Quantum Information Processing (2023) 22:409
https://doi.org/10.1007/s11128-023-04143-6

Disturbance rejection in pattern recognition: a realization of quantum neural network

Xiaobo Hu[1,2] · Jianbo Su[2] · Jun Zhang[1]

Received: 19 February 2023 / Accepted: 26 September 2023

Abstract
In the field of artificial intelligence, pattern recognition is widely used to extract the abstract information in those high dimensional inputs of image, voice, or video. However, the interpretability of pattern recognition still remains understudied. The incomplete features extracted from system input still limit the recognition performance. To reject the disturbance of feature incompleteness, an error compensation is realized into the pattern recognition model under a quantum computation framework. The quantum-based recognition system fulfills the information transmission from input to output with the transformation of quantum states. Then, a compensation for the quantum state is used to reject those intermediate errors in the pattern recognition task. The experiment results in this paper indicate an effectiveness of the proposed method, with which the compensated Quantum Neural Network obtains a better performance. The proposed method brings a more robust recognition system under unknown disturbances.

Keywords Disturbance rejection · Error compensation · Pattern recognition · Quantum neural network

✉ Jun Zhang
zhangjun12@sjtu.edu.cn

Xiaobo Hu
xinaidd520@sjtu.edu.cn

Jianbo Su
jbsu@sjtu.edu.cn

1 Joint Institute of UM-SJTU, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, Shanghai, China

2 Research Center of Intelligent Robotics, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, Shanghai, China

Published online: 15 November 2023

1 Introduction

Pattern recognition has been widely used in human daily life. A large amount of academic and industrial researches are taken in the field [1–3]. Under the development of quantum computation, the technologies in pattern recognition, which are previously modeled in a classic computing architecture, are successfully realized toward the quantum state [4–8]. Besides, the open source platform *Tensorflow* develops the dedicated quantum packages for the coding community [9]. The major operations of a quantum-based pattern recognition model are shown in Fig. 1, including data preprocessing, quantum encoding, feature extraction, and classification. In this paper, a realization of disturbance rejection is taken toward this model.

The major disturbance rejections for pattern recognition in the current researches are taken by a more robust data preprocessing [10], a deeper feature extraction [11], etc. One of the main challenges in the current researches is the constraint of interpretability, which leads to an unavoidable modeling error [12]. Furthermore, it brings the complete feature set unavailable [13]. Under the incompleteness constraint and the unavoidable interferences in practical application, the output of open loop recognition system is always different from the desired one, resulting in a *recognition error*. Little research has been done on what kinds of error the system produces as well as how to compensate for these errors.

Some of the solutions on this challenge are taken by the large-scale datasets [14–16]. If the training dataset is large enough, then the disturbances in all application scenarios are included in the training stages, so that the recognition system is robust enough to achieve an excellent performance [11]. However, it is almost impossible to ergodic all kinds of disturbance in a limited training dataset. The sample space is usually an infinity space. Although the sample space for a recognition task should be smaller than the whole universe in a physic view, it still be a heavy computation workload for the machine intelligence. In some current works, the recognition systems need to be simplified to constraint the computation cost [17]. Besides, the classical pattern recognition needs a series of nonlinear operation to fulfill the feature extraction

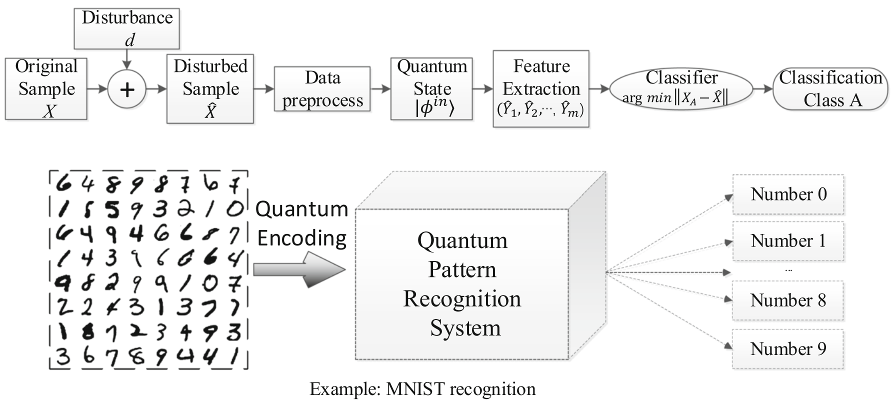

Fig. 1 A framework of the quantum-based pattern recognition model

process, which brings more difficulties on disturbance rejection. In contrast, under the quantum computation framework, these nonlinear operations are simplified into a linear unitary transformation. As a result, the forward recognition as well as the backward propagation become a more concise form.

In quantum computation, all messages are represented by the quantum state. An adequate quantum state space should completely represent the information transmission from input to output. Therefore, the disturbance rejection work in this paper is taken on the quantum state space, with which the information transmission is realized by *Feature* [6]. The robust feature is a key element in pattern recognition. To compensate for those passive and unavoidable intermediate error from an active point of view, the disturbance rejection control is applied on features in Hu et al. [18], in which a compensation method is used. The core principle of disturbance rejection control is a model-free error compensation under the internal and external disturbance, so as to stabilize a controllable system [19–21]. In this paper, the compensation principle is further applied into a feature extraction process for the quantum-based pattern recognition task.

Section 2 summarizes the related aspects about quantum-based pattern recognition. Then, Sect. 3 illustrates a disturbance rejection based on feature compensation in the quantum-based pattern recognition model. Furthermore, the error convergence is analyzed in a Markov sense by Sect. 4. The experiments are given in Sect. 5, followed by conclusions in Sect. 6.

2 Related works

2.1 Disturbance rejection in pattern recognition

The core principle for disturbance rejection in pattern recognition is a group of not only content descriptive but also semantic representative features [18, 22, 23]. Without loss of generality, a m-steps feature extraction for the *Input* X^{in} is taken by the following hierarchy process, i.e.,

$$\begin{cases} \widehat{Y}_1 = F_1(X^{in}) \\ \widehat{Y}_2 = F_2(\widehat{Y}_1) \\ \quad \cdots \\ \widehat{Y}_m = F_m(\widehat{Y}_{m-1}) \end{cases}, \tag{1}$$

in which *Operation* $F_i(\cdot)$, $i = 1 \sim m$, can be realized by a linear full connected layer, a nonlinear convolution layer, a non-trainable pooling layer, and so forth [11].

For the ideal features of an objective *Class* A, expressed as: $A^* = (Y^*_{1,A}, Y^*_{2,A}, ..., Y^*_{m,A})$, and its disturbed sample, expressed as: $\widehat{A} = (\widehat{Y}_{1,A}, \widehat{Y}_{2,A}, ..., \widehat{Y}_{m,A})$, a stability condition is given by Lemma 1.

Lemma 1 *To guarantee a correct recognition, an error convergence needs to be satisfied for distance* $\|\widehat{A} - A^*\|$, *i.e.,*

$$\|\widehat{A} - A^*\| \leq \varepsilon\ , \tag{2}$$

in which the stability margin is limited by ε.

More demonstration and application about Lemma 1 can be found in Jordan and Mitchell [24]. Furthermore, for the *Class B* (B≠A)—which is expressed as: $B^* = (Y^*_{1,B}, Y^*_{2,B}, ..., Y^*_{m,B})$, a mis-recognition occurs when the following condition is satisfied:

$$\|\widehat{A} - A^*\| \geq \|\widehat{A} - B^*\| \ . \tag{3}$$

The main disturbance rejection works are taken by: (1) The more complicated and robust recognition model, such as CNN [25], FPN [26], and Transformer [27]; (2) the more descriptive modeling of recognition method, such as Gabor [28], SIFT [29], and MFCC [30]; and (3) the larger scale datasets which try to cover every application scenario, such as MNIST [14], VOC [15], and COCO [16].

2.2 Pattern recognition with quantum

In a quantum-based pattern recognition task, the *Input* X^{in} is encoded by a quantum state $|\phi^{in}\rangle$ [8]. Then, it can be extended into a matrix form with $X^{in} = |\phi^{in}\rangle\langle\phi^{in}|$. In this paper, the quantum realization of feature extraction, i.e., $F_i(\widehat{Y}_{i-1})$ in Eq. (1), is taken by a series of unitary transformation referring to the Deep Quantum Neural Network proposed by Beer et al. [6]. Hence,

$$F_i(\widehat{Y}_{i-1}) = \mathrm{tr}_{i-1}\left(U^i \left(\widehat{Y}_{i-1} \otimes |0\cdots0\rangle_i\langle0\cdots0|\right) U^{i\,\dagger}\right), \tag{4}$$

in which i corresponds to the layer index. All these transformation are trained with the loss function based on a specific training dataset, e.g., $(|\phi^{\mathrm{in}}_n\rangle, |\phi^{\mathrm{out}}_n\rangle)$, $n = 1, 2, ..., N$. The loss function for the whole pattern recognition model is defined by *fidelity*, i.e.,

$$L = \frac{1}{N}\sum_{n=1}^{N}\langle\phi^{\mathrm{out}}_n|\widehat{Y}_{m,n}|\phi^{\mathrm{out}}_n\rangle \ . \tag{5}$$

in which the sample amount is N. Due to the constraint on computation resource, the current methods usually focus on a restricted case, e.g., $|\phi^{\mathrm{out}}\rangle = V|\phi^{\mathrm{in}}\rangle$ [31]. The unitary operation V corresponds to the linear regression in classical pattern recognition. With respect to the complex task in the current booming artificial intelligence field, many large-scale datasets have not yet been transferred into the quantum physic [8].

2.3 Disturbance rejection in qubit

The error convergence between $\widehat{Y}_A$ and Y^*_A for quantum-based recognition task is given by a fidelity limit of $\|1 - \mathcal{F}(\widehat{Y}_A, Y^*_A)\| \leq \varepsilon$, in which $\mathcal{F}(\cdot,\cdot)$ represents a fidelity calculation [4]. In this paper, a $\mathcal{N}$-dimensional quantum qubit system is used to construct the feature space for $\widehat{Y}_A$ and Y^*_A. Dimension $\mathcal{N}$ corresponds to the amount of objective classes in a recognition task. Then, the disturbance acting on a recognition process and the respective disturbance rejection are realized in a qubit sense. Some specified quantum states, just like the EPR states [32], can be designed manually to

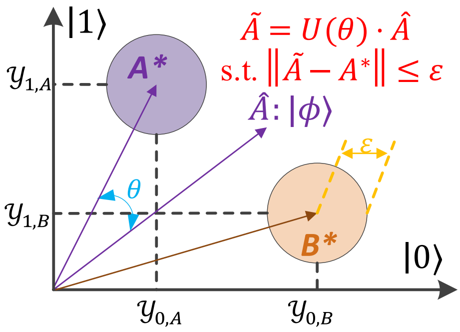

Fig. 2 Feature error in qubit

represent the ideal situation. In contrast, once an ideal qubit $|\phi\rangle = a|0\rangle + b|1\rangle$ is disturbed, it becomes $|\phi'\rangle = U^d \cdot |\phi\rangle = a'|0\rangle + b'|1\rangle$, shown in Fig. 2. Under the basic assumption for quantum mechanics and quantum information [33], the disturbed quantum state, which stores the information embedded in the system input, should be compensated to return back to its correct location in the feature space. As a result, the correct quantum state can be used to fulfill the recognition task successfully [5, 7, 8]. This disturbance rejection principle is realized in details by the following Sect. 3

3 Disturbance rejection in quantum-based recognition system

When the quantum-based recognition system is disturbed, the qubit, which stores the intermediate feature, will deviate from its original error-free state. Therefore, the disturbance rejection is taken by a compensation on the quantum state in this paper. The compensation principle is shown by Fig. 3.

3.1 Quantum representation of a recognition task

Firstly, this paper modifies a quantum-based recognition task based on the classical one-hot label [24]. In a classification toward $\mathcal{N}$ kinds of objective classes, the ideal

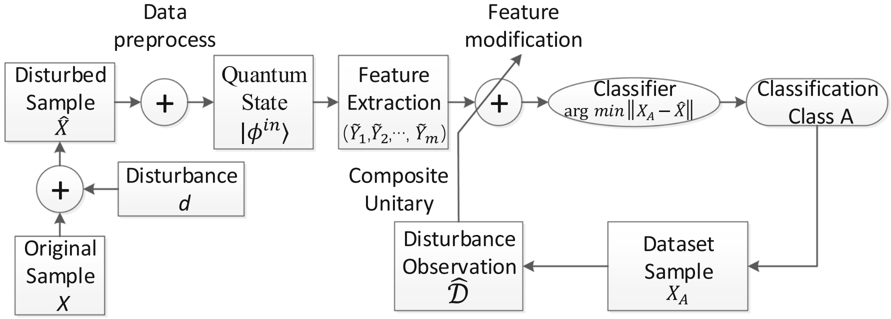

Fig. 3 A quantum-based pattern recognition model with compensation

Bayes probability estimation for the *Class A*, $A = 1, \cdots, \mathcal{N}$, should be a $P_A = 0$ or $P_A = 1$. Therefore, the ideal label in this paper is defined by $|\phi^{out}_{A,*}\rangle = 0 \cdot |0\rangle + 1 \cdot |1\rangle$, i.e., $a^2 = 0,\ b^2 = 1$ for a single qubit $a|0\rangle + b|1\rangle$. Otherwise, the negative label is $|\overline{\phi}^{out}_{A,*}\rangle = 1 \cdot |0\rangle + 0 \cdot |1\rangle$. A constraint needs to be satisfied with $a^2 \geq 0,\ b^2 \geq 0$, so that the classification task is meaningful in probability sense. Then, the A^{th} dimension of any *Sample* X^{in} is given by $|\phi_A\rangle = a|0\rangle + b|1\rangle$. For example, in the MNIST recognition task, $\mathcal{N} = 10$, so that the quantum states for an input is prepared with $X^{in} = |\phi_1\ \phi_2\ ...\ \phi_{10}\rangle$. Hence, a transformation between the classical one-hot and the quantum one-hot is realized as follows:

$$(\overbrace{0, ..., 0}^{\mathcal{N}-1}, 1) \Leftrightarrow \overbrace{|0\rangle, ..., |0\rangle}^{\mathcal{N}-1}, |1\rangle\ . \tag{6}$$

Besides, the solid label can be replaced by a soft one, e.g., $a^2 = 1, b^2 = 0 \Rightarrow a^2 = 0.99, b^2 = 0.01$. The simple deduction is omitted. The last but not least, the final recognition result in Eq. (1) is given by the maximum likelihood based on a Bayes probability estimation of $P_A = \langle\phi^*_A|\widehat{Y}_m|\phi^*_A\rangle$. With the quantum one-hot label, a Quantum Neural Network can be used to fulfill an objective recognition task, which will be tested in Sect. 5.2.

3.2 Modeling of error system

Disturbance always exists in a working system, so comes the error [34]. Once different kinds of internal and external disturbances act on the recognition system, errors will be generated in all intermediate operations in Eq. (1) [18]. Therefore, the error system is built by a hierarchy fidelity computation as follows:

$$\begin{cases} e_1 = \mathcal{F}^e_1(X^{in}, |\phi^{in*}\rangle) \\ e_2 = \mathcal{F}^e_2(\widehat{Y}_1, |\phi^*_1\rangle) \\ \quad\cdots \\ e_m = \mathcal{F}^e_m(\widehat{Y}_m, |\phi^*_m\rangle) \end{cases}. \tag{7}$$

The intermediate errors are obtained with a fidelity calculation between the features extracted by a specific input and the corresponding *Cluster Centers* $|\phi^*_{i,A}\rangle$, $(A = 1, \cdots, \mathcal{N})$, i.e., $\mathcal{F}^e_i = \langle\phi^*_{i,A}|\widehat{Y}_i|\phi^*_{i,A}\rangle$. The Cluster Center can be designed manually, or with a Center Loss optimization [35].

3.3 Disturbance compensation on feature

The extracted feature $\widehat{Y}_i$, whose message is stored in a qubit $|\phi_i\rangle$, is disturbed from its error-free state, i.e., a bias from $|\phi^*_i\rangle$. Therefore, under the Postulate 2 in quantum mechanics, it always can be compensated from $|\phi^*_i\rangle = U_e|\phi_i\rangle$ [33].

Remark 1 The closeness of the quantum-based recognition system is left for future study. Whether it is close or not does not affect the effectiveness of the unitary compensation. The Postulate 1 in quantum mechanics ensures the existence of the ideal quantum state in the feature space. What an effective compensation needs to do is the stabilization of the whole system under unknown disturbances [36].

Therefore, the compensation on feature, proposed in [18] is taken toward the intermediate errors in Eq.(7). Then, the compensated feature is used to fulfill the objective recognition task. In this paper, the compensated feature extraction is given as follows:

$$\begin{cases} \widetilde{Y}_1 = \Delta F_1(F_1(X^{in})) \\ \widetilde{Y}_2 = \Delta F_2(F_2(\widetilde{Y}_1)) \\ \quad \dots \\ \widetilde{Y}_m = \Delta F_m(F_m(\widetilde{Y}_{m-1})) \end{cases}, \tag{8}$$

in which the compensation operator $\Delta F_i(\cdot)$ should realize an error convergence between $\widetilde{Y}_m$ and Y_m^* with Lemma 1. For the ideal qubit $|\phi_i^*\rangle$, who stores the information of ideal feature Y_m^*, it is estimated by a compensation unitary U_e, i.e., $|\phi_i^*\rangle \approx U_e|\phi_i\rangle$. Therefore, the intermediate feature, which is expanded by $\widehat{Y}_i = |\phi_i\rangle\langle\phi_i|$, is compensated by:

$$\widetilde{Y}_i = U_e|\phi_i\rangle\langle\phi_i|U_e^\dagger = U_e\widehat{Y}_i U_e^\dagger \quad . \tag{9}$$

Ideally, the compensation unitary can be easily derived into $U_e^* = |\phi_i^*\rangle\langle\phi_i|$. However, the selection of $|\phi_i^*\rangle$ is taken by the open loop recognition result. It may be mis-recognized as another class. Therefore, the approximation of U_e needs a more systematic analysis.

4 Error convergence in quantum-based pattern recognition

4.1 Markov model of pattern recognition

In this paper, the Markov characteristic of a pattern recognition process is defined by the tuple of $< \Omega, E, Pr\{\cdot\} >$. The *Feature Space* is marked with Ω. The *Class Set* E includes *Class* A, *Class* B, etc. The definition of Ω and E correspond to the Sample Space and Event Space in a general Markov Chain [37]. Besides, the probability map is marked with $Pr\{\cdot\} \in [0, 1]$. The conditional probability $Pr\{Y_m = A|Y_{m-1} = A, Y_{m-2} = B, ..., Y_1 = C, X^{in} = A\} = Pr\{Y_m = A|Y_{m-1} = A\}$ introduces a Markov Chain between all objective state in a pattern recognition task. The symbol $=$ means *belong to* in these formulas. Besides, the Markov characteristic has nothing to do with whether $B = A$ and $C = A$ or not. In a practical scenario, if B or C is equal to A, the following step of feature extraction will become redundant. The depth of m just need to guarantee that all potential input can be separated, so that $Y_m = A$ [11]. A smaller m can lighten the whole recognition system [38]. The Assumption 1 models the recognition system in Eq.(1) into a Countable-state Markov Chain, i.e.,

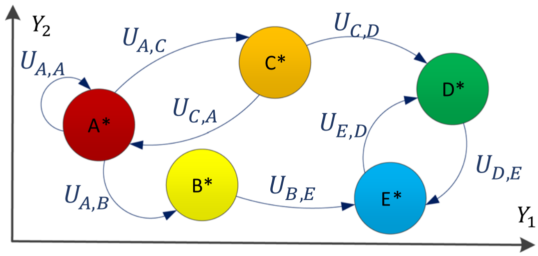

Fig. 4 The separation of different states

Assumption 1 The feature space in Eq. (1), i,e, Y_i $(i = 1, \cdots, m)$, is griddled by $\mathcal{N}$ kinds of objective classes. Besides, the extracted features for Input X^{in} in Y_i and Y_{i+1}, i.e., $\widehat{Y}_i$ and $\widehat{Y}_{i+1}$, satisfy: (1) $\widehat{Y}_i$ and $\widehat{Y}_{i+1}$ belong to the corresponding regions with respect to the Class set E, e.g., $\widehat{Y}_i \in$ Class A and $\widehat{Y}_{i+1} \in$ Class B; and (2) the belonging for $\widehat{Y}_i$ and $\widehat{Y}_{i+1}$ are independent with each other, e.g., $X^{in} \in$ Class C, $A \neq B \neq C$.

The Markov Chain is originally used to model a set of separated states [37]. Then, it is developed into the hierarchy feature extraction process in this paper, no matter for the quantum-based pattern recognition or the classical one. Then, the definitions of *Recurrent*, *Transient*, *Period*, and *Aperiodic* for a Markov Chain are transferred toward the pattern recognition task.

Definition 1 **Recurrent**: A State A is Recurrent if for each State B, shown in Fig. 4, there exist $m < \infty$ such that $Pr\{Y_m = A|X^{in} = A > 0\}$. **Transient**: If it is not Recurrent. **Period**: The greatest common divisor of m such that $P^m_{AA} = Pr\{Y_m = A|X^{in} = A\} > 0$. **Aperiodic**: If the Period is 1. **Ergodic**: In a finite-state Markov Chain, if a state is both Recurrent and Aperiodic, it is defined as Ergodic.

Lemma 2 *Either all states in a class are Transient, or all are Recurrent for Countable-state Markov Chain.*

More demonstration and application about Definition 1 and Lemma 2 can be found in Jordan and Mitchell [39]. When the error in a recognition process is convergent, Theorem 1 will be satisfied based on Lemma 2, e.g., with $X^{in} \in$ Class A, then $\widehat{Y}_m \in$ Class A, a correct recognition result is obtained.

Theorem 1 *All of the states in the Markov Chain of a stable recognition are Recurrent.*

Theorem 1 is proposed based on a consistent of the input sample and its recognition result. The system is stable iif Recurrent for any state in the objective recognition task. Otherwise, some states would not be returned when disturbance exists. Furthermore, two adjacent steps of feature extraction may obtain the same belonging for $\widehat{Y}_i$ and $\widehat{Y}_{i+1}$. So, Theorem 2 can be easily obtained in a common sense.

Theorem 2 *All of the states in the Markov Chain of a recognition task, no matter it is stable or not, are Aperiodic.*

With Theorems 1 and 2, the characteristic of Ergodicity in Definition 1 will guarantee a stability of the recognition system. To further analyze the convergence of a pattern recognition Markov Chain, the Drift Operator in a quantum sense is defined in this paper.

Definition 2 Consider a Countable-state Chain with Feature Space Ω and Transition Unitary $U_{\mathrm{A,B}}$, the Quantum Drift Operator Δ is defined for any nonnegative function $\mathcal{V}$: $\Omega \rightarrow \infty$, i.e., $\forall |\phi_{\mathrm{A}}\rangle \in \Omega$:

$$\Delta \mathcal{V}(|\phi_{\mathrm{A}}\rangle) = \sum_{\mathrm{B}} \mathcal{V}(U_{\mathrm{A,B}}|\phi_{\mathrm{A}}\rangle) - \mathcal{V}(|\phi_{\mathrm{A}}\rangle) \ , \tag{10}$$

in which A and B are donated for the class belonging in the Feature Space.

4.2 Stability condition

For any states in Ω, the Euclidean distance always satisfies $\| |\phi_i\rangle - |\phi_i^*\rangle \| \geq 0$, i.e., $\|\Delta|\phi'\rangle\| \geq 0$. An easy extension for the fidelity in Eq. (5) can also be derived. Under the Lyapunov criterion, the recognition Markov Chain is always Transient, at least in a small set, corresponding to a *Cluster* [11]. Then, the Ergodic Theorem for Countable-state Chain is used to prove a recognition system stability [39].

Lemma 3 *Ergodic Theorem: A Countable-state Chain with State Space Ω is Ergodic if there exists a Lyapunov function $\mathcal{V} : \Omega \rightarrow \mathbb{R}_{\geq 0}$ and a set $S = \zeta \in \Omega : \mathcal{V}(\zeta) \leq r$, in which $r < \infty$ such that:*

1. *(Drift Condition) $\Delta \mathcal{V}(\zeta) \leq -1 + b\mathbb{I}_C(\zeta)$ for all $\zeta \in \Omega$.*
2. *(Small-set Condition) There exists $m < \infty$ such that for all $\zeta \in S$ and all $\zeta' \in S$, $P^m_{\zeta\zeta'} > 0$.*

The probability $P^m_{\zeta\zeta'}$ means that the initial *State* ζ ends up with another *State* ζ' after m steps. A simple Lyapunov function can be defined as $\mathcal{V}(|\phi\rangle) = \| |\phi\rangle - |\phi^*_{\mathrm{A}}\rangle \|_p, \forall |\phi\rangle \in \Omega_A$. The neighbor set Ω_A is equivalent to a Cluster of $|\phi^*_{\mathrm{A}}\rangle$. The Drift Condition guarantees a positive recurrence. The stability condition points to a bounded drift operator, so that $b < \infty$. Therefore, the qubit who stores the intermediate feature should be close enough to the ideal one. With respect to the disturbance compensation in Eq. (8), the compensated quantum state, i.e., $|\widetilde{\phi_i}\rangle = U_e|\widehat{\phi_i}\rangle$, should guarantee the following convergence condition:

$$\max_{|\widehat{\phi_i}\rangle} \|(U_e - |\phi^*\rangle\langle\widehat{\phi_i}|)|\widehat{\phi_i}\rangle\| \leq \varepsilon \ , |\widehat{\phi_i}\rangle \in \Omega_A . \tag{11}$$

The approximated U_e should be close enough to the ideal unitary of $|\phi^*\rangle\langle\widehat{\phi_i}|$ under a limit ε. All normalized quantum states $|\widehat{\phi_i}\rangle \in \Omega_A$ should satisfy this condition.

4.3 Training strategy

A loss optimization is used in this paper to approach the approximation of U_e. Firstly, a decomposed 2×2 unitary matrix is constructed, i.e.,

$$u_e = e^{i\kappa} \begin{bmatrix} e^{-i\frac{\beta}{2}} & 0 \\ 0 & e^{i\frac{\beta}{2}} \end{bmatrix} \begin{bmatrix} \cos\frac{\gamma}{2} & -\sin\frac{\gamma}{2} \\ \sin\frac{\gamma}{2} & \cos\frac{\gamma}{2} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\delta}{2}} & 0 \\ 0 & e^{i\frac{\delta}{2}} \end{bmatrix} , \tag{12}$$

Table 1 Quantum representation of pattern recognition (Examples)

Sample	Input $\|\phi^{in}\rangle$			Label $\|\phi^{out}\rangle$
Qubits	$\|\phi_3\rangle$	$\|\phi_6\rangle$	$\|\phi_9\rangle$	$\|\phi_3\phi_6\phi_9\rangle$
3	$0.447\|0\rangle + 0.894\|1\rangle$	$0.949\|0\rangle + 0.316\|1\rangle$	$0.949\|0\rangle + 0.316\|1\rangle$	$\|100\rangle$
6	$0.995\|0\rangle + 0.100\|1\rangle$	$0.316\|0\rangle + 0.948\|1\rangle$	$0.954\|0\rangle + 0.300\|1\rangle$	$\|010\rangle$
9	$0.949\|0\rangle + 0.316\|1\rangle$	$0.775\|0\rangle + 0.632\|1\rangle$	$0.707\|0\rangle + 0.707\|1\rangle$	$\|001\rangle$

in which κ, β, γ, and δ are real-valued. Then, the *Feature Error* e_i in Eq. (7) is substituted into κ, β, γ, and δ, e.g., $\kappa \leftarrow \kappa \cdot e_i$. Then, a diagonal expansion, $U_e = diag(u_e, \mathbb{I})$, is used to match the intermediate dimension. The parameters κ, β, γ, and δ can be trainable or not. In the former case, they can be optimized with the training process of loss function in Eq. (5) referring to [6]. The trainable parameters are updated as follows: Firstly, the open loop is trained independently, then the close one.

The training tuples are $(|\phi_n^{in}\rangle, |\phi_n^{out}\rangle)$, $n = 1, 2, \cdots, N$. With the feature compensated, i.e., $\widetilde{Y}^{out} = U_e \widehat{Y}^{out} U_e^\dagger$, the gradient of loss function, i.e., $L = 1/N \sum_{n=1}^{N} \langle \phi_n^{out} | \widetilde{Y}_n^{out} | \phi_n^{out} \rangle$, is derived as follows:

$$
\begin{aligned}
\frac{\partial L}{\partial e} =& \frac{1}{N} \sum_{n=1}^{N} \langle \phi_n^{out} | \frac{\partial \widetilde{Y}_n^{out}}{\partial e} | \phi_n^{out} \rangle \\
=& \frac{1}{N} \sum_{n=1}^{N} \left\langle \phi_n^{out} \left| \left(\frac{d\, U_e}{de} \widehat{Y}_x^{out} U_e^\dagger + U_e \widehat{Y}_x^{out} \frac{d\, U_e^\dagger}{de} \right) \right| \phi_n^{out} \right\rangle
\end{aligned}
. \tag{13}
$$

The gradients for the parameters κ, β, γ, and δ all follow this form.

5 Experiments and analysis

5.1 Datasets and evaluation metrics

The experimental devices for quantum computation are still far from being applied into a large-scale dataset. Therefore, it is necessary to make a size scaling to test the system performance. The widely used validation datasets, MNIST and Cifar-10, are used in this paper. If all those objective $\mathcal{N} = 10$ classes are used, the unitary matrix in the intermediate feature extraction needs to be designed with a 2^{10} size. Then, the whole model needs an incredible RAM. As a result, a simplified test is taken. In Table 1, an input X_3^{in} belongs to *Class 3* with $P_3 = 0.8,\ P_6 = 0.1,\ P_9 = 0.1$. In contrast, another X_6^{in} belongs to *Class 6* with $P_6 = 0.9,\ P_3 = 0.01,\ P_9 = 0.09$. The left one is X_9^{in}, which belongs to *Class 9* with $P_9 = 0.5,\ P_3 = 0.1,\ P_6 = 0.4$. Bayes probability estimation P_A is prepared with a classical neural network for an input sample X^{in} belongs to *Class A*. The three qubits are obtained by $|\phi_i\rangle = \sqrt{(1 - P_A)}|0\rangle + \sqrt{P_A}|1\rangle$. Then, the quantum label $|\phi_3\phi_6\phi_9\rangle$ are obtained.

Table 2 Baseline performances on MNIST

Model information	Performance	
Selected classes	Average fidelity	Success rate (%)
6 & **9**	0.75	98.88
1 & **7**	0.62	97.97
2 & **5**	0.62	99.12
3 & **6** & **9**	0.52	97.72
0 & **1** & **7**	0.47	98.70
2 & **4**& **8**	0.41	95.72

The Bayes probability is estimated by a Multi-Layer Perceptron(MLP) trained by the vanilla MNIST [14], or a Vision-Transformer(VIT) trained by the vanilla Cifar-10 [27]. The disturbed MNIST in Sect. 5.4 is obtained by a Random Affine with the degree range of $(5^\circ, 45^\circ)$. Besides, the disturbed Cifar-10 is obtained by a compound transformation of Random Rotation in $(0^\circ, 30^\circ)$ and Random Perspective in distortion scale of 0.6. The disturbed MNIST brings the preprocessed MLP obtain a success rate of 70.33%, while the vanilla MNIST of 96.19%. Besides, a disturbed 76.52% and a vanilla 98.98% for VIT on Cifar-10. Except the fidelity loss in Eq. (5), a classical success rate can also be used as evaluation metrics for the recognition system.

5.2 Validation on quantum representation

Several simple tests on MNIST are taken to verify the quantum representation. The baseline is designed referring to [6]. The depth of pattern recognition model is set with the amount of objective classes, e.g., $m = 2$ in Eq. (1) for a recognition task for **6** and **9**. The system performance is shown in Table 2. The success rate of recognition system is basically equivalent to that of the classical one. But the average fidelity still needs to be improved to approach an ideal **1**, which is left as a future study. Then, these several baselines are used in the following ablation and overall tests.

5.3 Ablation studies on soft label

The hard label (marked with H) and the soft label (marked with S) are tested in the same baseline, respectively. The soft label is realized with the same quantum encoding for X^{in}, but a $P = 0.99$ for the positive $|1\rangle$, and $\overline{P} = 0.01$ for the negative. The soft label brings a better performance in Table 3, especially for the average fidelity, so that it is used as default.

5.4 Validation on feature compensation in quantum neural network

The system performances are compared between the baseline as well as the compensated model, shown in Table 4. The vanilla datasets are marked with M for MNIST and C for Cifar-10. Then, to validate the performance under disturbance, the disturbed datasets, marked with M-d for MNIST and C-d for Cifar-10, are used to test again

Table 3 Comparison between hard and soft label

Model information		Performance	
Selected classes	Label style	Average fidelity	Success rate (%)
6 & **9**	*S*	0.75	98.88
6 & **9**	*H*	0.63	98.93
1 & **7**	*S*	0.62	97.97
1 & **7**	*H*	0.62	97.87
2 & **5**	*S*	0.62	99.12
2 & **5**	*H*	0.62	97.35
3 & **6** & **9**	*S*	0.52	97.72
3 & **6** & **9**	*H*	0.50	95.77
1 & **4** & **7**	*S*	0.54	98.22
1 & **4** & **7**	*H*	0.43	65.50

for the open loop and close loop recognition systems. The preprocessing mentioned in Sect. 5.1 is used to obtain the datasets of *M-d* and *C-d*. The model marked with **O** is used in the open loop system, and **C** for close.

Several representative improvements are marked in bold. The quantum-based recognition system almost obtains an equivalent performance on the success rate compared with the classical system. The success rates of the compensated system are obviously boosted. Besides, most of the fidelity performances of the compensated system are also improved, although they can not chase the excellent **1** in the linear datasets [6]. A future study should be taken for the nonlinear operations in quantum computation, so that the recognition system can fulfill some more complicated recognition tasks. In spite of this imperfect performance, these improvements validate the effectiveness of the compensation on features.

5.5 Discussion on quantum resource cost

The previous validation tests show that those simple operations and a limited number of additional qubits can bring an effective compensation on features. On one hand, the resource cost to calculate those intermediate feature error in Eq. (7) is $\mathcal{O}(m)$ times the one for a fidelity loss corresponding to the baseline, e.g., the vanilla Quantum Neural Network in Beer et al. [6]. The linear complexity $\mathcal{O}(m)$ is related to the depth of feature extraction as well as the accuracy requirement for a quantum-based computation. On the other hand, the compensated recognition system reuses the open loop features as well as the information extracted from the training dataset, e.g., the Cluster Centers. The additional qubits are only used to calculate and store the compensated features. As a result, the necessary amount of qubits in the compensation loop is $\mathcal{O}(\mathcal{N}m)$, in which $\mathcal{N}$ is the amount of objective classes in a recognition task of Eq. (6). Besides, the compensation unitary U_e in Sect. 4.3 are constructed with a series of simple operations, e.g., the compound rotation in Eq. (12) and a diagonal expansion. With all these additional quantum operations, the open loop recognition system is compensated.

Table 4 Comparison on open and close loop system

Model information			Performance comparison	
Dataset	Selected classes	Model	Average fidelity	Success rate (%)
M	**1** & **7**	O	0.62	97.97
M	**1** & **7**	C	0.66 (**+0.04**)	98.61
M-d	**1** & **7**	O	0.55	83.77
M-d	**1** & **7**	C	0.58 (**+0.03**)	85.99 (**+2.22**)
M	**2** & **5**	O	0.62	99.12
M	**2** & **5**	C	0.62	99.01
M-d	**2** & **5**	O	0.55	85.19
M-d	**2** & **5**	C	0.56	85.14
M	**3** & **6** & **9**	O	0.52	97.72
M	**3** & **6** & **9**	C	0.52	97.95
M-d	**3** & **6** & **9**	O	0.45	79.95
M-d	**3** & **6** & **9**	C	0.45	80.79
M	**0** & **1** & **7**	O	0.47	98.70
M	**0** & **1** & **7**	C	0.42	98.89
M-d	**0** & **1** & **7**	O	0.42	87.69
M-d	**0** & **1** & **7**	C	0.39	90.84 (**+3.15**)
M	**1** & **4** & **7**	O	0.54	98.22
M	**1** & **4** & **7**	C	0.49	98.98
M-d	**1** & **4** & **7**	O	0.45	81.56
M-d	**1** & **4** & **7**	C	0.42	86.20 (**+4.64**)
C	*Automobile1* & *Horse7*	O	0.65	99.80
C	*Automobile1* & *Horse7*	C	0.68 (**+0.03**)	99.85
C-d	*Automobile1* & *Horse7*	O	0.60	83.25
C-d	*Automobile1* & *Horse7*	C	0.62 (**+0.02**)	84.20
C	*Cat3* & *Dog5*	O	0.68	97.95
C	*Cat3* & *Dog5*	C	0.68	97.90
C-d	*Cat3* & *Dog5*	O	0.58	77.30
C-d	*Cat3* & *Dog5*	C	0.59	77.45
C	*Deer4* & *Horse7*	O	0.64	99.80
C	*Deer4* & *Horse7*	C	0.66 (**+0.02**)	99.80
C-d	*Deer4* & *Horse7*	O	0.56	81.90
C-d	*Deer4* & *Horse7*	C	0.58 (**+0.02**)	82.70
C	*Deer4* & *Frog6* & *Horse7*	O	0.44	99.67
C	*Deer4* & *Frog6* & *Horse7*	C	0.45	99.77
C-d	*Deer4* & *Frog6* & *Horse7*	O	0.39	80.20
C-d	*Deer4* & *Frog6* & *Horse7*	C	0.39	80.80
C	*Airplane0* &*Bird2* & *Dog5*	O	0.51	66.63

Table 4 continued

Model information			Performance comparison	
Dataset	Selected classes	Model	Average fidelity	Success rate (%)
C	*Airplane0 &Bird2 & Dog5*	C	0.50	99.10 (**+32.47**)
C-d	*Airplane0 &Bird2 & Dog5*	O	0.46	62.13
C-d	*Airplane0 &Bird2 & Dog5*	C	0.43	80.23 (**+18.10**)

In this paper, the preparation of quantum state as well as unitary is taken in MATLAB, i.e., a classical simulation. Moreover, the training dataset $(|\phi^{\text{in}}\rangle, |\phi^{\text{out}}\rangle)$ is naturally assumed to be prepared with sufficient samples. The precise preparation of a quantum state, especially for multi-qubits, as well as the objective compensation unitary, is still an interesting topic in the quantum field [40]. However, no matter the preparation is precise or not, a probability bias always exists in pattern recognition field anyway, e.g., the noise effect in an image or a segment of voice [11, 18, 24]. These internal and external disturbances bring a quantum state, which store the useful information, far from an ideal state, resulting at the disturbance rejection problem to be solved in this paper. With the validation tests in the previous subsections, the proposed compensation on features successfully solve the disturbance rejection problem in a quantum-based pattern recognition system.

6 Conclusion

In this paper, the disturbance rejection principle on feature extraction is applied into a quantum-based pattern recognition model. The unitary transformation in quantum computation simplifies the operations in a pattern recognition task, so that the compensation on features can be modeled in an interpretable form. The hierarchy feature extraction process is modeled into a Markov Chain. Then, the system stability is proved. The modeling of stability proof and the experiment results validate the effectiveness of feature compensation in a quantum-based recognition system theoretically and practically, although the system performances still need to be improved to chase the excellent ones in classical artificial intelligence field.

Acknowledgements This work was partially financially supported by the Shanghai Cross-disciplinary Research Fund under Grants JYJC202214, and the National Natural Science Foundation of China under Grants 61533012, 91748120, and 52041502.

Data Availability The raw datasets used and analyzed during the current study are available in the open source MNIST repository http://yann.lecun.com/exdb/mnist/ and Cifar repository http://www.cs.toronto.edu/kriz/cifar.html.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Fan, P., Zhou, R.-G., Hu, W., Jing, N.: Quantum image edge extraction based on classical Sobel operator for NEQR. Quantum Inf. Process. (2019). https://doi.org/10.1007/s11128-018-2131-3
2. Arthur, D., Date, P.: Balanced K-means clustering on an adiabatic quantum computer. Quantum Inf. Process. (2021). https://doi.org/10.1007/s11128-021-03240-8
3. Li, J., Lin, S., Yu, K., Guo, G.: Quantum K-nearest neighbor classification algorithm based on hamming distance. Quantum Inf. Process. (2022). https://doi.org/10.1007/s11128-021-03361-0
4. Schuld, M., Sinayskiy, I., Petruccione, F.: An introduction to quantum machine learning. Contemp. Phys. **56**(2), 172–185 (2015). https://doi.org/10.1080/00107514.2014.964942
5. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**, 195–202 (2017). https://doi.org/10.1038/nature23474
6. Beer, K., Bondarenko, D., Farrelly, T., Osborne, T.J., Wolf, R.: Training deep quantum neural networks. Nat. Commun. **11**(1), 808 (2020). https://doi.org/10.1038/s41467-020-14454-2
7. Pesah, A., Cerezo, M., Wang, S., Volkoff, T., Sornborger, A.T., Coles, P.J.: Absence of barren plateaus in quantum convolutional neural networks. Phys. Rev. X **11**, 041011 (2021). https://doi.org/10.1103/PhysRevX.11.041011
8. Alchieri, L., Badalotti, D., Bonardi, P., Bianco, S.: An introduction to quantum machine learning: from quantum logic to quantum deep learning. Quantum Mach. Intell. (2021). https://doi.org/10.1007/s42484-021-00056-8
9. Broughton, M., Verdon, G., Mccourt, T., Martinez, A.J., Mohseni, M.: TensorFlow quantum: a software framework for quantum machine learning. Arxiv (2020). https://doi.org/10.48550/arXiv.2003.02989
10. Alexandre, G.R., Soares, J.M., Pereira The, G.A.: Systematic review of 3D facial expression recognition methods. Pattern Recogn. **100**, 107108 (2020). https://doi.org/10.1016/j.patcog.2019.107108
11. Yann, L., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
12. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018). https://doi.org/10.1016/j.patcog.2017.10.013
13. Su, J.: Evaluation and transformation of unrealizable tasks for robot systems in representation space. IEEE Access **7**, 81532–81541 (2019). https://doi.org/10.1109/ACCESS.2019.2923817
14. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
15. Everingham, M., Winn, J.: The Pascal visual object classes challenge 2012 (VOC2012) development kit. Pattern analysis, statistical modelling and computational learning, Tech. Rep **8** (2011)
16. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context. In: European Conference on Computer Vision, pp. 740–755 (2014). https://doi.org/10.1007/978-3-319-10602-1_48
17. Dong, Y.-Z., Guo, S.-M., Yang, K.: A face recognition algorithm based on compressive sensing and wavelets transform. In: International Computer Conference on Wavelet Active Media Technology and Information Processing, pp. 58–61 (2013). https://doi.org/10.1109/ICCWAMTIP.2013.6716600
18. Hu, X., Xu, W., Gan, Y., Su, J., Zhang, J.: Towards disturbance rejection in feature pyramid network. IEEE Trans. Artif. Intell. (2022). https://doi.org/10.1109/TAI.2022.3178062
19. Guo, B.-Z., Wu, Z.-H., Zhou, H.-C.: Active disturbance rejection control approach to output-feedback stabilization of A class of uncertain nonlinear systems subject to stochastic disturbance. IEEE Trans. Autom. Control **61**(6), 1613–1618 (2016). https://doi.org/10.1109/TAC.2015.247181
20. Zhao, Z.-L., Guo, B.-Z.: A novel extended state observer for output tracking of MIMO systems with mismatched uncertainty. IEEE Trans. Autom. Control **63**(1), 211–218 (2018). https://doi.org/10.1109/TAC.2017.272041
21. Zhao, Zhi-Liang., Guo, Bao-Zhu.: A nonlinear extended state observer based on fractional power functions. Automatica **81**, 286–296 (2017). https://doi.org/10.1016/j.automatica.2017.03.00
22. Xu, W., Gan, Y., Su, J.: Bidirectional matrix feature pyramid network for object detection. In: International conference on pattern recognition (2020). https://doi.org/10.1109/ICPR48806.2021.9412229
23. Gan, Y., Xu, W., Su, J.: SFPN: semantic feature pyramid network for object detection. In: International conference on pattern recognition (2020). https://doi.org/10.1109/ICPR48806.2021.9412639

24. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245SI), 255–260 (2015). https://doi.org/10.1126/science.aaa8415
25. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2016). https://doi.org/10.1109/TPAMI.2016.2577031
26. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: IEEE conference on computer vision and pattern recognition, pp. 2117–2125 (2017). https://doi.org/10.1109/CVPR.2017.106
27. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: International Conference on Neural Information Processing Systems, pp. 6000–6010 (2017). https://doi.org/10.5555/3295222.3295349
28. Liu, C.J., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Process. **11**, 467–476 (2002). https://doi.org/10.1109/TIP.2002.999679
29. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
30. Kamil, A., Al-Ali, H., Dean, D., Senadji, B., Naik, G.R.: Enhanced forensic speaker verification using a combination of DWT and MFCC feature warping in the presence of noise and reverberation conditions. IEEE Access **5**(99), 15400–15413 (2017). https://doi.org/10.1109/ACCESS.2017.2728801
31. Li, Z., Liu, X., Xu, N., Du, J.: Experimental realization of a quantum support vector machine. Phys. Rev. Lett. **114**, 140504 (2015). https://doi.org/10.1103/PhysRevLett.114.140504
32. Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777–780 (1935). https://doi.org/10.1103/PhysRev.47.777
33. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, New York (2010)
34. Tsien, H.S.: Engineering Cybernetics. McGraw-Hill, New York (1954)
35. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision (2016). https://doi.org/10.1007/978-3-319-46478-7_31
36. Han, J.: From PID to active disturbance rejection control. IEEE Trans. Industr. Electron. **56**(3), 900–906 (2009). https://doi.org/10.1109/TIE.2008.2011621
37. Ephraim, Y., Merhav, N.: Hidden Markov processes. IEEE Trans. Inf. Theory **48**(6), 1518–1569 (2002). https://doi.org/10.1109/TIT.2002.1003838
38. Chen, X., Su, J., Zhang, J.: A two-teacher framework for knowledge distillation. In: International Symposium on Neural Networks, pp. 58–66 (2019). https://doi.org/10.1007/978-3-030-22796-8_7
39. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer, London (1993)
40. Evans, E.N., Wang, Z., Frim, A.G., DeWeese, M.R., Theodorou, E.A.: Learning quantum-state feedback control with backpropagation-free stochastic optimization. Phys. Rev. A **106**, 052405 (2022). https://doi.org/10.1103/PhysRevA.106.052405