Adaptive Quantum Autoencoders for Violence Detection in Surveillance Videos

Aden Mobusher, Muhammad Hamza, Abrar Ul Haque, and Hyundong Shin
Department of Electronics and Information Convergence Engineering, Kyung Hee University, Korea
Email: hshin@khu.ac.kr

***Abstract*—This work presents a hybrid quantum-classical machine learning approach for automated violence detection in surveillance videos. The proposed pipeline integrates a deep neural network for visual feature extraction, a classical preprocessing module to prepare data for quantum input, a quantum autoencoder for feature compression and transformation, and a final classical classifier for decision-making. Experimental evaluations demonstrate that the hybrid approach outperforms conventional PCA-based dimensionality reduction, achieving superior accuracy, efficient compression, and enhanced robustness. These results highlight the framework's potential for scalable, real-time surveillance systems aimed at improving public safety.**

I. Introduction

Violence detection is an important problem in applications such as intelligent surveillance, public safety monitoring, and automated video analysis. Conventional methods often rely on handcrafted motion features or purely deep learning (DL). However, these methods tend to generate high-dimensional feature spaces that demand substantial memory and computational resources, making real-time deployment challenging. In addition, handcrafted approaches often fail to capture complex spatio-temporal dependencies, while traditional DL architectures can struggle with scalability and robustness when exposed to diverse environments or low-quality video streams. Quantum computing, particularly quantum machine learning (QML), offers a compelling path forward for efficient processing of such data by exploiting quantum parallelism and entanglement to reduce computational overheads and enhance representational power [1], [2]. Recent studies have also shown the potential of quantum-inspired feature processing and clustering for efficient pattern recognition [3], [4]. This emerging paradigm has the potential to bridge the gap between accuracy and efficiency, paving the way for next-generation intelligent surveillance systems. Furthermore, quantum autoencoder (QAE) provide an effective mechanism for compressing redundant features while retaining critical information, thereby improving both storage efficiency and downstream classification performance [5]. The synergy between DL feature extractors and quantum-enhanced representations also opens possibilities for improved generalization, noise resilience, and adaptability to real-world surveillance scenarios.

II. Proposed Methodology

We propose a hybrid quantum-classical (HQC) framework for violence detection in continuous video frames, integrating pretrained convolutional neural network (CNN) feature extraction with quantum encoding and classical classification. The pipeline consists of four stages.

A. Classical Feature Extraction

High-level classical features are extracted from custom built video dataset, using a pretrained ResNet50 convolutional layers and the network is truncated before the fully connected layers. Each input video frame is transformed into a 2048-dimensional feature vector which is the output of convolutional block of ResNet50:

$$\mathbf{f}_{\text{CNN}} \in \mathbb{R}^{2048} \tag{1}$$

B. Classical Preprocessing

To interface with the quantum layer, the 2048-dimensional feature vectors are reduced to a length of $2^{n_{\text{qubits}}}$ via a feedforward neural network (NN) [2]:

$$\mathbf{f}_{\text{classical}} = \text{NN}(\mathbf{f}_{\text{CNN}}) \in \mathbb{R}^{2^{n_{\text{qubits}}}} \tag{2}$$

This network has layer sizes $2048 \to 512 \to 64 \to 2^{n_{\text{qubits}}}$, with ReLU activations applied between hidden layers.

C. Quantum Feature Encoding

The vector output from the NN is encoded into the quantum state using amplitude embedding :

$$|\psi\rangle = \sum_{i=0}^{2^{n_{\text{qubits}}}-1} f_{\text{classical}}[i]|i\rangle \tag{3}$$

Quantum embedding has a potential advantage over classical methods like PCA by enabling representation in a Hilbert space that captures richer feature interactions through entanglement and superposition [5]. Unlike PCA, which performs linear dimensionality reduction, quantum embeddings can encode non-linear correlations more compactly and with fewer parameters, which enhances expressivity and generalization.

A parameterized quantum circuit, constructed with basic entangled layers, applies trainable rotations and entanglement operations to the encoded state $|\psi\rangle$. The quantum autoencoder (QAE) achieves a compression ratio of $2^n : (n/2)$, which scales exponentially with the number of features. In practice, this is realized by measuring the Pauli-Z expectation values on the first $n_{\text{qubits}}/2$ qubits, thereby producing a compressed quantum feature vector [6].

This parameterized quantum circuit is implemented in the QAE as:

$$U(\theta) = \prod_{i=1}^{L} \left(\prod_{j=1}^{4} R_j(\theta_{i,j}) \right) \cdot \mathrm{CNOT}_{\mathrm{entangling}} \tag{4}$$

where R_j are rotation gates and L is the number of layers.

D. Classical Postprocessing and Classification

The quantum output vector is passed to a classical post-processing layer for binary classification of Violence vs Non-violence actions.

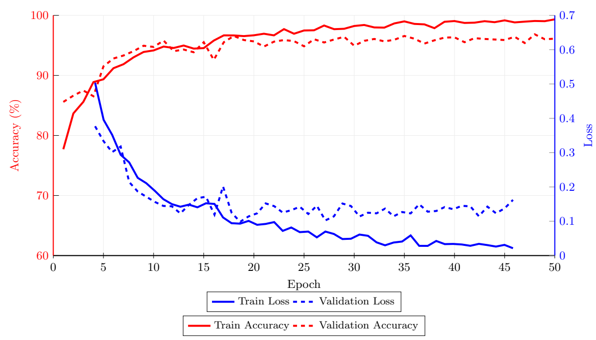

Figure 1. Loss and accuracy trends of the proposed model.

III. Results

The proposed HQC framework was implemented for binary classification of violent and non-violent actions. ResNet50 was employed as a feature extractor to generate 2048-dimensional embeddings, which were subsequently reduced to 16 dimensions for a 4-qubit system, and the model was trained and evaluated on a custom dataset of recorded surveillance videos captured under varying lighting conditions. These compressed features were encoded via amplitude embedding into a quantum layer, followed by entangling operations and classical post-processing for final classification.

The training and validation curves (Fig. 1) demonstrate stable and consistent convergence of the hybrid framework. The training accuracy rapidly improved, surpassing 90% within the first 10 epochs and converging towards 98–99% by epoch 50. Validation accuracy followed a similar trend, stabilizing around 95–96% with only minor fluctuations, indicating strong generalization capability. On the other hand, the training loss decreased steadily from an initial value above 0.5 to below 0.05, while the validation loss converged near 0.15, further confirming robust convergence.

These results verify that the integration of quantum feature embedding and entanglement operations preserves classification performance while providing efficient compression of high-dimensional CNN features. The hybrid model consistently achieved validation accuracy comparable to the original ResNet50 backbone while requiring significantly fewer dimensions for downstream classification, as summarized in Table I. This indicates that HQC models can exploit both classical feature hierarchies and quantum representational power to achieve high accuracy, robustness, and scalability for real-time surveillance applications.

Table I
Comparative Evaluation of Classical and Quantum-Classical Models for Violence Detection

Model	Training Acc.	Validation Acc.	Compression
Original ResNet50	99.6%	97.3%	1:1
Classical PCA	85.5%	84.3%	512:2
Quantum-Classical	99.03%	96.29%	512:2

IV. Conclusion

This work demonstrates the successful integration of a DL backbone (ResNet50) with a quantum-enhanced classifier for image-based violence detection. Beyond performance gains, the proposed HQC design highlights the potential of quantum-assisted compression to reduce computational load without compromising accuracy, which is critical for real-time surveillance. These findings indicate that HQC approaches can move beyond proof-of-concept and towards scalable, deployable systems in intelligent monitoring. Future extensions may consider larger qubit systems, temporal modeling of video sequences, and deployment on near-term quantum hardware to further validate the practical benefits of this paradigm.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under RS-2025-00556064, by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2021-II212046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and by a grant from Kyung Hee University in 2023 (KHU-20233663).

References

[1] S. M. A. Rizvi, U. I. Paracha, U. Khalid, K. Lee, and H. Shin, "Quantum machine learning: Towards hybrid quantum-classical vision models," *Mathematics*, vol. 13, no. 16, p. 2645, Aug. 2025.

[2] M. Plesch and Č. Brukner, "Quantum-state preparation with universal gate decompositions," *Phys. Rev. A*, vol. 83, no. 3, p. 032302, 2011.

[3] T. Bonny and A. Haq, "Emulation of high-performance correlation-based quantum clustering algorithm for two-dimensional data on fpga," *Quantum Information Processing*, vol. 19, no. 179, 2020. [Online]. Available: https://doi.org/10.1007/s11128-020-02683-9

[4] B. Narottama, A. U. Haq, J. A. Ansere, N. Simmons, B. Canberk, S. L. Cotton, H. Shin, and T. Q. Duong, "Quantum deep reinforcement learning for digital twin-enabled 6g networks and semantic communications: Considerations for adoption and security," *IEEE Transactions on Network Science and Engineering*, pp. 1–25, 2025.

[5] J. Romero, J. P. Olson, and A. Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," *Quantum Science and Technology*, vol. 2, no. 4, p. 045001, 2017.

[6] H. Ma, C.-J. Huang, C. Chen, D. Dong, Y. Wang, R.-B. Wu, and G.-Y. Xiang, "On compression rate of quantum autoencoders: Control design, numerical and experimental realization," *Quantum Information Processing*, vol. 19, no. 9, p. 306, 2020.