还采用分类应用的不同评估指标（精确率、召回率和F1分数）评估混合量子神经网络，说明其与实际应用的相关性。最后，我们通过考察混合量子神经网络主要组成部分（数据编码、拟设表达能力以及纠缠的引入/移除）的不同约束情形，阐明所提框架的重要性，并为参数的最优组合提供建议。

需要注意的是，文献中已经对各个独立性能参数（例如数据编码、拟设表达能力和纠缠）进行了数学分析 [43,54,55,59]。因此，各种性能参数的数学表述.....



（a）深度神经网络（DNN）　　（b）量子神经网络

图 2：DNN 和量子神经网络（QNN）的示意图

关于可训练性的性能参数不在本文讨论范围内。然而，针对实际应用，尚未探究混合量子神经网络的所有这些非平凡组成部分（性能参数）的综合影响。缺乏此类统一分析的一个可能原因是，很难在单一框架中同时从理论上分析所有这些概念彼此之间的影响。不过，实验研究可以为此类统一分析提供助力。因此，我们尝试从实践角度，通过实验/经验方式分析它们的联合（整体或同时）影响。

1.5. 结构安排

论文其余部分安排如下：第2 节介绍混合量子神经网络的必要背景及其主要组成部分。第3节讨论混合量子神经网络中贫瘠高原问题的潜在解决方案和分析的最新进展。第4节介绍混合量子神经网络分析框架的详细开发方法。第5节讨论实验详情，包括所执行的实验列表。第6节基于所得结果分析所提出的框架。第7 节阐述所提出框架对于混合量子神经网络各种设计约束的意义。最后，第8节总结全文。

2. 混合量子神经网络

本节将混合量子神经网络类比于经典深度神经网络进行讨论。此外，与本文所用内容相对应的混合量子神经网络主要组成部分，即数据编码、可观测量测量和成本函数的详细信息，分别在第2.1节和第2.2节中讨论。最后，拟设的可训练性和表达能力将在第2.3节中讨论。

在典型的经典深度神经网络（DNNs）设置中（图 2a），第一步是通过特征嵌入层将输入数据映射到特征空间 $F_x(.)$。随后，通过全连接神经元层对嵌入后的数据进行训练 $\prod_l W_l(.)$，以学习特定数据集输入与输出之间的内在关系。每一层的层数和神经元数量均可根据目标应用的复杂程度进行自定义。

DNNs 的量子对应形式（即 QNNs）近期因 DNNs 的巨大成功以及 QNNs 可能提供的增强计算能力而受到广泛关注。

[62]。量子神经网络的结构与深度神经网络相似，如图 2b 所示。与其他量子机器学习算法类似，量子神经网络也利用参数化量子电路，即依赖于参数（可通过经典方法优化）的量子电路。将这两种架构相结合即可得到混合量子神经网络。

混合量子神经网络分五个步骤运行 [63]：**(1)** 输入降维：混合量子神经网络运行的第一步是进行输入降维，以应对含噪中等规模量子设备的局限，主要是量子比特数量方面的限制。**(2)** 特征映射：经典数据点 x 被映射为 $n-$量子比特量子态 $|\psi\rangle$ ，由方程1表示，其中 $S(x)$为映射函数。**(3)** 训练：通过拟设 U对制备好的量子态进行处理和训练，如方程2所示，该拟设由一系列单量子比特和多量子比特幺正算子构成。拟设中的旋转角由向量 θ参数化。**(4)** 测量：随后测量量子态，并获得相应的测量可观测量本征值 O ，如方程3所示。**(5)** 对可观测量进行经典后处理：量子态测量会产生经典值，因此可以由经典设备处理。测量结果通常通过经典神经元层进行后处理。这也允许应用熟悉的非线性激活函数（如 SoftMax），并使用优化程序最小化成本函数。上述步骤$2-4$ 构成了量子神经网络的典型架构，如图2b所示。因此，在典型的混合量子神经网络架构中，量子神经网络被完整复制，使得对混合量子神经网络中量子部分（上述步骤 $2-5$ ）的分析同样适用于量子神经网络。

$$|\psi(x)\rangle = S(x)\,|0\rangle^{\otimes n} \tag{1}$$

$$|\phi(x,\theta)\rangle = U(\theta)\,|\psi(x)\rangle \tag{2}$$

$$f(x,\theta) = \langle\phi(x,\theta)|\,O\,|\phi(x,\theta)\rangle \tag{3}$$

一般而言，在量子机器学习中，由于量子计算的概率性，对于输入 x 和参数向量 θ ，底层参数化量子电路会被迭代执行，以近似其期望值。在量子神经网络中，该期望值通常被视为网络的输出 [63]。在接下来的小节中，我们将结合本文采用的方法，简要概述混合量子神经网络的不同组成部分。

2.1. 数据编码

在量子神经网络中，如何以量子电路能够处理的方式输入数据一直是一项紧迫的挑战，这通常称为数据编码。数据编码可以视为一种特征映射，将输入特征 x 映射到量子系统的希尔伯特空间，从而创建量子态 $|\psi_x\rangle$ [64]。编码过程是设计量子算法时的关键步骤，并且会显著影响其性能[27,64–66]。在实践中，这种变换（$x \longrightarrow |\psi_x\rangle$）通常通过幺正变换($S_x$)实现，并由一个变分量子电路执行，其参数取决于所编码的输入数据 [27,64,67]。

量子电路（S_x）随后作用于初始态 $|\psi\rangle$，该状态通常为基态，即$|\psi\rangle = 0^{\otimes n}$。随后按照方程 4 实现编码。

$$x \mapsto E(x) = S_x |\phi\rangle \langle\phi| S_x^\dagger = |x\rangle \langle x| =: \rho_x \tag{4}$$

然而，变换量子电路 S_x 必须具备硬件高效性，以适应 NISQ 时代所带来的限制。近期已有多种编码策略被提出 [64,66,67]。然而，在混合量子神经网络中，常用的编码策略是振幅编码和角度编码。

2.1.1. 振幅编码

在振幅编码中，数据被编码为量子态振幅。对于 $x \in \mathbb{R}^n$，振幅编码将 $x \longrightarrow E(x)$ 映射为一个 $n-$量子比特量子态的振幅，如下方方程所示：

$$|\psi_x\rangle = \sum_{i=1}^{N} x_i |i\rangle \tag{5}$$

其中，$N = 2^n$是 x 的 i^{th} 元素，$|i\rangle$ 是 i^{th} 计算基态。对于具有 M 个样本和 N 个特征的经典数据集 $\mathcal{D}$ ，如下方程所示：

$$\mathcal{D} = \{x^{(1)}, \ldots x^{(m)}, \ldots x^{(M)}\} \tag{6}$$

其中，$x^{(m)}$是 $m = 1$的 $N-$维特征向量，$\ldots M$。随后，可以通过将 $x^{(m)}$ 拼接成一个向量来轻松理解振幅，如下所示：

$$\alpha = \mathcal{C}_{norm} = \{x_1^{(1)}, \ldots, x_N^{(1)}, x_1^{(2)}, \ldots, x_N^{(2)}, x_1^{(M)}, \ldots, x_N^{(M)}\} \tag{7}$$

因子 $\mathcal{C}_{norm}$ 是归一化因子，必须进行归一化，使得$|\alpha|^2 = 1$。然后，可以使用以下方程在计算基中表示输入：

$$|\mathcal{D}\rangle = \sum_{i=1}^{2^n} \alpha_i |i\rangle \tag{8}$$

其中，α_i 表示振幅向量元素（α），$|i\rangle$ 为计算基态。编码的振幅总数为 $N \times M$。一个由 $n-$个量子比特组成的系统可以编码 2^n 个特征。因此，振幅编码需要$n \geq log_2（NM）$。如果要编码的特征总数大于 2^n [65]，则可以对常数进行填充。

2.1.2. 角度编码

角度编码，也称为量子比特编码，会将输入数据特征编码为量子比特的旋转角度，并已用于各种量子机器学习算法 [65,68,69]。对于一个

特征向量 $x = [x_1, x_2, \ldots x_N]^T \in \mathcal{X}^N$，以下方程通常表示角度编码。

$$|x\rangle = \bigotimes_{i=1}^{N} cos(x_i)\,|0\rangle + sin(x_i)\,|1\rangle \tag{9}$$

角度编码将 N 特征编码到一个 $n-$量子比特系统中。量子比特编码中每个量子比特对应一个特征的态制备幺正算子可由以下方程表示:

$$S_{x_j} = \bigotimes_{i=1}^{N} U_i \;\; where \;\; U_i := \begin{bmatrix} \cos(x_j^{(i)}) & -\sin(x_j^{(i)}) \\ \sin(x_j^{(i)}) & \cos(x_j^{(i)}) \end{bmatrix} \tag{10}$$

角度编码方法也可以推广为在单个量子比特中编码两个特征；这称为密集量子比特编码 [67]。不过，在本研究中，我们采用简单的量子比特编码。

2.2. 可观测量测量与成本函数

量子比特可以在不同的测量基中进行测量。在本研究中，我们使用 σ^z 的本征基来计算参数化量子电路的期望值。对于一个 $n-$量子比特系统，在 σ^z 基中的可观测量测量可描述为 n个泡利 Z 矩阵的张量积，即 $O = \sigma^{\otimes n}$。该 σ^z 可观测量对于奇数量子态返回 -1，对于偶数量子态返回 1，从而使参数化量子电路的整体期望值保持在范围$[-1, 1]$内。

混合量子神经网络的量子层训练需要找到参数向量θ，使其在每次训练迭代后都能最小化损失。由于我们考虑的是一个多类分类问题，因此采用稀疏分类交叉熵作为成本函数，如下方程所示:

$$Cost = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \tag{11}$$

其中，y_i 为真实标签，$\hat{}y_i$ 为预测值。经典优化技术（如梯度下降）可用于成本函数的优化；该方法只需计算每个参数的偏导数，并确定下一步的最小值方向。然而，PQC 的输出，即测量可观测量的期望值，需要针对变分拟设中的每个参数分别求导。根据参数移位规则（方程 12），可以分别对 PQC 的期望值关于每个参数求导。该规则最初由 [31] 针对量子机器学习算法提出，并由 [70]扩展。

$$\frac{df}{d\theta_i} = \frac{f(\theta_i + s) - f(\theta_i - s)}{2} \tag{12}$$

其中， s 为宏观位移，由相应门的特征值决定，而该门由 θ_i 参数化。

2.3. 拟设的可训练性与表达能力

拟设可以被视为由单量子比特参数化门组成的参数化量子电路。根据目标问题的不同，它可能包含也可能不包含多量子比特门（纠缠门）。这些参数化门取决于可调参数。在混合量子神经网络（或一般的量子机器学习）中，这些参数会在数据驱动任务中进行训练，这与经典神经网络的情况类似。只要优化算法能够在每次训练迭代中最小化损失（根据定义的成本函数），就可以说该网络是可训练的。当参数的梯度无法获取以进行进一步优化，且优化算法无法达到最优解时，网络的可训练性便会受到影响。

如果能够利用 PQC 均匀地探索酉群，那么就可以认为它具有表达能力 $\mathcal{U}(d)$。因此，可以用以下超算子来定义 PQC 的表达能力 [54]。

$$\mathcal{A}_{\mathbb{U}}{}^{(t)}(.) := \int_{\mathcal{U}(d)} d\mu(V) V^{\otimes t}(.)(V^{\dagger})^{\otimes t} - \int_{\mathbb{U}} dU U^{\otimes t}(.)(U^{\dagger})^{\otimes t} \tag{13}$$

其中， $d\mu(V)$是哈尔测度的体积元素， dU 则是对应于 $\mathbb{U}$上均匀分布的体积元素。如果对于所有算子 X，$\mathcal{A}_{\mathbb{U}}{}^{(t)}(X) = 0$ ，那么对 $\mathbb{U}$ 中各元素取平均，与对哈尔分布取平均在 t阶矩上相一致。在这种情况下， $\mathbb{U}$ 构成一个t-设计。

3. 相关工作

本节将详细介绍近期最先进的解决方案（主要受经典神经网络中的BP问题启发），这些方案旨在克服量子神经网络中的BP问题，以提升其可训练性。此外，我们还将探讨量子神经网络训练景观中BP的潜在来源，并从底层拟设的角度分析BP问题。

量子神经网络中的BP现象最早在 [14], 中得到研究：研究者对随机参数化量子电路进行初始化，随后计算偏导数的方差，即$var[\partial C] = \langle(\partial C^2)\rangle - \langle(\partial C)\rangle^2$。作者随后表明，对于阶数为 $poly(n)$的深层电路，由于该电路构成2-设计（对希尔伯特空间中的所有酉算子进行采样），方差会随量子比特数增加而呈指数级衰减，即$var[\partial C] = \frac{1}{2^n}$。

不同于经典情形，在经典情形中，基于梯度的反向传播算法可以解决这一可训练性问题，而其量子类比方案则难以实现[71]。在混合量子神经网络中，参数化量子电路在量子设备上运行，而其优化则在经典计算机上进行。两种本质不同的计算方法的结合，使得在量子神经网络中实现反向传播算法颇具挑战。近年来，量子神经网络中的贫瘠高原问题受到了广泛关注。一些研究提出了可能克服贫瘠高原问题的解决方案，另一些研究则分析量子神经网络的基本结构，以确定导致贫瘠高原问题的参数。在接下来的章节中，我们将介绍一些近期旨在解决和分析量子神经网络中贫瘠高原问题的最先进研究成果。

3.1. BP 的潜在解决方案

近期提出了若干可能克服量子神经网络中贫瘠高原问题的解决方案。在 [68],量子电路的一小部分被随机初始化，而其余参数则经过精心选择，以使整体实现恒等操作。这种方法仅能避免在第一次训练步骤中初始化于高原区域，而在后续训练迭代中，可学习性仍会受到影响。类似地，在 [72], 有研究提出了一种通过强制将量子电路中的多个参数设为相同值来减少需要训练的参数总数，从而避免贫瘠高原问题的策略。然而，这种方法将优化过程限制在特定的参数集合上，并最终增加了达到收敛所需的电路深度。

受到经典神经网络逐层训练方法的启发——研究表明，该方法有望避免经典神经网络中由随机初始化导致的 BP [73], ，近期也有研究将类似方法用于量子神经网络 [71]。量子神经网络中的逐层训练方法会逐步增加量子电路深度，并在每次训练迭代中仅训练一小部分参数。由于与训练完整量子电路相比，参与训练的参数数量较少，因此梯度幅值更大。

近期， [74],提出了一种缓解量子神经网络中 BP 的新方法：将经典神经网络中的残差方法应用于量子神经网络。研究表明，在量子神经网络中引入残差方法能显著提升其训练表现。

由于 BP 本质上是梯度消失问题，无梯度优化方法似乎能够帮助解决这一问题。然而，近期已有研究证明，即使采用无梯度优化，也无法摆脱量子神经网络中的 BP 问题 [75]。事实上，研究表明，在 BP 情况下，代价函数差异（这是无梯度优化方法中决定优化决策的因素）会受到指数级抑制。

3.2. 贫瘠高原分析

贫瘠高原是与代价函数景观相关的问题，其中参数的偏导数会随着系统规模增大而呈指数级变平。如前所述，初步研究探讨了深层拟设中贫瘠高原的出现，而人们通常认为深层拟设具有更强的表达能力。

一项近期研究探讨了浅层电路与深层电路中贫瘠高原（BP）的存在情况，详见 [76]。作者表明，通过使贫瘠高原依赖于成本函数，可以将其概念扩展至浅层电路。为此，他们研究了两种成本函数，即：局部代价函数（在多量子比特系统中测量单个量子比特）和全局代价函数（在多量子比特系统中测量所有量子比特）。随后，作者得出结论：对于全局代价函数，无论底层参数化量子电路（PQC）的深度如何，量子神经网络（QNNs）都会经历贫瘠高原（$\mathcal{O}(1)$ $\mathcal{O}(log(n))$ $\mathcal{O}(poly(log(n)))$ $\mathcal{O}(poly(n))$）。另一方面，对于局部代价函数，梯度最坏情况下仅呈多项式级消失，因此在深度达到 ($\mathcal{O}(log(n))$) 量级之前仍然可以训练。当深度达到 ($\mathcal{O}(poly(n))$) 量级时，局部代价函数开始出现贫瘠高原，而在这两个区域之间，存在一个过渡区域，梯度在此区域内从多项式衰减转变为指数衰减。在 [77], 最近开展的一项研究中，研究人员进一步考察了现实应用中成本函数的全局性与局部性。该研究指出，在多分类任务中，与局部代价函数相比，使用全局代价函数能够显著提升性能。

成本函数。然而，对于二分类，全局和局部代价函数都表现出相似的效果。这些发现表明，在为多类别场景设计分类模型时，必须仔细考虑成本函数的选择；而在二分类的情况下，任一种方法都同样可行。

近期研究 [54] 分析了拟设表达能力与梯度幅度之间的关系，并表明拟设的表达能力越强，越有可能出现贫瘠高原。另一方面，表达能力相对较低的拟设可以将贫瘠高原推迟到一定深度，但原则上，表达能力强的拟设仍然更可取，因为与针对特定问题的表达能力较弱的拟设相比，它们可能为多个问题提供解决方案。因此，深层拟设具有更强的表达能力，也就更容易在训练过程中遭遇贫瘠高原。

量子神经网络隐藏层中的过度纠缠也会导致贫瘠高原，并阻碍学习过程 [55]。利用 [55]，中量子热力学的体积定律，作者证明了在基于梯度和无梯度优化方法下，代价函数景观中均存在贫瘠高原。这一观察结果适用于前馈量子神经网络和量子玻尔兹曼机。此外，数据编码到量子神经网络中的方式也可能导致贫瘠高原 [43]。

本文并不旨在为贫瘠高原问题提出解决方案。相反，我们重点按照第3.2节的内容，对混合量子神经网络的各个组成部分进行分析。

4. 方法论

本节详细介绍了如何获得所提出分析的相关结果。我们首先介绍用于训练混合量子神经网络的数据集详情，以及如何对其进行预处理，以应对含噪中等规模量子设备的局限。随后，我们将介绍量子神经网络的构建细节，从量子比特初始化到最终测量。最后，我们将介绍量子神经网络结果的经典后处理细节，以此结束对方法的讨论。

本文用于统一分析数据编码、拟设表达能力和纠缠对可训练性的联合影响的混合量子神经网络典型架构如图3所示。该混合量子神经网络架构完全复现了量子神经网络的

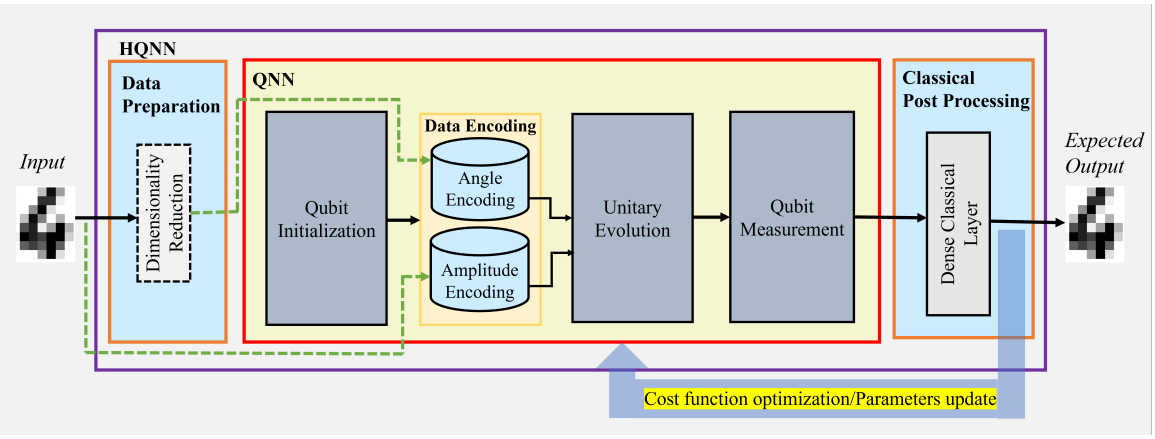


图 3：用于混合量子神经网络统一分析的拟议方法论

图中的架构 2b。我们将该架构（本文所用）称为混合型

由于两个原因；1）优化是在经典计算机上执行的，这在通常的独立量子神经网络中也是如此（处于 NISQ 时代）。其次，我们在量子层的末尾添加了一个经典神经元层。经典层的优点是可以应用传统机器学习中熟悉的非线性激活函数。此外，它还支持在真实世界数据集上进行实验。本文使用的混合量子神经网络包含三个主要组成部分，即：数据准备、量子神经网络构建和经典后处理。下面介绍本文所用混合量子神经网络的典型工作流程，具体说明从输入到输出的每个步骤。

4.1. 数据准备

本文所用混合量子神经网络的工作流程中，第一步是准备数据，使其能够高效地编码到量子态中，以便进行后续处理。此处的数据准备需考虑含噪中等规模量子设备的限制（主要是量子比特数量）。根据输入数据的大小和编码方案的选择，有时需要降低输入特征的维度。在这种情况下，输入（在本文中为图像）首先会经过降维算法处理，然后再进行编码。我们还在其中一种编码方案中使用降维，以应对含噪中等规模量子时代的限制；相关细节将在下一节介绍，届时我们将说明本文在量子神经网络构建中使用的数据编码。

4.2. 量子神经网络构建

数据准备完成后，混合量子神经网络工作流程中的下一项重要步骤是构建量子神经网络，其中包含四个要素：量子比特初始化、数据编码、幺正演化和量子比特测量。

4.2.1. 量子比特初始化

这是量子神经网络构建的第一步，通常用于定义量子比特的总数，换言之，也就是量子层的宽度。量子比特可以初始化为任意随机状态。不过，我们将所有量子比特初始化为默认基态，即 $|\psi\rangle^{\otimes n} = |0\rangle^{\otimes n}$，这是一种较为常见的做法。

4.2.2. 数据编码

定义好量子比特后，下一步是进行数据编码（经典到量子的特征映射），以便后续量子层 [2] 能够处理输入数据。我们使用量子神经网络中两种常用的数据编码策略，即振幅编码和角度编码。如上所述，含噪中等规模量子设备中的量子比特数量有限，因此必须减少输入特征维度，而这取决于输入特征的大小以及数据被编码到量子态中的方式。我们所使用的数据集的输入特征维度为 64（详见第 5节）。结合本文采用的数据编码方式，每当使用角度编码时，我们都会降低输入特征维度，因为角度编码需要 n 个量子比特来编码 n 个特征（参见第 2.1节）。

[2]本文中，我们交替使用量子层、参数化量子电路和拟设这几个术语。

对于振幅编码，不会对输入特征进行降维，因为振幅编码可以将 2^n 个特征编码到 n 个量子比特中（见第2.1节）。

4.2.3. 幺正演化

幺正演化主要由单量子比特参数化（依赖参数）门和多量子比特门（用于量子比特纠缠）组成。对于幺正演化中门的组合，并没有严格固定的规则；针对给定问题，通常会启发式地选用不同的门组合。我们使用两种相似的拟设结构（纠缠拟设和无纠缠拟设）来实现幺正演化，两者的区别仅在于是否引入纠缠。本文所用量子层从数据编码、幺正演化到测量的完整细节如下：

4.2.3.1. 纠缠拟设**..** 纠缠拟设结构由单量子比特参数化幺正门（$R_y(\theta)$）以及用于量子比特纠缠的双量子比特幺正门（受控非门）组成。考虑到含噪中等规模量子设备的局限性，本文仅考虑最近邻纠缠，即底层系统中的最后一个量子比特被视为与第一个量子比特相邻，如图4所示。

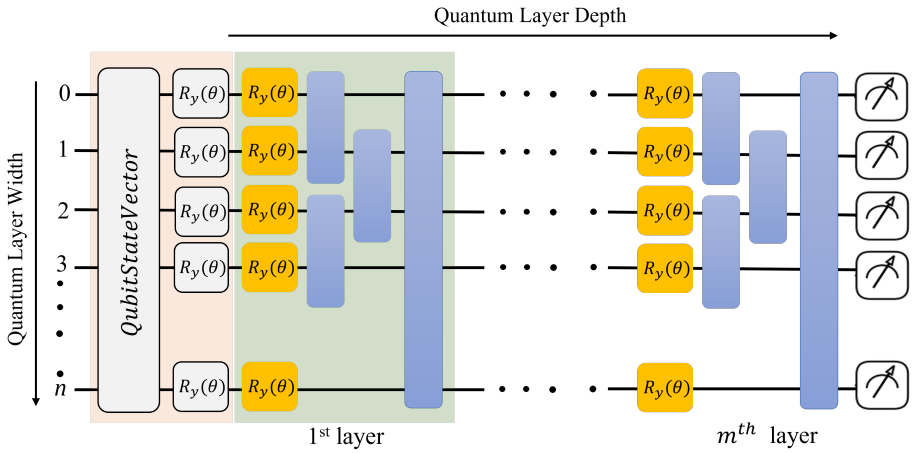


图4：带有纠缠的量子层拟设结构。浅粉色高亮区域为数据编码部分，具体使用量子比特态矢量还是 $R_y(\theta)$ 旋转门，取决于所采用的编码技术。振幅编码情况下使用前者，角度编码情况下使用后者。绿色阴影区域为训练中实际使用的量子拟设。黄色方框中的参数化旋转酉算子是可训练量子参数，蓝色竖条表示双量子比特酉算子

4.2.3.2. 无纠缠拟设。 纠缠是量子力学中的一项重要性质，并有望增强包括量子机器学习在内的各种应用。现在，我们更新了拟设结构，其中仅包含无纠缠的单量子比特参数化酉变换，如图 5 所示。之所以排除纠缠，是因为在量子神经网络中，可训练参数仅为单量子比特酉变换，这些参数会在训练过程中得到优化。

我们试图回答这样一个问题：无纠缠拟设是否有可能延缓贫瘠高原问题，从而提高混合量子神经网络的可训练性，因为纠缠本身也可能被视为贫瘠高原的潜在来源？我们展示了纠缠在混合量子神经网络中是否重要，取决于数据编码的类型。对于角度编码，纠缠并不能带来更好的性能；事实上，在完全不存在纠缠的情况下，整体性能反而有所提升。另一方面，对于振幅编码，结果则显示出相反的情况。

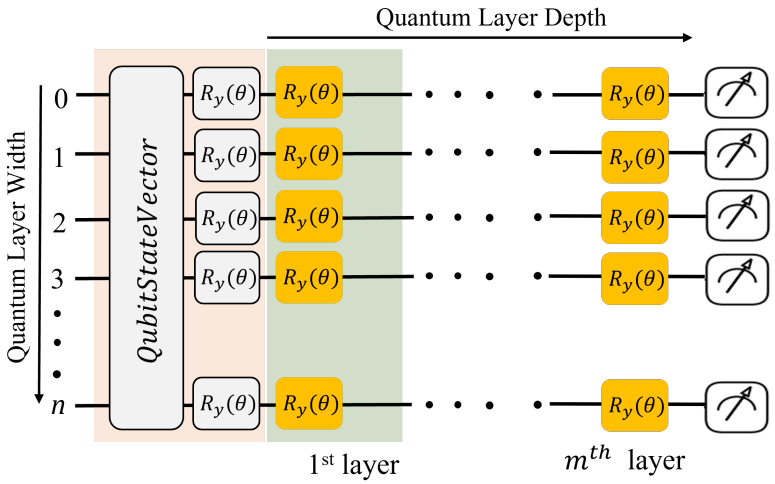


图5：无纠缠的量子层结构。浅粉色和灰色高亮区域表示与图4

4.2.4. 量子比特测量

对于量子比特测量，可以使用不同的测量基；然而，最常用的是默认的计算基，即 σ^z 的本征基。我们也使用计算基来获取量子神经网络的输出。

4.3. 经典后处理

混合量子神经网络（HQNN）工作流程的最后一步是对量子神经网络（QNN）进行经典后处理，因为量子比特测量的结果是经典值，而非量子态。为了对 QNN 的结果进行后处理，我们使用一个经典全连接神经元层（每个神经元都与参数化量子电路（PQC）中所有量子比特的测量输出相连接）。随后，根据给定问题定义成本函数，优化器以此为依据寻找问题的解。在每次训练迭代中，优化器都会更新可训练参数，并使用更新后的参数重新训练 PQC。重复这一过程，直到找到最优解。

在本文中，混合量子神经网络的经典配置（输入、末端带有非线性函数的神经元层，以及经典优化）保持不变，主要关注对封装其中的量子神经网络（量子层）进行实验。至此，本文所用混合量子神经网络的所有构成要素均已介绍。下一节将介绍实验细节，包括混合量子神经网络中使用的经典组件。

5. 实验设置

本文所用混合量子神经网络的架构细节已在上一节（第4节）中介绍。本节将说明所提出的混合量子神经网络架构的训练与评估实验细节。具体而言，还将介绍训练所用数据集、混合量子神经网络的经典超参数，以及为获得所提分析结果而开展的实验列表。

5.1. 数据预处理

我们考虑一个多类别（图像）分类问题。我们使用 sklearn 手写数字数据集来训练和评估混合量子神经网络，其中包含手写数字的图像 [78]。与 MNIST 等其他类似但同样流行的数据集相比，选择该特定数据集的原因在于其较小的输入特征规模，更适合含噪中等规模量子时代。该数据集共包含 10 个类别，每个类别有 180 个样本，总计约 $\cong 1797$ 个样本。此外，每个数据点都是一幅 8×8 的图像，因此特征维度为 64。此外，75% 的样本用于训练，其余 25% 用于测试混合量子神经网络。在每个实验中，数据分别采用振幅编码和角度编码进行编码。如第 4 节所述，在将输入传递给参数化量子电路之前，我们会缩减输入特征维度；对于角度编码，输入特征必须等于量子比特的总数。对于输入维度缩减，我们使用 sklearn 库（用于 Python 编程语言的常用机器学习库）中的主成分分析（PCA），该方法可以在不丢失数据集中最重要信息的情况下，将大型数据集缩减为较小的数据集。

5.2. 超参数规格

鉴于目标问题（多分类）的性质，我们使用 分类交叉熵 作为成本函数。对于成本函数的经典优化，我们使用*Adam* 优化器 [79]，初始学习率为 0。01。由于我们使用的数据集总共包含 10 个类别，因此混合量子神经网络中的最后一个经典神经元层共有 10 个神经元。此外，使用的非线性激活函数是 $SoftMax$，这是多分类问题中的常见选择。

本文所用的模型设置为最多训练 100 次训练迭代（训练轮次）。然而，为避免过拟合，我们使用 Keras 库（一种提供 Python 接口、用于人工神经网络的开源软件库）中的早停法，在训练过程中调整学习率。早停法会连续三次迭代监测验证损失；如果没有改善，学习率便会降低至原来的 0.1 倍，并将新的学习率设为：$newlearningrate = previouslearningrate \times 0.1$。如果验证损失连续四次训练迭代没有改善，早停法会强制终止训练，以避免过拟合。所有实验的输入训练数据均以 16 个样本为一批传入。最后，我们使用 pennylane（一种用于量子计算机可微编程的跨平台 Python 库）[6]来训练混合量子神经网络。

5.3. 实验列表

我们提出的分析所用的训练实验列表如图 6所示。由于量子层的结构和规模是我们分析的重点，我们对量子层采用了不同的宽度（n）和深度（m）进行实验。 n 表示量子比特的数量，而 m 表示测量前量子层的周期性重复次数。我们将最大宽度限制为 $n = 14$ ，最大深度限制为 $m = 10$ ，原因有二：1）当 n 和 m 增大时，总体准确率开始下降；我们推测，进一步增大n 和 m 会因所谓的贫瘠高原现象而导致准确率进一步下降；2）由于我们使用经典机器模拟量子比特系统，即使是本文所考虑的简单数据集，采用最大宽度（ $70 - 80$ ）和最大深度（$n = 14$）进行训练也需要大约 $m - 10$小时。

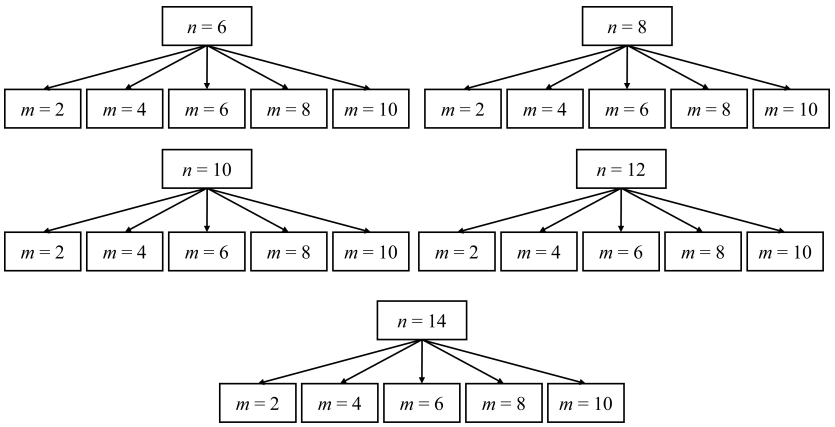


图6：本研究针对不同的 n 和 m所开展的实验列表。

鉴于在具有足够随机性和表达能力的参数化量子电路中，独立量子神经网络存在贫瘠高原现象 [14], 我们首先通过实证分析混合量子神经网络中是否存在贫瘠高原现象。我们通常分析添加经典神经元层在多大程度上能够缓解贫瘠高原现象。为此，我们针对底层参数化量子电路的不同宽度（n）和深度（m）训练混合量子神经网络。较大的 n 和 m 意味着更多的可训练参数，通常也被认为具有更强的拟设表达能力，反之亦然。改变参数化量子电路的 n 和 m 有助于分析混合量子神经网络中贫瘠高原现象的存在情况。我们分别针对两种编码方式以及两种拟设结构（纠缠和非纠缠）分析贫瘠高原问题。这有助于全面了解哪种编码策略与哪种拟设结构搭配效果更好。在观察到贫瘠高原现象对拟设深度（表达能力）的依赖关系后，我们进一步开展可训练性与表达能力之间的关系分析，同样针对每种拟设结构分别考察两种编码方式。这项分析全面揭示了拟设表达能力在贫瘠高原现象出现过程中所起的作用，以及在给定宽度下，拟设可以达到多深才会开始出现性能下降。此外，这也有助于确定 n和 m 之间是否存在某种关系，以实现相对更好的整体性能。随后，我们比较两种编码方式下两种拟设结构的性能，以了解纠缠在混合量子神经网络中的作用。最后，我们通过简要展示混合量子神经网络在分类任务中的应用，将其与其他重要的评估指标相结合，即精确率、召回率和F1分数，从而总结本研究的分析。

6. 结果与讨论

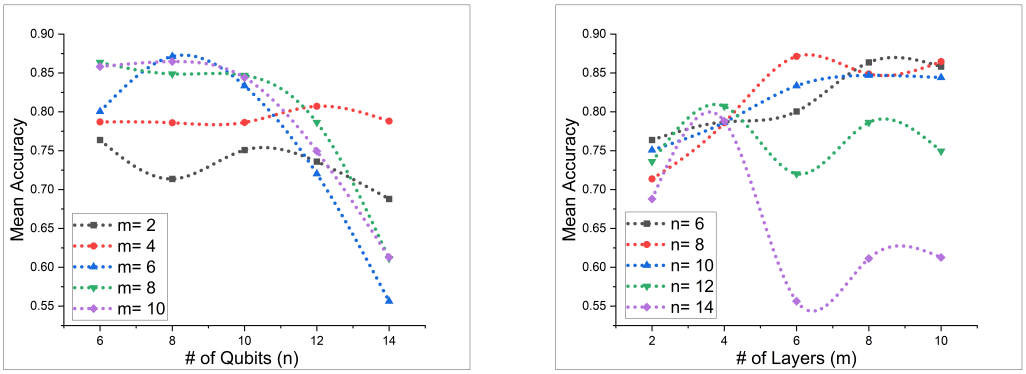
在本节中，我们展示实验分析。首先，我们分别针对两种拟设结构，使用本文采用的两种编码方式，证明混合量子神经网络中存在贫瘠高原。根据分析，我们认同贫瘠高原取决于混合量子神经网络中量子层的表达能力。随后，我们对混合量子神经网络进行可训练性与表达能力分析，即探究量子层的深度和宽度如何与贫瘠高原相关，以及可能如何影响网络的整体性能。可训练性与表达能力分析表明，底层拟设中的纠缠也会影响混合量子神经网络的整体性能。接着，我们简要比较本文采用的两种拟设结构和两种编码方式，以了解纠缠的作用。最后，为了使分析更加全面，我们根据其他面向应用的分类任务评估指标对混合量子神经网络进行评估。

6.1. 混合量子神经网络中贫瘠高原存在性的演示

在本节中，我们分析混合量子神经网络（HQNN）的训练景观中是否存在贫瘠高原（BP）。我们以混合量子神经网络的平均准确率作为分析贫瘠高原的基准。针对纠缠拟设结构和无纠缠拟设结构，分别对两种编码方式进行独立的贫瘠高原分析。贫瘠高原的存在会阻碍混合量子神经网络的学习过程，最终导致准确率降低。因此，在我们的分析中，较高的准确率本质上意味着模型尚未完全受到贫瘠高原的影响，反之亦然。

6.1.1. 振幅编码纠缠拟设的贫瘠高原演示

纠缠拟设包含单量子比特酉变换和最近邻量子比特纠缠，如图4所示。要编码为量子态的原始输入特征维度为 64，如第5节所述。因此，采用振幅编码时，编码这些特征所需的最少量子比特数为 $n = 6$；考虑到含噪中等规模量子时代，这对于量子层来说是一个合理的宽度。因此，使用振幅编码时，我们不采用主成分分析进行特征降维，而是直接编码所有输入特征。随后，我们根据图 n 和 m 中的设置调整6，以分析贫瘠高原的出现情况。所有实验的平均准确率如图7所示。



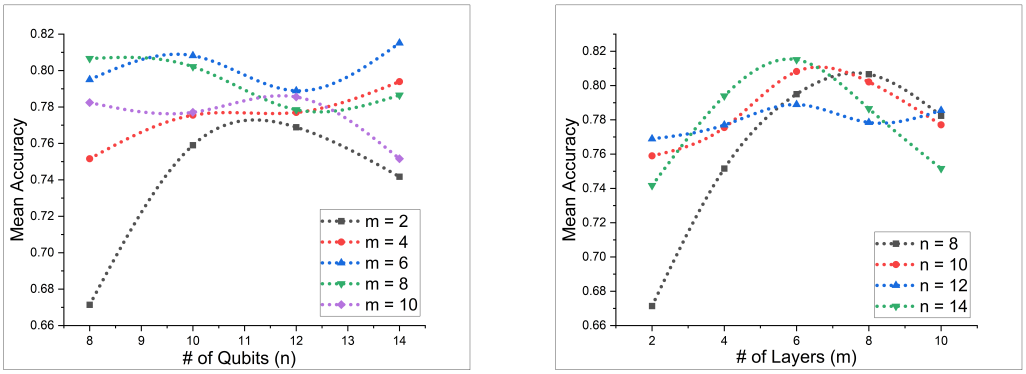
(a) 平均准确率随 n变化的函数依赖关系　　(b) 平均准确率随 m的变化函数

图 7：纠缠拟设：采用振幅嵌入时，不同 m 和 n 的平均准确率。

为分析宽度（n）而不考虑深度（m）的影响，图7a分别绘出了固定 m、改变 n 时不同训练实验的平均准确率。图7a中的总体准确率趋势呈下降态势。随着量子比特数（n）增加，性能下降，这表明训练景观中可能存在贫瘠高原。另一方面，为了观察深度（m）而不考虑宽度（n）的影响，图7b分别绘出了固定 n 、改变 m时的平均准确率。我们观察到，量子层的深度和宽度之间存在一种权衡，需在两者之间取舍以获得相对较好的性能。这里的权衡本质上是指，较小的 n 允许相对较大的 m，反之亦然。这种权衡的一种可能原因是过参数化现象；在本文使用的混合量子神经网络中，过参数化是 n 和 m增加的结果。过参数化对于经典深度学习至关重要，因为现代应用十分复杂，而且它通常有助于学习输入数据中最复杂的关系。然而，混合量子神经网络并不适合过度参数化网络，必须确定最优参数数量（最优 n 和 m），才能获得相对较好的性能。

6.1.2. 角度编码下纠缠拟设的贫瘠高原演示

在角度编码中，若要成功进行经典到量子的特征映射，量子比特数量或量子层的深度（n）必须与输入特征维度相等，如第 2.1 节所述。由于我们的数据输入特征大小为 64，因此需要至少 64 个量子比特才能将数据直接编码为量子态。如此多的量子比特不适用于含噪中等规模量子时代。因此，我们采用主成分分析（PCA）降低输入特征维度，以实验角度编码方法。降维后的维度与图 n 中量子层的宽度（n）相符6。不过，考虑到含噪中等规模量子设备的限制，以及降维时尽量避免信息损失，角度编码的最小宽度为 $n = 8$。与振幅编码类似，我们针对不同的 n 和 m进行了多项实验。所有实验的平均准确率如图 8所示。



(a) 平均准确率作为 n 的函数

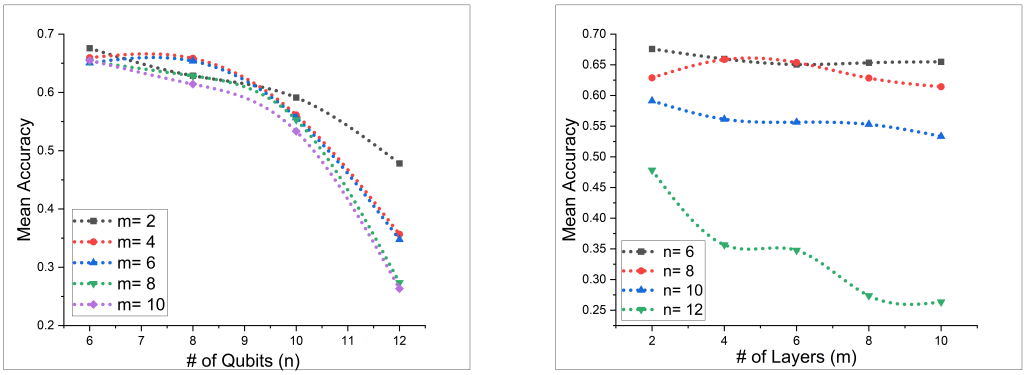
(b) 平均准确率随 m的变化

Figu图 8：纠缠拟设：使用角度编码时不同 m 和 n 下的平均准确率 g.

在经典机器学习中，人们通常推测，在底层模型具有足够复杂度的前提下，增加用于训练模型的输入特征能够提升整体性能。在角度编码中，输入特征的大小与 n的大小成正比。这通常意味着，每增加一个 n ，输入特征维度就会增大。因此，按照传统机器学习的经验，较宽的量子层应能带来更好的性能，因为与宽度较小的量子层相比，它们接收的输入特征维度更高。然而，混合量子神经网络并非如此：如图8所示，增加 n 会使准确率在达到某个 m后下降。为分析量子层宽度的影响，我们绘制了平均准确率随 n变化的曲线，如图 8a所示。可以观察到，随着 n 增加，准确率趋于下降，尤其是在深度较大时（$m = 8$ 和 10）。深度较小时（$m = 4$、6），随着 n 增加，准确率提升微乎其微，表明存在贫瘠高原。为分析量子层深度的影响，我们绘制了平均准确率随 m变化的曲线，如图 8b所示。一般而言，对于某个 n，准确率会提升至某个 m，随后开始下降。因此，在角度编码的情况下，也存在 权衡，需要在 n 和 m 之间取得平衡，才能获得更好的整体性能。

6.1.3. 振幅编码下无纠缠拟设的贫瘠高原演示

对于振幅编码，输入特征维度保持不变，即 64。我们从量子层的最小所需宽度（$n = 6$）开始，并完成图 6中的所有实验。为简化实验，我们跳过了针对 $n = 14$的实验，因为我们已经观察到贫瘠高原（BP）从 $n = 6$ 变为 $n = 12$。所有训练实验的平均准确率如图 9所示。



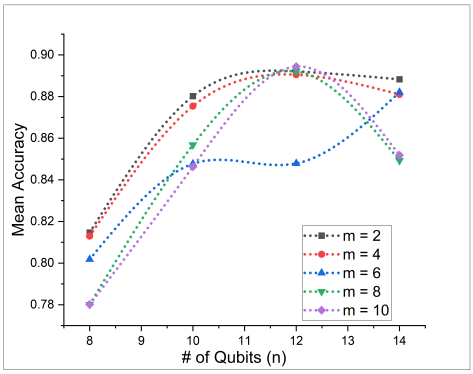
(a) 平均准确率随 n的变化　　(b) 平均准确率随 m的变化

图 9：无纠缠拟设：采用振幅编码时，不同 n 和 m 下的平均准确率。

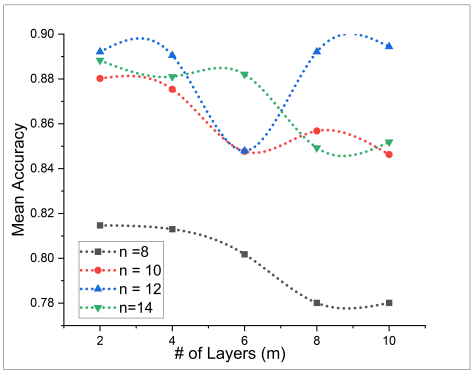
所有训练实验的平均准确率绘制为关于 n 在 9a中的函数，以分析量子比特数量的增加如何影响量子层中不包含纠缠时混合量子神经网络的性能。可以观察到，与纠缠拟设的情况类似，无论 n 如何增加，准确率都会开始下降，而不受 m的影响。准确率下降清楚表明了贫瘠高原的出现。此外，在无纠缠拟设中，随着 n 的增加，准确率下降十分明显，从而得出结论：量子层中不存在纠缠会使混合量子神经网络更容易受到贫瘠高原的影响。类似地，所有训练实验的平均准确率也绘制为关于 m （图 9b）的函数，这有助于分析 m 在固定 n的混合量子神经网络中的作用。我们观察到，与振幅编码下纠缠拟设的情况不同，在该情况下，允许的电路深度 m 会随不同的 n 而变化；而在无纠缠拟设的情况下，增加 m 会导致所有 n的准确率进一步降低（或仅有可忽略的提升）。随着 m 增加，性能下降在更大的 n下更加明显，由此得出结论：在无纠缠拟设的情况下，构建量子层时，较小的n 搭配较小的 m 更为合适。较大的 n 和m 会使无纠缠量子层更容易受到贫瘠高原的影响。

6.1.4. 采用角度编码的无纠缠拟设的贫瘠高原演示

与采用角度编码的纠缠拟设的贫瘠高原分析类似，输入特征维度经过缩减，以匹配量子层的合理宽度。我们从最小宽度 $n = 8$开始。所有实验的平均准确率如图10所示。



(a) 平均准确率随 n的变化



(b) 平均准确率随 m的变化

图10：无纠缠拟设：角度编码的平均准确率。

从贫瘠高原的角度分析 n 的影响，可以看到准确率总体有所提高，如图10a所示。这与纠缠拟设的结果相反（图8a）：后者的准确率在达到一定的 m后开始下降，尽管输入特征数量随 n的增加而增加。由此可得出结论：在将数据编码到量子比特旋转角时，与纠缠拟设相比，无纠缠拟设较不易受到贫瘠高原的影响。图10b的结果有助于分析数据采用角度编码时 m 对混合量子神经网络性能的影响。可以观察到，准确率总体上随着 m的增加而下降。对于除 $n = 12$以外的所有 n，最小深度（$m = 2$）的表现优于所有其他 m。不过，即使在这种情况下（$n = 12$），从 $m = 2$ 到 $m = 10$ 的准确率提升也并不显著。

6.2. 混合量子神经网络的可训练性与表达能力

对上一节结果进行更仔细的分析后发现，性能下降与量子层深度 m之间并不存在一致的关系（如二次关系、指数关系等），而是取决于量子层的深度。因此，深入分析在给定宽度下，量子层在遭遇贫瘠高原并最终导致性能下降之前可以达到多深，将具有重要意义。本节将再次分别针对两种拟设结构及两种编码方式，对混合量子神经网络的可训练性与表达能力进行实证分析。我们以模型的总体准确率和损失收敛情况作为评估指标，分析可训练性与表达能力。

6.2.1. 振幅编码下纠缠拟设的可训练性与表达能力

混合量子神经网络的可训练性与表达能力分析揭示了：对于特定宽度，在性能更优或至少相当的情况下，给定拟设能够具有多强的表达能力（在梯度开始消失之前）。我们针对不同的 n 和 m进行了不同的训练实验，如图6所示。模型根据总体准确率和收敛情况进行评估。结果如图11所示。

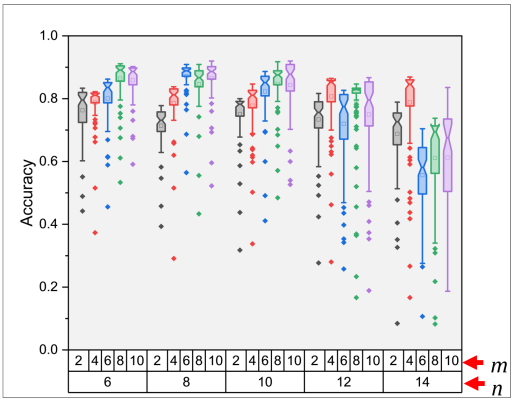


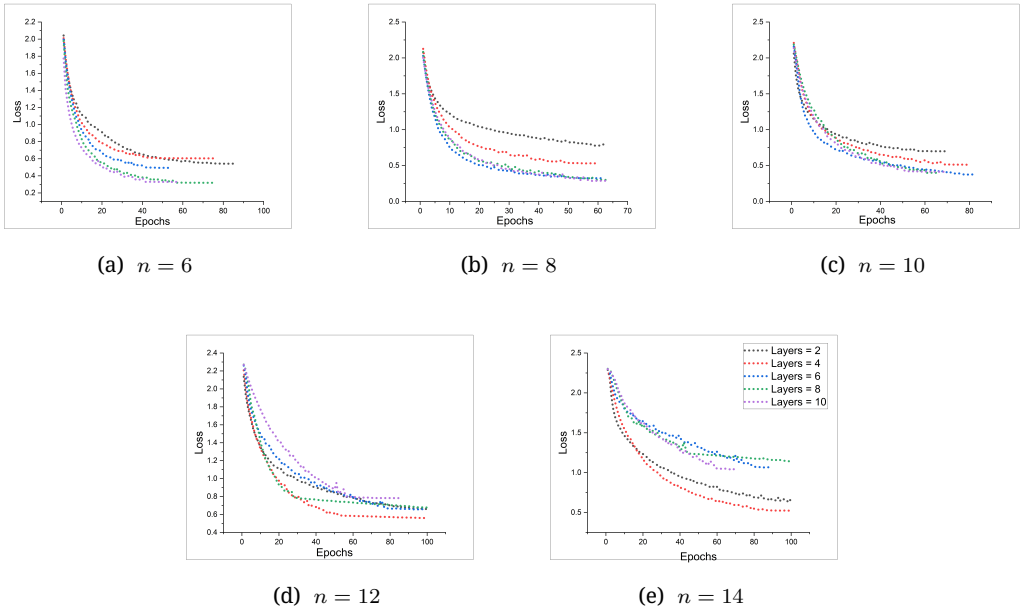
图 11：纠缠拟设：振幅编码下所有 n 和 m 的准确率趋势。

在图 11 中，对于 $n = 6$ ，较深的量子层，即 $m = 8$ 和 $m = 10$ ，在准确率方面表现更好；然而，如图 12a 所示，在 10 层的情况下收敛性更好。当我们将 n 增加到 8 个和 10 个量子比特时， $m = 6$、8、10 的量子电路深度具有相对更高的准确率。然而，在这两种情况下（$n = 8$ 和$n = 10$），如图 12b 和 12c 所示，当 $m = 6$ 时，模型分别能够相对更快地收敛；当使用更多量子比特时，量子层的适当深度随之减小。当 n增加到 12 个量子比特时，适当的量子电路深度进一步减小为 $m = 4$，在准确率和收敛性方面均如此，如图 11 和 12d 所示。然而，当我们进一步将 n增加到 14 个量子比特时，量子层深度的相应缩减变得更加明显，因为如图 12e 所示，对于 $m = 4$ ，模型显然收敛得更快。

考虑到11 和 12 中展示的结果，我们认同，混合量子神经网络中的贫瘠高原现象不仅与量子比特的数量有关（梯度会随着量子比特数量的增加而呈指数级消失，并且随着量子比特数量的增加，网络将变得无法训练），还取决于量子层的表达能力。如果我们整体分析图11中的准确率，我们会发现，随着 n 的增加，模型性能开始下降。尽管我们观察到 n 与 m 之间存在权衡关系（更大的 n 往往会降低 m ，反之亦然），但对单独准确率进行仔细分析后发现，较小的n 和相对较大的 m 更适合实现更好的性能，因为其达到的准确率更高，且分布偏向较高的一侧。例如，对于$n = 6$ 和 $m = 8$ ，近 75% 的准确率高于 $n = 14$ 和 $m = 4$所达到的最高准确率。此外，较大的 n 相比于较小的 n具有更大的准确率方差（准确率分布更加分散），这表明在较宽的量子层情况下，学习过程缺乏鲁棒性。

6.2.2. 角度编码下纠缠拟设的可训练性与表达能力 g

现在，我们分析不同的数据编码策略，即角度编码（将输入数据特征编码到量子比特旋转角中），如何影响特定情况下量子神经网络的可训练性 n 和 m。我们将数据编码为 RY 旋转，然后传入量子层。对于角度编码，如第 2.1 节所述，编码的特征数量应等于量子比特的数量。我们的输入特征是大小为 8×8 的图像，因此总共有 64 个特征。我们需要 64 个量子比特来编码 64 个特征，这在含噪声中等规模量子



(a) $n = 6$ (b) $n = 8$ (c) $n = 10$

(d) $n = 12$ (e) $n = 14$

Fig图 12：纠缠拟设：振幅编码下所有 n 和 m 的损失趋势 .

时代。因此，我们应用主成分分析，将输入特征维度降至 8、10、12 和 14 个特征。量子层的宽度（n）等于输入特征的维度；我们还对具有不同深度（$m = 2$、4、6、8、10）的量子层进行了实验，与振幅编码分析类似，结果如图 6所示，但 $n = 6$除外，因为我们倾向于在应用主成分分析时保持 n 较大，以免丢失过多信息。所有实验的准确率和损失分别如图 13 和 14所示。

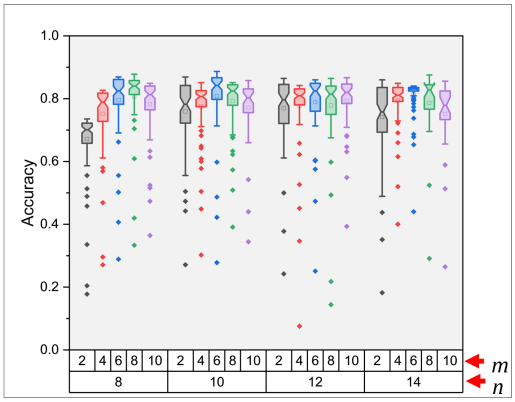
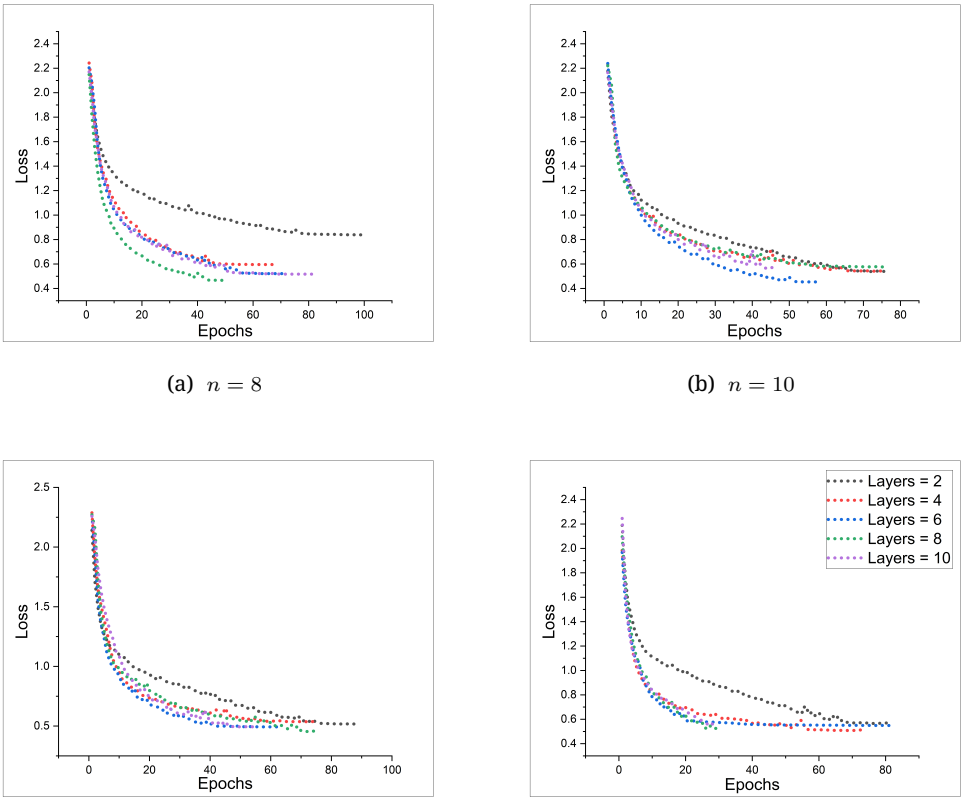


图13：纠缠拟设：角度编码下所有 n 和 m 的准确率趋势。

根据获得的结果，我们观察到，对于 $n = 8$，深度为 $m = 8$ 的量子电路在准确率（图 13 和收敛（图 14a）方面表现更好，而深度最小的量子层（$m = 2$）在 m的所有不同取值中表现最差。当量子电路宽度增加到 $n = 10$ 时，性能更好的电路深度

在准确率和收敛性方面，$m = 6$ 均优于从 $m = 8$降低的情况。尽管输入特征维度也有所增加（从 8 增至 10），但电路深度往往会降低，主要是因为梯度幅值变小。对于 $n = 12$，除 $m = 2$ 外，几乎所有层在准确率方面的表现都相当，但 $m = 6$ 的收敛率略优于其他层。然而，单独的准确率从 $n = 8$ 和 $n = 10$ 降低了。类似地，对于 $n = 14$，$m = 14$ 的电路深度在准确率和收敛率方面都更好。与振幅编码不同，当数据编码为量子比特旋转角时，较大的 n 所允许的电路深度更大，这是因为与较小的 n 相比，需要训练的输入特征数量更多。



(a) $n = 8$　(b) $n = 10$

(c) $n = 12$　(d) $n = 14$

图14：纠缠拟设：角度编码下所有 n 和 m 的损失趋势。

6.2.3. 振幅编码下无纠缠拟设的可训练性与表达能力

现在，我们展示在使用量子比特态矢量对输入数据进行编码时，无纠缠拟设的可训练性与表达能力分析。如前所述，对于我们的数据集，振幅编码无需进行输入维度缩减，因此不应用主成分分析。实验设置与纠缠拟设相同，即按照图6所示，改变量子层的宽度和深度。根据获得的结果，我们观察到，随着 n增加，性能持续下降，因此我们跳过了针对 $n = 14$ 及其所有对应 m的实验。无纠缠拟设的准确率和损失趋势如图 15 所示，以及

16，分别。

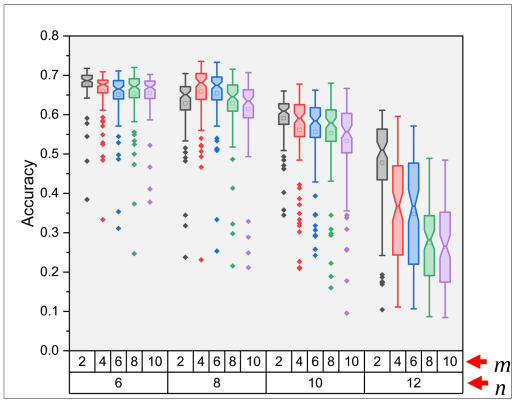
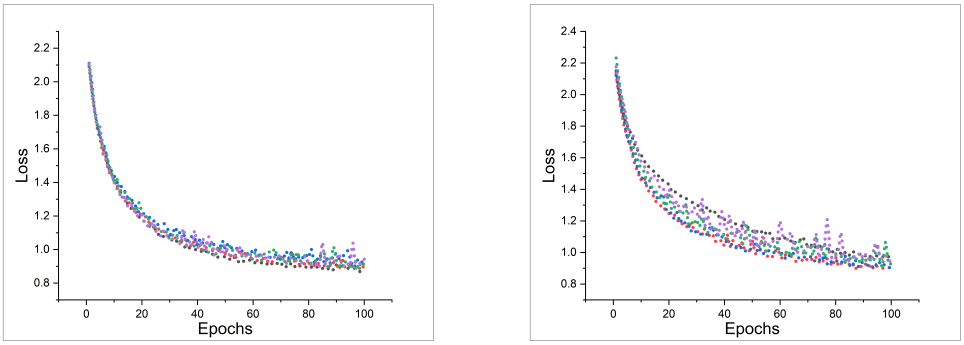


图 15：无纠缠拟设：振幅编码下所有 n 和 m 的准确率趋势。

我们观察到，与纠缠拟设中的振幅编码情况不同（在这种情况下，较小的宽度 n 对应的允许电路深度 m 相对较大），在无纠缠拟设中，即移除纠缠后，即使在较小的 n 下，允许的电路深度也趋于降低。对于 $n = 6$，几乎所有对应的深度 m 在准确率和损失收敛方面都表现出大致相同的性能，如图15和图16a所示。然而，与纠缠拟设中 $m = 8$ 和 10 在 $n = 6$ 的两项性能指标方面表现更好（图 11 和图 12a）的情况相反，较小的深度（$m = 2$）略优于其他 m。类似地，对于 $n = 8$，相对较小深度的量子层，通常为 $m = 4$ 和 6，在准确率和收敛方面均略优于其他深度，如图15和图16b所示。这再次与纠缠拟设形成对比：在纠缠拟设中，$n = 8$ 在最大深度（$m = 10$）下表现最佳，而在最小深度（$m = 2$）下表现最差，如图 11 和图 12b 所示。当 n 进一步增加到 10 和 12 时，整体性能会有所下降（图 15）。然而，与纠缠拟设中较大 n 的结果类似，相对较小的深度（通常为 $m = 2$ 和 4）对应的量子层表现优于深度更大的量子层。

此外，随着 n 增加，两种拟设结构中所有 m 的单独表现都变得更加不一致，如图 15 和 11所示。不过，对于采用振幅编码的混合量子神经网络，无纠缠拟设中的这种不一致性更为明显。此外，在量子层中无纠缠的情况下，当 n较大时，几乎所有电路深度都经历了波动，才达到代价函数景观的最小值，清楚地表明存在贫瘠高原（无法确定成本最小化方向）。还值得一提的是，对于无纠缠拟设中的所有 n 和 m，混合量子神经网络会陷入局部极小值，即使经过 100 次训练迭代也无法收敛，并且对正则化技术不敏感（本研究采用早停以避免过拟合）；而在纠缠拟设中，模型则更加稳健，收敛到代价函数景观最小值的过程相对更好、更平滑，对正则化的相对不敏感性也较低。



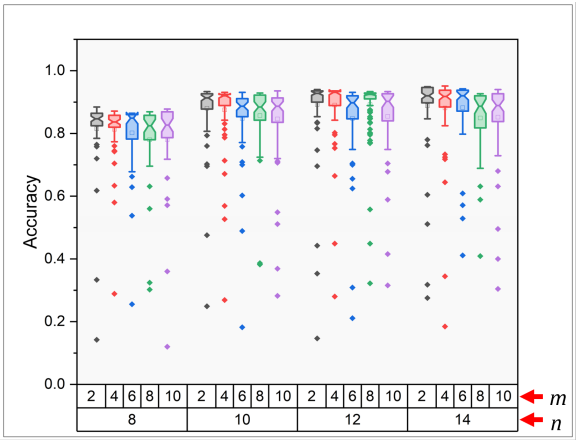
(a) $n = 6$ (b) $n = 8$

(c) $n = 10$ (d) $n = 12$

图 16：无纠缠拟设：振幅编码下所有 n 和 m 的损失趋势。

6.2.4. 角度编码下非纠缠试探态的可训练性与表达能力

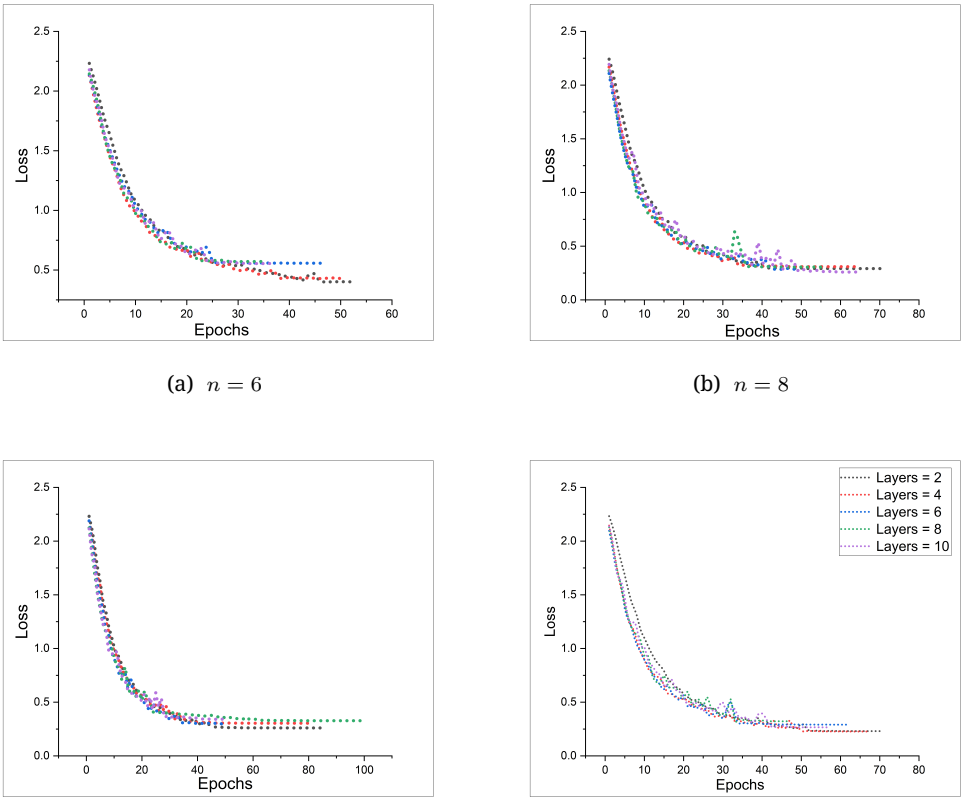
现在，我们展示数据编码到量子比特旋转角时，无纠缠拟设的可训练性与表达能力分析。由于角度编码要求输入数据的大小与量子比特数量相等，因此采用主成分分析来降低输入特征维度。实验从图 6开始，取值为$n = 8$。所有实验对应的准确率和损失收敛结果分别如图 17 和 18所示。



Figu关于图 17：无纠缠拟设：角度编码下所有 n 和 m 的准确率趋势 ng.

我们观察到，在角度编码情况下移除纠缠会影响混合量子神经网络的可训练性与表达能力。所有 n 和 m (图18)中的损失收敛情况都不太容易区分，因此我们通过准确率（图 17）来分析 n 和 m 对可训练性的影响。与纠缠拟设的角度编码不同，其表达能力往往会随着 n 增大而略微下降，即使输入特征有所增加；对于无纠缠拟设，可允许的量子电路深度更大，最终使隐藏量子层中拥有更多可训练参数，从而在相对较大的 n 和 m下取得更好的整体性能。此外，对于纠缠拟设，如图 $n = 8$所示，较窄的量子层（$m = 8$）比相对较宽的量子层（13）准确率更高；而对于无纠缠拟设，混合量子神经网络在几乎所有电路深度 17下的表现都相当。对于无纠缠拟设中较宽的量子层（$n = 14$），所有深度下的表现同样相当。不过，较小的深度（$m = 2$）似乎更可取，因为其表现略好，而且与更深的层相比，训练时间也更短。在传统机器学习中，人们通常认为，对于复杂度适中的模型，增加输入数据会带来更好的性能。在混合量子神经网络中，当数据被编码为旋转角时，如果需要增加量子层宽度，就必须增加输入特征维度。

根据本文给出的实验结果，我们认同，对于纠缠拟设，混合量子神经网络并未遵循（传统机器学习中）关于输入数据越多、性能越好的推测（图13），这是由于隐藏量子层宽度增加导致了所谓的贫瘠高原现象。然而，对于无纠缠拟设，我们观察到这一假设（更多数据 = 更高准确率）确实成立，增加输入特征维度确实能带来更好的整体性能。基于这些观察，我们可以肯定地说，在角度编码的情况下，去除纠缠可能会避免（或延缓）贫瘠高原现象。此外，与振幅编码下的无纠缠拟设不同，在这种情况下，随着 n 以及几乎所有相应的 m增加，模型难以找到成本最小化方向（图16），而在角度编码的情况下，代价函数景观中的收敛过程则更加平滑，如图18所示。最后，尽管角度编码下无纠缠拟设的学习过程比振幅编码更可靠，但它对正则化技术也表现出一定的不敏感性，尤其是在较大的 n 下，如图18c所示。不过，振幅编码情况下对正则化的不敏感性更为明显（图16）。



(a) $n = 6$ (b) $n = 8$

(c) $n = 10$ (d) $n = 12$

图18：无纠缠拟设：角度编码下所有 n 和 m 的损失趋势。

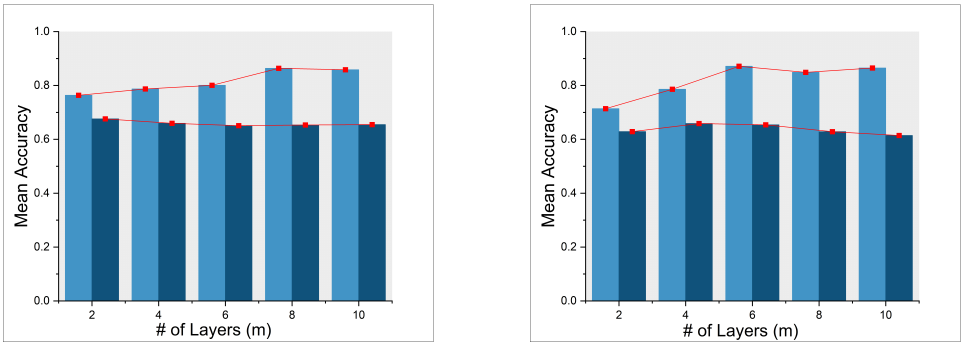
基于对纠缠拟设和无纠缠拟设的可训练性与表达能力分析，我们认同，纠缠会影响混合量子神经网络的训练，并最终影响其性能，而这种影响与底层编码策略相对应。这意味着，对拟设进行直接比较（有纠缠与无纠缠）有助于轻松理解纠缠在混合量子神经网络中的作用。比较结果将在下一节中给出。

6.3. 混合量子神经网络中的纠缠效应

前文各节介绍了 HQNN 中 BP 是否存在的分析，以及其可训练性与表达能力之间的关系。然而，该分析并未明确凸显底层量子层中引入/移除纠缠的重要性。纠缠是量子力学的一项重要基本属性，也是构建具有表达能力的 HQNN 参数化量子电路的关键。因此，理解纠缠在 HQNN 实际应用中的作用至关重要。本节比较两种拟设得到的结果，并分析纠缠是否会影响 HQNN 的整体性能。我们观察到，纠缠确实会影响 HQNN 的性能。然而，这种影响会提升还是降低性能，取决于数据的编码方式。因此，为了理解纠缠在 HQNN 中的作用，我们先比较两种拟设结构在振幅编码下的表现，再比较它们在角度编码下的表现。

6.3.1. 振幅编码中纠缠对混合量子神经网络的影响

当经典数据被编码到量子比特态矢量中，并使用参数化量子电路进行训练时，无纠缠拟设的量子层存在显著的梯度消失问题，详见第 6.1.3 节，导致混合量子神经网络的性能大幅下降。图 19 简要比较了两种拟设的性能。我们观察到，对于所有 n 及相应的 m，采用纠缠拟设量子层的底层混合量子神经网络性能显著优于采用无纠缠拟设的网络。还可以观察到，随着量子比特数量增加，性能会下降。根据所示并讨论的结果，我们认为，在量子比特振幅中编码数据时，构建混合量子神经网络的量子层并包含纠缠完全没有益处。因此，使用振幅嵌入方法时，建议仅使用单量子比特参数化酉变换，以获得更好的性能并缩短训练时间。



(a) $n = 6$ (b) $n = 8$

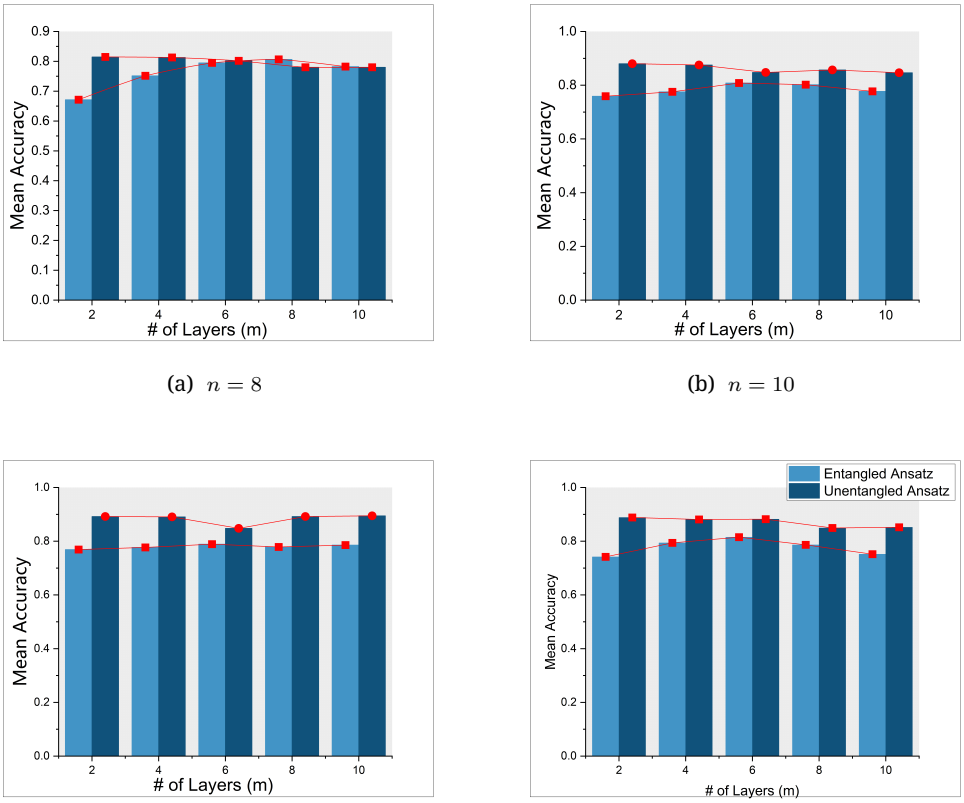
(c) $n = 10$ (d) $n = 12$

Figure 19：振幅编码下两种拟设结构的性能比较。

6.3.2. 纠缠对混合量子神经网络的影响 - 角度编码

当经典数据被编码为量子比特旋转角，并使用参数化量子层进行训练时，纠缠也会发挥作用。然而，与振幅编码不同，在角度编码的情况下，无纠缠拟设量子层能够提升性能。如前文图10a 和10b所示，在底层混合量子神经网络的量子层中不引入任何纠缠时，允许的电路深度

大于具有纠缠的量子层8a和8b。数据被编码为量子比特旋转角时，两种拟设的性能比较如图20所示。与振幅编码不同，在此情况下我们观察到，对于所有 n和 m，通常在量子层中不包含纠缠时，模型的表现更好。当 n 增大时，性能提升更加明显。根据这些结果，我们认为，当数据采用角度编码时，移除纠缠（无纠缠拟设）能够合理地提升模型性能。此外，无纠缠拟设量子层所允许的深度不仅大于纠缠拟设量子层，其训练时间也有所缩短。



(a) $n = 8$ (b) $n = 10$

(c) $n = 12$ (d) $n = 14$

Fi图 20：角度编码下两种拟设结构的性能比较 .

6.4. 混合量子神经网络在分类任务中的面向应用的评估

在前面的章节中，我们基于准确率和损失收敛讨论了所提出框架的结果。然而，对于机器学习中的分类任务（本文的目标应用），仅凭准确率不足以衡量模型性能，因为它只反映了正确预测值占预测值总数的比例。因此，分类问题需要使用更多样化的评估指标。

尽管本文讨论的是一个多类分类问题，但要解释多分类中的其他评估指标（精确率、召回率和 F1分数）相当棘手。因此，为简单起见，我们将说明为何二分类需要更多样化的评估指标，而这些指标也可直接应用于多-

类分类。在二分类中，有两个类别（正类和负类），机器学习模型的目标是预测正确的类别。通常，此时的准确率是指正确预测的类别数之和，而不考虑预测正确的是哪个类别。

当正类样本被分类为负类时，称为假阴性（FN）；当负类样本被预测为正类时，称为假阳性（FP）。当正类和负类样本均被正确预测为各自所属的类别时，则分别称为真阳性（TP）和真阴性（TN）。根据这些类别对分类性能进行评估，可以计算出其他重要指标，即精确率和召回率。精确率表示所有预测结果中真正属于正类的比例，而召回率表示所有实际正类样本中被正确分类的比例。另一个重要的评估指标是 F1 分数，它是精确率和召回率的调和平均数。从数学角度来看，所有这些指标均可通过以下方程计算得出。

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{F1-Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{16}$$

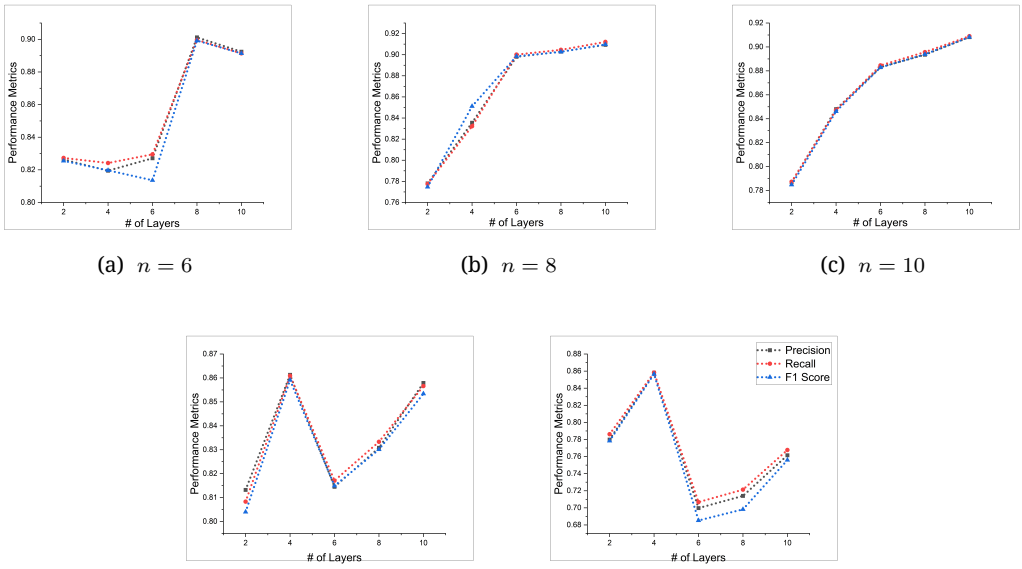
高精确率和召回率通常都是理想的，但在实践中，分类器容易出错，因此可能会产生不同的精确率和召回率。因此，可能需要在这两个指标之间进行权衡，而具体取舍在很大程度上取决于应用场景。例如，对于视频推荐系统，更高的精确率更为理想，以确保向用户推荐的都是潜在感兴趣的视频。同样，用于检测患者是否患有癌症的分类器需要较高的召回率，以便尽可能多的癌症患者都能得到正确诊断。F1分数通常更适合用于比较分类器的性能。例如，对于两个分类器，如果一个分类器的精确率更高，而另一个分类器的召回率更高，那么F1分数就是选出最佳分类器的合适指标。

现在，我们针对本文所用的拟设结构和数据编码，对混合量子神经网络在目标（分类）应用中的表现进行简要评估。根据第 6.2 节和 6.3节中讨论的结果，我们认为，采用振幅编码时，纠缠拟设结构的表现更好；而采用角度编码时，无纠缠拟设的表现更佳。因此，我们仅针对表现最佳且与相应编码搭配的拟设结构，展示混合量子神经网络面向应用的评估结果（振幅编码搭配纠缠拟设，角度编码搭配无纠缠拟设）。通常认为，精确率和召回率都较高（且接近总体准确率）的分类器表现良好。我们观察到，在本文进行的所有相关实验中，混合量子神经网络的精确率和召回率走势均与准确率一致，因此它们在广泛的应用场景中更加可靠。

6.4.1. 面向应用的评估 - 振幅编码

我们首先针对振幅编码和纠缠拟设进行面向应用的分析。对于所有 n，绘制精确率、召回率和 F1分数随 m 变化的曲线，如

如图所示21。



(a) $n = 6$ (b) $n = 8$ (c) $n = 10$

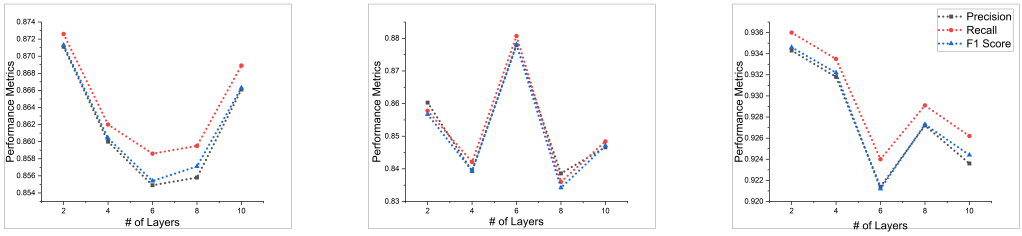
(d) $n = 12$ (e) $n = 14$

图21：振幅编码中，固定 n 和可变 m 时纠缠拟设的精确率、召回率和F1分数

根据图 21所示结果，我们观察到，对于所有 n 和 m，精确率、召回率及相应的 F1分数几乎以相同的比例升降，因此同样适用于对高精确率或高召回率有要求的应用。振幅编码下纠缠拟设的精确率、召回率和 F1分数总体趋势与准确率的趋势几乎一致（图11）。这意味着，当 n较小时，我们需要相对较大的 m （反之亦然），才能在相应的应用中取得更高的召回率和精确率。

6.4.2. 面向应用的评估 - 角度编码

由于无纠缠拟设在角度编码下表现更佳，因此在面向应用的分析中，我们仅考虑这些结果。固定 n 和可变 m 时的精确率、召回率和 F1 分数结果见图 22。所有性能指标均以相同比例同步升降，与振幅编码的情况类似，因此同样适用于高精确率或高召回率的应用。然而，在量子比特旋转角中对数据进行编码时，所有训练实验的召回率都略高于精确率，因此更适合高召回率应用。



(a) $n = 8$ (b) $n = 8$ (c) $n = 12$

图22：无纠缠拟设的角度编码在 n 固定、 m 可变时的精确率、召回率和F1分数

7. 我们框架的重要性

本文前几节提出并讨论了一个框架，重点分析混合量子神经网络的三个主要组成部分：数据编码、拟设表达能力，以及纠缠的引入/移除。换言之，这三个组成部分所受的约束会形成多种可能的情形。因此，在设计混合量子神经网络时，既要分别考虑这些组成部分各自受到的约束，也要考虑它们彼此之间的关系。本节根据第 6.2 节的结果，针对各种约束情形提出混合量子神经网络设计中不同参数的建议，以突出所提框架的重要性。各约束情形对应的建议摘要见附录A。

7.1. 无约束

在此情形下，我们假设设计者对任何参数都没有约束，并可自由选择任意类型的数据编码（振幅或角度）、任意的拟设宽度和深度（表达能力）。此外，是否包含纠缠也不受约束。根据我们提出的框架所得结果，在这种情形下，建议采用角度编码，并且底层量子层中的量子比特之间不包含纠缠。此外，综合考虑总体准确率和损失收敛，建议将拟设宽度设为 $n = 10$ ，深度设为 $m = 4$ ，如图 17 和 18b所示。这是我们框架所获得的最佳结果。

7.2. 单参数约束

在此情形下，我们考虑量子层的数据编码、拟设表达能力和纠缠这几项设计参数中，某一项受到约束的情况。

7.2.1. 数据编码约束。

这里，我们假设所使用的数据编码方案受到某种约束。如果要求将数据编码到量子比特振幅中，那么纠缠就不再是】【。

建议将其纳入底层量子层。此外，量子层的最优宽度和深度为 $n = 6$ 和 $m = 10$，如图11 和12a 所示。另一方面，如果限定使用角度编码，则可选择与第7.1节讨论的相同一组混合量子神经网络。

7.2.2. 拟设表达能力的约束。

拟设表达能力有两个影响因素：拟设宽度和拟设深度。如果约束针对拟设宽度，那么对于较小的宽度（$n = 6$，8），如图11 和12b所示，采用中等深度（$m = 6$）的振幅编码更适合用于提高准确率并促进收敛。此外，建议加入纠缠。另一方面，对于较大的宽度（$n = 10$，12，14），角度编码是更好的选择。在此情形下，其他参数的选择可参照第7.1 节的讨论。

如果约束针对量子层的深度，那么在深度较小（$m = 2$，4）时，角度编码搭配无纠缠拟设效果更好。此外，量子层允许采用更大的宽度，即 $n = 10$，12，14。不过，考虑到训练时间较短，使用 $n = 10$ 会更合适。同样，对于中等深度至较深的量子层（$m = 6$，8，10），振幅编码（结合纠缠）和角度编码（不结合纠缠）在准确率方面表现相当。然而，使用振幅编码时，量子层允许的宽度相对较小（$n = 6$）；而角度编码允许量子层具有更大的宽度（$n = 10$，12，14），因此比振幅编码更不易受到贫瘠高原的影响。

7.2.3. 关于纠缠的约束。

对于要求量子比特相互纠缠的约束情形，可以使用宽度和深度为 $n = 8$ 和 $m = 6$ 的振幅编码。另一方面，当约束要求量子比特之间不能相互纠缠时，角度编码的表现更好，并且可以采用第 7.1 节中讨论的相同规格。

7.3. 两个参数的约束

在这种情形下，我们考虑以下可能性：数据编码、拟设表达能力以及纠缠的引入/移除这三个因素中的任意两个受到约束。

7.3.1. 数据编码约束与拟设表达能力。

拟设表达能力有两个因素：宽度和量子层的深度。因此，我们分别考虑这两个表达能力因素的约束，以及数据编码的约束。我们首先考虑数据编码和拟设宽度的约束。在振幅编码和角度编码两种情况下，宽度较小时（$n = 8$,10），适当的深度较大（$m = 8$,10）；而宽度较大时（$n = 12$,14），适当的深度则降至 $m = 2$,4。类似地，在拟设深度和数据编码的约束下，拟设深度与宽度之间也存在权衡。此外，在这两种情况下（对拟设宽度或深度施加约束），当采用量子比特振幅编码时，建议将量子比特纠缠起来，而当采用角度编码时，

对于角度编码，纠缠量子比特并不能提升性能。

7.3.2. 数据编码与纠缠的约束。

现在我们考虑一种对数据编码以及是否包含纠缠施加约束的情况。如果约束要求数据必须编码到量子比特振幅中，那么无论量子比特是否纠缠，量子层允许的宽度都相对较小，即（$n = 6$,8）。更大的宽度会导致性能下降，表明存在贫瘠高原（BP）。然而，如果底层量子层中的量子比特存在纠缠，则允许的深度更大（$m = 8$,10）；而在量子比特不存在纠缠的情况下，允许的深度则会减小（$m = 2$,4）。

同样，当数据编码被限制为采用量子比特旋转角时，无论是否包含纠缠，合适的深度都大于振幅编码 ($n = 10$，12，14)，因此更不容易受到贫瘠高原的影响，并最终带来更好的性能。然而，在包含纠缠的情况下，量子层的允许深度更大 ($m = 8$，10)；而在无纠缠的情况下，允许深度则会降低 ($m = 2$，4)。

7.3.3. 表达能力与纠缠的约束。

我们首先考虑对表达能力的宽度因子施加约束、而深度可以任意设置的情形。此外，纠缠的引入/移除也受到约束。

当底层量子层中的量子比特被限定为相互纠缠，且要求量子层的宽度较小（$n = 8$，10）时，振幅编码效果更好，并且需要更深的量子层（$m = 10$）。然而，如果要求量子层的宽度较大（$n = 12$，14），同时引入纠缠，那么角度编码相对更好，底层量子层的适当深度较小（$m = 2$）。另一方面，当量子层中的纠缠被限定为移除时，无论宽度约束如何，对于本文采用的所有相应宽度，角度编码都优于振幅编码，且适当的深度通常较小（$m = 2$，4）。

现在我们考虑这样一种设置：深度因子表达能力受到约束，同时还需引入/移除纠缠。当约束要求深度较小（$m = 2$ 或 4），并且也要求引入纠缠时，两种编码方式（振幅或角度）的表现相近，且可使用更宽的量子层（$n = 10$）。然而，当约束要求量子层更深（$m = 6$,8 或 10），同时要求引入纠缠时，振幅编码的表现优于角度编码。此外，量子层的合适宽度也更大（$n = 8$）。当要求移除纠缠时，无论深度如何，角度编码都优于振幅编码，而且底层量子层的合适宽度通常更大（$n = 12$ 或 14）。

7.4. 三个参数的约束

在这种情况下，我们考虑 HQNN 的所有参数（数据编码、拟设表达能力和纠缠）均受到约束的情形。如前所述，拟设表达能力包含两个因素，即量子层的宽度和深度。

因此，我们分别考虑对这两个表达能力因素的约束，以及对其他数据编码和纠缠的约束。

7.4.1. 数据编码、拟设宽度和纠缠的约束。

当拟设宽度受到约束，同时数据编码和纠缠的引入/移除也受到约束时，若要求以量子比特振幅对数据进行编码、必须引入纠缠，且宽度较小（通常为$n = 6, 8, 10$），那么要获得相对较好的性能，底层量子层的深度应更大（通常为 $m = 10$）。然而，如果宽度被限制为较大值（$n = 12, 14$），那么要获得更好的性能，适当的深度应降至 $m = 4$。同样，如果要求移除纠缠，那么无论宽度受到何种约束，适当的深度通常都较小（$m = 2$）。

另一方面，如果要求以量子比特旋转角对数据进行编码、必须引入纠缠，且拟设宽度较小（通常为 $n = 8, 10$），那么适当的拟设深度相对较大（$m = 8$）。然而，当拟设宽度被限制为较大值（通常为 n =12，14）时，与振幅编码的情况类似，适当的拟设深度会减小（通常为 $m = 4$）。同样，如果要求移除纠缠，且要求将数据编码到量子比特角度中，那么无论宽度受到何种约束，使用较小的拟设深度（$m = 2$）都更合适。

7.4.2. 数据编码、拟设深度和纠缠的约束。

现在我们考虑拟设深度与数据编码以及纠缠的引入/移除同时受到约束的情况。在这种情况下，拟设宽度可以是任意值。当数据被约束为必须采用振幅编码时，也要求引入纠缠，并且拟设深度受到较小值的约束（通常为 $m = 2,4$），此时合适的拟设宽度稍大（$n = 10,12$）。然而，当拟设深度被约束为中等水平时，即 $m = 6$，合适的拟设宽度会降低至 $n = 8$；对于更深的拟设（$m = 8,10$），拟设宽度则进一步降低至 $n = 6$。类似地，当约束条件要求移除纠缠时，无论深度受到何种约束，采用中等拟设宽度（$n = 6$）都更有利于获得相对较好的性能。

另一方面，当约束要求将数据编码到量子比特的角度中时，需要包含纠缠，且拟设深度受限于较小的中等值（通常为 $m = 2, 4, 6$），此时适当的宽度相对较大（$n = 12, 14$）。然而，如果拟设深度受限于较大值（$m = 8, 10$），适当的宽度则会降至 $n = 8$。类似地，当约束要求移除纠缠，并要求将数据编码到量子比特的角度中时，无论深度受到何种约束，适当的拟设宽度都较大，即 $n = 12, 14$。

8. 结论

量子机器学习（QML）近年来已成为量子计算的一种潜在应用方向，旨在利用量子力学现象改进经典机器学习。在量子机器学习领域，量子神经网络（QNNs）受到了广泛关注，这主要归功于其经典版本所取得的非凡成功。

其对应形式，即神经网络（NNs）。然而，贫瘠高原（BP）现象对量子神经网络（QNNs）的实际应用构成了挑战。在这一现象中，随着系统规模的增大，参数的梯度会呈指数级减小，从而可能导致量子神经网络无法训练。为此，研究人员已将量子神经网络的主要组成部分，即数据编码、拟设表达能力以及量子比特之间的纠缠，确定为可能引发贫瘠高原的因素。尽管从贫瘠高原的角度出发，这些组成部分已分别得到独立研究，但在实际应用场景中，它们是同时存在的。因此，研究这些因素相互关联下的联合效应，对于实际应用具有重要意义。

本文提出了一个框架，用于实证研究上述量子神经网络各个组成部分对一项实际应用——多分类——的整体影响。在实际应用中，受含噪中等规模量子设备的限制，人们广泛使用混合量子神经网络（HQNNs）来探索量子神经网络中的潜在量子优势。由于混合量子神经网络完整复现了一般的量子神经网络架构（包含一些经典输入预处理与后处理），对混合量子神经网络中量子部分的分析可直接应用于量子神经网络。我们用于分析的混合量子神经网络包含以下一系列操作：1）输入维度缩减；2）量子比特初始化；3）数据编码（经典到量子的特征映射）；4）量子拟设（参数化量子电路）；5）量子比特测量；6）经典全连接神经元层，用于对量子比特测量结果进行后处理并获得输出。

我们的分析重点关注数据编码和量子层（其表达能力以及纠缠的引入与移除），它们是量子神经网络的主要组成部分。对于数据编码，我们采用两种常用的数据编码技术，即振幅编码和角度编码。对于拟设表达能力，我们改变量子层的宽度（n）和深度（m）。我们使用底层拟设为 n 和 m的混合量子神经网络进行训练。我们考虑两种相似的拟设结构：纠缠拟设（包含单量子比特参数化酉变换和最近邻纠缠）以及无纠缠拟设（仅包含单量子比特参数化酉变换）。首先，我们对训练实验的平均准确率进行基准测试，并证明了混合量子神经网络中贫瘠高原的存在。我们观察到，混合量子神经网络中的贫瘠高原与量子比特数量并不存在直接关系，而是取决于量子层的整体表达能力。随后，我们对混合量子神经网络的总体准确率和损失收敛进行基准测试，并开展全面的可训练性与表达能力分析。该分析从贫瘠高原的角度揭示了拟设表达能力如何影响混合量子神经网络的整体性能，以及对于每种编码，在出现贫瘠高原之前，给定宽度的拟设所能达到的最大深度。

此外，我们观察到，纠缠会对混合量子神经网络的训练景观产生影响，并且这种影响取决于编码类型。当数据采用量子比特态矢量进行编码时，纠缠拟设比无纠缠拟设能够获得更高的准确率，这表明在振幅编码的情况下，纠缠对混合量子神经网络的可训练性具有正向影响。相反，当数据通过量子比特旋转角进行编码时，无纠缠拟设比纠缠拟设能够获得更高的准确率，这表明纠缠对混合量子神经网络的可训练性具有负向影响。我们还简要评估了混合量子神经网络在分类应用中的表现，并考虑了分类问题的其他重要评估指标，即：精确率、召回率和F1分数。最后，我们通过针对不同约束情景（单独约束及组合约束），就底层量子层中的数据编码、拟设表达能力以及纠缠的引入/移除提供建议，阐明了所提出框架的重要意义。

参考文献

[1] Preskill J. 含噪中等规模量子时代及其未来的量子计算。Quantum。2018年8月；2：79。可访问：https://doi.org/10.22331/q-2018-08-06-79。[2] Arute F, Arya K, Babbush R, 等。利用可编程超导处理器实现量子霸权。Nature。2019年；574(7779)：505–510。[3] Zhong HS, Wang H, Deng YH, 等。利用光子实现量子计算优势。Science。2020年12月；370(6523)：1460–1463。可访问：http://dx.doi.org/10.1126/ science.abe8770。[4] Wu Y, Bao WSe。利用超导量子处理器实现强量子计算优势。Phys Rev Lett。2021年10月；127：180501。可访问：https://link. aps.org/doi/10.1103/PhysRevLett.127.180501。[5] Madsen LS, Laudenbach Fe。利用可编程光子处理器实现量子计算优势。Nature。2022年；606(7912)：75–81。可访问：https://doi.org/ 10.1038/s41586-022-04725-x。[6] Bergholm V, Izaac Je。Pennylane：量子-经典混合计算的自动微分；2018年。可访问：https://arxiv.org/abs/1811.04968。[7] Peruzzo A, McClean J, Shadbolt P, 等。在光子量子处理器上实现变分本征值求解器。《自然·通讯》。2014年7月；5(1)。可访问：https://doi.org/10.1038%2Fncomms5213。[8] LaRose R, Tikku A, O'Neel-Judy ´E, 等。变分量子态对角化。npj Quantum Information。2019年6月；5(1)。可访问：https://doi.org/10.1038% 2Fs41534-019-0167-6。[9] Anschuetz ER, Olson JP, Aspuru-Guzik A, 等。变分量子因式分解。arXiv。2018年；可访问：https://arxiv.org/abs/1808.08927。[10] Farhi E, Goldstone J, Gutmann S。量子近似优化算法。arXiv。2014年；可访问：https://arxiv.org/abs/1411.4028。[11] Klco N,Dumitrescu EF, McCaskey AJ, 等。利用量子计算机进行Schwinger模型动力学的量子-经典计算。Phys Rev A。2018年9月；98：032331。可访问：https://link.aps.org/doi/10.1103/PhysRevA.98.032331。[12] Klco N, Savage MJ。面向量子计算的标量场数字化。Phys Rev A。2019年5月；99：052335。可访问：https://link.aps.org/doi/10.1103/PhysRevA. 99.052335。[13] Sharma K, Khatri S, Cerezo M, 等。变分量子编译的噪声鲁棒性。新物理学杂志。2020年4月；22(4)：043006。可访问：https://doi.org/10. 1088%2F1367-2630%2Fab784c。[14] McClean JR, Boixo S, Smelyanskiy VN, 等。量子神经网络训练景观中的贫瘠高原。《自然·通讯》。2018年11月；9(1)。可访问：https:// doi.org/10.1038%2Fs41467-018-07090-4。[15] Grant E, Wossnig L, Ostaszewski M, 等。用于解决参数化量子电路中贫瘠高原问题的初始化策略。Quantum。2019年12月；3：214。可访问：https://doi.org/10.22331%2Fq-2019-12-09-214。[16] Cerezo M, Sone A, Volkoff T, 等。浅层参数化量子电路中依赖于成本函数的贫瘠高原。《自然·通讯》。2021年3月；12(1)。可访问：https://doi.org/10.1038%2Fs41467-021-21728-w。[17] McClean JR, Romero J, Babbush R, 等。变分量子-经典混合算法理论。新物理学杂志。2016年2月；18(2)：023023。可访问：https://doi.org/10.1088/1367-2630/18/2/023023。[18] Biamonte J。通用变分量子计算。Physical Review A。2021年3月；103(3)。可访问：https://doi.org/10.1103%2Fphysreva.103.l030401。[19] Goodfellow I, Bengio Y, Courville A。深度学习。MIT出版社；2016年。[20] K¨ubler JM, Muandet K, Sch¨olkopf B。概率分布的量子均值嵌入。Physical Review Research。2019年12月；1(3)。可访问：https://doi.org/10. 1103%2Fphysrevresearch.1.033159。

[21] Suzuki Y, Yano H, Gao Q, 等. 基于核的量子分类器特征映射的分析与合成。《量子机器智能》。2020年6月；2(1)。获取地址：https: //doi.org/10.1007%2Fs42484-020-00020-y。[22] Date P, Arthur D, Pusey-Nazzaro L. 用于训练机器学习模型的 QUBO 表述。《Scientific Reports》。2021年5月；11(1)。获取地址：https://doi.org/10.1038%2Fs41598-021-89461-4。[23] Date P, Potok T. 绝热量子线性回归。《Scientific Reports》。2021年11月；11(1)。获取地址：https://doi.org/10.1038%2Fs41598-021-01445-6。[24] Arthur D, Date P. 绝热量子计算机上的平衡 k 均值聚类。arXiv。2020年；获取地址：https://arxiv.org/abs/2008.04419。[25] Cerezo M, Arrasmith A, Babbush R, 等. 变分量子算法。《Nature Reviews Physics》。2021年8月；3(9)：625–644。获取地址：https://doi.org/10.1038% 2Fs42254-021-00348-9。[26] Farhi E, Neven H. 在近期处理器上使用量子神经网络进行分类；2018年。获取地址：https://arxiv.org/abs/1802.06002。[27] Havlı́ˇcek V, C´orcoles AD, Temme K, 等. 使用量子增强特征空间进行监督学习。《Nature》。2019年3月；567(7747)：209–212。获取地址：https://doi.org/ 10.1038%2Fs41586-019-0980-2。[28] Beer K, Bondarenko D, Farrelly T, 等. 训练深度量子神经网络。《自然·通讯》。2020年；11(1)：1–6。[29]Cong I, Choi S, Lukin MD. 量子卷积神经网络。《Nature Physics》。2019年8月；15(12)：1273–1278。获取地址：http://dx.doi.org/10.1038/ s41567-019-0648-8。[30] Du Y, Hsieh MH, Liu T, 等. 量子神经网络的可学习性。《PRX Quantum》。2021年11月；2：040337。获取地址：https://link.aps.org/doi/10.1103/PRXQuantum. 2.040337。[31] Mitarai K, Negoro M, Kitagawa M, 等. 量子电路学习。《Physical Review A》。2018年9月；98(3)。获取地址：http://dx.doi.org/10.1103/PhysRevA.98.032309。[32]Kashif M, Al-Kuwari S. 混合量子—经典神经网络的设计空间探索。《Electronics》。2021年；10(23)：2980。获取地址：https://www.mdpi.com/ 2079-9292/10/23/2980。[33] Kashif M, Al-Kuwari S. 展示混合量子神经网络在模型容量方面的量子优势。载于：2022 IEEE International Conference on Rebooting Computing (ICRC)；2022年。第36–44页。[34] Schuld M, Bocharov A, Svore KM, 等. 以量子电路为核心的量子分类器。《Phys Rev A》。2020年3月；101：032308。获取地址：https://link.aps.org/doi/10.1103/PhysRevA. 101.032308。[35] Zhang K, Hsieh MH, Liu L, 等. 探索量子神经网络的可训练性；2020年。获取地址：https://arxiv.org/abs/2011.06258。[36] Tacchino F, Barkoutsos P, Macchiavello C, 等. 人工前馈神经网络的量子实现。《量子科学与技术》。2020年10月；5(4)：044010。获取地址：https://doi.org/10.1088%2F2058-9565%2Fabb8e4。[37] Sharma K, Cerezo M, Cincio L, 等. 基于耗散感知器的量子神经网络的可训练性；2020年。获取地址：https://arxiv.org/abs/2005.12458。[38] Killoran N, Bromley TR, Arrazola JM, 等. 连续变量量子神经网络。《Phys Rev Research》。2019年10月；1：033063。获取地址：https://link.aps.org/ doi/10.1103/PhysRevResearch.1.033063。[39]Torrontegui E, Garcı́a-Ripoll JJ. 作为高效通用逼近器的酉量子感知器。《EPL (Europhysics Letters)》。2019年3月；125(3)：30004。获取地址：https://doi.org/10.1209%2F0295-5075%2F125%2F30004。[40] Wan KH, Dahlsten O, Kristj´ansson H, 等. 前馈神经网络的量子推广。npj Quantum Information。2017年9月；3(1)。获取地址：https://doi.org/ 10.1038%2Fs41534-017-0032-4。[41] Benedetti M, Lloyd E, Sack S, 等. 作为机器学习模型的参数化量子电路

模型。《量子科学与技术》。2019年11月；4(4)：043001。获取地址：https: // doi.org/10.1088/2058-9565/ab4eb5。[42] Hubregtsen T, Pichlmeier J, Stecher P, 等. 参数化量子电路的评估：分类准确率、表达能力与纠缠能力之间的关系；2020年。获取地址：https://arxiv.org/abs/2003.09887。[43] Abbas A, Sutter D, Zoufal C, 等. 量子神经网络的力量。《Nature Computational Science》。2021年6月；1(6)：403–409。获取地址：http://dx.doi.org/10. 1038/s43588-021-00084-1。[44] Rebentrost P, Mohseni M, Lloyd S. 面向大数据分类的量子支持向量机。《Physical Review Letters》。2014年9月；113(13)。获取地址：https://doi.org/ 10.1103%2Fphysrevlett.113.130503。[45] Biamonte J, Wittek P, Pancotti N, 等. 量子机器学习。《Nature》。2017年9月；549(7671)：195–202。获取地址：https://doi.org/10.1038%2Fnature23474。[46]Chia NH, Gily´en AP, Li T, 等. 用于去量子化量子机器学习的基于采样的次线性低秩矩阵运算框架。《Journal of the ACM》。2022年；69(5)：1–72。[47] Wang S, Fontana E, Cerezo M, 等. 噪声诱导的变分量子算法贫瘠高原。《自然·通讯》。2021年11月；12(1)。获取地址：https://doi.org/ 10.1038%2Fs41467-021-27045-6。[48] Wecker D, Hastings MB, Troyer M. 面向实用变分量子算法的进展。《Phys Rev A》。2015年10月；92：042303。获取地址：https://link.aps.org/doi/ 10.1103/PhysRevA.92.042303。[49]Bartlett RJ, Musia-l M. 量子化学中的耦合簇理论。《Rev Mod Phys》。2007年2月；79：291–352。获取地址：https://link.aps.org/doi/10.1103/ RevModPhys.79.291。[50] Lee J, Huggins WJ, Head-Gordon M, 等. 用于量子计算的广义酉耦合簇波函数。《Journal of chemical theory and computation》。2018年；15(1)：311–324。[51] Cao Y, Romero J, Olson JP, 等. 量子计算时代的量子化学。《Chemical Reviews》。2019年8月；119(19)：10856–10915。获取地址：https://doi.org/ 10.1021%2Facs.chemrev.8b00803。[52] Hadfield S, Wang Z, O'Gorman B, 等. 从量子近似优化算法到量子交替算子拟设。《Algorithms》。2019年；12(2)。获取地址：https://www.mdpi.com/1999-4893/12/2/34。[53] Kandala A, Mezzacapo A, Temme K, 等. 面向小分子和量子磁体的硬件高效变分量子本征求解器。《Nature》。2017年9月；549(7671)：242–246。获取地址：https://doi.org/10.1038%2Fnature23879。[54] Holmes Z, Sharma K, Cerezo M, 等. 将拟设表达能力与梯度幅值及贫瘠高原联系起来。《PRX Quantum》。2022年1月；3：010313。获取地址：https://link.aps.org/doi/10.1103/PRXQuantum.3.010313。[55] Marrero CO, Kieferov´a M, Wiebe N. 纠缠诱导的贫瘠高原；2020年。获取地址：https://arxiv.org/abs/2010.15968。[56]Maciejewski FB, Baccari F, Zimbor´as Z, 等. 读出噪声中串扰效应的建模与缓解及其在量子近似优化算法中的应用。《Quantum》。2021年6月；5：464。获取地址：https://doi.org/10.22331% 2Fq-2021-06-01-464。[57] Alam M, Ash-Saki A, Ghosh S. 真实噪声条件下超导量子比特上量子近似优化算法的分析；2019年。获取地址：https://arxiv. org/abs/1907.09631。[58] Xue C, Chen ZY, Wu YC, 等. 量子噪声对量子近似优化算法的影响；2019年。获取地址：https://arxiv.org/abs/1909.02196。[59] Schuld M, Sweke R, Meyer JJ. 数据编码对变分量子机器学习模型表达能力的影响。《Phys Rev A》。2021年3月；103：032430。获取地址：https://link.aps.org/doi/10.1103/PhysRevA.103.032430。

[60] Pedregosa F、Varoquaux G、Gramfort A 等。Scikit-learn：Python 机器学习。《机器学习研究杂志》。2011年；12：2825–2830。[61] LeCun Y。手写数字 MNIST 数据库。http://yann lecun com/exdb/mnist/。1998年；。[62] Du Y、Hsieh MH、Liu T 等。基于 Grover 搜索的分类量子学习方案。《新物理学杂志》。2021年2月；23(2)：023020。可查阅：https: //doi.org/10.1088/1367-2630/abdefa。[63] Arthur D、Date P。用于二分类的量子-经典混合神经网络架构；2022年。可查阅：https://arxiv.org/abs/2201.01820。[64] Schuld M、Killoran N。特征希尔伯特空间中的量子机器学习。Phys Rev Lett。2019年2月；122：040504。可查阅：https://link.aps.org/doi/10.1103/ PhysRevLett.122.040504。[65] Schuld M、Petruccione F。使用量子计算机进行监督学习。第17卷。施普林格；2018年。[66] Lloyd S、Schuld M、Ijaz A 等。用于机器学习的量子嵌入；2020年。可查阅：https://arxiv.org/abs/2001.03622。[67] LaRose R、Coyle B。适用于量子分类器的稳健数据编码。Physical Review A。2020年9月；102(3)。可查阅：https://doi.org/10.1103%2Fphysreva.102.032420。[68] Grant E、Benedetti M、Cao S 等。分层量子分类器。npj Quantum In- formation。2018年；4(1)：1–8。[69] Cao S、Wossnig L、Vlastakis B 等。参数化量子电路机器学习中的成本函数嵌入与数据集编码。Phys Rev A。2020年5月；101：052309。可查阅：https://link.aps.org/doi/10.1103/PhysRevA.101.052309。[70] Schuld M、Bergholm V、Gogolin C 等。在量子硬件上评估解析梯度。Physical Review A。2019年3月；99(3)。可查阅：https://doi.org/10. 1103%2Fphysreva.99.032331。[71] Skolik A、McClean JR、Mohseni M 等。量子神经网络的逐层学习。量子机器智能。2021年1月；3(1)。可查阅：https://doi.org/ 10.1007%2Fs42484-020-00036-4。[72] Volkoff T、Coles PJ。随机参数化量子电路中通过关联获得大梯度。《量子科学与技术》。2021年1月；6(2)：025008。可查阅：https: //doi.org/10.1088%2F2058-9565%2Fabd891。[73] Bengio Y、Lamblin P、Popovici D 等。深度网络的贪心逐层训练。神经信息处理系统进展。2006年；19。[74] Kashif M、Al-kuwari S。ResQNets：一种缓解量子神经网络贫瘠高原问题的残差方法；2023年。可查阅：https://arxiv.org/abs/2305.03527。[75] Arrasmith A、Cerezo M、Czarnik P 等。贫瘠高原对无梯度优化的影响。Quantum。2021年10月；5：558。可查阅：https://doi.org/10.22331% 2Fq-2021-10-05-558。[76] Cerezo M、Sone A、Volkoff T 等。浅层参数化量子电路中与成本函数相关的贫瘠高原。《自然·通讯》。2021年3月；12(1)。可查阅：https://doi.org/10.1038%2Fs41467-021-21728-w。[77] Kashif M、Al-Kuwari S。混合量子神经网络中成本函数全局性和局部性的影响：面向含噪中等规模量子设备。《机器学习：科学与技术》。2023年1月；4(1)：015004。可查阅：https://dx.doi.org/10.1088/2632-2153/acb12f。[78] Pedregosa F、Varoquaux G、Gramfort A 等。Scikit-learn：Python 机器学习。《机器学习研究杂志》。2011年；12：2825–2830。[79] Kingma DP、Ba J。Adam：一种随机优化方法；2014年。可查阅：https://arxiv.org/abs/1412.6980。

附录 A. 所有约束场景的设计指南

在此，我们根据框架的结果，对混合量子神经网络的设计参数提出详细建议。设计混合量子神经网络（通常是量子层）时，我们考虑所有可能的约束。由于我们提出的框架侧重于分析数据编码、拟设表达能力和纠缠，因此我们既分别考察这些混合量子神经网络设计要素的约束，也考察它们彼此之间的约束关系。需要注意的是，当针对单一宽度（或反之）推荐多个深度时，它们所达到的准确率几乎相同，即（±1%）。在这种情况下，我们会参考收敛率来选择最优宽度和深度。如果所有推荐的宽度和深度都以相同的速率收敛，那么选择最小者会更合适，因为其训练速度相对更快。

A.1. 无约束

推荐 数据编码	最优 宽度 (n)	最优 深度（米）	纠缠 包含	备注
角度编码	10	2,4	No	$m = 4$ 收敛更快
	12	2,4,8	No	
	14	2,4,6	No	

表 A1：对任何设计参数均无约束时的建议

A.2. 单参数约束

约束	建议			备注
	最优 宽度 (n)	最优 深度 (m)	纠缠 包含	
振幅编码	6	8, 10	Yes	$m = 10$ 收敛更快
	8	6、8、10	Yes	$m = 6$ 收敛更快
角度编码	**请参阅表格A1**			

表格 表 A2：仅对数据编码类型有约束时的建议

约束 试探态宽度	建议			备注
	最优 深度（米）	Data 编码	纠缠 包含	
8	6,8,10	振幅	Yes	$m = 6$ 收敛更快
10	2,4	角度	No	$m = 4$ 收敛更快
12	2,4,8	角度	No	
14	2、4、 6	角度	No	

表 A3：仅对拟设宽度施加约束时的建议

约束	建议			备注
拟设深度	最优宽度 (n)	Data 编码	纠缠包含	
2	10,12,14	角度	No	相似收敛 for 所有推荐的宽度
4	10,12,14	角度	No	
6	8	振幅	Yes	-
	14	角度	No	-
8	6	振幅	Yes	-
	12	角度	No	-
10	6,8	振幅	Yes	$n = 6$ 收敛更快
	12	角度	No	$n = 12$ 收敛更快

表格 A4：仅对拟设深度有约束时的建议

约束	建议			备注
	Data 编码	最优宽度 (n)	最优深度 (m)	
纠缠必需	振幅	6	8,10	$m = 10$ 收敛更快
		8	6,10	$m = 6$ 收敛更快
纠缠不需要	角度	10	2,4	$n = 12$ 收敛更快
		12	2,4,8	
		14	2,4,6	

表格 A5：仅对纠缠施加约束时的建议

A.3. 对两个参数的约束

约束		建议		备注
数据编码	拟设宽度 (n)	最优深度（m）	纠缠包含	
振幅	8	6,10	Yes	$m = 6$ 收敛更快
	10	6,8,10	Yes	$m = 8$ 收敛更快
	12	4	Yes	-
	14	2	Yes	-
角度	8	6,8	No	$m = 8$ 收敛更快
	10	6,8	No	$m = 6$ 收敛更快
	12	4,6	No	$m = 6$ 收敛更快
	14	4	No	-

表格 A6：存在约束数据编码和试探态宽度时的建议

约束		建议		备注
数据编码	拟设深度 (m)	最优宽度（n）	纠缠包含	
振幅	2	6,10	Yes	$n = 6$ 收敛更快
	4	12	Yes	-
	6	8	Yes	-
	8	6,8,10	Yes	$n = 6$ 收敛更快
	10	6,8	Yes	$n = 6$ 收敛更快
角度	2	10,12,14	No	$n = 6$ 收敛更快
	4	10,12,14	No	
	6	14	No	-
	8	12	No	-
	10	12,14	No	$n = 14$ 收敛更快

表格 A7：存在约束数据编码和拟设深度时的建议

约束		建议		备注
数据编码	纠缠包含	最优宽度 (n)	最优深度（m）	
振幅	Yes	6	8,10	$m = 10$ 收敛更快
		8	6,10	$m = 6$ 收敛更快
振幅	No	6	2	-
		8	4,6	相似收敛率 for 两种推荐深度
角度	Yes	8	8	-
		12	10	-
		14	6	-
角度	No	10	2,4	$n = 12$ 收敛更快
		12	2,4,8	
		14	2,4,6	

表格 A8.：存在约束数据编码和纠缠时的建议

约束		建议		备注
纠缠包含	拟设宽度 (n)	最优深度 (m)	Data 编码	
Yes	8	6	振幅	-
Yes	10	8,10	振幅	$m = 10$ 收敛更快
Yes	12	4	振幅	-
Yes	14	4,6	角度	$m = 6$ 收敛更快
No	8	2,4,6	角度	相似收敛所有 reco- 的速率推荐宽度，但 $n = 12$ 收敛略快
No	10	2,4	角度	
No	12	2,4,8	角度	
No	14	2,4,6	角度	

表格 A9.：存在约束纠缠和试探态宽度时的建议

约束		建议		备注
纠缠包含	拟设深度（m）	最优宽度（n）	Data 编码	
Yes	2	10	振幅	-
		10,12	角度	$n = 10$ 收敛更快
Yes	4	12	振幅	-
		14	角度	-
Yes	6	8	振幅	-
Yes	8	8,10	振幅	$n = 8$ 收敛更快
Yes	10	8	振幅	-
No	2	10,12,14	角度	几乎相同的收敛率，适用于所有宽度，但 $n = 12$ 收敛稍快
No	4	12,14	角度	

表 A10：纠缠和拟设深度受约束时的建议

A.4. 对全部三个参数的约束

约束			建议	备注
数据编码	拟设深度（米）	纠缠包含	最优宽度 (n)	
振幅	2	Yes	6,10	$n = 10$ 收敛更快
	4	Yes	12	-
	6	Yes	8	-
	8	Yes	6	-
	10	Yes	6,8	$n = 6$ 收敛更快
	2	No	6	-
	4	No	6,8	$n = 6$ 收敛更快
	6	No	6,8	
	8	No	6	-
	10	No	6	-
角度	2	Yes	10,12	$n = 10$ 收敛更快
	4	Yes	14	-
	6	Yes	10,14	$n = 10$ 收敛更快
	8	Yes	8,10	$n = 8$ 收敛更快
	10	Yes	8,12	$n = 12$ 收敛更快
	2	No	10,12,14	所有宽度几乎都有相同的收敛速率 但 $n = 12$ 会收敛稍快一些
	4	No	12,14	
	6	No	14	
	8	No	12	
	10	No	10,12,14	

表 A11：存在数据编码、拟设深度和纠缠约束时的建议

约束			建议	备注
数据编码	拟设宽度（n）	纠缠包含	最优深度（m）	
振幅	6	Yes	8,10	$m = 10$ 收敛更快
	8	Yes	6,10	$m = 6$ 收敛更快
	10	Yes	8,10	$m = 10$ 收敛更快
	12	Yes	4	-
	14	Yes	4	-
	6	No	2	-
	8	No	4,6	$m = 4$ 收敛更快
	10	No	2	-
	12	No	2	-
角度	8	Yes	6,8	$m = 8$ 收敛更快
	10	Yes	6,8	$m = 6$ 收敛更快
	12	Yes	2,4,6,8,10	$m = 6$ 收敛更快
	14	Yes	4,6,8	$m = 8$ 收敛更快
	8	No	2,4,6	所有宽度几乎都相同的收敛速率但 $m = 2$ 将会更好，因为训练时间更短训练时间
	10	No	2,4	
	12	No	2,4,8	
	14	No	2,4,6	

表 A12：存在数据编码、试探态宽度和纠缠约束时的建议