Journal of King Saud University Computer and Information Sciences (2025) 37:247
https://doi.org/10.1007/s44443-025-00274-1

ORIGINAL PAPER

Quantum autoencoder implementation of high-dimensional steganographic encoding for arbitrary quantum states

Chaolong Hao[1] · Quangong Ma[1] · Yaqi Chen[1] · Hao Zhang[1] · Dan Qu[1]

Received: 21 August 2025 / Accepted: 5 September 2025

Abstract

While classical steganography achieves maturity in digital media, hiding arbitrary quantum states ($\alpha|0\rangle + \beta|1\rangle$) has emerged as an intriguing frontier. To address this problem, we establish a formal model of controllable random perturbation unitaries for single/multi-stego state tasks. We progressively explore Quantum Autoencoder (QAE) structures through three stages: starting from single-state scenarios without perturbation, advancing to perturbed conditions, and finally extending to multi-state tasks. We design two perturbation-based encoding schemes using Quantum Autoencoders (QAE): the simple scheme (QAE-DD) leverages the inverse application of encoding–decoding modules, while the improved scheme (QAE-OSP) incorporates orthogonal projection routing and parallel subnetworks to restructure the hidden-layer architecture. In 3-qubit entangled-state simulations with data scales $n \leq 10$ and perturbation strengths $\varepsilon \in [0, 1]$, QAE-DD performs well under low perturbation, whereas QAE-OSP maintains higher fidelity between the carrier and secret states under high perturbation conditions (e.g., $n = 5, \varepsilon = 0.6$), with fidelity values $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right) = 0.91$ / $F\left(\rho_S, \tilde{\rho}_S\right) = 0.84$ providing a reference for network design. Finally, we extend the single-carrier ("$1 + 1$") task to the multi-carrier ("$1 + N$") scenario by constructing a "centroid" state training set based on the principal component of carrier-state groups and validating the applicability of both models. Under the conditions $n = 5$ and $\varepsilon = 0.6$, the QAE-OSP model successfully improves the average fidelity between multiple secret states and carrier states from 0.68 to 0.90, demonstrating its capability to aggregate multiple carriers to enhance overall concealment. Although the present study covers only small-scale data and networks, it lays the groundwork for a neural network framework that covertly embeds arbitrary quantum states into high-dimensional quantum states, providing a basis for future exploration.

Keywords Quantum information hiding · Arbitrary states · High-dimensional entangled state · Quantum autoencoder · Orthogonal projection splitting structure

1 Introduction

Information hiding technology, as an important means of ensuring information security and privacy, has evolved in close alignment with the development of information carriers. From ancient steganographic methods—such as tattooing on skin and using invisible ink—to modern digital information hiding techniques (Katzenbeisser and Petitcolas 2016; Petitcolas et al. 1999; Moulin and O'Sullivan 2003), this technology has continuously expanded its application boundaries. In the current field of digital multimedia, the primary applications are digital watermarking and steganography, with carriers encompassing images, audio, video, and text. Classical approaches include least significant bit (LSB) substitution and spatio/transform-domain embedding for multimedia (Cheddad et al. 2010; Subhedar and Mankar 2014; Potdar et al. 2005), as well as synonym substitution, syntactic modification, or semantic perturbation for text (Majeed et al. 2021). These methods are all dedicated to achieving covert embedding and reliable extraction of information within specific carriers. The carriers of information

✉ Quangong Ma
quangongma@163.com

✉ Dan Qu
qudan_xd@163.com

Chaolong Hao
hcl_xdspeechlab@aliyun.com

Yaqi Chen
chyaqi163@163.com

Hao Zhang
haozhang0126@163.com

[1] School of Information Systems Engineering, Information Engineering University, Zhengzhou, China

Published online: 13 October 2025

hiding are closely related to the development of the underlying techniques. Recent research has further extended it to texts generated by large language models (Liu et al. 2024), with representative approaches including soft green-list biasing (Kirchenbauer et al. 2023) and semantic space mapping (Liu et al. 2024).

With the rapid development of quantum information technology, quantum states—particularly arbitrary quantum states $\alpha|0\rangle + \beta|1\rangle$—have been extensively studied as a concept distinct from classical bits (0/1). Quantum properties such as superposition, entanglement, and the no-cloning theorem have, in theory, demonstrated significant advantages, as exemplified by Shor's algorithm (Shor 1994) and the BB84 protocol (Bennett and Brassard 1984). From the perspective of information hiding, this naturally raises the question: can a quantum state itself be embedded into a quantum carrier as the secret information? Most existing studies follow the classical paradigm—constructing quantum multimedia models and achieving embedding by modifying the least significant qubits of quantum images/audio (Hao et al. 2024; Xing et al. 2024; Dong and Yan 2024; Sun et al. 2022, 2023). Although such methods leverage quantum properties (e.g., superposition or entanglement) to enhance security during the secret embedding process, their essence still lies in a quantum adaptation of the classical paradigm—referred to as Classical-Paradigm Quantum Information Hiding. In this approach, secret qubits are independent of carrier qubits, and the concealment relies on the redundancy of perturbations at the media level collectively exhibited by multiple particles (see Fig. 1). However, do such models fully capture the unique potential of quantum information—particularly quantum entanglement?

A highly promising direction for exploration is how to embed an arbitrary quantum state into a high-dimensional entangled state, so that the secret information is diffusely distributed across the global structure of the entangled state, making it difficult for an eavesdropper to recover the complete secret state via nonlocal operations. This paradigm can be referred to as Quantum-Logic Based Information Hiding (see Fig. 1). Compared with the classical paradigm, the quantum-logic paradigm offers the following advantages: (1) the secret information itself is a quantum state, and the embedding process is naturally aligned with the intrinsic properties of quantum information, with the extracted secret state directly usable in quantum tasks; (2) by leveraging the nonlocal nature of quantum entanglement, security guaranteed by the principles of quantum mechanics can be more effectively achieved through approaches such as multipartite distribution and collaboration. A representative technique of this quantum paradigm is quantum information masking, which aims to conceal quantum information within quantum correlations such that each local subsystem

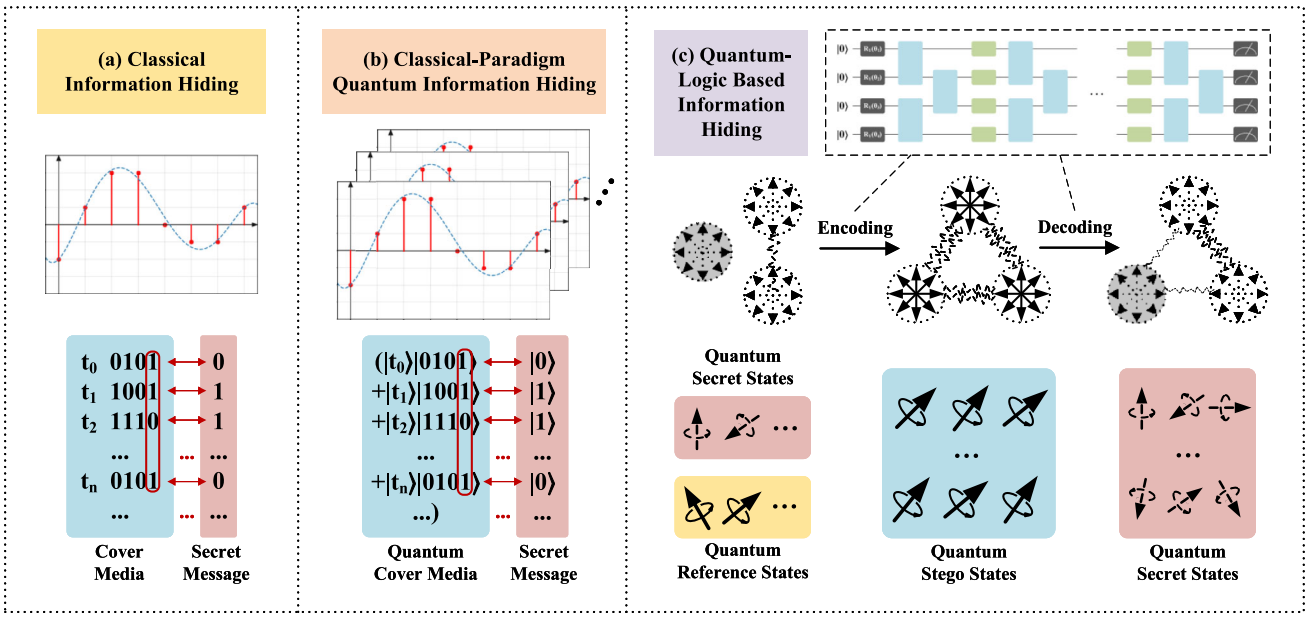

Fig. 1 Schematic comparison between classical and quantum information hiding paradigms. Subfigure (a) illustrates the classical LSB information hiding method; Subfigure (b) shows the classical-paradigm quantum information hiding method, where, once the time state $|t\rangle$ is measured and collapses to a fixed value, the qubits of the quantum multimedia carrier become independent, allowing secret information embedding via classical bitwise substitution; Subfigure (c) depicts quantum-logic information hiding, in which the secret state and the carrier's entangled state form a tightly integrated whole. Extraction of the secret quantum state requires specific encoding and decoding operations (entanglement and disentanglement), which can be implemented by manually designing high-dimensional unitary operators, or—as in this work—by using quantum neural networks

only exhibits a mixed state. However, due to the quantum no-masking theorem (Kavan et al. 2018), the set of maskable quantum states and the admissible operations are not arbitrary. Consequently, research has focused on identifying feasible sets and operations (Sheng and Ling 2018; Shi et al. 2021; Shen et al. 2023; Bai et al. 2025). In Table 1, we provide a conceptual comparison of classical, classical-quantum, and quantum logical information hiding methods, including quantum information masking.

In an entangled state, a secret quantum state cannot be embedded by simply replacing a single qubit, as in the classical LSB method (see Fig. 1). Instead, it is necessary to construct a specific global encoding operation $\mathcal{E}_S$ that fuses the secret state with a reference state through entanglement, while extraction requires executing the corresponding global decoding operation $\mathcal{D}_S$ to disentangle and retrieve the secret state. However, implementing such encoding and decoding via conventional unitary operations presents two major challenges: (1) to enhance encoding security—especially in high-dimensional scenarios—the entangling unitary becomes highly complex, and its resource consumption typically grows exponentially with system size; (2) fixed, pre-defined unitary operations lack adaptability to practical noise environments (e.g., quantum gate errors or channel disturbances), making it difficult to dynamically optimize for maintaining the fidelity and extractability of the embedded information.

It is worth noting that, in recent years, artificial intelligence (AI) technologies have provided a highly adaptive and general framework for information hiding (Wang et al. 2023), surpassing the limitations of traditional methods in terms of robustness and security (Kandi et al. 2017; Zeng et al. 2024). Recent research has also explored the integration of deep learning with quantum cryptography, demonstrating how AI can enhance the performance of quantum key distribution (QKD) and quantum secure communication by optimizing protocols and improving security (Pasupuleti 2024; Decker et al. 2025; Purohit and Vyas 2025). The nonlinear transformation capability of deep neural networks enables them to autonomously learn optimal embedding strategies, which can be tailored to customized objectives (e.g., capacity–imperceptibility trade-offs) through task-specific training. This approach not only enhances resistance to reverse-engineering attacks but also improves robustness against interference and imperceptibility, without the need to repeatedly design dedicated algorithms for each scenario (Zhu et al. 2018).

Inspired by this, this work adopts the QAE and its variants as a universal encoding–decoding architecture. Parameterized quantum circuits (PQCs) are employed to dynamically optimize parameters, enabling the quantum neural network to learn and approximate the encoding and decoding operations $\mathcal{E}_S$ / $\mathcal{D}_S$, thereby achieving the hiding of arbitrary secret states within high-dimensional entangled states.

The main contributions of this work are as follows:

1. We establish a formalized model of controllable random perturbation encoding operations for embedding arbitrary quantum states into high-dimensional entangled states in single/multi-carrier scenarios. A security analysis is conducted for cases where an eavesdropper accesses either the local or the global system, showing that security improves with both perturbation strength ε and entanglement dimension d_E.
2. We design two QAE-based implementations of the perturbation encoding operation: the basic scheme (QAE-DD), which uses encoding–decoding modules in reverse, and the improved scheme (QAE-OSP), which reconstructs the hidden-layer structure via parallel sub-networks with orthogonal projection routing. Simulations on 3-qubit systems with data size $n \leq 10$ and perturbation strength $\varepsilon \in [0, 1]$ indicate that QAE-DD performs well under low perturbation, while QAE-OSP maintains higher fidelity under strong perturbations. In a representative case ($n = 5$, $\varepsilon = 0.6$), QAE-OSP achieves a carrier-state fidelity of 0.91 and secret-state fidelity of 0.84, offering a reference for practical network design.
3. We extend the single-carrier ("1 + 1") task to a multi-carrier ("1 + N") scenario by constructing a "centroid" state training set based on the principal component of carrier-state groups, and validate both QAE-DD and QAE-OSP models. Under conditions $n = 5$ and $\varepsilon = 0.6$, the QAE-OSP model raises the average fidelity between multiple secret states and carrier states from 0.68 to 0.90, demonstrating its ability to aggregate multiple carriers for enhanced concealment.

2 Preliminaries

In this section, we provide some essential background on quantum information to help readers without prior knowledge in the field gain a basic understanding.

2.1 Quantum States and Quantum Operations

The fundamental unit of quantum information (Nielson and Chuang 2010) is described by either a quantum state vector $|\psi\rangle$ or a density matrix (density operator) ρ, where the density matrix provides a more complete description for mixed states or subsystems. The density matrix is defined as:

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| \tag{1}$$

Here, $\langle\psi_j|$ denotes the conjugate transpose of $|\psi_j\rangle$, and p_j represents the probability that the system is in the pure state

Table 1 Comparison of classical and quantum information hiding methods

Item	Classical Information Hiding	Classical Paradigm Quantum Information Hiding	Quantum Logic Information Hiding—Single State Information Hiding—Ouantum Information Masking	Quantum Logic Information Hiding—Single State Information Hiding—General entanglement related information hiding	Quantum Logic Information Hiding—Multiple states Information Hiding
Secret Info Type	Classical bits (0, 1)	Quantum basis states ($\|0\rangle$, $\|1\rangle$)	Arbitrary states $\alpha\|0\rangle + \beta\|1\rangle$	Arbitrary states $\alpha\|0\rangle + \beta\|1\rangle$	Arbitrary states $\alpha\|0\rangle + \beta\|1\rangle$
Carrier Type	Classical multimedia	Quantum multimedia	Single multi-body entangled state	Single multi-body entangled state	Multiple quantum entangled states
Embedding Method	LSB method	LSB method	Specifically designed unitary operations	General unitary/non-unitary encoding operations	General non-unitary encoding operations
Concealment Principle	Human perception	Post-measurement human perception	Local indistinguishability with strong security, global distinguishability possible	Local inability to obtain full information	Global indistinguishability
Consumption for Hiding 1 bit/qubit of Secret Info	$O(L)$, where L is the quantization level of the carrier, $L = 8, 16, 32, \ldots$	$O(L)$, where L is the quantization level of the carrier, $L = 8, 16, 32, \ldots$	$O(N)$, where N is the number of auxiliary states, $N \geq 3$ (Kavan et al. 2018)	$O(N)$, where N is the number of auxiliary states, generally $N \geq 3$	$O(N)$, where N is the number of auxiliary states, generally $N \geq 3$
Security Enhancement Mechanism	Classical cryptography	Quantum cryptography	Quantum cryptography	Quantum cryptography	Quantum cryptography

$|\psi_j\rangle$. The density matrix has the following properties: Hermiticity: $\rho = \rho^\dagger$; positivity: $\langle\phi|\rho|\phi\rangle \geq 0, \ \forall|\phi\rangle$; and unit trace: $\mathrm{Tr}(\rho) = 1$. Here, $\mathrm{Tr}(\cdot)$ denotes the trace operation. The state of a subsystem can be described by the reduced density matrix, which is obtained via the partial trace. The distance between two quantum states can be quantified by the fidelity:

$$F(\rho, \sigma) = \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}\right)^2 \tag{2}$$

Quantum operations describe the dynamical evolution of quantum systems and can be classified into two categories: unitary and non-unitary operations.

(1) Unitary operations correspond to the reversible evolution of closed quantum systems and are represented by a unitary operator U (satisfying $UU^\dagger = U^\dagger U = I$). The evolution of a quantum state in the density matrix formalism is given by:

$$\rho \longrightarrow U\rho U^\dagger \tag{3}$$

(2) Non-unitary operations primarily describe the irreversible evolution of open systems or artificially designed transformations. These mainly include:

(a) Quantum measurement: implemented by a set of measurement operators $\{M_m\}$ (satisfying $\sum_m M_m^\dagger M_m = I$), where the state collapses to the outcome corresponding to m as:

$$\rho \longrightarrow \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}\left(M_m \rho M_m^\dagger\right)} \tag{4}$$

(b) Physical generalized evolution (e.g., noise or decoherence): modeled by a set of Kraus operators $\{K_\mu\}$ (satisfying $\sum_\mu K_\mu^\dagger K_\mu = I$), with the evolution given by:

$$\rho \longrightarrow \sum\nolimits_\mu K_\mu \rho K_\mu^\dagger \tag{5}$$

(c) Artificially designed operations (e.g., neural network transformations): Within a quantum–classical hybrid computing framework, a parameterized non-unitary mapping $\mathcal{N}_\theta$ can be constructed, whose abstract form is given by:

$$\rho \longrightarrow \mathcal{N}_\theta(\rho) \tag{6}$$

Here, θ denotes the trainable parameters. Such operations can be embedded into variational quantum algorithms to perform data feature extraction or state reconstruction.

2.2 Parameterized quantum circuits

A PQC is a quantum circuit composed of a series of parameterized quantum gates and fixed non-parameterized gates connected in a specific structure (Fig. 2). PQCs are commonly used to implement quantum neural networks (Wecker et al. 2015; Chen et al. 2025). Input data, either in classical or quantum form, is loaded onto the initial state of the qubits, which then evolves through the PQC to produce the output state.

The core of a PQC lies in its parameters θ, which are trainable. By defining a loss function that depends on the measurement outcomes (or intermediate quantum states) and using a classical optimizer (e.g., gradient descent) to adjust θ so as to minimize the loss, a PQC can learn to perform a specific task. This hybrid framework—where the quantum processor performs computations while the classical processor optimizes the parameters—is a mainstay in current fields such as quantum machine learning, quantum optimization, and quantum simulation.

2.3 Quantum autoencoder architecture

A QAE is a type of quantum neural network architecture (Romero et al. 2017) inspired by the classical autoencoder, designed to learn efficient representations of quantum data. It can be applied to tasks such as quantum data compression, denoising, error correction, and circuit verification (Wu et al. 2024; Huang et al. 2020; Bondarenko and Feldmann 2020; Pepper et al. 2019; Zhang et al. 2021; Locher et al. 2023; Hao et al. 2025. Similar to its classical counterpart, a QAE consists of an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$, both of which can be implemented using parameterized quantum circuits. The formal model of the QAE structure is expressed as:

$$\rho_{\text{in}} \xrightarrow{\mathcal{E}} \rho_{\text{latent}} \xrightarrow{\mathcal{D}} \rho_{\text{out}} \tag{7}$$

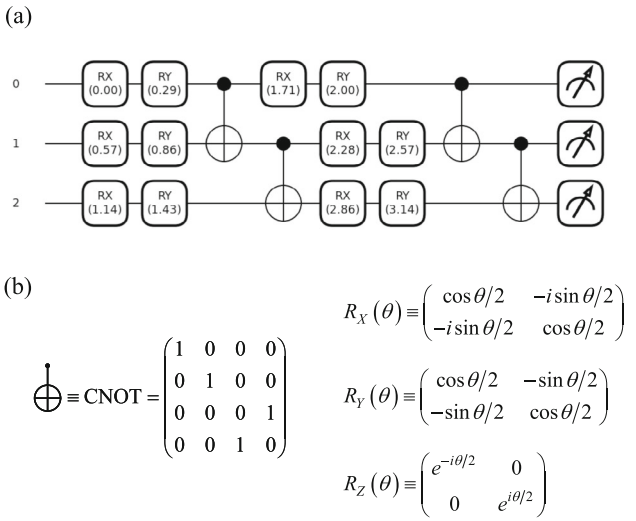

Fig. 2 Schematic diagram of a PQC. Subfigure (a) illustrates a circuit constructed from various quantum gates, where $R_X(\theta)$, $R_Y(\theta)$, and $R_Z(\theta)$ are rotation gates, and the CNOT gate is used to generate entanglement between different qubits. Subfigure (b) shows the matrix forms of the corresponding quantum gates (Nielson and Chuang 2010; Watrous 2018; Kaye et al. 2006)

As shown in Fig. 3, the encoder $\mathcal{E}$ functions to encode the input quantum state into a feature state with fewer qubits (decoupled from the reference state). The decoder $\mathcal{D}$ — which may be the inverse of the encoder $\mathcal{E}$ or trained independently — reconstructs the original state from the latent space with the aid of the reference state, and can also output partial states as required. For a mixed-state input ρ_{in}, let its rank be $r = \text{rank}(\rho_{\text{in}})$. The necessary condition for lossless compression in a QAE is:

$$r \leq 2^{n_L} \tag{8}$$

Here, n_L denotes the number of qubits in the latent space (excluding the reference state) (Ma et al. 2023). Furthermore, studies have shown that the representational capacity of the QAE can be enhanced through techniques such as noise injection in the decoder and network depth optimization (Sim et al. 2019; Cao and Wang 2021).

3 Formal analysis of quantum state information hiding tasks

As stated in the introduction, unlike the classical paradigm, embedding a secret state into a high-dimensional entangled state is not a simple bit-by-bit replacement. Instead, it requires applying a global quantum operation to the secret state and the auxiliary state to securely embed the secret state within the entangled state. In this section, we formally analyze such quantum operations and their security from the perspective of single and multiple carrier-state tasks.

3.1 Single carrier-state task

The single carrier-state task focuses on the encoding behavior of a single secret state. Formally, it is described as embedding an arbitrary quantum state $\rho_S \in \mathcal{H}_S$ into a high-dimensional entangled state $\rho_E \in \mathcal{H}_E$ within a larger Hilbert space (satisfying $d_E > d_S$, where $d = \dim(\mathcal{H})$ denotes the dimension of the space).Such high-dimensional carrier-entangled states can support subsequent quantum state secure transmission tasks, such as multi-party distribution and collaborative decoding. As described in Sect. 2.1, the embedding operation can be expressed as:

$$\rho_E = U\left(\rho_R \otimes \rho_S\right) U^{\dagger} \tag{9}$$

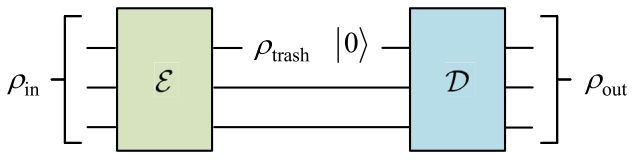

Fig. 3 QAE architecture diagram. Note: In some architectures, the position of the reference state in the latent layer is measured or traced out; in the architecture used in this paper, the reference state must be preserved

Here ρ_R is the introduced reference state (i.e., ancilla state) used to extend the system dimension. U is the global unitary operator acting on the composite system $\mathcal{H}_S \otimes \mathcal{H}_R$, implementing the entanglement functionality in $\mathcal{H}_E$ (with $d_E = d_S d_R$).

Ideally, the embedding operation U is randomly drawn from the Haar measure, such that, in a statistical sense, an eavesdropper can only obtain a maximally mixed state, thereby achieving theoretical unconditional security. However, when the Hilbert space dimension is large, the resources required to exactly implement a Haar-random unitary operation grow exponentially with the system size (Emerson et al. 2003; Ma and Huang 2025).

It is worth noting that, for secret state information hiding, since legitimate communicating parties can share the operation U, only a single copy or very few copies are required to ensure the retrieval of the secret state in its quantum form (note that this does not mean obtaining the classical parameters of its state vector components). In this case, we may adopt a more practical approach by choosing operations that are easier to implement instead: first, randomly select a fixed unitary operation V from the set of unitary operators $\mathcal{U}_{\text{Haar}}$ corresponding to the Haar measure; then, apply a small perturbation to it to enhance its randomness and robustness. This can be formally expressed as:

$$U = e^{i\varepsilon H} V \tag{10}$$

Here, H is randomly generated from the Gaussian Unitary Ensemble (GUE) and represents the Hamiltonian of the random perturbation, while ε denotes the perturbation strength coefficient, satisfying $0 \leq \varepsilon \leq 1$. From an application perspective, this model not only reflects an active perturbation operation introduced to enhance randomness, but also incorporates passive perturbations arising from the inherent imperfections of high-dimensional unitary operations—particularly in the Noisy Intermediate-Scale Quantum (NISQ) era.

We now analyze the security of this model:

(1) Suppose the eavesdropper (Eve) intercepts a subsystem B of ρ_E, with the reduced density matrix given by $\rho_B = \text{Tr}_A(\rho_E)$, where $\mathcal{H}_E = \mathcal{H}_A \otimes \mathcal{H}_B$ (A is the portion held by the legitimate party Alice, and B is the portion held by Eve). Since ρ_E is an entangled state, ρ_B is therefore a mixed state, and its form depends on the overall entanglement structure:

$$\rho_B = \text{Tr}_A\left(U\left(\rho_R \otimes \rho_S\right) U^{\dagger}\right) \tag{11}$$

Equation (11) shows that ρ_B depends simultaneously on ρ_S, ρ_R, and U. Under the action of entanglement, the information of the secret state ρ_S is "diffused" across the entire system. Eve cannot reconstruct ρ_S from ρ_B alone because the

partial trace operation removes quantum correlations with the unmeasured part. Moreover, for a fixed d_B, the larger the total dimension d_E, the stronger the overall entanglement, making ρ_B closer to the maximally mixed state and leaving the eavesdropper with less accessible information.

(2) Suppose Eve is able to intercept the entire carrier state ρ_E. Since she only possesses a single copy or very few copies, in order to obtain the secret state ρ_S, she would need to guess U in order to disentangle ρ_E. On the one hand, $U = e^{i\varepsilon H}V$, where V is drawn from $\mathcal{U}_{\text{Haar}}$ and has about d_E^2 free parameters, making the search space enormous. On the other hand, if Eve acquires information about the fixed V through other means, then without additional protection, she could easily obtain the secret state. But here, U is perturbated on the basis of V by the term $e^{i\varepsilon H}$, which generates the expected deviation (the detailed derivation can be found in Appendix A):

$$\mathbb{E}\left(\|U - V\|_1\right) \sim O(\varepsilon) \cdot d_E^{\frac{3}{2}} \tag{12}$$

This indicates that as the perturbation strength ε and the total entanglement dimension d_E increase, U also exhibits significant random deviations from V, making it difficult for Eve to accurately obtain U.

As the perturbation strength ε and the total entanglement dimension d_E increase, U exhibits significant random deviations from V, making it difficult for Eve to precisely determine U.

3.2 Multi-carrier quantum state task

For multiple secret states, we not only expect each of them to be secure within the high-dimensional entangled states, but also aim to aggregate the encoded carrier states so that they are difficult to distinguish as a whole, thereby improving the overall concealment (note: in connection with the analysis in Sect. 3.1, Eve can only proceed to crack an individual carrier state if she can first distinguish between different carrier states). We refer to this as the multi-carrier quantum state task.

Specifically, consider a set of random secret states $\left\{\rho_S^{(k)}\right\}$, with corresponding encoded states $\left\{\rho_{\text{stego}}^{(k)}\right\}$. If $\Delta(\{\rho\})$ denotes the divergence of this set of quantum states, then the task objective is to make the distribution of the secret states more concentrated after being encoded into the carrier states, i.e., $\Delta\left(\left\{\rho_{\text{stego}}^{(k)}\right\}\right) < \Delta\left(\left\{\rho_S^{(k)}\right\}\right)$. Moreover, the closer the carrier states are to each other, the better. Expressed in terms of fidelity:

$$F\left(\rho_{\text{stego}}^{(j)}, \rho_{\text{stego}}^{(k)}\right) = 1 - \delta, \quad j \neq k, \quad 0 < \delta \ll 1 \tag{13}$$

According to the Helstrom bound (Helstrom 1969), the minimum error probability for an eavesdropper (Eve) to distinguish between $\rho_{\text{stego}}^{(j)}$ and $\rho_{\text{stego}}^{(k)}$ $(j \neq k)$ is:

$$P_{\text{err}}^{\min}(j,k) = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\rho_{\text{stego}}^{(j)} - \rho_{\text{stego}}^{(k)}\right\|_1\right) \tag{14}$$

Here, $\|\cdot\|_1$ denotes the trace norm. According to the Fuchs–van de Graaf inequality (Nielson and Chuang 2010), the trace norm and the fidelity satisfy:

$$\left\|\rho_{\text{stego}}^{(j)} - \rho_{\text{stego}}^{(k)}\right\|_1 \leq 2\sqrt{1 - F\left(\rho_{\text{stego}}^{(j)}, \rho_{\text{stego}}^{(k)}\right)} \tag{15}$$

Combining (13), (14), and (15), we have:

$$P_{\text{err}}^{\min}(j,k) \geq \frac{1}{2}\left(1 - \sqrt{\delta}\right) \tag{16}$$

Equation (16) indicates that the closer the stego states are to each other ($\delta \to 0$), the larger the lower bound on the eavesdropper's probability of error in distinguishing them. When $\delta = 0$, we have $P_{\text{err}}^{\min}(j,k) \geq \frac{1}{2}$, which is equivalent to random guessing.

Then, naturally, another question arises: for a set of quantum states $\{\rho_k\}$, is it possible to find a quantum state that is closest to all of them? We call such a quantum state the 'centroid' of $\{\rho_k\}$ (denoted as ρ_{cen}), which represents the center of this set of quantum states. Once the 'centroid' state is found, we can aim to move the encoded states $\left\{\rho_{\text{stego}}^{(k)}\right\}$ towards the centroid state. This naturally minimizes the cost and makes the states more concentrated. Below, we adopt the idea of Principal Component Analysis and present a method to construct the 'centroid state' that maximizes the sum of the fidelities squared, with the proof provided in Appendix B.

Theorem 1 *Define the operator $M = \sum_k \rho_k$, and let $|\beta_{\max}\rangle$ be the eigenvector corresponding to the maximum eigenvalue of M (normalized to a unit vector). Define the 'centroid' state (pure state):*

$$\rho_{cen} = |\beta_{\max}\rangle\langle\beta_{\max}| \tag{17}$$

Then, ρ_{cen} is able to maximize the objective function:

$$J(\sigma) = \sum_k F^2(\rho_k, \sigma) \tag{18}$$

At the end of this subsection, we will discuss the category of quantum operations used to implement the encoding of the secret state mentioned above. We discuss the classes of quantum operations that can implement the above secret state

encoding. In the single-stego-state scheme of Sect. 3.1, we use a perturbed global unitary operation for encoding; however, for the aggregation task in the multi-stego-state scenario, unitary operations are no longer suitable. This is because a unitary operation does not change the distance between secret states (unless the reference state ρ_R is transformed simultaneously):

$$\begin{aligned}\left\|\rho_E^{(j)}-\rho_E^{(k)}\right\|_1 &= \left\|U\left(\rho_R\otimes\left(\rho_S^{(j)}-\rho_S^{(k)}\right)\right)U^\dagger\right\|_1 \\ &= \left\|\rho_S^{(j)}-\rho_S^{(k)}\right\|_1\end{aligned} \tag{19}$$

Therefore, the implementation of multi-stego state aggregation will rely on non-unitary operations. Combining the analyses in Sect. 3.1 and Sect. 3.2, in order to establish a unified framework for both types of operations and to circumvent the inherent complexity of high-dimensional unitary operations, the following chapters will introduce quantum neural network approaches based on QAE and its variants. In fact, research in multiple fields has shown that quantum neural networks—particularly variational quantum circuits—demonstrate significant potential in efficiently approximating target unitary operations (Li et al. 2024; de Oliveira et al. 2024), reducing resource overhead (Zhang et al. 2022; Wu et al. 2025; Ma et al. 2024), and enhancing implementation robustness (Huang et al. 2022), owing to their hybrid classical–quantum architecture and trainable parameter properties. Based on this, the following sections will focus on how to employ neural networks to realize the random perturbed unitary operations in single-stego state encoding and the aggregation operations in multi-stego state tasks.

4 A simple scheme for single-sego state tasks: reverse application of QAE functionality

4.1 Design concept

In the conventional QAE structure introduced in Sect. 2.3, the encoder $\mathcal{E}$ maps the input state ρ_{in} to a latent state $\rho_{\text{latent}} = \rho_{\text{info}} \otimes \rho_{\text{trash}}$, where ρ_{info} carries the useful compressed information, and ρ_{trash} is regarded as a "trash state" containing redundant information, which is measured or directly discarded (Romero et al. 2017). The decoder $\mathcal{D}$ (which can be $\mathcal{E}^\dagger$ (Locher et al. 2023; Cao and Wang 2021) or trained independently (Wang et al. 2025)) then reconstructs the input state using ρ_{value} together with auxiliary states (such as $|0\rangle^{\otimes k}$).

For the task in this study (embedding a low-dimensional secret state into a high-dimensional entangled state), we can implement it by reversing the function of the QAE. Specifically, we redefine the role of the traditionally so-called "trash state" ρ_{trash} in the latent layer to serve as the secret state ρ_S, while the original "information state" ρ_{info} is redefined as the reference auxiliary state ρ_R required for the expanded space. At the same time, we employ the trained QAE decoder $\mathcal{D}$ (taken as the inverse of the encoder, $\mathcal{D} = \mathcal{E}^\dagger$) as the encoding operation $\mathcal{E}_S$ for the information-hiding task (i.e., $\mathcal{E}_S = \mathcal{D} = \mathcal{E}^\dagger$). We denote this model structure as QAE-DD, as shown in Fig. 4.

After the network training is completed, the usage procedure is as follows: (1) The sender of the secret state (Alice) possesses the reference state ρ_R and the secret state ρ_S, and inputs them into the trained QAE decoder $\mathcal{D} = \mathcal{E}^\dagger$. At this point, $\mathcal{D}$ essentially functions as the encoding operation for information hiding, denoted as $\mathcal{E}_S$, whose output is the stego state ρ_{stego} (ideally, $\rho_{\text{stego}} \approx \rho_C$, indicating that the neural network has effectively learned the encoding unitary operation, though practical training introduces distortions). (2) The receiver (Bob), upon obtaining ρ_{stego}, applies the secret extraction operation $\mathcal{D}_S = \mathcal{E}$ (that is, the QAE's encoder $\mathcal{E}$) to retrieve the secret state ρ_S. Specifically, the result of $\mathcal{E}\rho_{\text{stego}}\mathcal{E}^\dagger$ is the tensor product of the reference state and the secret state, $\rho_R \otimes \rho_S$.

4.2 Simulation experiments

To validate this concept, we conducted preliminary simulation experiments (In our experiments, all models are implemented in Python 3 using PyTorch (Paszke et al. 2019) and PennyLane (Bergholm et al. 2022), same below):

1. **Data:** Randomly generate 3-qubit mixed states ρ_C satisfying the theoretical constraint (rank $\leq$ 4, see (8)). The secret state ρ_S is generated using PennyLane's Rot initialization: $\rho_S = U_{\text{Rot}}|0\rangle\langle 0|U_{\text{Rot}}^\dagger$, where $U_{\text{Rot}} =$

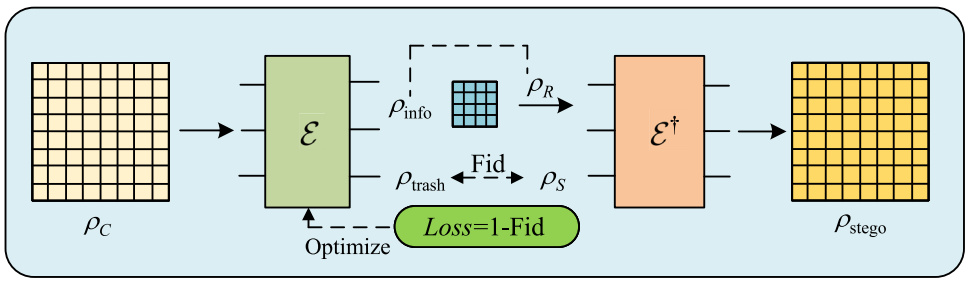

Fig. 4 Framework diagram of using QAE in reverse for secret state encoding (QAE-DD). ρ_C serves as the network input ρ_{in}, and the goal is for the output "trash state" ρ_{trash} to approximate ρ_S

$\mathrm{Rot}(\theta, \phi, \lambda)$ with parameters sampled uniformly: $\theta \sim \mathcal{U}(0, \pi)$, $\phi, \lambda \sim \mathcal{U}(0, 2\pi)$. The 2-qubit reference state ρ_R is extracted from the latent layer after network training. We generate n data pairs $(\rho_C^{(k)}, \rho_S^{(k)})$, $k = 0, \ldots, n-1$, which corresponds to training the network to learn n encoding operations $\mathcal{E}_S^{(k)}$.

2. **Model structure:** We construct a QAE network where the encoder $\mathcal{E}$ consists of a 10-layer variational quantum circuit.
3. **Model training:** Using the Hamiltonian corresponding to ρ_S to optimize the "trash" qubits, the loss function is defined by the fidelity:

$$\mathcal{L} = 1 - F(\rho_{\text{trash}}, \rho_S) \tag{20}$$

Here, the fidelity is given by $F(\rho_{\textbf{trash}}, \rho_S) = (\mathrm{Tr}\sqrt{\sqrt{\rho_S}\,\rho_{\textbf{trash}}\,\sqrt{\rho_C}})^2$. When $\mathcal{L} \to 0$, we have $\rho_{\text{trash}} \to \rho_S$, which can be vividly described as an optimization that "drags" the state toward ρ_S.

Algorithm 1 Workflow of the QAE-DD Algorithm.

Require: Number of qubits N_A, N_B, total $N = N_A + N_B$;
Training states $\{\rho_C^i\}_{i=1}^m$;
Secret states $\{\rho_S^i\}_{i=1}^m$;
Encoder parameters $\theta \in \mathbb{R}^d$ (randomly initialized);
Learning rate α, Training epochs T
Ensure: Optimized parameters θ
Initialize encoder parameters θ
for $t = 1$ to T **do**
  total_loss $\leftarrow 0$, total_fid $\leftarrow 0$
  for each training sample $i = 1 \ldots m$ **do**
    Normalize secret state $\rho_S^i \leftarrow \rho_S^i / \mathrm{Tr}(\rho_S^i)$
    Construct encoder unitary $U_E(\theta)$ via parametrized circuit
    Encode input: $\rho_{BA} \leftarrow U_E \rho_C^i U_E^\dagger$
    Obtain subsystems: $\rho_{\text{info}} \leftarrow \mathrm{Tr}_A(\rho_{BA})$, $\rho_{\text{trash}} \leftarrow \mathrm{Tr}_{B,A}(\rho_{BA})$
    Compute loss: $L_i \leftarrow 1 - \mathrm{Tr}(\rho_{\text{trash}} \cdot \rho_S^i)$
    Reconstruct joint state: $\rho_{\text{de-in}} \leftarrow \rho_S^i \otimes \rho_{\text{info}}$
    Decode output: $\rho_{\text{stego}} \leftarrow U_E^\dagger \rho_{\text{de-in}} U_E$
    Fidelity evaluation: $F_i \leftarrow \mathrm{Fid}(\rho_C^i, \rho_{\text{stego}})$
    Accumulate: total_loss $\leftarrow$ total_loss $+L_i$; total_fid $\leftarrow$ total_fid $+F_i$
  end for
  Compute averages: $L_{\text{avg}} \leftarrow$ total_loss $/m$, $F_{\text{avg}} \leftarrow$ total_fid $/m$
  Backpropagate and update θ using Adam optimizer
  Record L_{avg} and F_{avg}
end for
return θ, fidelity curves

In the simulation experiments, we first trained QAE-DD using a single data pair to verify the basic capability of applying QAE in reverse. Then, we increased the number of data pairs to evaluate the model's generalization ability. The training procedures of the QAE-DD is shown in Algorithm 1. As shown in Fig. 5(a), the simulation results indicate that by redefining the functional role of the QAE, it is feasible to embed secret quantum states into high-dimensional mixed entangled states using a neural network. The mixed entangled states in this case are more complex than those analyzed in Sect. 3.1, but they provide greater security benefits. However, when the data size n increases (especially when $n \geq 5$), the average fidelity $F(\rho_C, \rho_{\text{stego}})$ decreases significantly (see Fig. 5(b)), showing that the network's ability to learn multiple information-hiding encoding–decoding operations $\mathcal{E}_S^{(k)}$ is limited. This limitation arises because the network structure in this scheme is unitary, while $\mathcal{E}_S^{(k)}$ differs for different data. Therefore, when applying this scheme, the task should be adapted to match its structure, ensuring that the network learns only a single $\mathcal{E}_S$ or one with small perturbations.

5 Improved scheme for single secret state tasks: hidden-layer variants based on the QAE structure

5.1 Design concept

As described in Sect. 3, the unitary operations used to implement high-dimensional information-hiding encoding can suffer from passive distortion due to hardware limitations, and in some cases, active scrambling is required to enhance security. However, the simulations in Sect. 4 show that a conventional QAE performs well only when learning a single static $\mathcal{E}_S$. The underlying reason is that the simple scheme optimizes the latent-layer trash state ρ_{trash} toward the target secret state ρ_S (1-qubit) via Hamiltonian optimization in a one-way manner, which theoretically cannot cover the two-dimensional Hilbert space with a single dimension.

To address this, we consider using a branching structure that optimizes the target simultaneously in two orthogonal directions, followed by synthesis. This module is implemented in the latent layer of the QAE, where two parallel networks optimize the latent representation from two orthogonal directions, aiming to obtain a more complete feature representation. Specifically, we explore two strategies:

(1) Soft constraint – embed an orthogonal basis optimization term into the loss function, driving the latent state to converge toward the directions corresponding to their Hamiltonians simultaneously: $H_{E_0}^{(\text{trash})} \to |0\rangle\langle 0|$, $H_{E_1}^{(\text{trash})} \to |1\rangle\langle 1|$.

(2) Hard decomposition – use projection operators to force the latent layer state to decompose into the secret subspace and its orthogonal complement, and learn features separately within the complementary space: $H_{E_0}^{(\text{info})} = (I_2 \otimes \langle 0|)\,\rho_{E_0}\,(I_2 \otimes |0\rangle)$, $H_{E_1}^{(\text{info})} = (I_2 \otimes \langle 1|)\,\rho_{E_0}\,(I_2 \otimes |1\rangle)$.

Afterward, a synthesis module is constructed to output the stego state ρ_{stego}, and the decoder $\mathcal{D}$ is trained independently to extract the secret state ρ_S. The implementation of this network structure is shown in Fig. 6. Note that this structure introduces operations such as tracing out, decomposition, and summation, making it overall non-unitary.

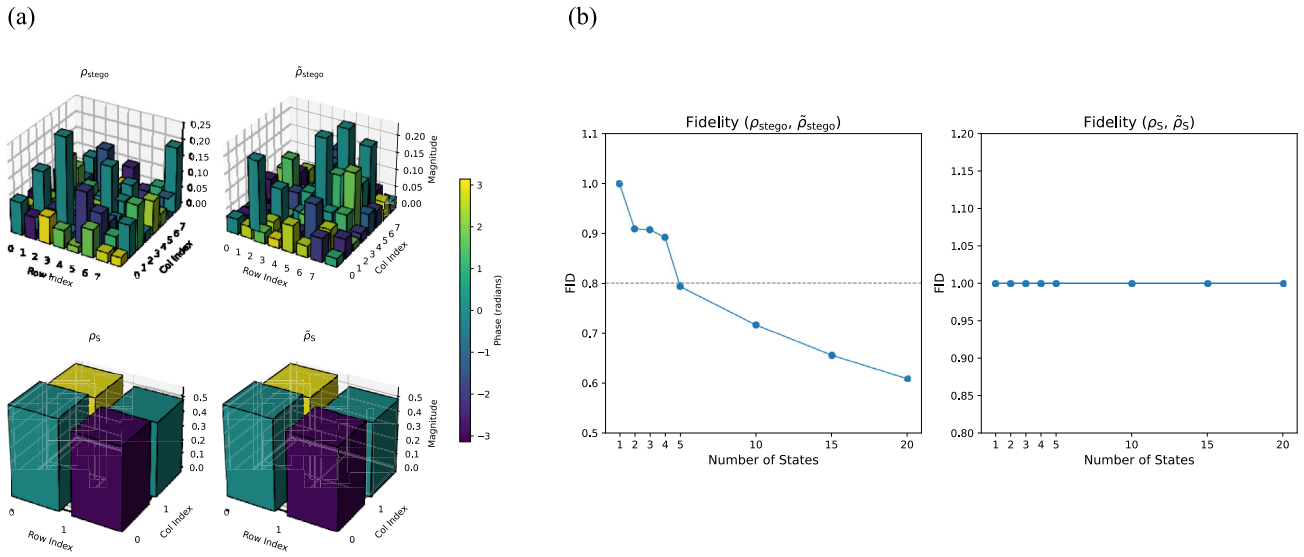

Fig. 5 Training results of the simple QAE-DD model. Subfigure (a) shows the comparison of density matrices for the stego state (top) and the secret state (bottom) before and after training the network with a single data pair. Subfigure (b) shows the fidelity curves during training with multiple data pairs, where the left plot presents the fidelity variation of the stego states and the right plot presents that of the secret states. Since the QAE-DD framework embeds the secret state into the stego state by reversing the input to the network, the fidelity of the secret state remains 1 throughout

The application of the above model in the information-hiding task is as follows: (1) Alice trains the network using all possible secret states to be transmitted, and after training, she shares the decoder $\mathcal{D}$ with Bob. (2) Alice can obtain the matrix elements of the reference states $\{\rho_{E_0}, \rho_{E_1}\}$ through simulation, and prepare them in a real quantum system (ρ_{E_0} and ρ_{E_1} have lower dimensions than ρ_{stego}, reducing the difficulty of preparation). She then uses the network's forward inference to obtain $\tilde{\rho}_{\text{stego}}$. (3) After receiving $\tilde{\rho}_{\text{stego}}$, Bob applies the decoder $\mathcal{D}$ to extract the secret state $\tilde{\rho}_S$.

5.2 Simulation experiments

The basic experimental setup is as follows:

1. **Data**: The sizes are still set as a 1-qubit secret state ρ_S and a 2-qubit reference state ρ_R. The generation method for ρ_S is the same as in the simple scheme. The reference state ρ_R is generated from a diagonal density matrix of a mixed state: $\rho_R = WDW^\dagger$ where $D = \text{diag}(p_1, p_2, p_3, p_4)$, $\sum_k p_k = 1$ and W is chosen

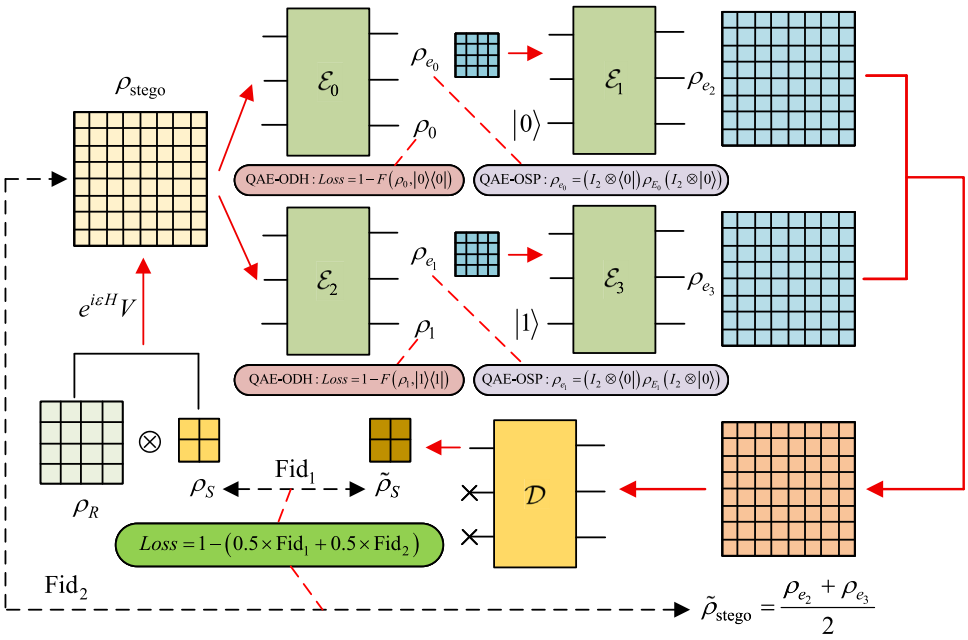

Fig. 6 Framework diagram of the orthogonal-branch latent-layer variant of QAE for secret state encoding. The two strategies, QAE-ODH and QAE-OSP, are both illustrated in the latent layer (although, in the experiments, they are implemented as two distinct network architectures; they are shown together in the schematic only for comparison purposes)

Algorithm 2 Workflow of the QED-OSP Algorithm.

Require: Number of qubits N_A, N_B, total $N = N_A + N_B$;
Training states $\{\rho_R^i\}_{i=1}^m$;
Secret states $\{\rho_S^i\}_{i=1}^m$;
Predefined unitary transforms $\{V_1^i\}_{i=1}^m$;
Encoder parameters $\theta_0, \theta_1, \theta_2, \theta_3$;
Decoder parameters ϕ;
Learning rate α, Training epochs T
Ensure: Optimized parameters $\theta_0, \theta_1, \theta_2, \theta_3, \phi$
Initialize $\theta_0, \theta_1, \theta_2, \theta_3, \phi$ randomly
for $t = 1$ to T **do**
  total_loss $\leftarrow 0$, total_fid_in $\leftarrow 0$, total_fid_secret $\leftarrow 0$
  for each sample $i = 1 \ldots m$ **do**
    Construct augmented input: $\rho_{\text{stego}} \leftarrow V_1^i(\rho_S^i \otimes \rho_R^i)V_1^{i\dagger}$
    Pass through encoder-0: $\rho_{enc_0} \leftarrow \mathcal{E}_0(\rho_{\text{stego}}, \theta_0)$
    Extract subsystem $\rho_{e_0} \leftarrow \text{Tr}_{\text{label}=0}(\rho_{enc_0})$
    Pass through encoder-1: $\rho_{enc_1} \leftarrow \mathcal{E}_1(\rho_{\text{stego}}, \theta_1)$
    Extract subsystem $\rho_{e_1} \leftarrow \text{Tr}_{\text{label}=1}(\rho_{enc_1})$
    Construct inputs for encoder-2,3: $\rho_{\text{in2}} \leftarrow \rho_{e_0} \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|$, $\rho_{\text{in3}} \leftarrow \rho_{e_1} \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|$
    Apply encoder-2: $\rho_{e_2} \leftarrow \mathcal{E}_2(\rho_{\text{in2}}, \theta_2)$
    Apply encoder-3: $\rho_{e_3} \leftarrow \mathcal{E}_3(\rho_{\text{in3}}, \theta_3)$
    Merge outputs: $\tilde{\rho}_{\text{stego}} \leftarrow \frac{1}{2}(\rho_{e_2} + \rho_{e_3})$, normalize
    Decode: $\tilde{\rho}_S \leftarrow \mathcal{D}(\tilde{\rho}_{\text{stego}}, \phi)$
    Compute fidelities: $F_1^i \leftarrow \text{Fid}(\rho_S, \tilde{\rho}_S)$, $F_2^i \leftarrow \text{Fid}(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}})$
    Loss: $L_i \leftarrow 1 - \left(0.5 \times F_1^i + 0.5 \times F_2^i\right)$
    Accumulate: total_loss $+= L_i$, total_fid_in $+= F_1^i$, total_fid_secret $+= F_2^i$
  end for
  Compute averages: $L_{\text{avg}} \leftarrow$ total_loss$/m$, $F_1 \leftarrow$ total_fid_in$/m$, $F_2 \leftarrow$ total_fid_secret$/m$
  Backpropagate and update $\theta_0, \theta_1, \theta_2, \theta_3, \phi$ with Adam optimizer
  Record $L_{\text{avg}}, F_{\text{in}}, F_s$
end for
return optimized $\theta_0, \theta_1, \theta_2, \theta_3, \phi$

from a Haar-random unitary matrix (fixed after sampling), $W \sim \text{U}_{\text{Haar}}^{4\times4}$. The stego state is given by:

$$\rho_{\text{stego}} = e^{i\varepsilon H} V \left(\rho_R \otimes \rho_S\right) V^\dagger e^{-i\varepsilon H} \tag{21}$$

The meanings of the variables are the same as in (10). The stego state and secret state output by the model are denoted as $\tilde{\rho}_{\text{stego}}$ and $\tilde{\rho}_S$, respectively. Due to hardware limitations, the dataset size is restricted to $n \leq 10$, denoted as n groups of "1 + 1" data (representing pairs $\left(\rho_S^{(k)}, \rho_{\text{stego}}^{(k)}\right)$).

2. **Model structure:** In addition to the two latent-layer improvement strategies described in Sect. 5.1 (Fig. 6), we also include the QAE-DD structure from the simple scheme in Sect. 4. We denote the model structure corresponding to the orthogonal-direction Hamiltonian optimization strategy ("soft constraint") as QAE-ODH, and the model structure corresponding to the orthogonal-subspace projection strategy ("hard decomposition") as QAE-OSP. Each submodule in the network is implemented using variational quantum circuits (the encoder uses a 6-layer parameterized quantum circuit, and the decoder uses an 8-layer parameterized quantum circuit).
3. **Model training:** QAE-DD still uses the loss function defined in (20); the loss function for QAE-ODH is:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{E_0} + \mathcal{L}_{E_1} + \mathcal{L}_{\rho_{\text{stego}}} + \mathcal{L}_{\rho_S} \\ &= (1 - F(\rho_0, |0\rangle\langle 0|)) + (1 - F(\rho_1, |1\rangle\langle 1|)) \\ &\quad + \left(1 - F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)\right) + (1 - F(\rho_S, \tilde{\rho}_S))\end{aligned} \tag{22}$$

The roles of $\mathcal{L}_{E_0}$ and $\mathcal{L}_{E_1}$ are to perform Hamiltonian optimization along orthogonal directions. In contrast, QAE-OSP enforces a projection onto orthogonal subspaces:

$$\begin{cases}\rho_{e_0} = (I_2 \otimes \langle 0|)\, \rho_{E_0}\, (I_2 \otimes |0\rangle) \\ \rho_{e_1} = (I_2 \otimes \langle 1|)\, \rho_{E_1}\, (I_2 \otimes |1\rangle)\end{cases} \tag{23}$$

Here, ρ_{E_0} and ρ_{E_1} denote the outputs (3-qubits) of the parallel encoding modules E_0 and E_1, respectively, while ρ_{e_0} and ρ_{e_1} represent the latent-layer features. The loss function of QAE-OSP is defined as:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\rho_{\text{stego}}} + \mathcal{L}_{\rho_S} \\ &= \left(1 - F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)\right) + (1 - F(\rho_S, \tilde{\rho}_S))\end{aligned} \tag{24}$$

The number of training epochs is set to 150, and the learning rate is set to 0.2. The training procedures of the QAE-OSP and QAE-ODH are shown in Algorithm 2 and Algorithm 3.

The simulation results are shown in Figs. 7, 8, 10, and 11, with the analysis as follows:

(1) Figures 7 and 8 present the training curves of the three models for multiple datasets under $\varepsilon = 0$ and $\varepsilon = 0.6$. It can be seen that, within 10 datasets, the QAE-OSP model achieves fidelities for reconstructing ρ_{stego} and extracting ρ_S that are close to those of the simpler QAE-DD scheme; in contrast, the performance of QAE-ODH is quite limited, and its fidelity decreases significantly as the data size increases (Fig. 9).

(2) Figures 10 and 11 show two representative cases: the variation of fidelity with ε under fixed data size $n = 5$, and the variation of fidelity with data size n under fixed $\varepsilon = 0.6$. The fixed $\varepsilon = 0.6$ setting is chosen to balance hiding strength and network capability (maintaining fidelity). The results indicate that, for smaller ε and smaller n, the QAE-DD scheme performs better, as the neural network tends to simulate a static, single information-hiding encoding operation $\mathcal{E}_S$ more

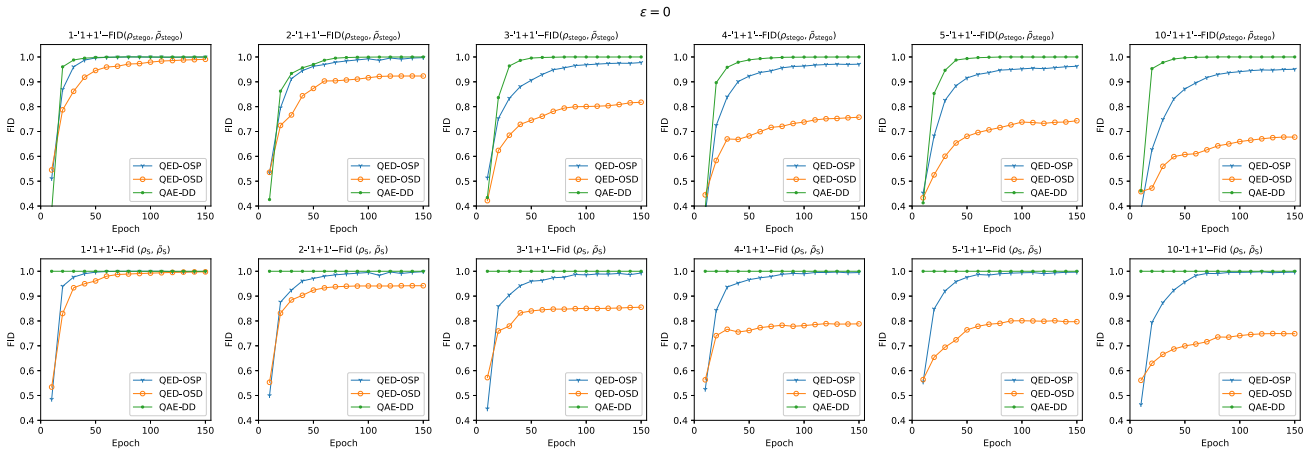

Fig. 7 Fidelity training curves for $\epsilon = 0$ with data sizes of 1, 2, 3, 4, 5, and 10. The first row shows $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)$, and the second row shows $F\left(\rho_S, \tilde{\rho}_S\right)$

effectively and with lower resource consumption. However, for larger ε and larger n, which demand more dynamic and disturbance-resistant capabilities from the neural network, the QAE-OSP model outperforms, with stronger reconstruction capability of ρ_{stego} compared to QAE-DD. This validates the advantage of orthogonal projection branching with parallel subnetworks. The QAE-ODH model consistently yields low numerical results, indicating that the "soft constraint" strategy of Hamiltonian optimization in orthogonal directions has limited effectiveness for this task.

(3) Additionally, in the QAE-DD model, $F\left(\rho_S, \tilde{\rho}_S\right) \equiv 1$ because, in this scheme (see Sect. 4), $\tilde{\rho}_{\text{stego}}$ is obtained by applying the decoder $\mathcal{D}$ to the new reference state $\tilde{\rho}_R$ (extracted by the encoder $\mathcal{E}$ of QAE) together with ρ_S. Applying $\mathcal{D}^\dagger$ perfectly decouples and extracts the secret state ρ_S. The trade-off, however, is that with larger perturbations and larger data sizes, the reconstruction fidelity of ρ_{stego} decreases, reducing the network's ability to simulate $\mathcal{E}_S$.

6 Multi-stego state task scheme

6.1 Design concept

We still employ the QAE and its variant architectures. As described in Sect. 3.2, implementing multi-stego-state information hiding tasks requires the construction of non-unitary operations. We adopt the previously introduced QAE-DD and QAE-OSP models for this purpose (in QAE-DD, the reference state ρ_R changes with the secret state during operation, and can therefore be regarded as globally non-unitary). Unlike the single-stego-state case, the focus here lies in the construction of training data, specifically in achieving the objective of making stego states close to each other.

The specific design is as follows: For a set of n random secret states $\left\{\rho_S^{(k)}\right\}$, each secret state $\rho_S^{(k)}$ is tensor-producted with the same reference state ρ_R, and then encoded into a stego state $\left\{\rho_{\text{stego}}^{(k)}\right\}$ using (21). According to the analysis in Section 3.2, calculate the 'centroid' state of this

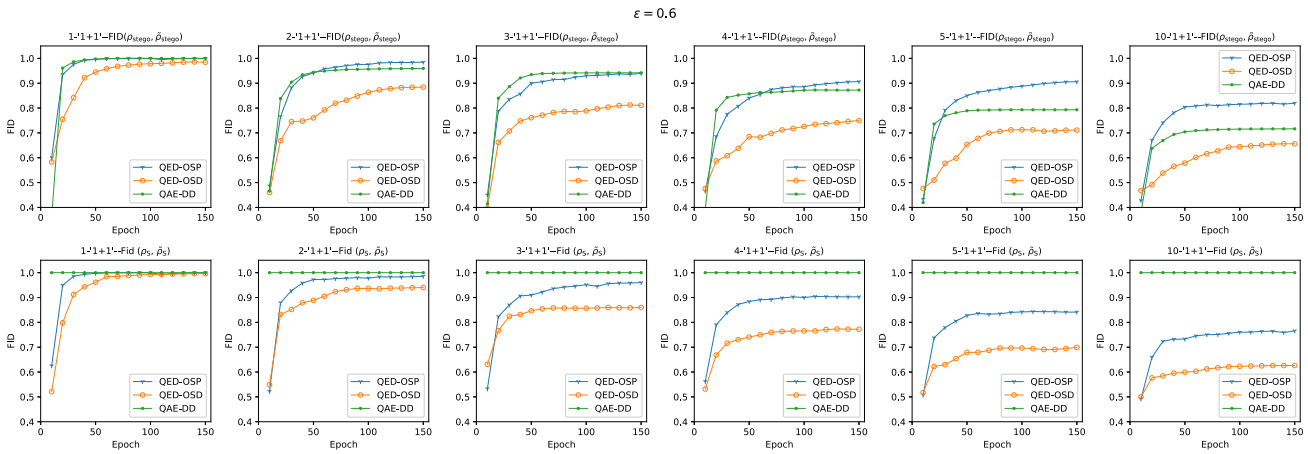

Fig. 8 Fidelity training curves for $\epsilon = 0.6$ with data sizes of 1, 2, 3, 4, 5, and 10. The first row shows $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)$, and the second row shows $F\left(\rho_S, \tilde{\rho}_S\right)$

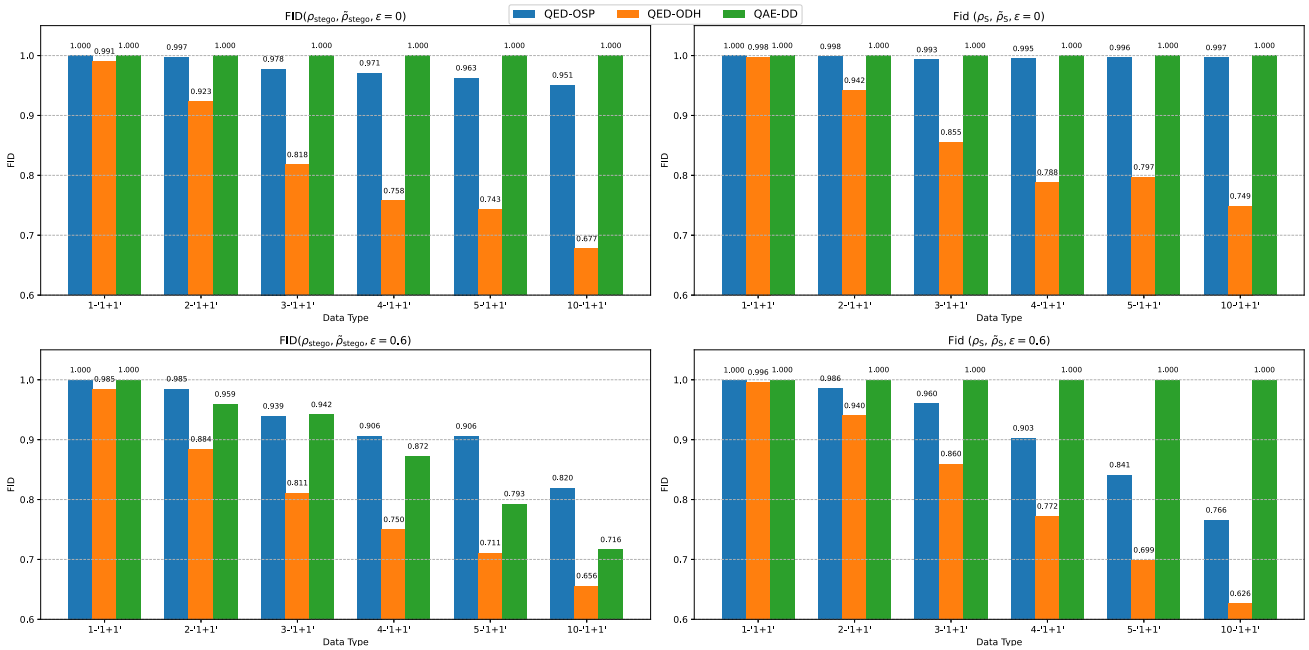

Fig. 9 Fidelity comparison across different data types after model convergence ($\epsilon = 0$ and $\epsilon = 0.6$ corresponding to Figs. 7, 8)

stego state set, $\rho_{\text{stego}}^{\text{cen}}$, and form the training data pairs $\left\{\left(\rho_{\text{stego}}^{(0)}, \rho_{\text{stego}}^{\text{cen}}\right), \left(\rho_{\text{stego}}^{(1)}, \rho_{\text{stego}}^{\text{cen}}\right), \ldots\right\}$.

Once the network is trained, the information hiding task can be performed in the manner described in Sect. 4 or Sect. 5. From the perspective of an eavesdropper, these highly similar stego states would be mistaken for multiple copies of the same entangled state, with noise errors introduced during the preparation or transmission process.

6.2 Simulation Experiments

The experimental setup is generally the same as in Sect. 4 and Sect. 5:

1. **Data**: According to the construction method in Sect. 6.1, we build $'1 + n'$ training data pairs (i.e., one $\rho_R + n\rho_S$ states, specifically $\{'1 + 1', \ldots, '1 + 5', '1 + 10'\}$).

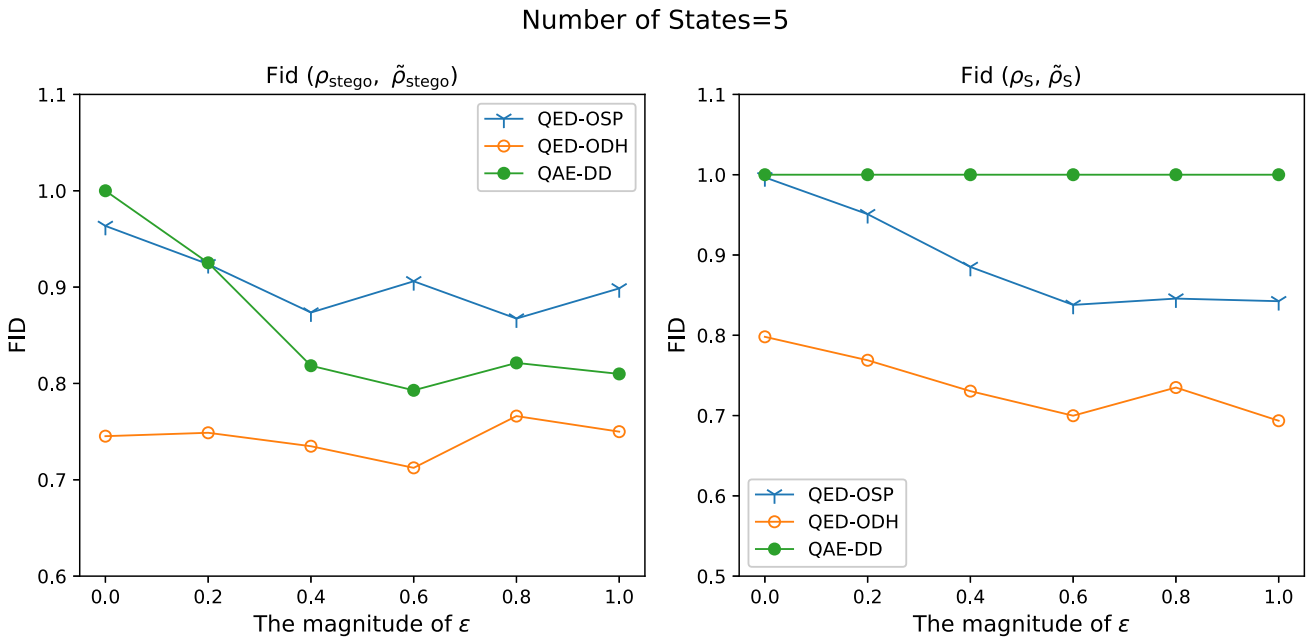

Fig. 10 Fidelity curves versus ϵ after model training with a data size of 5. The left figure shows $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)$, and the right figure shows $F\left(\rho_S, \tilde{\rho}_S\right)$

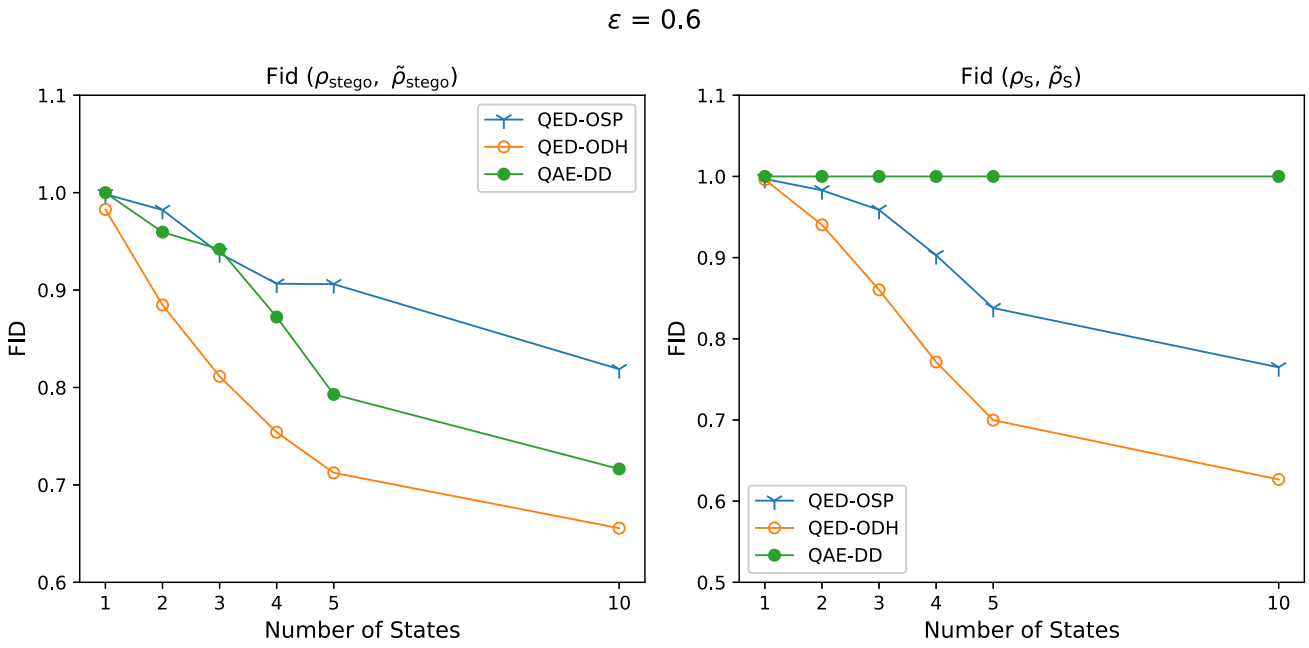

Fig. 11 Fidelity curves versus data size after model training with $\epsilon = 0.6$. The left figure shows $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right)$, and the right figure shows $F\left(\rho_S, \tilde{\rho}_S\right)$

2. **Model Structure**: We choose the QAE-DD and QAE-OSP models. Note again that in the use of the QAE-DD model, the intermediate reference state $\tilde{\rho}_R$ extracted by the encoder $\mathcal{E}$ is different from the ρ_R in the training data (see Sect. 4 for details).
3. **Model Training**: The settings are the same as in Sect. 4 and Sect. 5. The number of training epochs is set to 150, and the learning rate is set to 0.2.

The simulation results are shown in Figs. 12, 13, and 14, where: (1) Figures 12 and 13 respectively present the fidelities of ρ_{stego} reconstruction and ρ_S extraction for the two models under multiple datasets when the perturbation strengths are $\varepsilon = 0$ and $\varepsilon = 0.6$. The results are similar to those in Sect. 5: the QAE-DD model performs better under low perturbation, while the QAE-OSP model performs better under high perturbation. Simulation data indicate that, for $n \leq 10$, the fidelities of both models are close to or exceed 0.9 (except in the case of $n = 4$, where this anomaly suggests that the "centroid" of the stego states is not well balanced with the number of training data groups). (2) Figure 14 shows the comparison between the distance distributions of the original secret states and those of the encoded stego states for the two models under $n = 5, 10$ datasets and $\varepsilon = 0.6$ perturbation

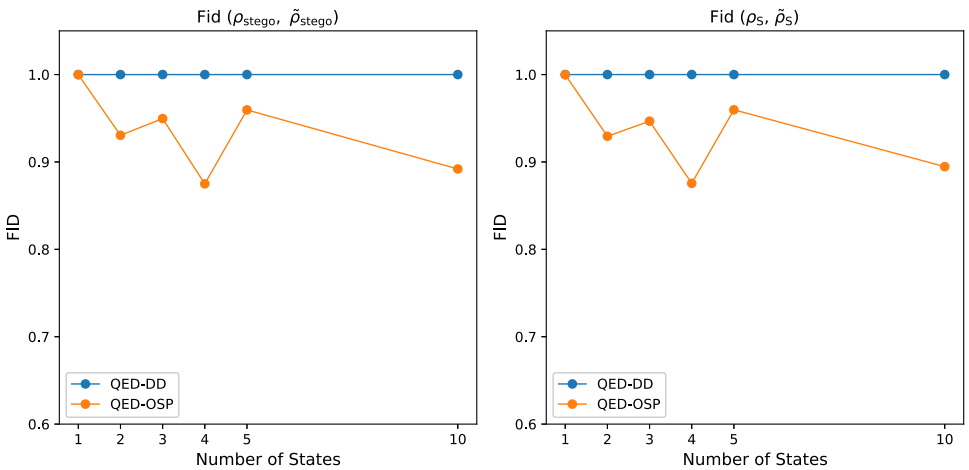

Fig. 12 Fidelity curves of the QAE-DD and QAE-OSP models versus data size after model training with $\epsilon = 0$

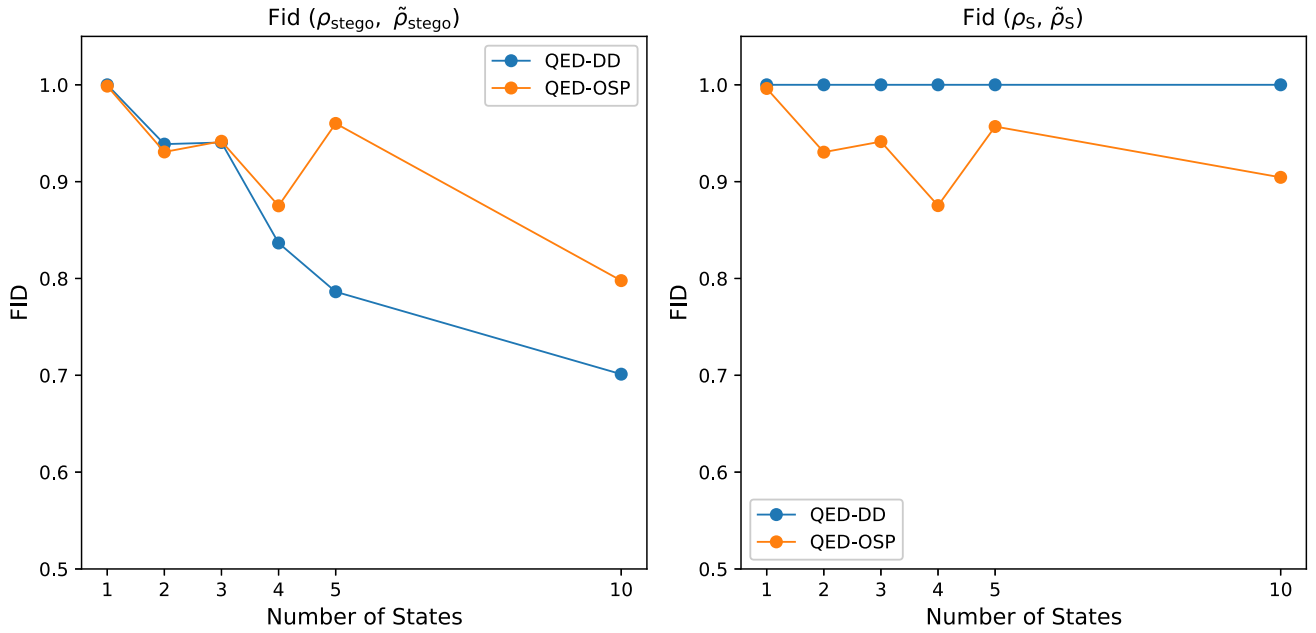

Fig. 13 Fidelity curves of the QAE-DD and QAE-OSP models versus data size after model training with $\epsilon = 0.6$

strength. As shown, after processing by the neural network, the originally scattered random secret states become aggregated once encoded into high-dimensional stego entangled states. In particular, for the QAE-OSP model, the average inter-state fidelity from secret states to stego states increases from 0.68 to 0.91 when $n = 5$, and from 0.49 to 0.62 when $n = 10$.

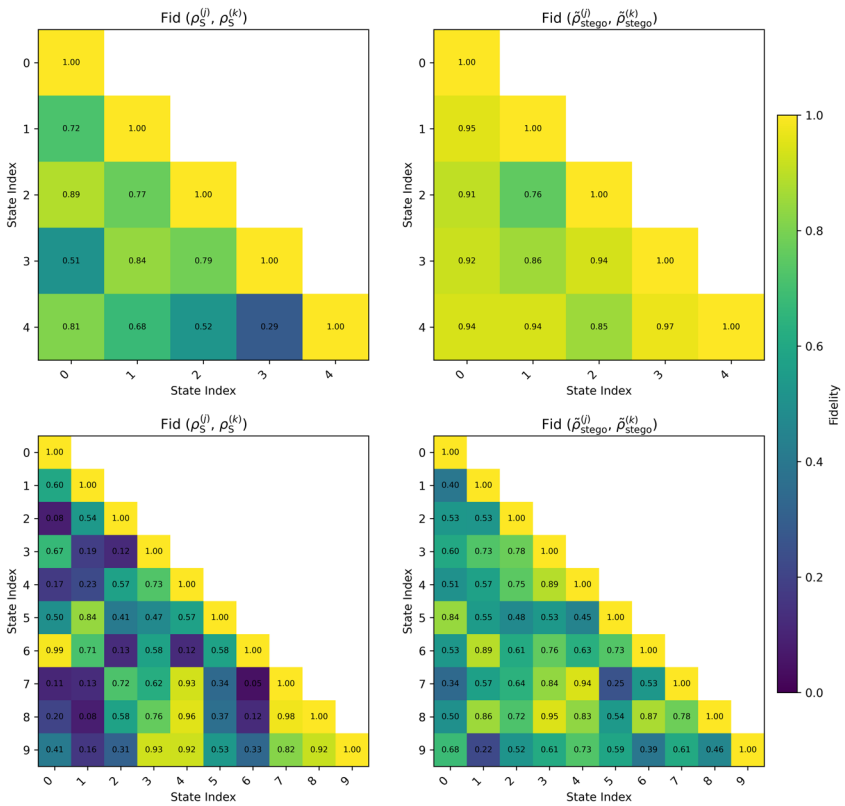

Fig. 14 Cross-group fidelity results for the QAE-OSP model with $\epsilon = 0.6$ and data sizes of 5 and 10. The plots include $F\left(\rho_S^{(j)}, \rho_S^{(k)}\right)$ and $F\left(\tilde{\rho}_{\text{stego}}^{(j)}, \tilde{\rho}_{\text{stego}}^{(k)}\right)$

7 Conclusion

This study presents a preliminary exploration into the emerging field of quantum information hiding, proposing a quantum state diffusion encoding paradigm based on high-dimensional entangled states, whereby secret quantum states are concealed within the global structure of multi-particle entangled systems. On the theoretical level, we establish an initial controllable random perturbation unitary operation model. Through formal analysis, we clarify the positive correlation between hiding strength and the perturbation parameters as well as entanglement dimension, thereby providing theoretical guidance for the design of information hiding codes. In terms of implementation, we explored two QAE-based neural network schemes: QAE-DD, which adapts to low-disturbance scenarios through the inverse use of encoding and decoding functions, and QAE-OSP, which introduces an orthogonal projection branching structure to optimize performance under high-disturbance conditions. The capabilities of both models were verified in a 3-qubit system with limited-scale data ($n \leq 10$). For example, under high perturbation conditions (e.g., $n = 5, \varepsilon = 0.6$), the QAE-OSP model maintains high fidelity between the carrier and secret states, with $F\left(\rho_{\text{stego}}, \tilde{\rho}_{\text{stego}}\right) = 0.91$ and $F\left(\rho_S, \tilde{\rho}_S\right) = 0.84$. Furthermore, we extended the single-stego-state task to a multi-stego-state setting, constructing the training set based on the idea of principal component analysis and evaluating the adaptability of both models to varying disturbance strengths. Under the representative conditions of $n = 5$ and $\varepsilon = 0.6$, the QAE-OSP model was validated to have the capability of aggregating multiple stego states, with the average inter-state fidelity from multiple secret states to stego states improving from 0.68 to 0.91. These exploratory works provide a potential implementation pathway for arbitrary quantum state information hiding.

Nevertheless, as an early exploration in this field, this study remains a preliminary proof-of-concept work with evident limitations: (1) Due to hardware constraints, the experiments were conducted only on small-scale systems with $n \leq 10$ data and 3 qubits, and large-scale scalability has not yet been verified. (2) The principles for setting the disturbance parameter ε were not discussed, and no adaptive mechanism has been implemented. (3) The construction of the "centroid" state relies on classical principal component analysis, without a fully quantum feature extraction process, which may become a computational bottleneck for multi-stego-state extensions.

In subsequent studies, we plan to deepen our research in three directions: (1) Enhance the representational capacity of QAE through methods such as Generative Adversarial Networks (Ma et al. 2025) and Quantum Architecture Search (Ma et al. 2024), and further expand the data scale. (2) Investigate dynamic perturbation enhancement mechanisms for the encoding operations, enabling real-time adaptation to data and channel conditions. (3) Develop a fully quantum processing pipeline, attempting to replace classical PCA with a variational quantum feature solver. These improvements are expected to advance quantum information hiding from theoretical validation toward preliminary applications, though practical deployment will still require long-term exploration.

Recent research in information security demonstrates that cross-disciplinary approaches are increasingly vital for enhancing robustness across diverse application domains. In image and video protection, studies have combined compression with nonlinear dynamics (Lin et al. 2024) and explored GAN-based anti-forensics (Wang et al. 2025), underscoring how hybrid physical–generative models improve resilience. Similarly, chaos-based cryptography has advanced through novel hyperchaotic systems and optimization strategies (Şimşek et al. 2025; Kocak et al. 2024), while cryptanalysis-oriented work has emphasized high-dimensional mappings and hardware-feasible implementations (Gao et al. 2025a, b, c). Parallel efforts in video encryption (Gao et al. 2024a; Gao et al. 2024b) further highlight the role of intelligent architectures in managing security-critical multimedia data. At the same time, steganography and watermarking studies (Li et al. 2024, 2022) focus on imperceptibility and robustness, offering valuable insights for covert communication. These works, particularly in exploring information hiding, provide significant inspiration for our research on quantum state embedding. Building on this foundation, by situating our QAE-driven framework in this broader landscape, we emphasize that its hybrid quantum–classical paradigm continues the trajectory of cross-domain innovations, extending them into the quantum regime to enable secure embedding of arbitrary quantum states and laying a foundation for future quantum steganography and cryptography.

Appendix A: Deviation calculation of *U* and *V*

Since $U = e^{i\varepsilon H} V$, and the trace norm is invariant under unitary transformations (i.e., $\|AV\|_1 = \|A\|_1$ for any unitary matrix V), we have:

$$\|U - V\|_1 = \|(e^{i\varepsilon H} - I)V\|_1 = \|e^{i\varepsilon H} - I\|_1 \tag{A1}$$

The problem now becomes to calculate $\|e^{i\varepsilon H} - I\|_1$.

The Taylor expansion of $e^{i\varepsilon H}$ is:

$$e^{i\varepsilon H} = I + i\varepsilon H - \frac{\varepsilon^2}{2}H^2 + O(\varepsilon^3) \tag{A2}$$

Therefore,

$$e^{i\varepsilon H} - I = i\varepsilon H - \frac{\varepsilon^2}{2}H^2 + O(\varepsilon^3) \tag{A3}$$

For small perturbation strength ε, the first-order term $i\varepsilon H$ dominates the deviation:

$$\|e^{i\varepsilon H} - I\|_1 \approx \varepsilon \|H\|_1 \tag{A4}$$

where $\|H\|_1 = \mathrm{tr}\,|H| = \sum_{k=1}^{d_E} |\lambda_k|$ (where λ_k are the eigenvalues of H).

The eigenvalue distribution of H follows the Wigner semicircle law (Anderson et al. 2009). For large dimension d_E, the probability density function of the eigenvalue λ is:

$$\mu(\lambda) = \frac{1}{2\pi d_E}\sqrt{4d_E - \lambda^2}, \quad \text{for } |\lambda| \le 2\sqrt{d_E} \tag{A5}$$

The expectation of the trace norm is:

$$\mathbb{E}[\|H\|_1] = d_E \cdot \mathbb{E}[|\lambda|] \approx d_E \int_{-2\sqrt{d_E}}^{2\sqrt{d_E}} |\lambda|\,\mu(\lambda)\,d\lambda \tag{A6}$$

The symbol μ in (A5) and (A6) represents the probability density function.

Due to the symmetry of the function:

$$\begin{aligned}\mathbb{E}[\|H\|_1] &\approx 2d_E \int_0^{2\sqrt{d_E}} \lambda\,\rho(\lambda)\,d\lambda \\ &= 2d_E \int_0^{2\sqrt{d_E}} \lambda \cdot \frac{1}{2\pi d_E}\sqrt{4d_E - \lambda^2}\,d\lambda \\ &= \frac{1}{\pi}\int_0^{2\sqrt{d_E}} \lambda\sqrt{4d_E - \lambda^2}\,d\lambda\end{aligned} \tag{A7}$$

Let $u = 4d_E - \lambda^2$, so $du = -2\lambda\,d\lambda$. Changing the limits of integration: when $\lambda = 0$, $u = 4d_E$; when $\lambda = 2\sqrt{d_E}$, $u = 0$.

$$\begin{aligned}\int_0^{2\sqrt{d_E}} \lambda\sqrt{4d_E - \lambda^2}\,d\lambda &= \int_{4d_E}^0 \sqrt{u}\left(-\frac{du}{2}\right) \\ &= \frac{1}{2}\int_0^{4d_E} u^{1/2}\,du \\ &= \frac{1}{2}\cdot\frac{2}{3}u^{3/2}\Big|_0^{4d_E} \\ &= \frac{1}{3}(4d_E)^{3/2} = \frac{8}{3}d_E^{3/2}\end{aligned} \tag{A8}$$

Substituting gives:

$$\mathbb{E}[\|H\|_1] \approx \frac{1}{\pi}\cdot\frac{8}{3}d_E^{3/2} = \frac{8}{3\pi}d_E^{3/2} \tag{A9}$$

$$\mathbb{E}[\|U - V\|_1] \approx \varepsilon \cdot \frac{8}{3\pi}d_E^{3/2} \tag{A10}$$

Appendix B: Proof of the optimality of the centroid state

Proof For any pure state $\sigma = |\psi\rangle\langle\psi|$, the fidelity is:

$$F^2(\rho_k, |\psi\rangle\langle\psi|) = \langle\psi|\rho_k|\psi\rangle. \tag{B11}$$

Objective function:

$$J(\sigma) = \sum_k \langle\psi|\rho_k|\psi\rangle = \langle\psi|\left(\sum_k \rho_k\right)|\psi\rangle = \langle\psi|M|\psi\rangle. \tag{B12}$$

Since the operator M is derived from ρ_k, and is a Hermitian matrix, it can be expressed as:

$$M = \sum_j \alpha_j |\beta_j\rangle\langle\beta_j|. \tag{B13}$$

where α_j are real eigenvalues, and $|\beta_j\rangle$ are the corresponding eigenvectors.

According to the Rayleigh-Ritz theorem (Wang et al. 2004), for any pure state $|\psi\rangle$, we have:

$$\langle\psi|M|\psi\rangle \le \alpha_{\max}, \tag{B14}$$

where $\alpha_{\max}$ is the largest eigenvalue of M, and it is achieved when $|\psi\rangle$ is the eigenvector corresponding to $\alpha_{\max}$.

Thus, $|\psi\rangle = |\beta_{\max}\rangle$, so:

$$\max J(\sigma) = \alpha_{\max}. \tag{B15}$$

The maximum value is obtained when $\sigma = |\beta_{\max}\rangle\langle\beta_{\max}| = \rho_{\mathrm{cen}}$.

Appendix C: Detailed algorithms for QED-ODH

Algorithm 3 Workflow of the QED-ODH Algorithm.

Require: Number of qubits N_A, N_B, total $N = N_A + N_B$;
Training states $\{\rho_R^i\}_{i=1}^m$; Secret states $\{\rho_S^i\}_{i=1}^m$; Predefined unitary transforms $\{V_1^i\}_{i=1}^m$; Encoder parameters $\theta_0, \theta_1, \theta_2, \theta_3$; Decoder parameters ϕ; Weights w_{in}, w_s; Learning rate α, Training epochs T
Ensure: Optimized parameters $\theta_0, \theta_1, \theta_2, \theta_3, \phi$
Initialize: $\theta_0, \theta_1, \theta_2, \theta_3, \phi$ randomly
for $t = 1$ to T **do**
  Initialize accumulators: total_loss $\leftarrow 0$, total_loss_e0 $\leftarrow 0$, total_loss_e1 $\leftarrow 0$, total_fid_in $\leftarrow 0$, total_fid_s $\leftarrow 0$
  for each sample $i = 1 \ldots m$ **do**
    Build stego input: $\rho_{\text{stego}} \leftarrow V_1^i \left(\rho_S^i \otimes \rho_R^i\right) V_1^{i\dagger}$
    Encoder-0 forward: $\rho_{\text{enc0}} \leftarrow \mathcal{E}_0(\rho_{\text{stego}}, \theta_0)$
    Subsystem extraction (E0): $\rho_{e_0} \leftarrow (I_2 \otimes \langle 0|)\, \rho_{\text{enc0}}\, (I_2 \otimes |0\rangle)$
    Encoder-1 forward: $\rho_{\text{enc1}} \leftarrow \mathcal{E}_1(\rho_{\text{stego}}, \theta_1)$
    Subsystem extraction (E1): $\rho_{e_1} \leftarrow (I_2 \otimes \langle 0|)\, \rho_{\text{enc1}}\, (I_2 \otimes |0\rangle)$
    Build inputs for E2/E3: $\rho_{\text{in2}} \leftarrow \rho_{e_0} \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|$, $\rho_{\text{in3}} \leftarrow \rho_{e_1} \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|$
    Encoder-2 forward: $\rho_{e_2} \leftarrow \mathcal{E}_2(\rho_{\text{in2}}, \theta_2)$
    Encoder-3 forward: $\rho_{e_3} \leftarrow \mathcal{E}_3(\rho_{\text{in3}}, \theta_3)$
    Merge and normalize: $\tilde{\rho}_{stego} \leftarrow \frac{1}{2}\left(\rho_{e_2} + \rho_{e_3}\right)$; $\tilde{\rho}_{stego} \leftarrow \tilde{\rho}_{stego}/\text{Tr}(\tilde{\rho}_{stego})$
    Decoder: $\tilde{\rho}_S \leftarrow \mathcal{D}(\tilde{\rho}_{stego}, \phi)$
    Fidelities: $F_{\text{stego}}^i \leftarrow \text{Fid}(\rho_{\text{stego}}, \tilde{\rho}_{stego})$, $F_S^i \leftarrow \text{Fid}(\rho_S^i, \tilde{\rho}_S)$
    Encoder-specific losses: $L_{e0}^i \leftarrow 1 - \text{Tr}(\rho_{\text{enc0}}^{\text{trash}} \cdot |0\rangle\langle 0|)$, $L_{e1}^i \leftarrow 1 - \text{Tr}(\rho_{\text{enc1}}^{\text{trash}} \cdot |1\rangle\langle 1|)$ ▷ $\rho_{\text{enc}*}^{\text{trash}}$ denotes the discarded subsystem orthogonal to ρ_{e_*}
    Global loss: $L_i \leftarrow 1 - \left(0.5 \times F_{\text{stego}}^i + 0.5 \times F_s^i\right)$
    Accumulate: total_loss $+= L_i$; total_loss_e0 $+= L_{e0}^i$; total_loss_e1 $+= L_{e1}^i$; total_fid_in $+= F_{\text{stego}}^i$; total_fid_s $+= F_s^i$
  end for
  Averages: $L_{\text{avg}} \leftarrow$ total_loss$/m$; $L_{e0}^{\text{avg}} \leftarrow$ total_loss_e0$/m$; $L_{e1}^{\text{avg}} \leftarrow$ total_loss_e1$/m$; $F_{\text{stego}}^{\text{avg}} \leftarrow$ total_fid_in$/m$; $F_s^{\text{avg}} \leftarrow$ total_fid_s$/m$
  Backpropagate: update θ_0 with L_{e0}^{avg}; update θ_1 with L_{e1}^{avg}; update θ_2, θ_3, ϕ with L_{avg}
end for
return optimized $\theta_0, \theta_1, \theta_2, \theta_3, \phi$

Acknowledgements The authors would like to thank Qing Mu for valuable discussions and constructive suggestions that greatly improved the quality of this work.

Author Contributions CLH and QGM designed the experiments, developed the idea and framework. CLH was responsible for deriving the theoretical framework and coding. DQ, YQC, and HZ contributed to the review, editing, and evaluation of the results. The manuscript was written with input from all authors, and all authors reviewed and approved the final version.

Funding This work is supported by the National Natural Science Foundation of China under Grant No. 62171470, Henan Province Central Plains Science and Technology Innovation Leading Talent Project (No. 234200510019), Natural Science Foundation of Henan (No. 232300421240), Natural Science Foundation of Henan (252300420990).

Data Availability Data will be made available on request.

Code Availability Code will be made available on request.

Declarations

Competing interests The authors declare no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for Publication The Author confirms: (1) that the work described has not been published before; (2) that it is not under consideration for publication elsewhere.

References

Anderson GW, Guionnet A, Zeitouni O (2009) An introduction to random matrices. Cambridge University Press, Cambridge, U.K

Bai C-M, Wang M-Y, Zhang S-J, Liu L (2025) Masking quantum information in multipartite systems via fourier and hadamard matrices. Commun Theor Phys 77:101–111

Bennett CH, Brassard G (1984) Quantum cryptography: public key distribution and coin tossing. Proc. IEEE Int. Conf. Comput. Syst. Signal Process, New York, NY, USA, pp 175–179

Bergholm V, Izaac J, Schuld M, Gogolin C, Ahmed S, Ajith V, Alam MS, Alonso-Linaje G, AkashNarayanan B, Asadi A, Arrazola JM, Azad U, Banning S, Blank C, Bromley TR, Cordier BA, Ceroni J, Delgado A, Matteo OD, Dusko A, Garg T, Guala D, Hayes A, Hill R Ijaz A, Isacsson T, Ittah D, Jahangiri S, Jain P, Jiang E, Khandelwal A, Kottmann K, Lang RA, Lee C, Loke T, Lowe A, McKiernan K, Meyer JJ, Montañez-Barrera JA, Moyard R, Niu Z, O'Riordan LJ, Oud S, Panigrahi A, Park C-Y, Polatajko D, Quesada N, Roberts C, Sá N, Schoch I, Shi B, Shu S, Sim S, Singh A, Strandberg I, Soni J, Száva A, Thabet S, Vargas-Hernández RA, Vincent T, Vitucci N, Weber M, Wierichs D, Wiersema R, Willmann M, Wong V, Zhang S, Killoran N (2022) PennyLane: automatic differentiation of hybrid quantum-classical computations. https://arxiv.org/abs/1811.04968

Bondarenko D, Feldmann P (2020) Quantum autoencoders to denoise quantum data. Phys Rev Lett 124:130502

Cao C, Wang X (2021) Noise-assisted quantum autoencoder. Phys Rev Appl 15:054012

Cheddad A, Condell J, Curran K, Mc Kevitt P (2010) Digital image steganography: survey and analysis of current methods. Signal Process 90(3):727–752

Chen Y, Hou Y, Wang Z, Wang T, Wu Z, Li Z, Peng X (2025) Enhanced natural parameterized quantum circuit. Phys Rev Res 7(1):013221

Decker T, Gallezot M, Kerstan SF, Paesano A, Ginter A, Wormsbecher W (2025) Quantum key distribution as a quantum machine learn-

ing task. npj Quant Inf 11. https://doi.org/10.1038/s41534-025-00709-0
Dong Y, Yan R (2024) A new integrated steganography scheme for quantum color images. J Supercomput 80(16):24758–24780
Emerson J, Weinstein YS, Saraceno M, Lloyd S, Cory DG (2003) Pseudo-random unitary operators for quantum information processing. Science 302(5653):2098–2100
Gao S, Iu HH-C, Mou J, Erkan U, Liu J, Wu R, Tang X (2024) Temporal action segmentation for video encryption. Chaos Solitons Fract 183:114958. https://doi.org/10.1016/j.chaos.2024.114958
Gao S, Liu J, Iu HH-C, Erkan U, Zhou S, Wu R, Tang X (2024) Development of a video encryption algorithm for critical areas using 2d extended Schaffer function map and neural networks. Appl Math Model 134:520–537. https://doi.org/10.1016/j.apm.2024.06.016
Gao S, Ding S, Iu HH-C, Erkan U, Toktas A, Simsek C, Wu R, Xu X, Cao Y, Mou J (2025) A three-dimensional memristor-based hyperchaotic map for pseudorandom number generation and multi-image encryption. Chaos 35(7):073105. https://doi.org/10.1063/5.0270220
Gao S, Zhang Z, Iu HH-C, Ding S, Mou J, Erkan U, Toktas A, Li Q, Wang C, Cao Y (2025) A parallel color image encryption algorithm based on a 2-d logistic-rulkov neuron map. IEEE Internet Things J 12(11):18115–18124. https://doi.org/10.1109/JIOT.2025.3540097
Gao S, Ho-Ching I, H., Erkan, U., Simsek, C., Toktas, A., Cao, Y., Wu, R., Mou, J., Li, Q., Wang, C. (2025) A 3d memristive cubic map with dual discrete memristors: design, implementation, and application in image encryption. IEEE Trans Circuits Syst Video Technol 35(8):7706–7718. https://doi.org/10.1109/TCSVT.2025.3545868
Hao C, Yang X, Ma Q, Qu D, Wang R, Zhang T (2024) Quantum audio lsb steganography with entanglement-assisted modulation. Quantum Inf Process 23(3):106
Hao C-L, Ma Q-G, Si N-W, Liu B-Y, Qu D (2025) Neural-enabled quantum information hiding with error-correcting codes: a novel framework for arbitrary quantum state embedding. EPJ Quant Technol 12:88
Helstrom CW (1969) Quantum detection and estimation theory. J Stat Phys 1(2):231–252
Huang C-J, Ma H, Yin Q, Tang J-F, Dong D, Chen C, Xiang G-Y, Li C-F, Guo G-C (2020) Realization of a quantum autoencoder for lossless compression of quantum data. Phys Rev A 102:032412
Huang Y, Li Q, Hou X, Wu R, Yung M-H, Bayat A, Wang X (2022) Robust resource-efficient quantum variational ansatz through an evolutionary algorithm. Phys Rev A 105(5)
Kandi H, Mishra D, Gorthi SRS (2017) Exploring the learning capabilities of convolutional neural networks for robust image watermarking. Comput Sec 65:247–268
Katzenbeisser S, Petitcolas F (2016) Information hiding. Artech House, Norwood, MA
Kavan M, Kumar AP, Aditi DS, Sen U (2018) Masking quantum information is impossible. Phys Rev Lett 120:230501
Kaye P, Laflamme R, Mosca M (2006) An introduction to quantum computing. Oxford University Press, Oxford, U.K. https://doi.org/10.1093/oso/9780198570004.001.0001 . Online edition published 12 Nov. 2020, Accessed on 21 Aug 2025
Kirchenbauer J, Geiping J, Wen Y, Katz J, Miers I, Goldstein T (2023) A watermark for large language models. In: International conference on machine learning. PMLR, pp 17061–17084
Kocak O, Erkan U, Toktas A, Gao S (2024) Pso-based image encryption scheme using modular integrated logistic exponential map. Exp Syst Appl 237 Part A:121452. https://doi.org/10.1016/j.eswa.2023.121452
Li Q, Wang X, Ma B, Wang X, Wang C, Gao S, Shi Y (2022) Concealed attack for robust watermarking based on generative model and perceptual loss. IEEE Trans Circuits Syst Video Technol 32(8):5695–5706. https://doi.org/10.1109/TCSVT.2021.3138795
Li Y, Hao X, Liu G, Shang R, Jiao L (2024) Qea-qcnn: optimization of quantum convolutional neural network architecture based on quantum evolution. Memetic Comput 16(3):233–254
Li Q, Ma B, Wang X, Wang C, Gao S (2024) Image steganography in color conversion. IEEE Trans Circuits Syst II Express Briefs 71(1):106–110. https://doi.org/10.1109/TCSII.2023.3300330
Lin Y, Xie Z, Chen T, Cheng X, Wen H (2024) Image privacy protection scheme based on high-quality reconstruction dct compression and nonlinear dynamics. Expert Syst Appl 257:124891. https://doi.org/10.1016/j.eswa.2024.124891
Liu A, Pan L, Lu Y, Li J, Hu X, Zhang X, Wen L, King I, Xiong H, Yu P (2024) A survey of text watermarking in the era of large language models. ACM Comput Surv 57(2):1–36
Liu A, Pan L, Lu Y, Li J, Hu X, Zhang X, Wen L, King I, Xiong H, Yu P (2024) A survey of text watermarking in the era of large language models. ACM Comput Surv 57(2):1–36
Locher DF, Cardarelli L, Müller M (2023) Quantum error correction with quantum autoencoders. Quantum 7:942
Ma H, Huang C-J, Chen C, Dong D, Wang Y, Wu R-B, Xiang G-Y (2023) On compression rate of quantum autoencoders: control design, numerical and experimental realization. Automatica 147:110659
Ma Q-G, Hao C-L, Yang X-K, Qian L-L, Zhang H, Si N-W, Xu M-C, Qu D (2024) Continuous evolution for efficient quantum architecture search. EPJ Quant Technol 11:54
Ma Q, Hao C, Si N, Chen G, Zhang J, Qu D (2025) Quantum adversarial generation of high-resolution images. EPJ Quantum Technol 12:3
Ma F, Huang H-Y (2025) How to construct random unitaries. In: Proceedings of the 57th Annual ACM symposium on theory of computing. pp 806–809
Majeed MA, Sulaiman R, Shukur Z, Hasan MK (2021) A review on text steganography techniques. Mathematics 9(21):2829
Moulin P, O'Sullivan JA (2003) Information-theoretic analysis of information hiding. IEEE Trans Inf Theory 49(3):563–593. https://doi.org/10.1109/TIT.2002.808134
Nielson MA, Chuang IL (2010) Quantum computation and quantum information, 10th anniversary, edition. Cambridge University Press, Cambridge, U.K
Oliveira NM, Park DK, Araujo IF, Silva AJ (2024) Quantum variational distance-based centroid classifier. Neurocomputing 576:127356
Pasupuleti MK (2024) Quantum cryptography: combining quantum computing with machine learning algorithm. https://doi.org/10.62311/nesx/rb978-81-978755-9-5
Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: an imperative style, high-performance deep learning library. Curran Associates Inc., Red Hook, NY, USA
Pepper A, Tischler N, Pryde GJ (2019) Experimental realization of a quantum autoencoder: the compression of qutrits via machine learning. Phys Rev Lett 122:060501
Petitcolas FAP, Anderson RJ, Kuhn MG (1999) Information hiding-a survey. Proc IEEE 87(7):1062–1078. https://doi.org/10.1109/5.771065
Potdar VM, Han S, Chang E (2005) A survey of digital image watermarking techniques. In: INDIN '05. 2005 3rd IEEE international conference on industrial informatics, 2005. pp 709–716. https://doi.org/10.1109/INDIN.2005.1560462
Purohit K, Vyas AK (2025) Quantum key distribution through quantum machine learning: a research review. Front Quantum Sci Technol 4. https://doi.org/10.3389/frqst.2025.1575498

Romero J, Olson JP, Aspuru-Guzik A (2017) Quantum autoencoders for efficient compression of quantum data. Quant Sci Technol 2(4):045001

Shen Y, Zhang F-L, Chen Y-Z, Zhou C-C (2023) Masking quantum information in the Kitaev Abelian Anyons. XXPhys A 612:128503

Sheng ML, Ling YW (2018) Masking quantum information in multipartite scenario. Phys Rev A 98:062346

Shi F, Li M-S, Chen L, Zhang X (2021) k-uniform quantum information masking. Phys Rev A 104:032418

Shor PW (1994) Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th annual symposium on foundations of computer science. pp 124–134. https://doi.org/10.1109/SFCS.1994.365700

Sim S, Johnson PD, Aspuru-Guzik A (2019) Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Adv Quantum Technol 2:1900070

Şimşek C, Erkan U, Toktaş A, Lai Q, Gao S (2025) Hexadecimal permutation and 2d cumulative diffusion image encryption using hyperchaotic sinusoidal exponential memristive system. Nonlinear Dyn 113(13):17177–17208. https://doi.org/10.1007/s11071-025-11001-w

Subhedar MS, Mankar VH (2014) Current status and key issues in image steganography: a survey. Comput Sci Rev 13:95–113

Sun H, Qu Z, Sun L, Chen X, Xu G (2022) High-efficiency quantum image steganography protocol based on double-layer matrix coding. Quantum Inf Process 21(5):165

Sun J-Y, Wang W-T, Zhang H, Zhang J (2023) Color image quantum steganography scheme and circuit design based on dwt+ dct+ svd. Physica A Stat Mech Appl 617:128688

Wang Z, Byrnes O, Wang H, Sun R, Ma C, Chen H, Wu Q, Xue M (2023) Data hiding with deep learning: a survey unifying digital watermarking and steganography. IEEE Trans Comput Soc Syst 10(6):2985–2999. https://doi.org/10.1109/TCSS.2023.3268950

Wang H, Cheng X, Wu H, Luo X, Ma B, Zong H, Zhang J, Wang J (2025) A gan-based anti-forensics method by modifying the quantization table in jpeg header file. J Vis Commun Image Rep 110:104462. https://doi.org/10.1016/j.jvcir.2025.104462

Wang G, Warrell J, Emani PS, Gerstein M (2025) Quantum variational autoencoder utilizing regularized mixed-state latent representations. Phys Rev A 111(4):042416

Wang S, Shi J, Yu S, Wu M (2004) Linear models (in Chinese). University Mathematics Series No. 3, p. 308. Science Press, Beijing

Watrous J (2018) The theory of quantum information. Cambridge University Press, Cambridge, U.K. https://doi.org/10.1017/9781316848142

Wecker D, Hastings MB, Troyer M (2015) Progress towards practical quantum variational algorithms. Phys Rev A 92:042303. https://doi.org/10.1103/PhysRevA.92.042303

Wu J, Fu H, Zhu M, Zhang H, Xie W, Li X-Y (2024) Quantum circuit autoencoder. Phys Rev A 109(3)

Wu S, Song Y, Li R, Qin S, Wen Q, Gao F (2025) Resource-efficient adaptive variational quantum algorithm for combinatorial optimization problems. Adv Quantum Technol

Xing Z, Lam C-T, Yuan X, Im S-K, Machado P (2024) Mmqw: multi-modal quantum watermarking scheme. IEEE Trans Inf Forensics Secur 19:5181–5195. https://doi.org/10.1109/TIFS.2024.3394768

Zeng X, Feng B, Xia Z, Peng Z, Qin T, Lu W (2024) Robust image hiding network with frequency and spatial attentions. Pattern Recogn 155:110691

Zhang X-M, Kong W, Farooq MU, Yung M-H, Guo G, Wang X (2021) Generic detection-based error mitigation using quantum autoencoders. Phys Rev A 103:040403

Zhang Y, Cincio L, Negre CFA, Czarnik P, Coles PJ, Anisimov PM, Mniszewski SM, Tretiak S, Dub PA (2022) Variational quantum Eigensolver with reduced circuit complexity. NPJ Quantum Infor 8(1)

Zhu J, Kaplan R, Johnson J, Fei-Fei L (2018) Hidden: Hiding data with deep networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp 657–672