

Peer Review File

Quantum deep generative prior with programmable quantum circuits



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

See attached pdf file.

Reviewer #2 (Remarks to the Author):

This paper investigates the application of parameterized quantum circuits (PQCs), employing angle encoding and data re-uploading, to generate priors for various image reconstruction and enhancement tasks. The research builds upon the Deep Generative Prior (DGP) framework by Pan et al., wherein the generator $G(z)$ of a pre-trained GAN, along with the latent code z , is fine-tuned to yield the desired output. In this work, the latent code z is derived from the output of the PQC. The methodology is explored across several domains, including Ghost Imaging (GI), and four computer vision tasks: category transfer (CT), inpainting (I), colorization (C), and super-resolution (SR).

The authors claim that in GI, their method leads to better reconstruction and out-of-distribution generalization. They hypothesize this can be explained by the physics-informed nature of the model; they also highlight how PQCs can generate non-trivial latent codes, hard to simulate classically. Regarding the four computer vision tasks, they claim their method shows better or comparable performance with respect to its classical counterpart, evaluated by common similarity measures as SSIM and PSNR. They also show how their method is robust under quantum noise.

This paper finds a place among the quantum-enhanced algorithms, in which a part of a classical algorithm is delegated to a quantum circuit. In this regard, this approach is novel and interesting, and explores new ways to use PQCs to possibly enhance classical models. The results shown in this paper are promising, specifically regarding the ability of PQCs to better generalise to unseen data. The resistance to quantum noise makes this method particularly useful in practice.

On the downside, some of the claims made by the authors appear to lack solid grounding. For GI, CT, and C, it is not clear if the claims of better performance are based only on the samples shown in the paper. In such cases, such a limited dataset is insufficient to assert the superiority of a model. Moreover, the fact that the model is 'potentially' better, without robust statistical backing, does not constitute a compelling result. It is important to mention, however, that I and SR are evaluated on the ImageNet validation set (1k samples), hence providing a more robust claim. On top of this,

although explanations are always reported as hypothetical, it is not very clear how the physics informed model constitutes a good prior for image reconstruction, and a more in-depth explanation of this mechanism, if possible, would be helpful.

In conclusion, I consider this paper to be valid and of interest, but I believe the concerns, particularly regarding the limited number of samples, should be addressed before publication, or the claims should be moderated. If my interpretation is incorrect and more samples have indeed been used, it would be beneficial for the paper to articulate this more transparently, and report this quantitatively.

Some typos:

215: hat on the 'x'

225: parentheses in the formula

234: "conducte"

494: "does not directly responsible"

Reviewer #3 (Remarks to the Author):

Major Claims:

The Deep Generative Prior [48] is an existing architecture for image processing (denoising, colorizing, infilling...) that uses a large pretrained GAN as a starting point for image processing; to process an image, the model is given further per-image finetuning, both of the latent space and the model parameters, to produce a cleaned image. The main goal of this paper is to improve the Deep Generative Prior (DGP) by changing the latent space: instead of a latent space with a Gaussian distribution, the latent space is specified as the outputs of a parameterized quantum circuit (PQC).

The claim is that this circuit adds expressivity to the model through new parameters (the parameters of the quantum circuit), which leads to better quality image denoising. The quantum component could also, in principle, lead to latent space vectors that may be hard to find classically, allowing a quantum computer to lend some of its power to this algorithm. This appears to be a novel approach, and I expect it would most interest those working in quantum machine learning.

There is a secondary topic in this paper, which concerns the use of their model for Ghost Imaging (GI). Ghost Imaging is an imaging technique that uses quantum effects to enhance sensitivity in low-light environments, but it comes with its own noise model. GI forms an additional good, purpose-motivated test problem for denoising images, and it's great to explore how (Q)DGP performs on this noise model. But the fact that GI involves quantum effects has no connection to the quantum architecture of QDGP, no more than a GPU being made of silicon would make it more adept at denoising photographs of silicon. This double "quantumness" in the paper is potentially confusing or even misleading.

Are the claims convincing:

I seriously disagree with the statement that the PQC latent space improves the *expressiveness* of the model. In the BigGAN model underlying both DGP and QDGP, z is a latent vector of 120 parameters. The PQC used seems to be a circuit with $3NLk = 1440$ parameters (lines 135-138 of the supplementary material, although the precise number is not entirely clear). But, no matter how many parameters there are, these 1440 parameters get compressed to a 120-parameter vector again for feeding into the BigGAN. The PQC cannot add expressiveness, because in the end it is just a complicated formula to produce the parameters of the original model.

What is a real possibility is that the PQC improves the trainability of the model. The denoising process involves navigating a loss landscape towards a de-noised image that remains close to the original, and it's possible that the PQC changes the loss landscape to favor landing in better images. Indeed, the earlier DGP and DIP models were trying to do exactly this: DIP is the idea that the structure of a CNN leads to a better loss landscape, and DGP took that idea further by baking more knowledge of typical images into a GAN via pretraining.

But this would be extraordinary to actually occur here. CNNs, and even moreso pretrained GANs, are models specifically designed to capture structures of images. By design, they include information about what typical images look like, and that could shape the loss landscape for better training is both plausible and born out by experiment. PQCs, on the other hand, are quantum processes on small numbers of qubits. They have no connection to images, latent spaces, or anything that would make me expect an improved loss landscape.

The main effect I would expect is a different effective learning rate. Depending on how precisely the parameters are setup and the loss determined, training the PQC could lead to an effective change in z that's a few times faster or slower than what it was in the standard DGP training process. One would expect this to have a small but nonzero impact on accuracy. Since the original DGP paper did

not pick a separate learning rate for the latent space variables, it quite likely that some scaling of the learning rate (either up or down) would improve the accuracy -- it is unlikely to be optimal already.

The analysis presented (Figure 2; Table I; Table II) shows that the QDGP and DGP perform very similar, with the QDGP performing slightly better on most metrics. But I suspect this is primarily due to the different learning rate improving the training schedule mostly by accident. The metrics given also don't give any sense of the uncertainty in the quantities: is a difference of 16.95 vs 17.22 statistically significant (Table I)? Compared to other values given (in the authors' paper and ref [48]), this difference seems small.

How to strengthen the paper:

The main experiment I would want to see done is this.

- Measure the typical learning rates of the latent space variables in the DGP and the QDGP (2-norm travelled per iteration).
- Call the ratio between these two rates k .
- Run the DGP model where the latent space variables are scaled to be k times more sensitive. That is, $z = z_0 + k^*(dz)$, where dz are the new trainable parameters initialized at 0, and z_0 is in the initial latent vector (now held fixed).
- Compare this rescaled DGP model with the QDGP.

It seems likely that this one additional hyperparameter closes the gap between QDGP and DGP. If it doesn't, and QDGP outperforms DGP, then some error bars on the analysis and the accuracy will be needed to provide a convincing argument that this is actually improving upon the model somehow.

Separately, the Ghost Imaging problem as a denoising problem seems quite useful as a benchmark on these models, and the physics-forward noise model deserves more discussion. I think that could make a fine paper on its own, with or without the QDGP.

Technical notes:

In equation (5), two different loss terms (a total square error, and a total variation) are added together. While these do in principle have the same units and could be added, it seems surprising

to just add them with no weighting. I would expect something like $\|B-Hx\|^2 + \lambda * \|x\|_{TV}^2$ where λ is a hyperparameter. Looking at their code, it seems this was the actual form they used, with $\lambda = 1e-5$; also, the TV loss is the anisotropic TV loss, and without a squaring either. I believe it should read,

$$\|B-Hx\|^2 + 1e-5 * \|x\|_{TV, aniso}$$

or similar, with the coefficient and no exponent on the TV term.

As said in line 657, they state that the class label is set to a random embedding hidden vector. On line 668, they say they use 500 latent codes z and pick the one with lowest loss as the initial guess. These both match exactly the strategy in [48] and I think that citation should be repeated here. (The number 500 seems particular.)

The present work, that we find well written and presented, aims at improving on existing classical GANs by “quantizing” the latent space, whose samples are fed into a classical generator to synthesize fake images belonging to some desired distribution that depends on the task at hand. The quantization is done by using an input latent code $\mathbf{z}$ whose components z_i are obtained measuring the output state of trainable PQC U_θ as $z_i = \text{Tr}[\rho_0 U_\theta^\dagger Z_i U_\theta]$, where $\rho_0 = (|0\rangle\langle 0|)^{\otimes N}$, N is the number of qubits used, and Z_i is the Pauli Z operator acting on the i -th qubit. This quantum prior leads, as proved by the authors’ numerical studies, to results that outperform conventional, classical methods, which instead rely on sampling $\mathbf{z}$ from Gaussian distributions. This improvements can be attributed to the higher expressibility that quantum distributions have with respect to the latter, hence in principle justifying the use of a quantum device and adding to the always increasing family of hybrid quantum-classical ML techniques.

However, the paper does not consider the problem of scalability of the quantum prior, leading to our following major concern:

- Recent advancements [1] in the understanding of the long-known barren-plateaus phenomenon [2, 3] show how the hardware efficient ansatz used as the trainable agent responsible for the quantum prior, together with the latter being a collection of local quantum expectation values will not be trainable unless its depth is logarithmic in the number of qubits used. Sadly, in the trainable regime the use of a quantum device will be pointless, as the quantum prior could be as easily simulated on a classical machine. This is exactly the regime the authors find themselves working in, as they deal with $N = 20$ qubits and $L = 4$ layers of single qubit rotations followed by a ladder of entangling CZ gates. Particularly, their results could have been obtained by resorting to Tensor Network methods, representing the parameterized state as an MPS since the output bond dimension would have been at most $\chi_{\max} = 2^L = 16$ (4^L if more general 2-qubits gates were being used), allowing for the complete simulation of all the 120 qubits needed for sampling the BigGAN latent code.

The role played by the quantumness of the latent space is therefore at least unclear, if not even null. However, making the PQC deeper to explore more quantum codes would lead to its untrainability.

We are afraid that there is no way out of this issue with the present framework, and therefore we believe that the paper needs a major restyling before publication, probably parting ways from the parameterized quantum circuit approach.

Minor comments:

- Peak Signal-to-Noise Ration (PSNR) and Structural Similarity (SSIM) are not defined, consider adding a small paragraph defining them.
- The claim starting at line 163 needs a citation.
- The parentheses of the in-line equation at line 225 need fixing.
- In Fig. 5, the authors should set the colors’ α to some value less than 1, since the orange bars completely cover the blue ones and it is impossible to tell what the initial distributions looked like.
- While the main text is well written, we feel like the Supplementary Notes could use some polishing. Besides some typos: *Paulu-Z gate* in Fig. 1’s caption, D instead on N when defining the number of qubits in line 134; we feel like the Supplementary Notes read worse and could be presented better.

References

- [1] M Cerezo, Martin Larocca, Diego García-Martín, N L Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, et al. Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing. *arXiv preprint arXiv:2312.09121*, 2023.
- [2] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1–12, 2021.

- [3] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):1–6, 2018.

Response letter for paper *COMMSPHYS-23-1349*

Quantum deep generative prior with programmable quantum circuits,
Tailong Xiao, Xinliang Zhai, Jingzheng Huang, Jianping Fan, Guihua Zeng

Dear Editor and Reviewers:

First of all, we are very grateful to you for reviewing our manuscript entitled "Quantum deep generative prior with parameterized quantum circuits (previous)". We sincerely appreciate your comments and they really help to improve the quality of our article. According to the reviewers' comments, corresponding amendments have been made in the revised manuscript. We believe that the revised one is suitable for further consideration on *Communications Physics*.

In the following, we first highlighted each comment, then we provided a reply, and finally we briefly described how the manuscript was adjusted in the main text. Below, we respond to the referees' comments and summarize the main changes:

- We provide extensive experiments, including ghost imaging and classical image processing and manipulation tasks, to demonstrate the potential utility of quantum machine learning (QDGP) compared to DGP and other methods. For each task, we conduct repeated experiments to calculate the mean and standard deviation, and report these values in the revised manuscript. In the super-resolution task, although the difference between QDGP and DGP is small, the standard deviation is also small, indicating that the advantage is not due to statistical fluctuations. From the numerical results, it is evident that QDGP consistently demonstrates potential advantages over DGP. Therefore, the conclusions in the original manuscript are further strengthened and validated. Additionally, we provide a detailed analysis of the PQC-expectation model in characterizing universal continuous multi-mode distributions, highlighting the superior expressivity of PQC-induced latent codes over classical codes.
- We also analyze the numerical results when using matrix product states and state-vector-based parameterized quantum circuits (PQCs). The number of layers is increased from 6 to 10 to observe the Barren Plateau effect. The loss curves for different layers are recorded. We do not observe the Barren Plateau effect in deeper layers, but the performance decreases when using state-vector-based PQCs. In contrast, the performance does not decrease when using matrix-product state-based PQCs, as we restrict the entanglement of quantum states

by setting a relatively small bond dimension. Additionally, we revise the discussion on the quantum computational advantage of QDGP according to the suggestions of Reviewer 1, focusing on the utility of PQCs in enhancing out-of-distribution generalization ability.

- We provide descriptions of the PSNR and SSIM metrics in the main text. We add the recommended reference and relevant references in the comments into the main text. Additionally, we have revised the figures of latent code distribution by incorporating the parameter α . Various typos in the equations and text have been corrected according to the reviewers' suggestions. Furthermore, we have thoroughly proofread the entire document, including the main text and supplementary notes, to ensure that all typos and grammatical errors have been corrected.

We deliberate that the main changes have fully addressed the comments, including the potential applications and way out of the PQC-based QDGP method, the statistical errors in different image denoising tasks, the additional experiments on learning rate analysis, and all the typos and grammatical errors in the main text and supplementary notes. We hope the revised manuscript clarifies any confusions and meets your expectations. We look forward to hearing from you soon.

Best regards,

All authors

REVIEWER # 1

COMMENT # 1.1

The present work, that we find well written and presented, aims at improving on existing classical GANs by quantizing the latent space, whose samples are fed into a classical generator to synthesize fake images belonging to some desired distribution that depends on the task at hand. The quantization is done by using an input latent code z whose components z_i are obtained measuring the output state of trainable PQC U_θ as $z_i = \text{Tr}[\rho_0 U_\theta^\dagger Z_i U_\theta]$, where $\rho_0 = (|0\rangle\langle 0|) \otimes N$, N is the number of qubits used, and Z_i is the Pauli Z operator acting on the i -th qubit. This quantum prior leads, as proved by the authors numerical studies, to results that outperform conventional, classical methods, which instead rely on sampling z from Gaussian distributions. This improvements can be attributed to the higher expressibility that quantum distributions have with respect to the latter, hence in principle justifying the use of a quantum device and adding to the always increasing family of hybrid quantum-classical ML techniques.

Thank you for your positive feedback. One of the primary contributions of our work involves leveraging parameterized quantum circuits (PQCs) as the latent code generator, as you have described. In contrast to conventional Gaussian priors, the continuous variable distribution generated by PQCs may be significantly more versatile, as demonstrated by our numerical results showing benefits in untrained ghost imaging and classical image manipulation tasks. We wholeheartedly agree with your observation that these improvements can be attributed to the higher expressibility of quantum distributions compared to conventional Gaussian distributions. Indeed, this was our original motivation.

Moreover, while many current PQCs-based machine learning algorithms focus on exploring theoretical advantages and limitations such as expressibility [1, 2] and trainability [3-6], recent studies have shown that PQCs-based algorithms may possess parameter complexity advantages in function approximation when compared to classical deep neural networks [7]. The investigation of practical advantage of PQCs in real applications is very rare.

Nevertheless, our emphasis lies in the development of practical PQCs-based machine learning algorithms and techniques. We believe that practical, especially hybrid quantum-classical machine learning algorithms, may offer utility or advantages over conventional methods in the NISQ era. Unlike quantum linear algebra enhanced machine learning algorithms, PQCs-based algorithms can be readily deployed on noisy quantum devices. In reality, we find that PQCs-based supervised machine learning has potential advantage in blind ghost imaging in our previous work.

We maintain an optimistic view that PQCs-based machine learning can be integrated into classical machine learning algorithms to carve out their unique niche. While identifying a problem or application where PQCs-based algorithms demonstrate a definitive quantum advantage capable of entirely replacing classical algorithms may be challenging, hybrid quantum-classical algorithms can collaborate to enhance practical performance. Our algorithm is specifically designed with this consideration in mind, utilizing PQCs as the prior generator rather than the backbone of the machine learning model, making it highly suitable for implementation on real quantum processors.

Furthermore, PQC-induced quantum distributions can be seamlessly integrated into existing generative AI models such as variational quantum encoders (VAEs) to enhance performance.

- [1] Du, Yuxuan, et al. "Efficient measure for the expressivity of variational quantum algorithms." *Physical Review Letters* 128.8 (2022): 080506.
- [2] Haug, Tobias, Kishor Bharti, and M. S. Kim. "Capacity and quantum geometry of parametrized quantum circuits." *PRX Quantum* 2.4 (2021): 040309.
- [3] Wang, Samson, et al. "Noise-induced barren plateaus in variational quantum al-

gorithms." Nature communications 12.1 (2021): 6961.

[4] Cerezo, Marco, et al. "Cost function dependent barren plateaus in shallow parametrized quantum circuits." Nature communications 12.1 (2021): 1791.

[5] Heyraud, Valentin, et al. "Efficient estimation of trainability for variational quantum circuits." PRX Quantum 4.4 (2023): 040335.

[6] McClean, Jarrod R., et al. "Barren plateaus in quantum neural network training landscapes." Nature communications 9.1 (2018): 4812.

[7] Yu, Zhan, et al. "Provable advantage of parameterized quantum circuit in function approximation." arXiv preprint arXiv:2310.07528 (2023).

COMMENT # 1.2

However, the paper does not consider the problem of scalability of the quantum prior, leading to our following major concern:

- Recent advancements [1] in the understanding of the long-known barren-plateaus phenomenon [2, 3] show how the hardware efficient ansatz used as the trainable agent responsible for the quantum prior, together with the latter being a collection of local quantum expectation values will not be trainable unless its depth is logarithmic in the number of qubits used. Sadly, in the trainable regime the use of a quantum device will be pointless, as the quantum prior could be as easily simulated on a classical machine. This is exactly the regime the authors find themselves working in, as they deal with $N = 20$ qubits and $L = 4$ layers of single qubit rotations followed by a ladder of entangling CZ gates. Particularly, their results could have been obtained by resorting to Tensor Network methods, representing the parameterized state as an MPS since the output bond dimension would have been at most $\chi_{\max} = 2^L = 16$ (4^L if more general 2-qubits gates were being used), allowing for the complete simulation of all the 120 qubits needed for sampling the BigGAN latent code. The role played by the quantumness of the latent space is therefore at least unclear, if not even null. However, making the PQC deeper to explore more quantum codes would lead to its untrainability.

We are afraid that there is no way out of this issue with the present framework, and therefore we believe that the paper needs a major restyling before publication, probably parting ways from the parameterized quantum circuit approach.

[R1] M Cerezo, Martin Larocca, Diego García-Martín, N L Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, et al.

Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing. arXiv preprint arXiv:2312.09121, 2023.

[R2] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. Nature Communications, 12(1):112, 2021.

[R3] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. Nature Communications, 9(1):16, 2018.

Thanks very much for your constructive and fruitful comments. We want to reply your comments from two aspects: conceptual level and technique level.

From conceptual level:

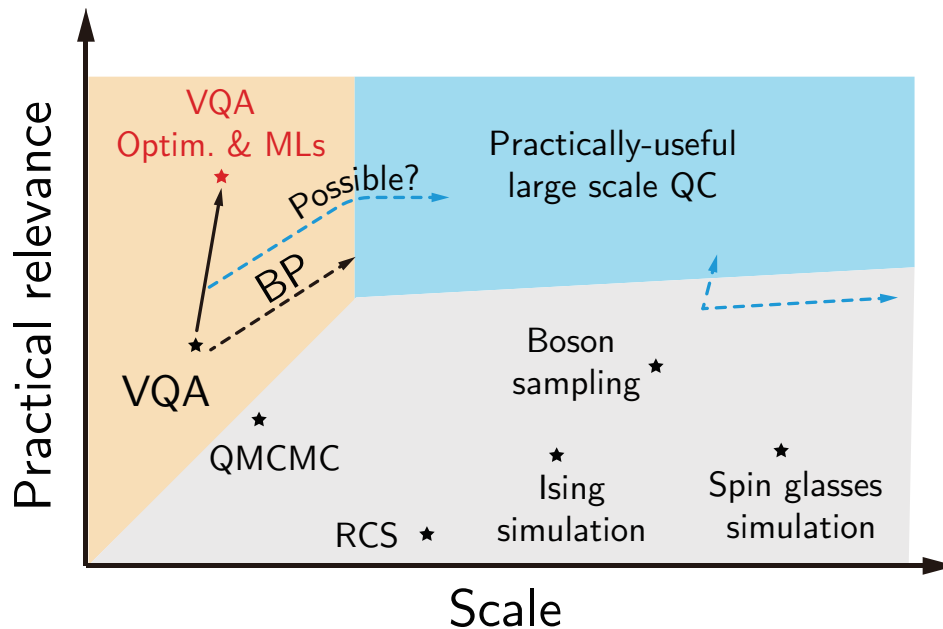


Figure 1: Gray shaded area has good scalability but has relative lower practical relevance. Yellow shaded area has good practical relevance but its scalability is seriously limited by the BP problem. QMC MC is quantum enhanced MCMC sampling [6], RCS is random circuit sampling [7], Ising simulation is the IBM's experiment [8], Boson sampling is the Xanadu's experiment [9], Spin glasses simulation is the D'wave experiment [10].

Seeking the quantum advantage of quantum computation is a long-standing challenge. Since quantum algorithms with provable and practically useful algorithms such as Shor and Grover have been proposed, few algorithms have been proposed that have the capability to show theoretical quantum advantages especially on noisy

quantum computers. Investigating quantum algorithm is extremely hard, which renders the studies alone for few quantum experts. As Cerezo et al. in Ref [R1] states that the arrival of variational quantum algorithms (VQAs) effectively democratized the world of quantum computing. VQAs can be readily implemented on current noisy quantum algorithms, therefore we can conduct experiments with real quantum hardware to investigate the quantum advantage. A most straightforward idea is that we can design a large-scale VQA with large qubits and depth so that cannot be simulated by classical computers, we can safely conclude that we achieve quantum advantage. Unfortunately, VQA has its own limitations for example, naive designed VQAs suffer from Barren Plateau [R3]. The potential of quantum computational advantage in VQAs is thus limited.

We also notice the impressive result in [R1], which states that shallow hardware-efficient VQAs with local observable as the loss function does not suffer from Barren Plateau but it can also be classically simulated by using matrix product states (MPS) with the overhead scales with 2^L . As you have stated that we cannot increase the depth of our VQA as it will induce the Barren Plateau problem. Therefore, the quantum computational advantage seems to be ruled out. To be frank, the starting point or motivation of our proposed PQC-based quantum machine learning algorithm is not seeking the quantum computational advantage. We argue that finding the quantum computational advantage in noisy quantum computers is hard.

Recently, IBM research team investigates the quantum simulation of quantum Ising model on 127 qubits quantum computer. After their claim of quantum advantage, Tindall et al. propose a tensor-network based classical algorithm that can efficiently simulate the quantum Ising model with the same scale with IBM's work [1]. Boson sampling experiments have also been analyzed that depth is less than quadratic scaling of lattice spacing, the highly efficient simulation is possible within exponentially small error. The merits of Boson sampling is its no demanding for variational training. Therefore, increasing the depth sufficiently deep, the quantum computational advantage is still possible [2]. Besides, Google's quantum supremacy experiments with 72 qubits in 2019 have again been dequantized also by tensor-network based classical algorithm [3, 4]. Going back even further, Tang et al. [5] propose a series of dequantized algorithms that can achieve the same speedup with quantum algorithms under appropriate theoretical assumptions. Consequently, although many efforts have been paid, it is still very hard to obtain a definite quantum advantage in current devices. However, it is undeniable that these results are beneficial for understanding the boundary of quantum and classical computation complexity. Besides, these promising quantum algorithms either designed for quantum simulation or dedicated optimization as shown in Figure 1. As the scale and quality of quantum computers continually improve, these dedicated quantum algorithms may be more practically-useful and simultaneously showing quantum computational advantage.

On the other hand, we try to find the practical utility of PQC-based quantum machine learning algorithms in classical applications when considering hybrid algorithmic architecture. Although our proposed hardware-efficient VQA may not be scaled to large depth as the efficient training limitation to possess the computational advantage. However, our proposed hybrid quantum machine learning algorithm can enrich the family of quantum machine learning algorithms and provides numerical evidence that PQCs-based quantum machine learning is still useful in noisy quantum computers. Although one may think that the only advantage of quantum computer is its computational power compared to classical computer, we think that PQCs-based quantum algorithms may provide a brand-new algorithmic method which can be integrated into classical models to improve performance. As we go deeper into VQA, it is still possible to design advanced optimization methods and architectures to overcome Barren Plateau. As we understand, the Barren Plateau problem of VQAs is deduced under the worse case situation. The expressiveness is large but the trainability is small. We may seek a trade-off between these two factors to further improve VQAs performance when considering specific problems. Recently, overparameterization [11] and correlated parameter learning [12] provide a promising route that tackle the problem of Barren Plateau. Besides, Park et al. propose that hardware-efficient quantum circuits can be trained without Barren Plateau in any depth [13-15]. Rudolph et al. claim that MPS synergistic pretraining can mitigate BP problem when the qubit size is up to 100 [17]. Holmes et al. find that a good warm start may benefit for VQA training [16]. In essence, this proposal finds a good way to balance the expressiveness and trainability. Therefore, PQC-based algorithm can be flexibly scaled and used for future fault tolerant quantum computation and be applied into practically useful applications such as optimization, machine learning and quantum chemistry [18].

From technique level:

In our algorithm, we make use of PQC to generate the latent space with 120-dimensions. Suppose we have a 120-qubits quantum computer, we can design 4 or slightly larger layers to generate the quantum latent code in a naive way. This PQC can be dequantized by MPS with appropriate bond dimension (< 20). However, using the methods in [13, 14], the depth may be even larger so that the overhead of classical simulation is also considerable. We note that increasing the depth is not always a good way in terms of practical performances. In addition, when the required dimension of quantum latent code is larger, the overhead of classical simulation will be further increased. We mention that although the exponential computational advantage does not exist in shallow depth, we could expect the polynomial advantage when estimating the loss function compared to quantum inspired parameterized loss function (MPS) [R1]. In classical setting, polynomial speedup sometimes is also notable. On the other hand, it has been shown that shallow quantum circuits are un-

conditionally more powerful than shallow classical circuits [19-21] in basis-sampling based problems.

In addition, PQC can also be used to generate universal continuous multivariate distributions [20]. As stated in [22], we find that PQC with specified number of qubits $n \in \Theta(M)$ and 1 observable spectrum norm can approximate any M -dimensional continuous distributions. One can also use dense encoding i.e. $n \in \Theta(\log M)$ with $k = \Theta(M)$ observable spectrum norm to generate universal M -dimensional continuous distributions. Although currently, the expectation sampling model can be simulated with classical algorithms such as MPS, it is still possible to construct stronger advantage arguments where we find the existence of an expectation value sampler such that its output distribution cannot be sampled by any polynomial-time randomized Turing machine (subject to standard assumptions) [20]. Therefore, from practical view, PQC has larger flexibility to approximate continuous distributions, which can be used in generative artificial intelligence.

As for your suggestion of probably parting ways from the parameterized quantum circuit approach, we indeed understand your intention. We will weaken the expressions of quantum computational advantage since it is not the focus of our work. Our contribution is to propose a novel hybrid quantum(-inspired if using MPS simulation) algorithm that has practical performance improvement. Furthermore, we also propose a physical forward enhanced quantum machine learning algorithm, which can also be called quantum untrained machine learning algorithm. Compared to VQE which minimizes the observable expectation variationally, our loss function is totally different. For ghost imaging and classical image manipulation, the loss function is calculated by physical model or the pretrained discriminator. To our knowledge, our proposed quantum machine learning algorithms have not been studied before. Therefore, it may provide new references and inspiration for algorithmic research in quantum machine learning. Besides, we revise the title of our manuscript. We change the parameterized quantum circuits into programmable quantum circuits which is a more general concept.

- [1] Tindall, Joseph, et al. "Efficient tensor network simulation of ibms eagle kicked ising experiment." PRX Quantum 5.1 (2024): 010308.
- [2] Oh, Changhun, et al. "Classical simulation of boson sampling based on graph structure." Physical Review Letters 128.19 (2022): 190501.
- [3] Pan, Feng, Keyang Chen, and Pan Zhang. "Solving the sampling problem of the sycamore quantum circuits." Physical Review Letters 129.9 (2022): 090502.
- [4] Pan, Feng, and Pan Zhang. "Simulation of quantum circuits using the big-batch tensor network method." Physical Review Letters 128.3 (2022): 030501.
- [5] Tang, Ewin. "Dequantizing algorithms to understand quantum advantage in machine learning." Nature Reviews Physics 4.11 (2022): 692-693.

- [6] Layden, David, et al. "Quantum-enhanced Markov chain Monte Carlo." *Nature* 619.7969 (2023): 282-287.
- [7] Arute, Frank, et al. "Quantum supremacy using a programmable superconducting processor." *Nature* 574.7779 (2019): 505-510.
- [8] Kim, Youngseok, et al. "Evidence for the utility of quantum computing before fault tolerance." *Nature* 618.7965 (2023): 500-505.
- [9] Arrazola, Juan M., et al. "Quantum circuits with many photons on a programmable nanophotonic chip." *Nature* 591.7848 (2021): 54-60.
- [10] King, Andrew D., et al. "Computational supremacy in quantum simulation." *arXiv preprint arXiv:2403.00910* (2024).
- [11] Larocca, Martin, et al. "Theory of overparametrization in quantum neural networks." *Nature Computational Science* 3.6 (2023): 542-551.
- [12] Volkoff, Tyler, and Patrick J. Coles. "Large gradients via correlation in random parameterized quantum circuits." *Quantum Science and Technology* 6.2 (2021): 025008.
- [13] Park, Chae-Yeun, and Nathan Killoran. "Hamiltonian variational ansatz without barren plateaus." *Quantum* 8 (2024): 1239.
- [14] Chae-Yeun Park, Minhyeok Kang, Joonsuk Huh. "Hardware-efficient ansatz without barren plateaus in any depth." *arXiv preprint arXiv:2403.04844* (2024).
- [15] Zambrano, Leonardo, et al. "Avoiding barren plateaus in the variational determination of geometric entanglement." *Quantum Science and Technology* (2023).
- [16] Drudis, Marc, Supanut Thanasilp, and Zoë Holmes. "Variational quantum simulation: a case study for understanding warm starts." *arXiv preprint arXiv:2404.10044* (2024).
- [17] Rudolph, Manuel S., et al. "Synergistic pretraining of parametrized quantum circuits via tensor networks." *Nature Communications* 14.1 (2023): 8367.
- [18] Cerezo, Marco, et al. "Variational quantum algorithms." *Nature Reviews Physics* 3.9 (2021): 625-644.
- [19] Bravyi, Sergey, David Gosset, and Robert König. "Quantum advantage with shallow circuits." *Science* 362.6412 (2018): 308-311.
- [20] Watts, Adam Bene, and Natalie Parham. "Unconditional quantum advantage for sampling with shallow circuits." *arXiv preprint arXiv:2301.00995* (2023).
- [21] S. Bravyi, D. Gosset, R. König, and M. Tomamichel, Quantum advantage with noisy shallow circuits, *Nature Physics* 16, 10401045 (2020).
- [22] Barthe, Alice, et al. "Expressivity of parameterized quantum circuits for generative modeling of continuous multivariate distributions." *arXiv preprint arXiv:2402.09848* (2024).

We conducted additional numerical experiments on the QDGP using deeper quantum layers with $L = 6, 8, 10$. Additionally, we performed optical experiments to collect the bucket signals, setting the number of measurements to 1024 and the image resolu-

tion to 64×64 . By fixing the sampling rates and resolution, we focused on analyzing the impact of quantum layers on reconstruction performance.

First, we plotted the loss curve of the optimization process and calculated the PSNR value of the QDGP under different quantum layers. We studied both randomly initialized and pretrained generators in QDGP. For comparison, we also performed experiments on QDGP using a matrix product state (MPS) with a relatively low bond dimension, set to 16. The circuit structure remained identical to the original PQC. The key difference is that MPS-based PQCs may not accurately characterize quantum states with complex entanglement. Evidence suggests that complex entanglement can induce the Barren Plateau phenomenon. Therefore, we restricted entanglement in the QDGP with an MPS backend to observe if the Barren Plateau occurs. The loss curve is shown in Figure2.

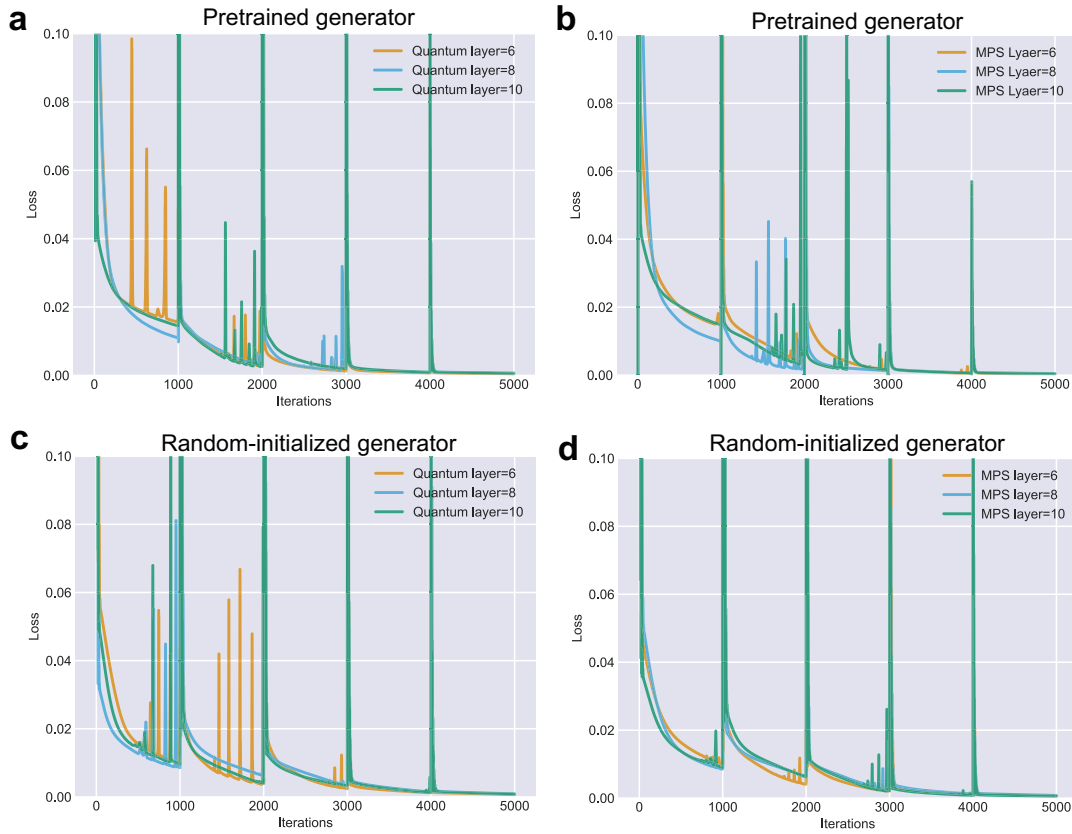


Figure 2: The loss curve of the optimization process both for MPS-based and state-vector simulator based PQC induced BigGAN model for ghost imaging. (a) The loss curve of QDGP with pretrained generator. (b) The loss curve of QDGP-MPS with pretrained generator. (c) The loss curve of QDGP with random initialized generator. (d) The loss curve of QDGP-MPS with random initialized generator.

Here is the polished version of your sentences:

We found that increasing the layers of PQCs, whether MPS-based or state vector-based, does not affect optimization convergence. However, with 10 layers, both types of PQCs exhibit a larger change in loss when the learning rate is adjusted during optimization. This is expected, as deeper PQCs have more quantum parameters, leading to greater variation in the induced distribution. When the learning rate changes, the quantum distribution induced by the PQC may vary more significantly than with fewer parameters. This behavior was observed in both MPS and state-vector-based PQCs. Importantly, our results indicate that QDGPs based on PQCs are not severely affected by the Barren Plateau problem.

Two potential reasons explain these results. First, the PQC is used to generate latent code for the subsequent BigGAN model, not as the backbone of the machine learning process. Second, our proposed algorithm employs a hybrid structure, where the generated quantum latent code is further processed by a classical neural network. Evidence shows that classical neural networks can reshape the loss landscape of quantum circuits [23, 24]. These factors may clarify why our hybrid deep quantum model is not affected by the Barren Plateau.

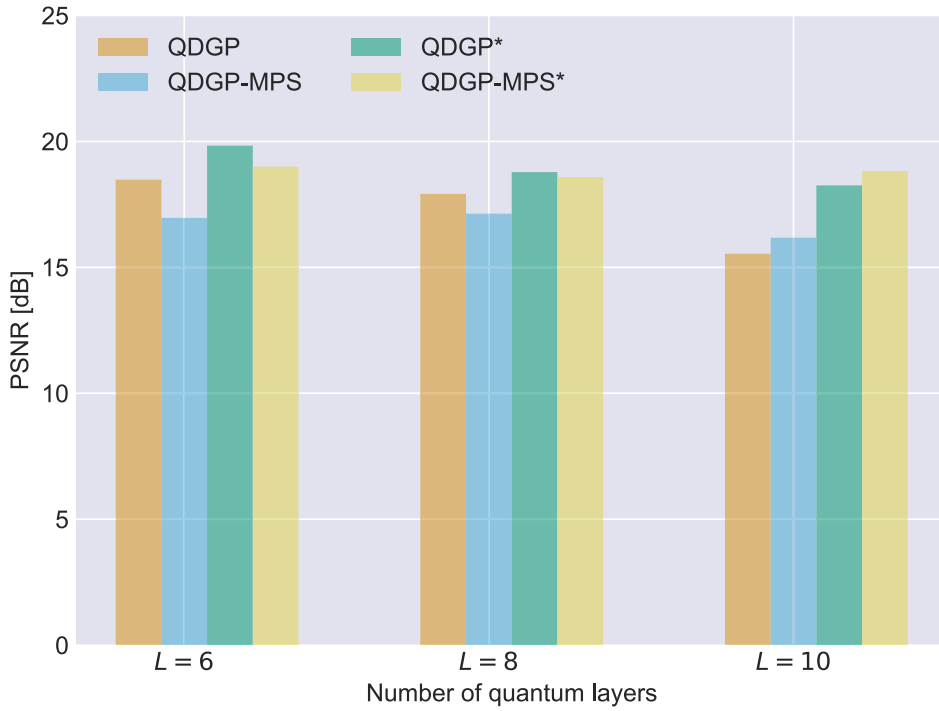


Figure 3: The PSNR value of images reconstructed by QDGP and QDGP-MPS algorithms with pretrained and random initialized generator.

Although the loss curves are similar, the MPS- and state-vector-based PQC-induced BigGAN models exhibit different reconstruction performances. Figure 3 presents the PSNR values of images reconstructed by different methods with various quantum

layers. As shown in Figure 3, increasing the layers of PQCs in QDGP slightly decreases the PSNR value. In contrast, the PSNR value of QDGP-MPS does not decrease with additional quantum layers. However, the PSNR value of the QDGP-MPS algorithm is slightly lower than that of QDGP when $L = 6$. When the quantum layers reach 10, QDGP-MPS surpasses the performance of QDGP. This indicates that deep quantum layers reduce the performance of QDGP, while QDGP-MPS has a higher layer threshold before showing the Barren Plateau phenomenon. Excessive entanglement can lead to Barren Plateau [25]. Since QDGP-MPS with a low bond dimension cannot accurately characterize quantum states with high quantum entropy, the Barren Plateau phenomenon does not occur in the QDGP-MPS algorithm.

When using a randomly initialized generator, the PSNR results are consistent with those found in the main text. Randomly initialized QDGP performs better than pre-trained models when reconstructing out-of-distribution samples. In this case, QDGP also shows performance decay with increasing quantum layers. Conversely, QDGP-MPS, by restricting the bond dimension and limiting the entanglement of quantum states, does not exhibit performance decay with additional quantum layers. These results suggest that, in practical use, programmable quantum circuits with appropriate quantum layers are preferable to deep quantum layers, as the latter are harder to train. MPS can be used to pretrain deep quantum circuits, which can then be fine-tuned using real quantum computers [26]. Additionally, MPS with proper bond dimensions can simulate the features of programmable quantum circuits, enhancing classical machine learning algorithms. MPS can also serve as an entanglement indicator. By analyzing performance with MPS at various quantum layers, one can infer the required quantum layers for real quantum computers based on the bond dimension. If an MPS-based PQC with many quantum layers does not exhibit Barren Plateau, it does not necessarily mean that a state-vector-based PQC with the same structure and layers will avoid Barren Plateau. Conversely, if a state-vector-based PQC does not exhibit Barren Plateau with many quantum layers, its respective MPS-based PQC will certainly not exhibit Barren Plateau.

[23] Friedrich, Lucas, and Jonas Maziero. "Avoiding barren plateaus with classical deep neural networks." *Physical Review A* 106.4 (2022): 042433.

[24] Rad, Ali, Alireza Seif, and Norbert M. Linke. "Surviving the barren plateau in variational quantum circuits with bayesian learning initialization." *arXiv preprint arXiv:2203.02464* (2022).

[25] Ortiz Marrero, Carlos, Mária Kieferová, and Nathan Wiebe. "Entanglement-induced barren plateaus." *PRX Quantum* 2.4 (2021): 040316.

[26] Rudolph, Manuel S., et al. "Synergistic pretraining of parametrized quantum circuits via tensor networks." *Nature Communications* 14.1 (2023): 8367.

We have added the above analysis into the Supplementary Note 3.

COMMENT # 1.3

Minor comments:

- Peak Signal-to-Noise Ration (PSNR) and Structural Similarity (SSIM) are not defined, consider adding a small paragraph defining them.
- The claim starting at line 163 needs a citation.
- The parentheses of the in-line equation at line 225 need fixing.
- In Fig. 5, the authors should set the colors' alpha to some value less than 1, since the orange bars completely cover the blue ones and it is impossible to tell what the initial distributions looked like.
- While the main text is well written, we feel like the Supplementary Notes could use some polishing. Besides some typos: Paulu-Z gate in Fig. 1s caption, D instead on N when defining the number of qubits in line 134; we feel like the Supplementary Notes read worse and could be presented better.

Thanks for your careful reading and useful advice. We have thoroughly revised the typos and grammar errors in the main manuscript. Besides, we also carefully rewritten the supplementary files. We believe the revised version can meet with your approval. We present the clarification in the below shaded box for you convenient reading. The revision is added into the main text.

As for the definition of PSNR and SSIM, we add them into the main text. The revised conent can be seen in the following:

(Main Text, Page 4, line 253-265)

... QDGP consistently outperforms DGP and DGI across common metrics such as Peak Signal-to-Noise Ration (PSNR) and Structural Similarity (SSIM). PSNR is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right),$$

where MSE is the mean square error between the noise-free image I and noisy image K with scale $m \times n$ and it can be written as

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2.$$

MAX_I is the maximum possible pixel value of the image. When the image are presented using pulse-code modulation with B bits, $MAX_I = 2^B - 1$. Besides, SSIM can be defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

where x, y are two images, μ_x, μ_y are the mean value of two images, σ_x^2, σ_y^2 are the variance of x, y , $\sigma_{x,y}$ is the covariance of x and y , $c_1 = (k_1 MAX_I)^2, c_2 = (k_2 MAX_I)^2$ are two variables to stabilize the division with weak denominator. Generally, we have $k_1 = 0.01, k_2 = 0.03$.

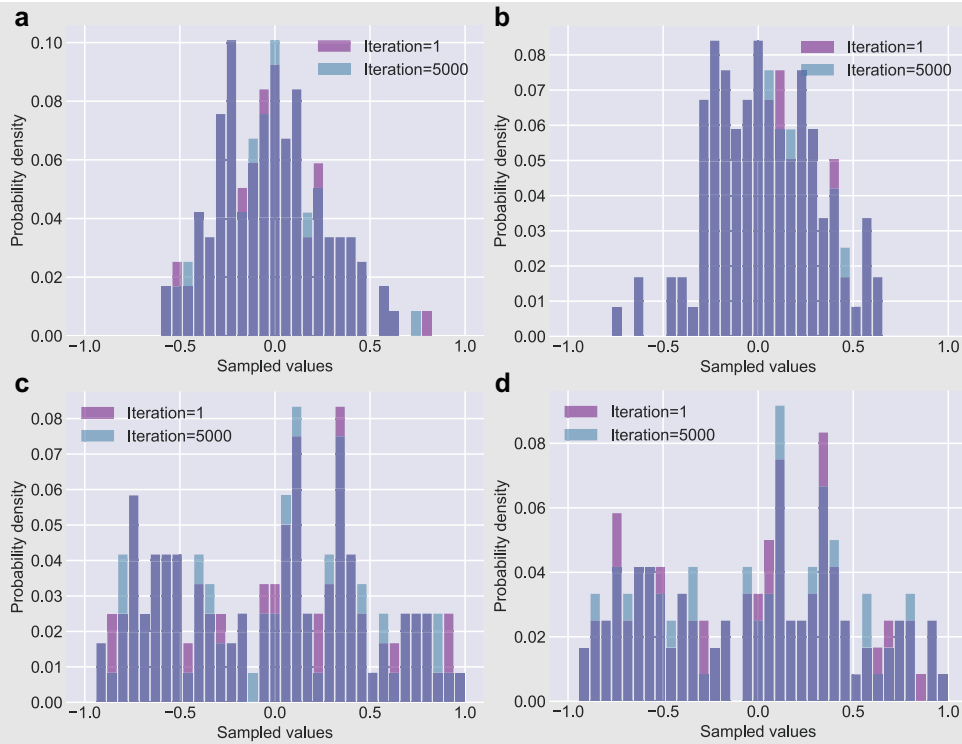
The equation in 225 has also been revised, which can be found in the following:

(Main Text, Page 3, line 225-228)

... where we define GAN reconstructed image as $\tilde{x} = G(\langle O(\theta) \rangle; \xi)$, λ is the coefficient hyperparameter. Upon convergence of the optimization process, we obtain $\tilde{x} = x^* = G(\langle O(\theta) \rangle; \xi)$

As for the third point, we thanks for the useful suggestion. We re-plot the distribution figure and set the color's α . The revised can clearly see the initial distribution and the optimized distribution, which can be found in the following:

(Main Text, Page 7, figure 5)



The probability density of DGP and QDGP with random and pretrained initialization. (a) The density of classical latent space for random initialized generator. (b) The density of classical latent space for pretrained generator. (c) The density of quantum latent space for random initialized generator. (d) The density of quantum latent space for pretrained generator.

As for the supplementary notes, we have rewritten the whole text. The pointed typos and grammar errors are revised. We have revised Paulu-Z gate into Pauli-Z gate. We have also defined the number of qubits as N rather than D .

We believe that our replies and corresponding revisions can explain your doubts clearly and meet with your satisfaction. We also expect your reply.

REVIEWER # 2

COMMENT # 2.1

This paper investigates the application of parameterized quantum circuits (PQCs), employing angle encoding and data re-uploading, to generate priors for various image reconstruction and enhancement tasks. The research builds upon the Deep Generative Prior (DGP) framework by Pan et al., wherein the generator $G(z)$ of a pre-trained GAN, along with the latent code z , is fine-tuned to yield the desired output. In this work, the latent code z is derived from the output of the PQC. The methodology is explored across several domains, including Ghost Imaging (GI), and four computer vision tasks: category transfer (CT), inpainting (I), colorization (C), and super-resolution (SR).

The authors claim that in GI, their method leads to better reconstruction and out-of-distribution generalization. They hypothesize this can be explained by the physics-informed nature of the model; they also highlight how PQCs can generate non-trivial latent codes, hard to simulate classically. Regarding the four computer vision tasks, they claim their method shows better or comparable performance with respect to its classical counterpart, evaluated by common similarity measures as SSIM and PSNR. They also show how their method is robust under quantum noise.

This paper finds a place among the quantum-enhanced algorithms, in which a part of a classical algorithm is delegated to a quantum circuit. In this regard, this approach is novel and interesting, and explores new ways to use PQCs to possibly enhance classical models. The results shown in this paper are promising, specifically regarding the ability of PQCs to better generalise to unseen data. The resistance to quantum noise makes this method particularly useful in practice.

First of all, we thanks very much for your positive feedback. The main contribution of our work the proposal of the hybrid quantum-classical quantum algorithm for classical tasks. As we have seen, many quantum machine learning algorithms have been proposed for quantum tasks, such as quantum phase classification, quantum compiling and quantum dynamic learning. These works focus on using quantum machine learning to tackle the important problems in quantum domain. However, quantum machine learning algorithms for large-scale classical tasks or real applications are still rare. This is exactly our starting point to study the hybrid quantum-classical machine learning for typical useful classical tasks. Considering the quality and scalability limitations in current quantum hardware, we do not make use of quantum circuits as the backbone of large-scale machine learnig models. However, PQCs are very suited for distribution learning such as quantum generative adversarial network (QGAN) [1]. Recently, we notice that PQCs-based sampling model is provably to be universal to approximate multivariate continuous distributions.

Therefore, we regard PQC as the quantum latent code generator to replace the original Gaussian classical prior. Under appropriate fine-tuning, PQC can be trained to match the prior of the specified downstream task. We deliberate that PQC and the pretrained generator model can be finetuned to conduct many useful applications. Besides, compared to Gaussian prior, we think quantum latent spaced enhanced generative model will perform better in ghost imaging and classical image manipulation tasks, particularly improving the out-of-distribution generalization.

In addition, since current quantum hardware is noisy, we consider the noise effects of quantum circuits on the final performance. We find that the performance is not sensitive to quantum noise. This result will render our proposed algorithm more feasible for current quantum computers as you have stated. Besides, considering the relative limitations of quantum computers, we highly agree with you on the hybrid quantum algorithm design. Current quantum algorithms need find their own place in machine learning or other applications. We do not expect that quantum algorithm can fully replace classical algorithms in near or even longer term since the latter is strong enough and also developed quickly. However, we do believe that PQC-based quantum machine learning algorithms have the unique feature that can be seamlessly integrated into classical models to improve their performance.

Finally, we note that PQCs cannot be simulated efficiently in large depth, since the overhead is increased exponentially as the depth linearly increases. This is a potential quantum computational advantage. Recently, there are proposes claim that PQC can be scaled to any depth without Barren Plateau, therefore it will be highly promising to scale the PQCs to show quantum computational advantage [2]. Although we do not claim that our proposed hybrid quantum algorithm shows quantum computational advantage, the study is still beneficial for enriching the family of quantum machine learning algorithms exactly as you have stated.

[1] Hu, Ling, et al. "Quantum generative adversarial learning in a superconducting quantum circuit." *Science advances* 5.1 (2019): eaav2761.

[2] Park, Chae-Yeun, Minhyeok Kang, and Joonsuk Huh. "Hardware-efficient ansatz without barren plateaus in any depth." *arXiv preprint arXiv:2403.04844* (2024).

COMMENT # 2.2

On the downside, some of the claims made by the authors appear to lack solid grounding. For GI, CT, and C, it is not clear if the claims of better performance are based only on the samples shown in the paper. In such cases, such a limited dataset is insufficient to assert the superiority of a model. Moreover, the fact that the model is potentially better, without robust statistical backing, does not constitute a compelling

result. It is important to mention, however, that I and SR are evaluated on the ImageNet validation set (1k samples), hence providing a more robust claim. On top of this, although explanations are always reported as hypothetical, it is not very clear how the physics informed model constitutes a good prior for image reconstruction, and a more in-depth explanation of this mechanism, if possible, would be helpful.

Thanks for your constructive suggestions. Indeed, for GI and colorization, we do not report the results of QDGP for large scale number of samples. Here, we present more numerical results for GI and colorization. As for category transfer, we may not provide the results on large number of samples. The category transfer task is performed by training on image of a certain category and then given different category labels style transfer is performed on different images via BigGAN model. In principle, there is not object metrics to evaluate the performance compared with other tasks where we can calculate the PSNR and SSIM through ground-truth images. The performance of category transfer can be evaluated based on the visual appearance. Under such situation, absolute performance comparison is not applicable. We think that both QDGP and DGP can show impressive performance in this task. Therefore, we only present the results of QDGP with large number of samples for GI and colorization.

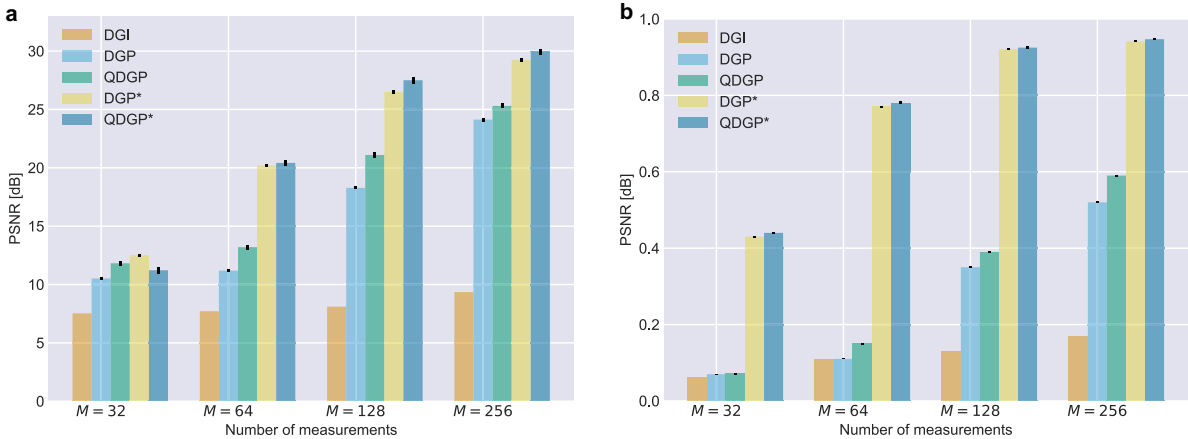


Figure 4: The numerical results of QDGP, DGP and DGI methods on ghost imaging for MNIST dataset. (a) PSNR results; (b) SSIM results. The error bars are the standard deviation of the repeated runs.

We randomly sample 200 images in MNIST test dataset and conduct optical ghost imaging experiments to collect the bucket signals and the respective illumination speckles. More samples requires more time to conduct optical experiments. The samples in MNIST dataset are handwritten digits which are not included in the ImageNet dataset. We also present the metrics of PSNR, SSIM under different measurements {32, 64, 128, 256} for DGI, DGP and QDGP methods. For each measurement,

QDGP and DGP are executed with 20 independent runs with different initialization to obtain the mean and standard deviation of the metrics. The numerical results are presented in Fig. 4.

From the results, it turns out that QDGP has a more obvious utility on the image reconstruction than DGP and DGI when using pretrained models. When using random initialized models, QDGP however achieves competitive results with DGP. The result of recovering the image from the pretrained BigGAN models further demonstrates that the potential of PQC-induced distributions for enhancing the out-of-distribution generalization ability of pretrained models.

We also evaluate the performance of colorization of QDGP and DGP on the 1k ImageNet validation set. The numerical results can be found in Table 1.

Table 1: Colorization performance. DGP and QDGP share the same hyperparamter settings and both use progressive reconstruction method. The mean and standard variance are calculated over independent 20 runs.

	DGP	QDGP
PSNR [dB]↑	21.93 (± 0.022)	23.52 (± 0.025)
SSIM [1e-2]↑	90.11 (± 0.026)	93.08 (± 0.031)

The numerical results on 1k validation set also showcase higher PSNR and SSIM value of QDGP over DGP method. The improved PSNR and SSIM can be attributed to quantum circuit induced latent distribution. Compared to classical latent code, quantum latent code has a natural distribution constraint. Classical latent code is a 120-dimensional free variables that can be optimized. However, quantum latent code is generated by a parameterized quantum circuits with many quantum parameters. In theoretical, the expectation value of the quantum circuits can approximate any complex probability distributions [1]. This distribution constraint in nature is unique that classical latent code cannot own. In addition, we also conduct repeated experiments to calculate the statistical error of QDGP and DGP on super-resolution and inpainting tasks. The result are presented as below. For inpainting tasks, we also repeat the algorithm for 20 independent runs and calculate the mean and s.t.d error which can be seen in the Table 2.

Table 2: Inpainting evaluation. DGP and QDGP share the same hyperparamter settings and both adopt feature and MSE loss. The reported mean and standard variance are calculated over independent runs.

	DIP	DGP	QDGP
PSNR [dB]↑	14.51 (± 0.02)	16.90 (± 0.03)	17.35 (± 0.04)
SSIM [1e-2]↑	29.23 (± 0.11)	45.76 (± 0.12)	47.21 (± 0.13)

For super-resolution results, the numerical results can be seen in the Table 3.

Table 3: Super-Resolution ($\times 4$) performance with different metrics. DGP and QDGP share the same hyperparameter settings and both adopt feature and MSE loss. The reported mean and standard variance are obtained from independent 20 runs.

	DIP	DGP	QDGP
NIQE $\downarrow$	6.10 (± 0.01)	4.98 (± 0.02)	5.03 (± 0.03)
PSNR $\uparrow$	22.71 (± 0.02)	23.02 (± 0.03)	23.40 (± 0.03)
RMSE $\downarrow$	17.83 (± 0.01)	17.58 (± 0.01)	17.50 (± 0.02)

From Table 1-3, we can find that QDGP can improve the performance of DGP when using a quantum circuit induced latent code. We note that although the improve is not significant than DGP over DIP, QDGP provides a new way to generate the latent code with inherent probability constraint. In fact, DGP has achieved impressive results on various image manipulation tasks. Here, we make use of PQC is to explore the utility of quantum advantage by replacing the original free classical latent code with quantum latent code with inherent distribution constraint. Here, we mainly aim to improve the generalization performance of DGP in some out-of-distribution samples. Also, by fine-tuning the latent code and the pretrained generator, we can generate more samples with more diversity with quantum latent code compared to classical latent code.

Finally, we want to give more explanation on the physical forward prior. The physical forward model is used in the ghost imaging tasks. Ghost imaging in our experimental setup can actually be called single-pixel imaging. The imaging process includes three key elements. The first key elements is the light modulation patterns H . Then the modulated light field propagation into the object and is detected by the single-pixel detector. We denote the detected signals as B . It is worth to point out that each pattern H_i will generate one signal B_i . The imaging processing can be modelled by matrix multiplication i.e. $HX = B$ where X denotes the imaging objects. Here we ignore the noise terms. In real applications, we can only reconstruct X with under-sampling bucket signals. Therefore, it is a inverse problem with under-sampling condition. To recover high-quality X , we reparameterize the inverse process with QDGP. That is, we can use QDGP to generate a $\hat{X}$ that is totally different from truth X at the first iteration. Then, the physical forward model H used in the real experiments are adopted to map the estimated image into the bucket signals $\hat{B}$. Then, we can calculate the mean square loss between experimental B and $\hat{B}$ to update parameters of QDGP. The physical forward model is vital here since it represents the true imaging process although the actual process may be noisy. From this perspective, the physical forward is an important prior to recover the image. Our QDGP models with physical

forward model can also be called physical-data driven machine learning model. It can be used to solve many inverse problems when the forward model is known. In many practical situations, forward model can be characterized accurately but the inverse process is hard to characterize [2].

The results discussed here are updated into the revised manuscript. For clarity, we do not present the revised details here. You can find the revision from the tables in the main text.

[1] Barthe, Alice, et al. "Expressivity of parameterized quantum circuits for generative modeling of continuous multivariate distributions." arXiv preprint arXiv:2402.09848 (2024).

[2] Qayyum, Adnan, et al. "Untrained neural network priors for inverse imaging problems: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.5 (2022): 6511-6536.

COMMENT # 2.3

In conclusion, I consider this paper to be valid and of interest, but I believe the concerns, particularly regarding the limited number of samples, should be addressed before publication, or the claims should be moderated. If my interpretation is incorrect and more samples have indeed been used, it would be beneficial for the paper to articulate this more transparently, and report this quantitatively.

Thanks again for your very positive evaluation. We agree with your viewpoint. To enhance the confidence of the conclusion, we have added more experiments on ghost imaging and colorization tasks. For ghost imaging, we conduct optical experiment to collect 200 handwritten digit samples and analyze the performance of QDGP and DGP in pretrained and randomly initialized conditions. We note that more samples requires longer optical experimental time. For colorization task, we also conduct numerical experiments on the ImageNet 1k validation samples and repeat 20 independent runs. For other tasks, we conduct the numerical experiments both on DGP and QDGP to calculate the mean and standard error of the metrics. The relevant results have been added into the revised manuscript. We believe that the additional results can enhance the confidence of our claim.

COMMENT # 2.4

Some typos:
215: hat on the x

225: parentheses in the formula
234: conducte
494: does not directly responsible

Thanks for your useful suggestions. We have thoroughly revised all the typos and mistakes in the original manuscript. We believe that the revised manuscript has no typos and grammar errors, to our knowledge. For example, the error of line 234 'conducte' is changed to 'conduct'. The grammar error of line 494 'does not directly responsible' is changed to 'is not directly responsible'.

We appreciate your warm work earnestly, which are valuable to improve the quality of our manuscript. We hope that the corrections will meet with approval. Once again, thank you very much for your comments and suggestions. We also look forward to your reply.

REVIEWER # 3

COMMENT # 3.1

The Deep Generative Prior [48] is an existing architecture for image processing (denoising, colorizing, infilling...) that uses a large pretrained GAN as a starting point for image processing; to process an image, the model is given further per-image finetuning, both of the latent space and the model parameters, to produce a cleaned image. The main goal of this paper is to improve the Deep Generative Prior (DGP) by changing the latent space: instead of a latent space with a Gaussian distribution, the latent space is specified as the outputs of a parameterized quantum circuit (PQC).

The claim is that this circuit adds expressivity to the model through new parameters (the parameters of the quantum circuit), which leads to better quality image denoising. The quantum component could also, in principle, lead to latent space vectors that may be hard to find classically, allowing a quantum computer to lend some of its power to this algorithm. This appears to be a novel approach, and I expect it would most interest those working in quantum machine learning.

First of all, we thank for your positive feedback. The starting point or motivation of this study is to design a hybrid quantum-classical machine learning algorithm. Then, we try to investigate the performance of quantum machine learning in ghost imaging and classic image processing tasks. In current noisy quantum device, variational quantum algorithm (VQA) is one of the most promising algorithms to demonstrate practical quantum advantage. Here, the practical quantum advantage does not only include computational advantage where classical algorithms are hard to simulate, it can also be referred to those algorithms that can perform better than pure classical counterpart. VQA as its easy to train and deploy has been widely studied in machine learning and combinatorial optimization [1]. However, many quantum machine learning can only process small-scale problems such as the downsampled MNIST or other low dimensional classification problems. One important reason is that those algorithms regard parameterized quantum circuits as the backbone of the machine learning. Since the bottleneck of classical to quantum encoding, the number of qubits is hard to scale up or implemented in practical quantum computers. Although those proposed algorithms have been demonstrated to show certain improvements, the practical utility is still limited. In our work, we intend to investigate the utility of quantum machine learning in ghost imaging or other image processing tasks. Hybrid quantum-classical algorithm can be viewed as a highly promising framework to fully exploit the practical utility of current noisy quantum devices.

Our hybrid quantum machine learning algorithm makes use of parameterized or programmable quantum circuits (PQC) as the quantum latent code generator and the pretrained big generative model as the image processor. The latent space does not

sample from measurement basis but from observables. The expectation of the observables can be estimated from projective measurements. The parameter optimization is also easier than bitstring-based quantum generative models [2, 3]. In noisy conditions, the expectation value can be further filtered through quantum error mitigation methods [4]. Therefore, our proposed quantum algorithm is also feasible for noisy quantum devices. Our proposed algorithm may provide a new reference and inspiration for quantum machine learning algorithms.

- [1] Cerezo, Marco, et al. "Variational quantum algorithms." *Nature Reviews Physics* 3.9 (2021): 625-644.
- [2] Liu, Jin-Guo, and Lei Wang. "Differentiable learning of quantum circuit born machines." *Physical Review A* 98.6 (2018): 062324.
- [3] Coyle, Brian, et al. "The Born supremacy: quantum advantage and training of an Ising Born machine." *npj Quantum Information* 6.1 (2020): 60.
- [4] Cai, Zhenyu, et al. "Quantum error mitigation." *Reviews of Modern Physics* 95.4 (2023): 045005.

COMMENT # 3.2

There is a secondary topic in this paper, which concerns the use of their model for Ghost Imaging (GI). Ghost Imaging is an imaging technique that uses quantum effects to enhance sensitivity in low-light environments, but it comes with its own noise model. GI forms an additional good, purpose-motivated test problem for denoising images, and it's great to explore how (Q)DGP performs on this noise model. But the fact that GI involves quantum effects has no connection to the quantum architecture of QDGP, no more than a GPU being made of silicon would make it more adept at denoising photographs of silicon. This double "quantumness" in the paper is potentially confusing or even misleading.

Thanks for your constructive comments and positive feedback on computation GI. Exactly as you have stated, GI is a very promising computational imaging method. GI uses the illumination speckles to modulate the light field and then uses a single pixel detector to measure the light field that interacts with the unknown object. GI has strong resilience to noise and has been widely studied in remote sensing [1,2] and biological imaging [3]. In our work, we investigate GI in real experiments and collect the bucket signals as the experimental data. Compared to other quantum machine learning algorithms using simulation or toy data, our works make use of the optical experimental data to recover the large-scale images in GI.

Besides, we highly agree with your criticism on the double "quantumness". In our

GI experimental setup, we do not use entanglement light and single-photon detector. Instead, we use the laser and a single pixel detector, which are pure classical equipment as we have stated in *Methods* section. The improvement of QDGP is indeed not come from the so called double "quantumness". In our revised manuscript, we will clarify and avoid the so called connection that may be confusing or potentially misleading.

[1] Gong, Wenlin, and Shensheng Han. "High-resolution far-field ghost imaging via sparsity constraint." *Scientific reports* 5.1 (2015): 9280.

[2] Erkmen, Baris I. "Computational ghost imaging for remote sensing." *JOSA A* 29.5 (2012): 782-789.

[3] Li, Wenwen, et al. "Single-frame wide-field nanoscopy based on ghost imaging via sparsity constraints." *Optica* 6.12 (2019): 1515-1523.

COMMENT # 3.3

Are the claims convincing:

I seriously disagree with the statement that the PQC latent space improves the *expressiveness* of the model. In the BigGAN model underlying both DGP and QDGP, z is a latent vector of 120 parameters. The PQC used seems to be a circuit with $3N_{lk} = 1440$ parameters (lines 135-138 of the supplementary material, although the precise number is not entirely clear). But, no matter how many parameters there are, these 1440 parameters get compressed to a 120-parameter vector again for feeding into the BigGAN. The PQC cannot add expressiveness, because in the end it is just a complicated formula to produce the parameters of the original model.

What is a real possibility is that the PQC improves the trainability of the model. The denoising process involves navigating a loss landscape towards a de-noised image that remains close to the original, and it's possible that the PQC changes the loss landscape to favor landing in better images. Indeed, the earlier DGP and DIP models were trying to do exactly this: DIP is the idea that the structure of a CNN leads to a better loss landscape, and DGP took that idea further by baking more knowledge of typical images into a GAN via pretraining.

But this would be extraordinary to actually occur here. CNNs, and even more so pretrained GANs, are models specifically designed to capture structures of images. By design, they include information about what typical images look like, and that could shape the loss landscape for better training is both plausible and born out by experiment. PQCs, on the other hand, are quantum processes on small numbers of qubits. They have no connection to images, latent spaces, or anything that would make me expect an improved loss landscape.

The main effect I would expect is a different effective learning rate. Depending on how precisely the parameters are setup and the loss determined, training the PQC could lead to an effective change in z that's a few times faster or slower than what it was in the standard DGP training process. One would expect this to have a small but nonzero impact on accuracy. Since the original DGP paper did not pick a separate learning rate for the latent space variables, it quite likely that some scaling of the learning rate (either up or down) would improve the accuracy – it is unlikely to be optimal already.

Thanks for your very insightful comments. First of all, we agree with your viewpoint on the statement that PQC may not enhance the expressiveness of the whole model. In fact, the expressiveness of the whole model is hard to evaluate in practice. However, we argue that PQC with trainable quantum parameters is universal and can approximate any multivariate continuous function as reference [1] have stated, as shown in Figure 5.

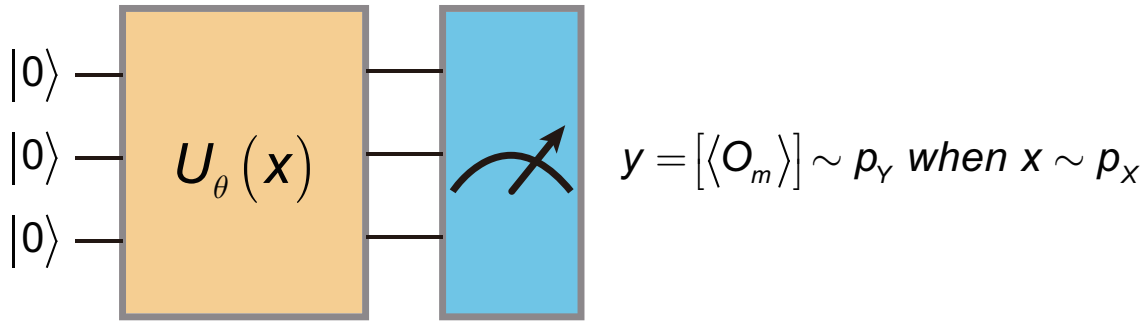


Figure 5: Expectation value sampling model. A random sampled vector is encoded into a quantum state using a parameterized quantum circuit. The expectation values of fixed observables are returned as another random vector.

Firstly, we present the main result of reference [1]. The definition of generalized polynomial chaos expansion (GPCE) is presented as follows:

Definition 1 A generalized chaos expansion is characterized by a probability density function p_X defined on the support $\mathcal{X} \subseteq \mathbb{R}^M$ with finite moments (usually chosen as standard distributions, such as Gaussian or uniform). This choice defines an inner product for functions in $\{f : \mathcal{X} \rightarrow \mathbb{R}\}$:

$$\langle f|g \rangle_{p_X} := \int_{\mathcal{X}} f^*(x)g(x)p_X(x)dx. \quad (1)$$

This choice of inner product comes with the choice of an order family of functions, usually polynomials, that are orthonormal with respect to the above inner product:

$$\Phi_{p_X} = \{\phi_l : \mathcal{X} \rightarrow \mathbb{R}, \forall (k, l), |\phi_l\rangle\langle\phi_k| = \delta_{k,l}\}. \quad (2)$$

A generalized chaos expansion is a representation of a random variable Y with probability density p_Y as a vector α in this Hilbert space, such that:

$$Y = \sum_{l=0}^{\infty} \alpha_l \phi_l(X) \sim p_Y, \quad X \sim p_X. \quad (3)$$

This provides a Hilbert space as a potential structure to study random variable mappings. In particular, one of the main results of polynomial chaos expansion is that they are universal generative models. When using data reuploading quantum circuits with integer-valued spectrum $U_\theta(x)$, PQC can be also viewed as a GPCE, therefore we have a theorem:

Theorem 1 *Any expectation value sampling model using a quantum reuploading model with integer-valued spectrum $U_\theta(x)$, together with the uniform distribution on $[0, 2\pi)^M$ outputs a random variable Y that has an exact finite polynomial chaos expansion for any choice of observables.*

Our proposed PQC also adopts a data reuploading encoding style with a integer-valued spectrum, thus it can also be viewed as GPCE, i.e. they are universal generative models. In contrast, The latent space of the DGP is a 120 independent trainable parameters and they are initialized or sampled from a standard Gaussian distribution. During training, 120 parameters are independently optimized. Since they are optimized without constraint from a parametric distribution, it may mismatch the pretrained manifold. Thus, in practice, one may add a L2 constraint to regularize the latent code such that it can maintain a specific distribution. However, PQC provides a brand-new method for distribution generating or learning. Since PQC is universal for generating multivariate continuous function, its generated latent code thus naturally has a intrinsic constraint of parametric distribution. We do not need add any constraint into the latent code. The multivariate distribution can be trained or optimized to match the manifold of the pretrained or random initialized BigGAN. From this point of view, PQC induced distribution can be viewed as a generalized distribution than 120 independent parameters induced distribution. More importantly, PQC induced distribution is a multivariate continuous distribution with intrinsic parametric model. Therefore, we can safely conclude that PQC can increase the expressiveness of the latent space model. The expressiveness of the whole model will not be largely influenced since the manifold of BigGAN has not been changed.

Moreover, the feature of PQC can allow a more universal distribution learning or modelling. On the one hand, using PQC as the latent code generator, it may increase the generalization ability of the generative model. On the other hand, PQC based sampling model can be combined with classical generative models to learn the data distribution.

As for your second viewpoint, firstly we thanks for your constructive analysis. Whether PQC can improve the trainability of the model is also a open problem, which is hard to be clearly answered. In fact, the classical manifold is maintained in our hybrid quantum-classical algorithm compared to DGP and DIP method, therefore the generative prior can also be used in our algorithm. The only different point is the trainable latent space code. As we have discussed above, compared to the original latent code with 120 independent trainable parameters, we have 1440 trainable parameters to generate a 120-dimensional distribution. Although as you have analyzed, they are the same in terms of the number of variables (120) used for the input. However, they are generated in a different manner. Our 120-dimensional distribution is generated from a PQC where each of the variable is correlated. However, 120 trainable parameters are principally independent. One of our original intention using PQC is to improve the loss landscape of the BigGAN to improve the denosing performance. From the numerical results, QDGP shows a better reconstruction performance compared to its classical counterpart although we are also uncertain that QDGP can improve the optimization loss landscape. It involves the interpretation of hybrid quantum machine learning. This area has yet been widely investigated. Therefore, you may provide a very good route for investigating the origins of performance improvement of hybrid quantum machine learning. We argue that PQC may be optimized under the physical or pretrained loss function to adapt to find a better latent space.

As for your third point about the learning rate, we firstly thanks for your very motivating discussion. We exactly are aware of your ideas that latent code is only different part between QDGP and DGP. From the in-depth analysis of the external phenomena, it can be concluded that the latent code $\mathbf{z}$ is optimized in different ways, showing different distributions, thus affecting the results. We argue that you may miss one possibility that $\mathbf{z}$ in QDGP and DGP may be optimized to show different patterns. Of course, as you analyzed, they may also be optimized faster or slower. We can easily check the performance of the algorithm with latent code $\mathbf{z}$ under different learning rate. However, we think that the latent code is optimized to different patterns is hard to evaluate the gap between QDGP and DGP as it involves deep interpretation. In the next comment, we will provide the detailed numerical results of QDGP and DGP with different learning rate for latent code. Besides, we notice that the original DGP paper also adopts a separate learning rate for latent code in several different tasks although they may not be optimal. In our supplementary, we have presented different learning rates for PQC and classical latent code. We think that you provided learning rate analysis is very beneficial to evaluate the performance increment of QDGP compared to DGP.

Finally, we want to highlight that exploring the potential advantage of quantum generative models in tackling the classical problems particularly pretrained models

is important for current quantum devices [2,3]. Since in the noisy intermediate-scale quantum (NISQ) era, PQC's are not provided with large scale quantum error correction as the limitation of qubits and coherence time [4]. Our proposed sampling model is a highly convenient test bed for NISQ devices to explore the utility. Meanwhile, combining with large scale pretrained models, it may be highly likely to verify the utility of NISQ devices to deal with real large scale classical problems such as ghost imaging and image manipulation.

[1] Barthe, Alice, et al. "Expressivity of parameterized quantum circuits for generative modeling of continuous multivariate distributions." arXiv preprint arXiv:2402.09848 (2024).

[2] Perdomo-Ortiz, Alejandro, et al. "Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers." *Quantum Science and Technology* 3.3 (2018): 030502.

[3] Huang, Hsin-Yuan, et al. "Power of data in quantum machine learning." *Nature communications* 12.1 (2021): 2631.

[4] Bharti, Kishor, et al. "Noisy intermediate-scale quantum algorithms." *Reviews of Modern Physics* 94.1 (2022): 015004.

COMMENT # 3.4

The analysis presented (Figure 2; Table I; Table II) shows that the QDGP and DGP perform very similar, with the QDGP performing slightly better on most metrics. But I suspect this is primarily due to the different learning rate improving the training schedule mostly by accident. The metrics given also don't give any sense of the uncertainty in the quantities: is a difference of 16.95 vs 17.22 statistically significant (Table I)? Compared to other values given (in the authors' paper and ref [48]), this difference seems small.

Thanks for your fruitful comment. In the current manuscript, the numerical results show that QDGP outperforms DGP by a noticeable margin. Compared to the margin that DGP outperforms earlier works, QDGP obtained margin over DGP seems relatively small. Firstly, we think that DGP is a very good algorithm and show impressive results in classical imaging manipulations. Outperforming DGP by a great margin is relatively hard to be accomplished. The similar result can also be found in comparison between DGP and DIP methods. Since DIP is the first algorithm exploiting generative prior, DGP is an improved algorithm by utilizing pretrained models and improved loss function and achieves state-of-the-art performance. We also can found that in super-resolution [1], the performance gap of DGP and QDGP is also small. From the perspective of classical machine learning, QDGP is improved based

on DGP by replacing the original latent code with PQC induced latent code. We can hardly expect that QDGP can achieve a performance leap compared to DGP while the numerical results reveal that QDGP can achieve competitive or better performance than DGP. From the perspective of quantum machine learning, it might be a good starting point to exploit noisy quantum computers to finish some classical and real applications. On the other hand, we want to emphasize that QDGP can potentially improve generalization ability of pretrained models. As for pretrained models, generalization is of great interest and importance. The foundation of current generative artificial intelligence is based on the big model pretraining and downstream fine-tuning. We expect that the pretrained models can adapt to different domains and achieve good performance in unseen situations. It may require the model having good out-of-distribution ability. The out-of-distribution of quantum machine learning has been studied in [2], demonstrating that the out-of-distribution generalization in the task of learning an unknown unitary. There are no works to investigate the out-of-distribution of quantum machine learning in classical tasks. Exploiting a good latent space using PQCs may be beneficial to improve generalization since its generated multivariate continuous distributions are universal. Therefore, PQCs-based sampling models can be integrated into other classical generative models such as variational auto encoder (VQE) [3] and diffusion models [4].

- [1] Pan, Xingang, et al. "Exploiting deep generative prior for versatile image restoration and manipulation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.11 (2021): 7474-7489.
- [2] Caro, Matthias C., et al. "Out-of-distribution generalization for learning quantum dynamics." *Nature Communications* 14.1 (2023): 3751.
- [3] Wang, Gaoyuan, et al. " ζ -QVAE: A Quantum Variational Autoencoder utilizing Regularized Mixed-state Latent Representations." *arXiv preprint arXiv:2402.17749* (2024).
- [4] Zhang, Bingzhi, et al. "Generative quantum machine learning via denoising diffusion probabilistic models." *Physical Review Letters* 132.10 (2024): 100602.

In the revised manuscript, we additionally provide the standard deviation of the achieve metrics to exclude the statistical fluctuations by repeating the experiments. We note that since the overhead the classical simulation of quantum circuits, we only repeat the QDGP and DGP for different tasks with 20 independent runs. The mean and standard variance are reported in the table of the main text.

COMMENT # 3.5

How to strengthen the paper:

The main experiment I would want to see done is this.

- Measure the typical learning rates of the latent space variables in the DGP and the QDGP (2-norm travelled per iteration).
- Call the ratio between these two rates k .
- Run the DGP model where the latent space variables are scaled to be k times more sensitive. That is, $z = z_0 + k \cdot (dz)$, where dz are the new trainable parameters initialized at 0, and z_0 is in the initial latent vector (now held fixed). - Compare this rescaled DGP model with the QDGP.

It seems likely that this one additional hyperparameter closes the gap between QDGP and DGP. If it doesn't, and QDGP outperforms DGP, then some error bars on the analysis and the accuracy will be needed to provide a convincing argument that this is actually improving upon the model somehow.

Thank you for your valuable suggestions. We conducted additional numerical experiments based on your proposed rule. This rule can be implemented using PyTorch code, where d_z is set as a trainable tensor and z_0 as a non-trainable tensor. Each time we run the model, we use $z = z_0 + k \times d_z$ as the latent code input. First, we measured the typical learning rates of QDGP and DGP, then adjusted the learning rates according to the ratio k of the two learning rates for each stage. Your rule suggests that the updating amplitude of the latent code can influence the final performance. In our hyperparameter settings, the learning rate of the PQC-induced latent code is significantly higher than that of the classical latent code. It is noteworthy that the performance of the imaging reconstruction can be influenced by the latent code to some extent. However, whether the learning rate of the latent code, i.e., the updating amplitude, determines the ultimate performance still requires further investigation.

To analyze the performance, including PSNR and SSIM, we conducted numerical experiments with a series of learning rate ratios between QDGP and DGP. We fixed the learning rates of QDGP and rescaled the learning rates of DGP to be k times more sensitive or insensitive. The results are presented in Figure 6.

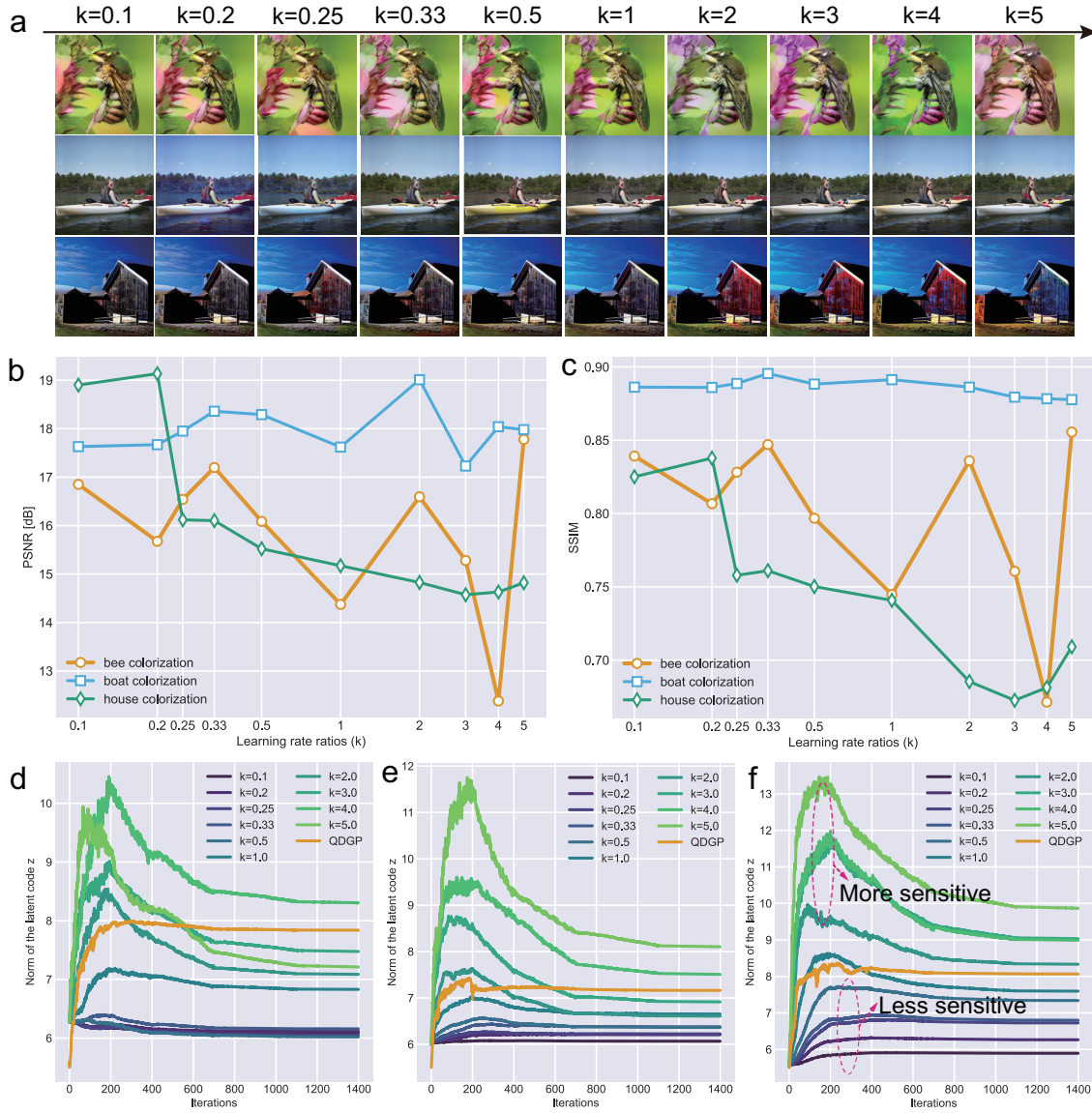


Figure 6: The hyperparameter analysis of learning rates for classical DGP. The learning rate ratio k is calculated between the typical learning rates of QDGP and DGP. Smaller k indicates more insensitivity to the latent code and vice versa. (a) Colorized images based on DGP with different learning rate ratios. (b,c) The measured PSNR and SSIM of the reconstructed images under different learning rate ratios. (d,e,f) The measured norm of the latent code during the optimization process with different learning rate ratios, where subfigure (d) is for bee, (e) for boat, and (f) for house.

The results indicate that rescaling the learning rate ratio k does not significantly improve the colorization performance of different images compared to the original learning rate setting. Notably, carefully tuning the learning rate of the latent code can achieve optimal performance, but it is time-consuming. The PSNR and SSIM metrics

oscillate with the learning rate ratio. Occasionally, a high learning rate ratio results in higher PSNR and SSIM values, as seen in bee colorization. However, for boat colorization, changing the learning rate does not greatly affect the performance, particularly the SSIM value. Additionally, for house colorization, increasing the learning rate ratio leads to decreased performance, indicating that a smaller learning rate ratio may result in better performance.

From Figure 6(d, e, f), the variation in the norm of the latent code with iterations is demonstrated. It is observed that with an increased learning rate ratio, the norm rises dramatically at the early stage of optimization. However, after a certain number of iterations, the norm stabilizes and converges to a steady value. In contrast, the yellow curves representing the quantum latent code show a different increase pattern. The norm initially rises to a certain value and remains stable without dropping. It is important to note that the norm metric indicates the global property of the latent code and does not characterize the distribution pattern. Although the norm curves for the quantum and classical latent codes differ, more significant differences may lie in their distribution patterns.

We also conducted experiments on the inpainting task, with additional results presented in Supplementary Note 4. In both ablative experiments, adjusting the learning rate ratio does not entirely close the performance gap between QDGP and DGP, although careful tuning of the learning rate hyperparameter improves performance.

In the revised manuscript, we performed extensive experiments to calculate the mean and standard deviation of the performance metrics to support our arguments. For each task, we conducted 20 independent runs, presenting the mean value with error bars in the revised tables. The results showcase the tangible advantage of quantum machine learning in enhancing image reconstruction performance. We also conducted time-consuming optical experiments for ghost imaging. The mean value and error bars of the PSNR and SSIM in this task are reported in an additional figure (see Figure 6 in the main text). We found that QDGP consistently outperforms DGP in image reconstruction and enhances out-of-distribution performance.

Given the substantial revisions, we do not detail them here. Instead, we provide a brief summary of the additional results, which can be found in the main text and supplementary notes. The additional learning rate analysis can be found in the Supplementary Note 4 which contains the detailed analysis for colorization and inpainting tasks. The repeated experiments for inpainting, colorization, super-resolution and ghost imaging can be found in the Table 1,2,3 and Figure 6 in the main text.

COMMENT # 3.6

Separately, the Ghost Imaging problem as a denoising problem seems quite useful as a benchmark on these models, and the physics-forward noise model deserves more discussion. I think that could make a fine paper on its own, with or without the QDGP.

Thank you for your positive and constructive comments. The first example of QDGP involves reconstructing images in a ghost imaging system. We propose using a physical forward model as the inductive bias to optimize the quantum learning model. First, we briefly explain the imaging process of ghost imaging, also known as single-pixel computational imaging, in our optical experiments. The light field is modulated by a randomized or prescribed pattern (denoted as P loaded onto a spatial light modulator (SLM)). The modulated light then propagates and interacts with the object, effectively modulating the light field. Finally, a single-pixel detector collects the optical intensities. The potential of ghost imaging lies in its ability to resist noise in the optical imaging process. However, in practical applications, the sampling rate is limited to enable quick image recovery, posing a significant challenge. Additionally, while ghost imaging can resist backscatter noise, electronic noise and other imperfections are also present, necessitating image recovery from noisy bucket signals. To address these challenges, we propose using QDGP with a physical forward model. By incorporating the inductive bias of the machine learning generator, it is possible to use fewer samples to recover high-quality images.

By combining the physical forward model H , we can calculate the estimated bucket signals $\hat{B}$ based on the estimated images $\hat{X}$ using $\hat{B} = H\hat{X}$. We then use the mean square loss between $\hat{B}$ and the experimental B from the real imaging process to update the parameters of QDGP. Compared to conventional optimization-based variational quantum algorithms, which optimize parameters based on the expectation of the Hamiltonian, our method combines the forward model and optimizes parameters using mean square error loss. Thus, it differs significantly from the conventional variational quantum eigensolver [1]. Our proposed method can be viewed as the first physics-data-driven hybrid quantum machine learning algorithm. Notably, we do not consider the light propagation process directly, instead using H as the imaging process. In ghost imaging, i.e., single-pixel computational imaging, the modulation pattern generates a prescribed light field that interacts with the object. A single-pixel detector then collects all the optical intensities, ignoring the spatial information of the light field, making it reasonable to use H as the imaging process.

Furthermore, in many scenarios, the forward model is known, but the inverse process is unknown and difficult to solve [2,3]. We can use quantum machine learning to pre-parameterize the inverse process and employ the physical forward model as the inductive bias to update the parameters. From this perspective, our proposed method has numerous applications in solving inverse problems.

- [1] Kandala, Abhinav, et al. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets." *nature* 549.7671 (2017): 242-246.
- [2] Zhang, Qihang, et al. "Extracting particle size distribution from laser speckle with a physics-enhanced autocorrelation-based estimator (PEACE)." *Nature Communications* 14.1 (2023): 1159.
- [3] Pei, Zongrui, et al. "Machine-Learning Microstructure for Inverse Material Design." *Advanced Science* 8.23 (2021): 2101207.

We have added more discussion on the QDGP with physical forward model. The revised manuscript can be found in the following:

Main Text, Page 10, line 601-614 ...

... Noteworthy, QDGP with the physical forward model H as the inductive bias can make fully use of the physics information to recover the images. Since using GI, we can simply regard H as the imaging process and all the information involving the spatial freedom are ignored. However, our method can also be applied into cases where the forward model is known but the inverse process is unknown and hard to solve [68, 69]. We can use quantum machine learning to pre-parameterize the inverse process and employ the physical forward model as the inductive bias to update the parameters. From this perspective, our proposed method has numerous applications in solving inverse problems. ...

COMMENT # 3.7

In equation (5), two different loss terms (a total square error, and a total variation) are added together. While these do in principle have the same units and could be added, it seems surprising to just add them with no weighting. I would expect something like $\|B - Hx\|^2 + \lambda \|x\|_{TV}^2$ where λ is a hyperparameter. Looking at their code, it seems this was the actual form they used, with $\lambda = 10^{-5}$; also, the TV loss is the anisotropic TV loss, and without a squaring either. I believe it should read, $\|B - Hx\|^2 + 10^{-5} * \|x\|_{TV, aniso}$ or similar, with the coefficient and no exponent on the TV term.

Thanks for your careful reading and checking. In our implementation, we have one additional coefficient $\lambda = 10^{-5}$ to control the TV constraint. In addition, we also calculate anisotropic TV loss without exponent on the TV term. In the revised manuscript, we have rewritten the equation (5). Thanks again for your careful checking.

COMMENT # 3.8

As said in line 657, they state that the class label is set to a random embedding hidden vector. On line 668, they say they use 500 latent codes z and pick the one with lowest loss as the initial guess. These both match exactly the strategy in [48] and I think that citation should be repeated here. (The number 500 seems particular.)

Thanks for your suggestion. We have added the citation of Ref. [48] in the revised manuscript.

We appreciate your warm work earnestly, which are valuable and insightful to improve the quality of our manuscript. We hope that the corrections will meet with approval. Once again, thank you very much for your comments and suggestions. We also look forward to your reply.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

We deeply appreciate the authors' comprehensive response to our comments. The manuscript has notably improved in both quality and clarity of exposition.

While some fundamental questions regarding the scalability and simulability of the quantum circuits in the proposed method remain, we recognize that addressing these questions is beyond the scope of the current work. This limitation makes it challenging to fully assess the potential impact of the paper. Therefore, we are willing to defer to the editor's judgment regarding its publication.

Reviewer #2 (Remarks to the Author):

The claims regarding Ghost Imaging (GI) and Colorization (C) are now accompanied by numerical results.

For ghost imaging: I disagree with the claim that “DGP shows a significant advantage”. The SSIM values are the same within the uncertainty bars. For PSNR, QDGP performs better than DGP for $M = 128$ and 256 , worse for $M = 32$, and the same (within error) for $M = 64$.

Note: The y-axis text in figure 6b should be “SSIM” instead of “PSNR”.

For colorization: results seem convincing.

For category transfer: The authors rely on visual inspection due to the lack of available quantitative metrics. Being this the case, I disagree with the claim that “DGP produces images less consistent with natural image standards than QDGP”, in fact I think Figure 7 suggests the contrary. Anyway, I think that the visual inspection of such a small number of images is insufficient to support such a claim.

Finally, I kindly ask the authors to address criticisms directly, avoiding redundant phrases that make it challenging to discern the answer. I found a 36-page letter overwhelming.

Reviewer #3 (Remarks to the Author):

I find that the authors have done a good job of responding to what I feel were the weaknesses of the paper. The error bars on the performance comparisons between DGP and their new QDGP make the improvement statistically significant. They clearly separated the quantumness of the Ghost-Imaging task from the quantum circuit. And the additional experiments they ran with varying learning rate have (Fig 6 in the response) have satisfied me that there is some nontrivial change to the learning dynamics due to the quantum latent code generator.

I still contend that, at least by the standard usage of the term, the "expressivity" of the QDGP and the standard DGP is unchanged. The authors have pointed out that quantum circuits are universal function approximators, which is certainly true; but with a fixed latent code dimension as the only input to the generator, the latent code forms an Information Bottleneck (IB) that limits the expressivity. Put it this way: if f is a function from $\mathbb{R}^{10}$ to $\mathbb{R}^{100}$, and g is a function from $\mathbb{R}^{20}$ to $\mathbb{R}^{10}$, then the composition $(f \circ g)$ is just as expressive as f : both of them can only express a 10-dimensional manifold within $\mathbb{R}^{100}$, even if g takes a 20-dimensional input. This is essentially the scenario here: their parameterized quantum circuit maps a ~ 1300 -dimensional space down to ~ 120 , and then this gets mapped into the very high dimensional space of images. This is still only able to express the same 120-dimensional manifold of the DGP.

But that's fine! A model doesn't have to have greater expressivity to be a better model. I think it would be reasonable to say that their model has better trainability or learns better or forms a better prior. Those are all mathematically possible, and reasonably borne out by the data.

Personally I am generally suspicious of claims of quantum advantage. But the model here outperforms previous models by some reasonably objective statistics; and a reasonable effort has been done to exclude the simplest explanation (that the training rate is effectively different). Some qualitatively different training dynamics have been demonstrated, which makes this interesting. I think it now well deserves to be published.

Response letter for paper *COMMSPHYS-23-1349-A*

Quantum deep generative prior with programmable quantum circuits,
Tailong Xiao, Xinliang Zhai, Jingzheng Huang, Jianping Fan, Guihua Zeng

Dear Editor and Reviewers:

First of all, we are very grateful to you for reviewing our manuscript entitled "Quantum deep generative prior with programmable quantum circuits". We sincerely appreciate your comments and they really help to improve the quality of our article. According to the reviewers' comments, corresponding amendments have been made in the revised manuscript. We believe that the revised one is suitable for further consideration on *Communications Physics*.

In the following, we first highlighted each comment, then we provided a reply, and finally we briefly described how the manuscript was adjusted in the main text. We deliberate that the main changes have fully addressed the comments. We hope the revised manuscript clarifies any confusions and meets your expectations. We look forward to hearing from you soon.

Best regards,
All authors

REVIEWER # 1

COMMENT # 1.1

We deeply appreciate the authors' comprehensive response to our comments. The manuscript has notably improved in both quality and clarity of exposition.

While some fundamental questions regarding the scalability and simulability of the quantum circuits in the proposed method remain, we recognize that addressing these questions is beyond the scope of the current work. This limitation makes it challenging to fully assess the potential impact of the paper. Therefore, we are willing to defer to the editor's judgment regarding its publication.

Thank you for your positive feedback on our revisions. In the previous revision, we increased the layers of the quantum circuit and found that a modest number of layers does not lead to a Barren Plateau. Recently, we discovered a paper [1] stating that the gradient-based trainability of the quantum model implies that the quantum circuit model does not exhibit a Barren Plateau. Additionally, the simulability of the

quantum circuit model can lead to dequantization, meaning the quantum learning model has no quantum advantage.

In our work, numerical simulations show that our hybrid model is trainable with the gradient descent method for up to 10 layers in an 18-qubit circuit. Although we cannot theoretically analyze the trainability and simulability of the proposed model, we leave these problems for future research. We have discovered that approximate quantum circuit simulation using matrix product states is likely to have lower performance compared to using real quantum circuit models. This result is consistent with the motivation in Ref. [2], where pretraining quantum circuits with matrix product states can find suitable parameter initialization, and further training of real quantum models can achieve higher performance.

[1] Gil-Fuster, Elies, et al. "On the relation between trainability and dequantization of variational quantum learning models." arXiv preprint arXiv:2406.07072 (2024).

[2] Rudolph, Manuel S., et al. "Synergistic pretraining of parametrized quantum circuits via tensor networks." *Nature Communications* 14.1 (2023): 8367.

In future work, we will utilize dynamical Lie algebra [3] to investigate the ansatz of variational quantum circuits. We hope this approach will help us find a balance between trainability and simulability.

[3] Ragone, Michael, et al. "A unified theory of barren plateaus for deep parametrized quantum circuits." arXiv preprint arXiv:2309.09342 (2023).

REVIEWER # 2

COMMENT # 2.1

The claims regarding Ghost Imaging (GI) and Colorization (C) are now accompanied by numerical results.

For ghost imaging: I disagree with the claim that “DGP shows a significant advantage”. The SSIM values are the same within the uncertainty bars. For PSNR, QDGP performs better than DGP for $M = 128$ and 256 , worse for $M = 32$, and the same (within error) for $M = 64$. Note: The y-axis text in figure 6b should be “SSIM” instead of “PSNR”.

For colorization: results seem convincing.

For category transfer: The authors rely on visual inspection due to the lack of available quantitative metrics. Being this the case, I disagree with the claim that “DGP produces images less consistent with natural image standards than QDGP”, in fact I think Figure 7 suggests the contrary. Anyway, I think that the visual inspection of such a small number of images is insufficient to support such a claim.

Finally, I kindly ask the authors to address criticisms directly, avoiding redundant phrases that make it challenging to discern the answer. I found a 36-page letter overwhelming.

Thank you for your constructive suggestions. Regarding ghost imaging (GI), as you pointed out, the SSIM values of QDGP* and DGP* are nearly the same when using randomly initialized methods. However, when using pretrained weights, QDGP performs better than DGP for $M = 64, 128, 256$ in terms of both PSNR and SSIM. For $M = 32$, QDGP with randomly initialized weights performs worse than DGP. Nonetheless, when using pretrained GAN, QDGP shows competitive performance with DGP. These results indicate that QDGP has better performance when using pre-trained weights, implying that PQC-induced distribution likely enhances the generalization ability of generative models.

For category transfer, we agree with your viewpoint. Since we rely on visual inspection of fewer samples due to the lack of available quantitative metrics, we have removed the statement that “DGP produces images less consistent with natural image standards than QDGP” as it cannot be supported by a large number of experiments.

In the revised manuscript, we have revised the statements and provided more rigorous descriptions of the performance of the QDGP and DGP methods. Additionally, we have revised the y-axis label of Fig. 6b (now Fig. 4b) to the SSIM metric.

Main text, Page 6, Line 359-372

..., both QDGP and DGP shows significant advantage over DGI in terms of imag-

ing reconstruction quality. When using pretrained GAN, QDGP performs better than DGP for $M = 64, 128, 256$ both in terms of PSNR and SSIM metrics. For $M = 32$, QDGP achieves nearly the same SSIM value with DGP. With randomly initialized GAN, QDGP weights performs worse than DGP for $M = 32$. For $M = 64, 128, 256$, QDGP shows competitive performance with DGP. Therefore, QDGP has no tangible improvement over DGP for different sampling rates when using randomly initialized GANs. ...

Overall, thank you for your valuable suggestions on the claims of QDGP over the DGP method, which have significantly improved the rigor of the paper. We have revised the statements accordingly and hope the revised version meets your approval.

REVIEWER # 3

COMMENT # 3.1

I find that the authors have done a good job of responding to what I feel were the weaknesses of the paper. The error bars on the performance comparisons between DGP and their new QDGP make the improvement statistically significant. They clearly separated the quantumness of the Ghost-Imaging task from the quantum circuit. And the additional experiments they ran with varying learning rate have (Fig 6 in the response) have satisfied me that there is some nontrivial change to the learning dynamics due to the quantum latent code generator.

Thank you again for your insightful suggestions to improve the paper. Specifically, the learning rate checking rules out the possibility of trivial improvement of the quantum latent code. Parameterized quantum circuits can generate multivariate continuous distributions, which can be used in generative models and potentially enhance trainability.

COMMENT # 3.2

I still contend that, at least by the standard usage of the term, the "expressivity" of the QDGP and the standard DGP is unchanged. The authors have pointed out that quantum circuits are universal function approximators, which is certainly true; but with a fixed latent code dimension as the only input to the generator, the latent code forms an Information Bottleneck (IB) that limits the expressivity. Put it this way: if f is a function from $\mathbb{R}^{10}$ to $\mathbb{R}^{100}$, and g is a function from $\mathbb{R}^{20}$ to $\mathbb{R}^{10}$, then the composition $(f \circ g)$ is just as expressive as f : both of them can only express a 10-dimensional manifold within $\mathbb{R}^{100}$, even if g takes a 20-dimensional input. This is essentially the scenario here: their parameterized quantum circuit maps a ~ 1300 -dimensional space down to ~ 120 , and then this gets mapped into the very high dimensional space of images. This is still only able to express the same 120-dimensional manifold of the DGP.

But that's fine! A model doesn't have to have greater expressivity to be a better model. I think it would be reasonable to say that their model has better `_trainability_` or `_learns better_` or `_forms a better prior_`. Those are all mathematically possible, and reasonably borne out by the data.

Thank you for the intuitive explanation. We agree with your viewpoint. Since we cannot utilize the parameter space of quantum circuits to evaluate the expressivity of the QDGP model, the number of dimensions of the latent code remains the same as the classical variables. The difference lies in that the conventional latent space is

generated with 120-dimensional free variables. However, by using quantum circuits with a high-dimensional parameter space, we can generate latent codes that obey an internal distribution constraint. This feature is likely to improve the trainability of the DGP model, as validated by the presented numerical results.

COMMENT # 3.3

Personally I am generally suspicious of claims of quantum advantage. But the model here outperforms previous models by some reasonably objective statistics; and a reasonable effort has been done to exclude the simplest explanation (that the training rate is effectively different). Some qualitatively different training dynamics have been demonstrated, which makes this interesting. I think it now well deserves to be published.

Thank you for your positive feedback. Demonstrating quantum advantage remains a significant challenge. In this paper, our goal is to showcase the utility of quantum machine learning models, which differs somewhat from achieving a clear quantum advantage. Generally, quantum advantage implies scenarios where classical computers struggle to efficiently compute a problem that quantum computers can handle effectively. Examples include random circuit sampling and Boson sampling, which have successfully demonstrated quantum advantage. Quantum utility in this work, on the other hand, focuses on enhancing the learning capabilities of classical models, which is also meaningful in practical applications. We aim for our method to offer valuable insights into demonstrating the quantum utility of quantum machine learning across various problem domains.

We appreciate your warm work earnestly, which are valuable and insightful to improve the quality of our manuscript. We hope that the corrections will meet with approval. Once again, thank you very much for your comments and suggestions. We also look forward to your reply.

REVIEWERS' COMMENTS:

Reviewer #2 (Remarks to the Author):

I appreciate the authors' prompt and precise response and thank them for it. It seems to me that all the points I raised were addressed, either by providing more in-depth explanations or by renouncing the unsubstantiated claims. With these changes, there are no more obstacles to publication from my side.