Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid Complex Quantum U-Net for Enhanced Medical Polyp Image Segmentation

Mario Jojoa-Acosta

mariofernando.jojoa@uva.es

Universidad Mariana

Cristian Castillo-Olea

Autonomous University of Baja California

Rubén M. Lorenzo

Universidad de Valladolid

Giovanni Hernandez

Universidad Mariana

Alfonso Bahillo

Universidad de Valladolid

Article

Keywords: Complex-valued neural networks, Quantum machine learning, Polyp segmentation, U-Net architecture, Medical image analysis

Posted Date: September 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10694115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid Complex–Quantum U-Net for Enhanced Medical Polyp Image Segmentation

Mario Jojoa-Acosta[1,3*], Cristian Castillo-Olea[2], Rubén M. Lorenzo[1], Giovanni Hernández[3], Alfonso Bahillo[1]

[1*]Department of Signal Theory and Communications, Universidad de Valladolid, Valladolid, 47011, Spain.
[2]Autonomous University of Baja California, Mexicali, Baja California, Mexico.
[3]Universidad Mariana, Pasto, Nariño, Colombia.

*Corresponding author(s). E-mail(s): mariofernando.jojoa@uva.es;
Contributing authors: castillo.cristian@uabc.edu.mx;

Abstract

In this work, a hybrid segmentation architecture is proposed that integrates real, complex, and quantum computation within a U-Net framework. The model combines a ResNet-50 encoder for initial feature extraction, a complex-valued bottleneck convolutional block to preserve phase and amplitude information, and a quantum variational layer with eight qubits to encode global statistics derived from complex activations. The proposed model is evaluated on a polyp segmentation task using the PolipSet, Kvasir-SEG, and CVC-ClinicDB datasets, with direct baseline comparisons against a pretrained U-Net and SegFormer reported on PolipSet. Experimental results demonstrate that the hybrid complex–quantum U-Net achieves the best PolipSet performance among the compared models, with mean Intersection over Union (IoU) of 79.36 %, while also obtaining 78.29 % on Kvasir-SEG and 85.07 % on CVC-ClinicDB; the corresponding Dice coefficients are 88.38 %, 87.51 %, and 91.63 %, respectively. On PolipSet, these values outperform baselines such as Attention U-Net with ResNet-50 (IoU 75.93 %, Dice 82.94 %) and SegFormer (IoU 72.97 %, Dice 80.86 %). These findings highlight the potential of integrating complex-valued and quantum layers within deep learning pipelines for medical image analysis.

Keywords: Complex-valued neural networks, Quantum machine learning, Polyp segmentation, U-Net architecture, Medical image analysis

1 Introduction

The segmentation of medical images is of critical importance in the development of computer-aided diagnostic systems. The accurate and robust delineation of anatomical structures or pathological regions enables clinicians to make informed decisions and contributes to early disease detection, personalised treatment and improved patient outcomes. Colon polyps are abnormal tissue growths that develop on the inner lining of the large intestine and are recognized as important precursors to colorectal cancer.

Their prevalence increases with age, diet, and certain genetic factors, making early detection critical for cancer prevention. According to the World Health Organization (WHO) [1], colorectal cancer ranks among the top three most common cancers worldwide, and a substantial proportion of cases are linked to the malignant transformation of adenomatous polyps. Statistical analyses compiled by WHO and the International Agency for Research on Cancer (IARC) indicate that routine screening and timely removal of polyps can reduce colorectal cancer incidence by up to 60 %. These global data highlight the necessity of accurate segmentation and detection methods for colon polyps to support effective screening programs and reduce disease burden. Polyps are a key focus in the field of colorectal cancer research due to their potential to contribute to the early diagnosis and prevention of this serious condition, which is one of the most prevalent and deadly cancers worldwide. Nevertheless, the automated segmentation of polyps [2] remains a challenging problem due to their diverse appearance, varying sizes, complex textures, and occlusions.

In recent years, convolutional neural networks (CNNs) [3], especially encoder–decoder architectures such as U-Net, have become the de facto standard for medical image segmentation. Variants of U-Net [4], frequently augmented with attention mechanisms, residual blocks, or pretrained encoders (e.g., ResNet), have attained state-of-the-art performance in various benchmarks. More recently, transformer-based models such as SegFormer have also been adopted, offering improved global context modelling. Despite their evident success, these models are confined exclusively to the real-valued domain, a limitation that curtails their capacity to capture certain aspects of signal structure, particularly those related to phase, frequency, and complex relationships in the data that are naturally expressed in the complex domain.

In order to address this limitation, researchers have initiated the exploration of complex-valued neural networks (CVNNs) [5], which facilitate the representation and processing of complex numbers by neurons and weights. The potential advantages of CVNNs are manifold. They are capable of modelling richer signal representations, preserving amplitude and phase information, and providing enhanced gradient flow in certain settings. In particular, complex convolutions have demonstrated considerable potential in domains such as radar processing, audio signal analysis, and biomedical imaging, where the phase carries critical semantic content. However, the potential of CVNNs [5] in the field of medical image segmentation remains largely untapped, and their integration into hybrid architectures continues to be a subject of ongoing research.

Concurrently, quantum machine learning (QML) [6] has emerged as a novel paradigm at the intersection of quantum computing and data-driven learning. Variational quantum circuits (VQCs) [7], in which trainable quantum gates operate on

encoded classical data, have been proposed as building blocks of quantum-enhanced neural networks. The theoretical potential of QML lies in its ability to model complex probability distributions, capture entangled correlations, and explore high-dimensional solution spaces with fewer parameters. While current quantum hardware remains limited, hybrid quantum–classical models simulated on classical machines offer a fertile ground for experimentation and innovation.

In this study, we explore the synergistic integration of three computational paradigms – real, complex and quantum – into a unified architecture for polyp segmentation. The proposed model is a novel hybrid complex–quantum U-Net that combines a ResNet-50 backbone for real-valued encoding, a complex-valued convolutional bottleneck to capture amplitude–phase relationships, and a variational quantum layer that encodes global statistics from the complex feature space. The quantum layer, implemented via an 8-qubit circuit, receives a compact summary of the magnitude and phase activations and returns a modulating signal that is re-injected into the bottleneck's feature map. This design facilitates the integration of local convolutional structures with global quantum-informed representations in a manner that is fully differentiable.

The efficacy of the proposed approach is evaluated using a polyp segmentation dataset, with the results being benchmarked against two strong baselines: a pretrained U-Net and a SegFormer model [8]. The findings demonstrate that the proposed hybrid architecture attains enhanced performance in both Intersection over Union (IoU) and Dice coefficient metrics on the primary benchmark, thereby substantiating the merits of integrating complex-valued processing and quantum encoding within a deep learning pipeline. This work contributes to the growing field of hybrid architectures that integrate classical and quantum learning paradigms for real-world applications. Our main contributions can be summarized as follows. First, we introduce a hybrid complex–quantum U-Net that combines a ResNet-50 encoder, a complex-valued bottleneck to preserve amplitude–phase information, and an 8-qubit variational quantum layer to modulate global context. Second, we propose a compact quantum input mapping that summarizes magnitude and phase statistics from complex activations into an 8-dimensional vector, enabling quantum processing with minimal overhead. Third, we evaluate the model on polyp segmentation and show improvements in IoU and Dice metrics over strong baselines such as Attention-U-Net and SegFormer on PolipSet, together with competitive cross-dataset performance on Kvasir-SEG and CVC-ClinicDB. Finally, we provide a detailed architectural description and training setup to facilitate reproducibility.

This study explores the integration of complex-valued convolutional processing with variational quantum circuits in a unified architecture for medical polyp segmentation. In contrast to conventional CNNs and transformer-based frameworks, the proposed hybrid design employs a complex-valued bottleneck to preserve amplitude–phase information and a quantum variational module to modulate global statistical descriptors. The methodological rigor of the work is supported by controlled ablation experiments, baseline comparisons on PolipSet, and cross-dataset evaluation on PolipSet, Kvasir-SEG, and CVC-ClinicDB. These elements emphasize the originality of the approach and its potential clinical relevance.

2 Related Work

2.1 U-Net and Transformer-Based Models in Medical Image Segmentation

The U-Net architecture, first introduced by Ronneberger et al. [9], has become a cornerstone in the field of medical image segmentation due to its ability to effectively capture context while preserving spatial resolution. A plethora of extensions have been posited with a view to enhancing U-Net's capabilities. Among these are the utilisation of more sophisticated encoders, such as ResNet [10], DenseNet [11] , and EfficientNet [12]. These models typically leverage transfer learning from large-scale image classification datasets, significantly improving convergence speed and accuracy in medical tasks.

Attention mechanisms have also been widely integrated into the U-Net structure to focus the model on relevant anatomical regions. Oktay et al. [13] introduced Attention U-Net, which selectively emphasizes target structures while suppressing irrelevant background information. More recently, transformer-based architectures have shown strong potential in medical image segmentation. Models such as TransUNet [14] and UNETR [15] leverage the global self-attention mechanism of transformers to capture long-range dependencies in the image. In particular, SegFormer [16] offers a lightweight and efficient framework by combining multi-scale transformers with a hierarchical decoder, achieving competitive performance across multiple benchmarks.

Despite their effectiveness, these models operate in the real-valued domain and may struggle to capture certain aspects of biomedical signals, such as phase or frequency content, which are crucial in various imaging modalities.

2.2 Complex-Valued Neural Networks for Image and Signal Processing

The concept of complex-valued neural networks (CVNNs) represents an extension of traditional neural networks by incorporating complex numbers in weights, activations, and operations. The theoretical underpinnings for convolutional neural networks (CVNNs) were initially established in early works, such asHirose [17] and Aizenberg et al. [18], which laid the foundation for complex-valued backpropagation and activation functions. Recent advancements have seen the adaptation of convolutional layers to operate on complex inputs, thereby enabling the representation of phase and amplitude in learned features.

Trabelsi et al. [19] proposed a comprehensive framework for deep complex networks, demonstrating enhanced performance in audio, radar, and MRI data. Their research demonstrated that the utilisation of complex representations can facilitate enhanced regularisation and more efficient feature encoding in scenarios where the input signal inherently incorporates phase information. In the domain of medical imaging, convolutional neural networks (CVNNs) have been employed in various applications, including magnetic resonance imaging (MRI) reconstruction [20], electroencephalography (EEG) analysis [21], and functional imaging [22]. In these applications, phase plays a critical role.

In the context of segmentation tasks, the utilisation of complex-valued architectures remains infrequent. However, recent endeavours, including the Complex U-Net model proposed by [23] and hybrid complex–real models by [24], have demonstrated encouraging outcomes. These architectures typically involve the replacement of the convolutional blocks in the bottleneck or decoder with complex variants, whilst preserving real-valued components in order to ensure compatibility. Nevertheless, the full potential of combining complex and additional paradigms (e.g., quantum computation) remains largely unexplored.

2.3 Quantum Neural Networks and Hybrid Quantum–Classical Architectures

Quantum machine learning (QML) is an emerging field that aims to leverage quantum phenomena to enhance learning models [25]. In this context, one of the most extensively studied architectures is the variational quantum circuit (VQC), which consists of parameterised quantum gates acting on qubit states (see reference [26] for further details). These models are particularly well-suited to hybrid quantum-classical systems, in which classical networks are combined with quantum subroutines to process data in high-dimensional Hilbert spaces.

Several works have explored the use of quantum circuits for image classification, such as Farhi et al.'s quantum classifier [17] and the Quanvolutional Neural Network (QNN) [27], which introduces a quantum analogue of convolution. Quantum layers have also been embedded in classical architectures for tasks such as graph classification [28], molecular modeling [29], and medical imaging [30]. Schuld and Killoran [31] provide a theoretical framework for encoding classical data into quantum circuits and differentiating through quantum nodes using backpropagation.

In the field of medical imaging, there is a paucity of research that has examined the impact of quantum layers in segmentation pipelines. A notable work in this area is Sankari et al. [32], which employs a hybrid quantum–classical model for retinal vessel segmentation. This work demonstrates that even shallow quantum circuits have the capacity to inject nontrivial features into a deep learning pipeline. Nevertheless, the integration of quantum layers with complex-valued representations remains a novel research direction.

2.4 Context-Aware Transformers for Medical Segmentation

While transformer-based models have demonstrated strong performance in general segmentation tasks, recent advancements have tailored these architectures to better capture anatomical context in medical images. Chen et al. [33] proposed MedT, a gated axial-attention transformer that combines local convolutional cues with long-range dependencies via attention gating mechanisms. Their model demonstrates that incorporating both spatial and contextual priors leads to improvements in segmenting low-contrast and irregularly shaped structures. Similarly, Cao et al. [34] introduced

Swin-UNet, which adapts the Swin Transformer architecture to a hierarchical U-Net-like design. These models further validate the effectiveness of multi-scale, attention-driven mechanisms in addressing the inherent complexity and variability of medical imaging datasets.

2.5 Recent Hybrid Neuro–Quantum Vision Models

In the broader field of quantum machine learning, there has been growing interest in combining quantum layers with deep vision models beyond toy datasets. Chen et al. [35] proposed a hybrid QCNN–ResNet architecture for image classification, in which feature maps from early ResNet layers are encoded into quantum circuits for enhanced feature representation. Meanwhile, Kuo et al. [36] explored the fusion of quantum circuits and attention-based modules for improved image reconstruction, emphasizing the complementary nature of classical and quantum representations. These recent developments underscore the feasibility of integrating quantum modules into deep learning pipelines and provide additional evidence that quantum processing can offer performance and robustness benefits, especially when dealing with sparse or noisy data—a common challenge in medical imaging.

2.6 Limitations of existing methods and our contribution

Despite notable progress, current polyp segmentation methods present important limitations. Classical real-valued convolutional neural networks often fail to encode the phase and amplitude information that is critical for accurately representing complex colonoscopic textures. Transformer-based architectures improve global context modeling but typically incur high computational costs and require large-scale training data, which may hinder deployment in clinical settings. Furthermore, hybrid CNN-transformer approaches rarely exploit the potential of quantum-inspired computation to enhance feature representation. Our proposed hybrid complex–quantum U-Net is specifically designed to overcome these limitations by incorporating a complex-valued bottleneck to preserve amplitude–phase relationships and a quantum variational layer to capture global statistical dependencies. This combination leads to improved Intersection over Union (IoU) and Dice coefficients on the primary benchmark, together with strong cross-dataset performance on PolipSet, Kvasir-SEG, and CVC-ClinicDB, indicating encouraging generalization and robustness.

3 Methodology

The following section details the materials and methodological framework employed in this study. Firstly, the dataset employed for training and evaluating the segmentation models is described. Secondly, the architecture of the proposed hybrid model integrating real-valued, complex-valued, and quantum-inspired components is detailed.

3.1 Dataset Description

In order to evaluate the effectiveness of state-of-the-art deep learning architectures in medical image segmentation, a custom dataset was introduced. This dataset was

named PolipSet. This dataset comprises high-quality colonoscopy images that have been manually collected and annotated to highlight polyp regions. Each sample comprises an RGB image and its corresponding binary segmentation mask, indicating the precise location of the polyp.

PolipSet was designed to capture clinical variability, including differences in image quality, lighting conditions, polyp sizes, and anatomical structures. The dataset under consideration is comprised of 246 images, which were divided into three sets: a training set, a validation set, and a testing set, with each set containing 60%, 20%, and 20% of the images, respectively. In order to circumvent the occurrence of data leakage, images pertaining to a single patient were allocated exclusively to one of the subsets. All images were resized to 512×512 pixels and normalised to the [0, 1] range. This was done to standardise the input dimensions and facilitate training convergence. All image annotations were performed manually by board-certified gastroenterologists with extensive experience in colonoscopy, ensuring clinical accuracy and consistency across the datasets. PolipSet was compiled at the Hospital de Tijuana following approval by the Comité de Ética y Evaluación of the Universidad Autónoma de Baja California (UABC; approval number CTJ-172023). The study was conducted in accordance with the Declaration of Helsinki and all applicable guidelines and regulations; informed consent was obtained from all participants and/or their legal guardians, as applicable.

To further evaluate generalization, we trained and validated the model on the publicly available Kvasir-SEG and CVC-ClinicDB datasets. Kvasir-SEG contains 1000 polyp images with corresponding masks, while CVC-ClinicDB includes 612 images. All images were resized to 512×512 and normalized to [0,1] following the same preprocessing pipeline as PolipSet.

Table 1 Details of the datasets used for training and evaluation

Dataset	**Images**	**Train**	**Val**	**Test**
PolipSet	246	146	50	50
Kvasir-SEG	1,000	600	200	200
CVC-ClinicDB	612	368	122	122

3.2 Proposed hybrid Model

This section introduces a novel hybrid neural architecture designed for medical image segmentation, specifically polyp delineation. The architecture fuses real-valued convolutional encoding, complex-valued bottleneck transformations, and quantum-based global modulation into a single U-Net-inspired framework.

3.2.1 Encoder: ResNet-50 Backbone

Let $x \in \mathbb{R}^{B\times 3\times 512\times 512}$ denote the input image batch, where B is the batch size and 512 the image resolution. The encoder $E(\cdot)$ is defined as a composition of layers from a pretrained ResNet-50:

$$E(x) = e_5 = \text{ResNet50}(x) \in \mathbb{R}^{B\times 2048\times 16\times 16} \tag{1}$$

This representation captures hierarchical real-valued features through residual connections and downsampling.

3.2.2 Bottleneck: Complex-Valued Convolutional Processing

To extend representational richness, we define the latent complex-valued tensor:

$$C = e_5 + i \cdot 0 \in \mathbb{C}^{B\times 2048\times 16\times 16} \tag{2}$$

A complex convolution is then applied as:

$$C' = \sigma\left(W_r * \Re(C) - W_i * \Im(C) + i(W_r * \Im(C) + W_i * \Re(C))\right) \tag{3}$$

Here, $W_r, W_i \in \mathbb{R}^{512\times 2048\times 3\times 3}$ are the real and imaginary parts of the convolutional kernels, $\Re(\cdot), \Im(\cdot)$ denote real and imaginary parts, and σ applies a non-linearity (e.g., ReLU) on the complex modulus.

We extract global spatial descriptors by averaging:

$$m = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}|C'_{b,:,i,j}|, \quad \phi = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\angle C'_{b,:,i,j} \tag{4}$$

where $H = W = 16$ are spatial dimensions, yielding magnitude $m \in \mathbb{R}^{B\times 512}$ and phase $\phi \in \mathbb{R}^{B\times 512}$.

3.2.3 Quantum Module

The quantum input vector is formed by compressing these global descriptors through learned linear projections:

$$q_{\text{in}} = [P_m m \,\|\, P_\phi \phi] \in \mathbb{R}^8 \tag{5}$$

This 8D vector parameterizes a variational quantum circuit (VQC):

$$q_{\text{out}} = f_{\text{VQC}}(q_{\text{in}}, \theta) \in \mathbb{R}^8 \tag{6}$$

where $P_m, P_\phi \in \mathbb{R}^{4\times 512}$ are trainable projections and $\|$ denotes concatenation. The VQC output is then projected back to the bottleneck channel dimension and expanded to match spatial resolution:

$$Q = (W_q q_{\text{out}}).\text{unsqueeze}(-1).\text{unsqueeze}(-1).\text{expand}(-1,-1,16,16) \tag{7}$$

and added to the real part of C':

$$\Re(C') := \Re(C') + Q \tag{8}$$

where $W_q \in \mathbb{R}^{512\times 8}$, so that $Q \in \mathbb{R}^{B\times 512\times 16\times 16}$ is dimensionally compatible with $\Re(C')$.

3.2.4 Variational quantum circuit (VQC)

Given an input vector $q_{\text{in}} \in \mathbb{R}^8$ and trainable parameters $\theta \in \mathbb{R}^{8\times p\times L}$, the circuit defines the map

$$q_{\text{out}} \;=\; f_{\text{VQC}}(q_{\text{in}};\theta) \;=\; \big(\langle Z_0\rangle,\ldots,\langle Z_7\rangle\big) \;\in\; [-1,1]^8,$$

where each component is obtained as an expectation value of a Pauli–Z observable,

$$q_{\text{out},i} \;=\; \langle Z_i\rangle \;=\; \langle 0|^{\otimes 8}\, U(q_{\text{in}},\theta)^\dagger\, Z_i\, U(q_{\text{in}},\theta)\,|0\rangle^{\otimes 8}, \qquad i=0,\ldots,7.$$

The circuit unitary is $U(q_{\text{in}},\theta) = U_{\text{var}}(\theta)\,U_{\text{enc}}(q_{\text{in}})$ with data encoding

$$U_{\text{enc}}(q_{\text{in}}) \;=\; \bigotimes_{i=0}^{7} R_Y\big(q_{\text{in},i}\big), \qquad R_Y(\alpha) = e^{-i\alpha Y/2},$$

and a variational part composed of L repeated layers without entangling gates,

$$U_{\text{var}}(\theta) \;=\; \prod_{\ell=1}^{L}\ \bigotimes_{i=0}^{7} \text{Rot}\big(\theta_{i,1}^{(\ell)},\,\theta_{i,2}^{(\ell)},\,\theta_{i,3}^{(\ell)}\big), \qquad \text{Rot}(\phi,\vartheta,\omega) = R_Z(\phi)\,R_Y(\vartheta)\,R_Z(\omega).$$

Here L denotes the number of repeated variational layers and p the number of trainable angles per qubit and per layer. With the Euler Z–Y–Z parameterization one has $p = 3$. In our implementation $L = 1$ and $p = 3$, which yields a total of $8\times p\times L = 24$ quantum trainable parameters. The circuit is simulated on PennyLane's `default.qubit` state-vector backend with PyTorch integration.

3.2.5 Decoder with Attention

Reconstruction proceeds via attention-gated upsampling blocks. Let x be the encoder skip feature and g the gating signal. The attention mechanism computes:

$$\alpha = \sigma(W_x x + W_g g + b), \quad x' = \alpha \cdot x \tag{9}$$

The decoder output is:

$$\hat{y} = \sigma(\text{Conv}_{1\times1}(U_5(U_4(U_3(U_2(U_1(\Re(C')))))))) \tag{10}$$

where U_i correspond to successive attention-gated upsampling blocks with skip connections from the encoder, progressively recovering the spatial resolution from 16×16 to 512×512.

3.2.6 Model architecture

The Figure 1 illustrates the architecture of the proposed Hybrid Complex-Quantum U-Net model. This architecture integrates a pretrained ResNet-50 backbone as encoder for hierarchical feature extraction, followed by a complex-valued convolutional module and a quantum-modulation layer. The decoder follows a U-Net structure with skip

connections and five progressive upsampling stages to recover the original spatial resolution. This hybrid combination results in enhanced segmentation performance on the primary benchmark compared to standard U-Net and SegFormer models.

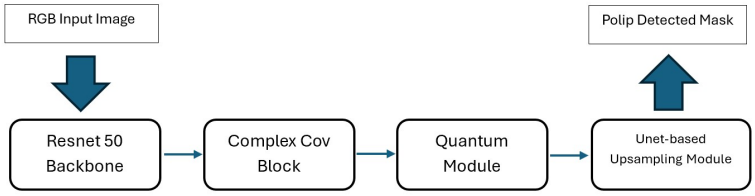


Fig. 1 Proposed hybrid architecture integrating ResNet encoder, complex bottleneck, and quantum module.

The architectural design of the proposed Hybrid Complex-Quantum U-Net is summarized in Table 2. The model accepts RGB images of resolution 512×512 as input and employs a pretrained ResNet-50 backbone as encoder to extract hierarchical real-valued features. These features are then transformed into a 512-channel complex-valued representation through a complex convolutional block. Subsequently, global magnitude and phase descriptors are compressed into an 8-dimensional vector, processed by the variational quantum circuit, projected back to 512 channels, and used to modulate the real part of the bottleneck representation. The decoder follows a U-Net-style design with progressive upsampling and skip connections, reconstructing the spatial resolution back to 512×512. Finally, a convolution with a 1×1 kernel reduces the number of channels to one, followed by a sigmoid activation to produce the segmentation mask.

Table 2 Layer-wise dimensions of the Hybrid Complex-Quantum U-Net with ResNet-50 encoder.

Phase	**Operation**	**Output Size**
Input	RGB image input	$B \times 3 \times 512 \times 512$
Encoder	ResNet-50 backbone (pretrained)	$B \times 2048 \times 16 \times 16$
Complex Block	Complex-valued convolution	$B \times 512 \times 16 \times 16$
Quantum Layer	VQC modulation projected to bottleneck width	$B \times 512 \times 16 \times 16$
Decoder	U-Net style decoder with upsampling	$B \times C \times 512 \times 512$
Output	Conv (kernel 1×1) to 1 channel + Sigmoid	$B \times 1 \times 512 \times 512$

4 Experimental Setup

The hyperparameters associated with the training configuration of each evaluated model are presented in Table 3. This includes the total number of epochs, learning rate (LR), encoder initialization (pretrained or not), number of quantum bits (Qbits) used in the quantum layer (if applicable), optimization algorithm, and batch size. These parameters were chosen to ensure a fair comparison among models and to maximize segmentation performance under consistent experimental conditions.

Table 3 Hyperparams of the trained models

Model	**Epochs**	**LR**	**Encoder Pretrained**	**Qbits**	**Optimizer**	**Batch Size**
Hybrid real-complex-quatum UNET	40	0.0001	Yes	8	ADAM	8
Attention UNET – Resnet 50	30	0.0001	Yes	0	ADAM	8
Segformer – Pretrained	35	0.0001	Yes	0	ADAM	32
Segformer – From Scratch	45	0.0001	Not	0	ADAM	32

The hyperparameters listed in Table 3 correspond to the final training configuration used for the reported experiments. They were selected empirically based on validation performance and computational feasibility, with special attention to segmentation accuracy (IoU and Dice) and convergence stability on PolipSet.

5 Results

The experimental results demonstrate that the proposed hybrid real-complex-quantum UNet outperforms both the attention-based ResNet50 UNet and the transformer-based SegFormer variants on the polyp segmentation task. Specifically, the hybrid model achieves the highest performance, with an Intersection over Union (IoU) of 79.36% and a Dice coefficient of 88.38%.

In comparison with the UNet baseline, the hybrid model demonstrates a relative enhancement of 3.43 percentage points in IoU and 5.44 points in Dice score. This finding indicates that the incorporation of complex-valued convolutions and quantum-aware modulation enhances the structural and boundary representation of the segmented polyps.

In contrast, the pretrained SegFormer model, despite its strength in semantic segmentation, exhibits suboptimal performance in comparison to the UNet variants, particularly with regard to IoU. The discrepancy is further accentuated when SegFormer is trained from the beginning, thereby validating the efficacy of transfer learning and attention mechanisms in scenarios where data is scarce.

The superior performance of the hybrid model can be attributed to its ability to capture phase information through complex operations that jointly contribute to improved spatial coherence and segmentation accuracy, as reflected in the final scores.

To assess the contribution of each architectural component, we performed ablation experiments by selectively removing the complex-valued bottleneck (Hybrid–no-complex) or the quantum module (Hybrid–no-quantum). Table 4 summarizes the results on PolipSet.

Table 4 Comparison of segmentation performance (IoU, Dice) and per-sample FLOPs (512×512).

Model	**IoU (%)**	**Dice (%)**	**FLOPs / sample (GFLOPs)**
Hybrid Quantum-complex UNET	**79.36**	**88.38**	**64.83**
Attention UNET – ResNet-50	75.93	82.94	≈72.0
SegFormer – Pretrained	72.97	80.86	≈22.0
SegFormer – From Scratch	70.24	78.76	≈22.0
Hybrid-no-complex	74.40	82.52	≈58.0
Hybrid-no-quantum	77.30	85.81	≈62.0

In line with established practice in medical image segmentation, the performance of the proposed model is primarily assessed using the Intersection over Union (IoU) and Dice coefficient metrics. These measures directly evaluate the spatial overlap between predicted masks and ground truth, which is the main objective in segmentation tasks. Attention-U-Net and SegFormer were selected as primary baselines because they represent two dominant and complementary segmentation paradigms: advanced CNN-based models and transformer-based architectures. Both have demonstrated strong and consistent performance across multiple medical imaging benchmarks, making them suitable and challenging references for evaluating the proposed hybrid complex–quantum U-Net on PolipSet. To further examine generalization, we additionally trained, validated, and tested the proposed model on the benchmark datasets Kvasir-SEG and CVC-ClinicDB. The obtained results are shown in Table 5.

Table 5 Performance of Hybrid Quantum-complex UNET. Testing with Kvasir-SEG and CVC clinic DB

Metric	**Kvasir-SEG**	**CVC Clinic DB**
IoU (%)	78.29	85.07
Dice (%)	87.51	91.63

As illustrated by Figure 2, the training and validation loss curves corresponding to the four segmentation models evaluated in this study demonstrate the effectiveness of the proposed approach. The vertical dashed lines represent the points at which early stopping was initiated, as determined by the validation performance. As was previously demonstrated, the Hybrid Complex-Quantum UNet displays a more stable and smoother convergence behaviour throughout the training process. This suggests that it possesses enhanced generalisation capacity and improved optimisation dynamics when compared to other architectures.

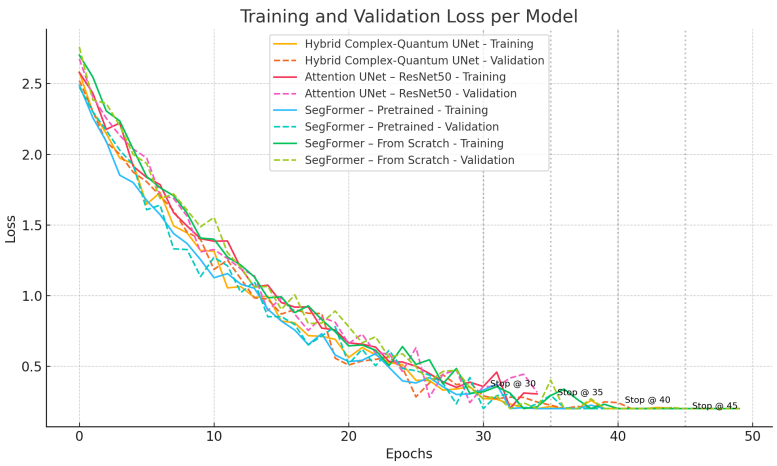


Fig. 2 Training and validation loss curves for the four segmentation models. Early stopping is indicated by the vertical dashed lines. The Hybrid Complex-Quantum UNet demonstrates a smoother and more stable convergence.

Overall, the proposed hybrid complex–quantum U-Net achieved the highest Intersection over Union (IoU) and Dice coefficients among the compared models on PolipSet and also showed strong cross-dataset performance on Kvasir-SEG and CVC-ClinicDB. These results support the effectiveness of combining complex-valued computation with a quantum variational layer for accurate and efficient medical polyp segmentation.

6 Discussion

The findings of this study underscore the efficacy of incorporating real-valued, complex-valued, and quantum components into a cohesive deep learning architecture for medical image segmentation. The proposed hybrid complex-quantum U-Net exhibited superior performance to the pretrained U-Net and SegFormer baselines on PolipSet across standard metrics such as Intersection over Union (IoU) and Dice coefficient, while also maintaining strong results on Kvasir-SEG and CVC-ClinicDB. These enhancements imply that the augmented representational capacity provided by the complex and quantum modules contributes substantially to segmentation performance.

The ablation results in Table 4 highlight the critical role of both the complex-valued bottleneck and the quantum module. Removing the complex-valued bottleneck (Hybrid–no-complex) led to a noticeable decrease in IoU and Dice scores, indicating that preserving amplitude–phase information is essential for accurate boundary representation and feature discrimination. Similarly, omitting the quantum module (Hybrid–no-quantum) resulted in a consistent but less pronounced performance drop, suggesting that global context modeling via quantum processing provides complementary benefits. These findings confirm that the two components contribute synergistically to the overall performance of the proposed architecture.

A fundamental element of this enhancement originates from the complex-valued bottleneck, which preserves both amplitude and phase information—features that are frequently disregarded in conventional real-valued networks. In the context of polyp segmentation, where subtle texture and boundary cues are critical, the ability to model phase relationships appears to offer a tangible advantage. This finding aligns with the observations made in other fields, such as radar or MRI, where complex-valued models have demonstrated an augmented capacity for discrimination.

The integration of the quantum layer serves to further amplify this effect. By compressing and encoding global statistics of the complex activation maps into an 8-qubit variational quantum circuit, the network applies an additional nonlinear transformation to a compact summary of the bottleneck activations before reinjecting the result into the decoder pathway. Whilst the quantum layer operates in simulation mode and, in the present implementation, does not include entangling gates, it still acts as a structured modulation mechanism that complements the classical feature extractor. The quantitative enhancements observed in our experimental investigations support the usefulness of this design choice.

This study emphasizes the conceptual novelty of the proposed hybrid complex–quantum U-Net, which combines complex-valued representations with quantum variational processing. Unlike conventional CNNs or transformer-based networks, this design introduces a different mechanism for modulating global bottleneck descriptors while preserving amplitude–phase relationships. The architecture is intentionally modular, enabling its quantum layer to be integrated into a variety of state-of-the-art backbones. This flexibility suggests that the approach is not limited to a single model family but rather opens a pathway to enrich diverse segmentation frameworks with quantum-informed feature modulation.

Furthermore, the hybrid design remains computationally feasible. The restriction of quantum processing to a concise, eight-dimensional synopsis of the bottleneck activations is a strategy employed to circumvent the exponential expense ordinarily linked with quantum simulations. This design choice renders the model trainable on standard GPU hardware while maintaining extensibility to future quantum backends. Furthermore, this approach paves the way for additional exploration of quantum feature injection in other components of the network, including attention maps and decoder blocks. However, it is important to acknowledge several limitations. Firstly, it is important to note that the present quantum implementation is simulated and does not account for hardware noise, decoherence, or circuit depth limitations that would arise in a real quantum processor.

In summary, the hybrid complex–quantum U-Net demonstrates that combining multiple computational paradigms in a coherent architectural framework can lead to meaningful performance gains in medical image segmentation. The findings of this study contribute to the growing body of research that advocates for hybrid models that go beyond classical deep learning. These findings suggest a promising direction for the integration of physics-inspired computation in clinical AI applications.

When compared with the selected baselines on PolipSet, the proposed hybrid complex–quantum U-Net demonstrated clear advantages but also revealed certain limitations. On the one hand, the model outperformed both the conventional U-Net and

the transformer-based SegFormer in terms of Intersection over Union (IoU) and Dice coefficient, as reported in Table 4. These gains can be attributed to the complex-valued bottleneck, which preserves amplitude–phase information often discarded in real-valued networks, and to the quantum variational module, which provides complementary global feature modulation. On the other hand, while U-Net remains simpler and computationally lighter, and SegFormer benefits from scalability and maturity of transformer-based architectures, our approach still relies on quantum simulation and does not yet account for hardware-related constraints such as noise or decoherence. This indicates that, although the hybrid design introduces a novel and effective paradigm that bridges gaps in amplitude–phase representation and global feature encoding, its current feasibility remains bounded by the limitations of simulated quantum processing. These trade-offs highlight both the promise and the challenges of integrating physics-inspired modules into medical image segmentation frameworks.

7 Conclusion

This study proposes a novel hybrid neural architecture that integrates real-valued encoding, complex-valued processing, and quantum computation into a unified framework for medical image segmentation. The model utilises the strengths of each computational paradigm, namely the stability and efficiency of real-valued convolutions, the expressive power of complex-valued representations to preserve amplitude and phase information, and the global feature encoding capacity offered by variational quantum circuits. The design was instantiated through a U-Net variant with a ResNet-50 encoder, a complex-valued bottleneck, and an 8-qubit quantum layer. Collectively, these elements constitute a coherent pipeline for end-to-end training.

The experimental results on polyp segmentation tasks demonstrated that this hybrid architecture significantly outperforms two widely used baselines on PolipSet: a pretrained U-Net and the transformer-based SegFormer model. Quantitative evaluations employing metrics such as Intersection over Union (IoU) and Dice coefficient showed clear improvements on the held-out test partition of PolipSet, together with strong transferability to Kvasir-SEG and CVC-ClinicDB. The findings indicate that quantum-based bottleneck modulation can provide useful global context, even when applied to low-dimensional summaries of activation maps. Furthermore, complex-valued convolutions appear to offer an advantage in capturing the nuanced spatial and textural patterns often observed in medical imaging data.

The Hybrid Quantum-complex U-Net achieved the best performance on the polyp dataset, reaching an IoU of 79.36% and a Dice of 88.38%, outperforming all baselines and ablation variants. Despite incorporating quantum–complex operations, its computational cost of approximately 64.8 GFLOPs per sample remains within the expected range for high-resolution (512×512 px) segmentation models and is only slightly higher than that of its simplified variants without quantum or complex components. In comparison, Attention U-Net–ResNet-50 and SegFormer obtained lower segmentation scores (IoU $\leq 75.9\%$, Dice $\leq 82.9\%$) even though their inference cost is similar or, for SegFormer, considerably lower. These findings demonstrate that combining quantum

and complex-valued modules provides a substantial accuracy gain without imposing a prohibitive increase in inference requirements.

In addition to empirical performance, the proposed approach provides a conceptual framework for the incorporation of quantum components into classical neural networks in a modular and interpretable manner. The restriction of quantum computation to a single layer that operates on condensed statistical features has two principal benefits. Firstly, it helps maintain the scalability and efficiency of the overall model. Secondly, it enables the study of quantum-informed feature modulation without redesigning the full network. Notably, the model maintains compatibility with contemporary GPU-based training, and the quantum layer can be executed in simulation, signifying a pragmatic advancement towards hybrid quantum–classical learning.

Nevertheless, several challenges and open questions remain. While the quantum circuits under consideration are differentiable and trainable in simulation, it is clear that real-world deployment on noisy intermediate-scale quantum (NISQ) devices will require careful design to handle decoherence, limited qubit counts, and gate fidelity. Furthermore, additional analysis is required to explore how quantum layers might adaptively modulate different stages of the network. The investigation of the applicability of this hybrid design in other clinical scenarios, including multi-organ segmentation, histopathology, and time-series biomedical data, represents a promising direction for future research.

The proposed architecture is model-agnostic by design. Its complex-valued bottleneck and variational quantum layer operate as independent modules that can be embedded into other high-performing medical image segmentation frameworks, including TransUNet, Swin-UNet, and UNETR. Such integration requires no fundamental structural changes and could potentially improve the representation of fine-grained boundaries and global dependencies in these architectures. Future work will systematically investigate these extensions to further validate the generality and scalability of the complex–quantum paradigm.

This work introduces a new pathway for enhancing deep learning in medical image analysis by fusing real, complex, and quantum computations in a single architecture. The consistent performance gains and the conceptual novelty of the design suggest that hybrid models can exceed the capabilities of purely classical systems. As quantum computing continues to evolve, it is anticipated that such hybrid paradigms will play an increasingly central role in the development of intelligent, adaptable, and explainable models for high-stakes domains such as healthcare.

Despite these promising results, several limitations must be acknowledged. First, the evaluation was conducted on datasets of limited size, which may constrain the generalizability of the findings. Second, the employed variational quantum circuit was restricted to separable rotations without entangling gates, which limits the expressive power of the quantum module. Third, the implementation was performed entirely in simulation, thus not accounting for hardware-related factors such as noise, decoherence, and limited qubit fidelity that would arise in near-term quantum devices. Future work will therefore focus on expanding the evaluation to larger and more diverse benchmarks, exploring entangling circuit architectures with higher expressive capacity, and testing the hybrid design on real quantum hardware to assess its practical viability.

An additional limitation of the present work lies in the configuration of the quantum variational circuit. In this study, we fixed the design to 8 qubits with a single layer ($L = 1$) and $p = 3$ Euler rotations per qubit, without entanglement. While this choice provided a practical balance between representational capacity and computational feasibility, we acknowledge that a systematic exploration of quantum-specific hyperparameters—including qubit count, rotation schemes, circuit depth, and the possible introduction of entanglement—remains an important direction for future research.

Data Availability

The PolipSet dataset generated and analyzed during this study is not publicly available because of ethical and confidentiality restrictions. However, it may be made available upon reasonable request, subject to the applicable ethical and confidentiality requirements. Requests should be submitted by email to Cristian Castillo-Olea at castillo.cristian@uabc.edu.mx. Kvasir-SEG and CVC-ClinicDB were not generated by the authors and are publicly available through their respective repositories.

Ethical Approval

The PolipSet database was compiled at the Hospital de Tijuana. The study was approved by the Comité de Ética y Evaluación of the Universidad Autónoma de Baja California (UABC) under approval number CTJ-172023. All procedures involving human participants were conducted in accordance with the Declaration of Helsinki and all applicable guidelines and regulations. Informed consent was obtained from all participants and/or their legal guardians, as applicable.

Funding

This research was partially supported by the Spanish Ministry of Science and Innovation under the Aginplace project (ref. PID2023-146254OB-C41), financed by MICIU/AEI/10.13039/501100011033 and FEDER, UE, and by the SWALU project (ref. CPP2022-010045).

References

[1] World Health Organization: Colorectal cancer. https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer. Accessed: 2025-09-26

[2] Qayoom, A., Xie, J., Ali, H.: Polyp segmentation in medical imaging: Challenges, approaches and future directions. Artificial Intelligence Review **58**(6), 169 (2025)

[3] Yao, W., *et al.*: From cnn to transformer: A review of medical image segmentation models. Journal of Imaging Informatics in Medicine **37**(4), 1529–1547 (2024)

[4] Yuan, Y., Cheng, Y.: Medical image segmentation with unet-based multi-scale context fusion. Scientific Reports **14**(1), 15687 (2024)

[5] Jojoa, M., Garcia-Zapirain, B., Percybrooks, W.: A fair performance comparison between complex-valued and real-valued neural networks for disease detection. Diagnostics **12**(8), 1893 (2022)

[6] Lu, W., *et al.*: Quantum machine learning: Classifications, challenges, and solutions. Journal of Industrial Information Integration **42**, 100736 (2024)

[7] Gong, L.-H., *et al.*: Quantum convolutional neural network based on variational quantum circuits. Optics Communications **550**, 129993 (2024)

[8] Makarov, I., *et al.*: A comparative analysis of segformer, fabe-net and vgg-unet models for the segmentation of neural structures on histological sections. Diagnostics **15**(18), 2408 (2025)

[9] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241. Springer, Cham (2015)

[10] Targ, S., Almeida, D., Lyman, K.: Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029 (2016)

[11] Zhu, Y., Newsam, S.: Densenet for dense flow. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 790–794. IEEE, ??? (2017)

[12] Koonce, B.: Efficientnet. In: Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization, pp. 109–123. Apress, Berkeley, CA (2021)

[13] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)

[14] Chen, J., Lin, C., Dou, Q., Chen, J., Yang, J., Zhou, Y., Heng, P.-A.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)

[15] Lu, Z., Wang, M., Xue, Y., Tan, H., Wang, G.: Condensed unetr: A lightweight medical image segmenter. In: Sixteenth International Conference on Digital Image Processing (ICDIP 2024), p. 541. SPIE, ??? (2024)

[16] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In:

Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 12077–12090 (2021)

[17] Hirose, A.: Complex-valued neural networks. IEEJ Transactions on Electronics, Information and Systems **131**(1), 2–8 (2011)

[18] Aizenberg, I.: Why we need complex-valued neural networks? In: Complex-Valued Neural Networks with Multi-Valued Neurons, pp. 1–53. Springer, Berlin, Heidelberg (2011)

[19] Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J.F., Krishnan, S., Rajaratnam, B., Bengio, Y.: Deep complex networks. arXiv preprint arXiv:1705.09792 (2017)

[20] Yang, Y., Li, Z., Chen, D., Cong, W., Wang, G.: Deep learning-based mri reconstruction with artificial fourier transform network (aftnet). Computers in Biology and Medicine **192**, 110224 (2025)

[21] Abadal, S., Galván, P., Mármol, A., Mammone, N., Ieracitano, C., Giudice, M.L., Salvini, A., Morabito, F.C.: Graph neural networks for electroencephalogram analysis: Alzheimer's disease and epilepsy use cases. Neural Networks **181**, 106792 (2025)

[22] Sunil, G., Gowtham, S., Bose, A., Harish, S., Srinivasa, G.: Graph neural network and machine learning analysis of functional neuroimaging for understanding schizophrenia. BMC neuroscience **25**(1), 2 (2024)

[23] Liang, G., Chen, Y., Wang, J., Li, Y., Qiu, L.: Enhanced noise resilience in passive tone detection via broad-receptive field complex-valued convolutional neural networks. The Journal of the Acoustical Society of America **155**(6), 3968–3982 (2024)

[24] Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., Liu, F.: Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. Bioengineering **11**(10), 1034 (2024)

[25] Khurana, S., Nene, M., Nene, M.: Quantum machine learning: unraveling a new paradigm in computational intelligence. Quantum **74**(1) (2024)

[26] Chen, S.Y.-C., Tseng, H.-H., Lin, H.-Y., Yoo, S.: Learning to measure quantum neural networks. In: 2025 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp. 1–5 (2025). IEEE

[27] Bhatia, A.S., Kais, S., Alam, M.A.: Federated quanvolutional neural network: a new paradigm for collaborative quantum learning. Quantum Science and Technology **8**(4), 045032 (2023)

[28] Zheng, J., Gao, Q., Lü, Y.: Quantum graph convolutional neural networks. In: 2021 40th Chinese Control Conference (CCC), pp. 6335–6340 (2021). IEEE

[29] Delgado-Granados, L.H., Krogmeier, T.J., Sager-Smith, L.M., Avdic, I., Hu, Z., Sajjan, M., Abbasi, M., Smart, S.E., Narang, P., Kais, S., *et al.*: Quantum algorithms and applications for open quantum systems. Chemical Reviews **125**(4), 1823–1839 (2025)

[30] Ajlouni, N., Özyavaş, A., Takaoğlu, M., Takaoğlu, F., Ajlouni, F.: Medical image diagnosis based on adaptive hybrid quantum cnn. BMC Medical Imaging **23**(1), 126 (2023)

[31] Schuld, M., Sweke, R., Meyer, J.J.: Effect of data encoding on the expressive power of variational quantum-machine-learning models. Physical Review A **103**(3), 032430 (2021)

[32] Sankari, V.R., Umapathy, U., Alasmari, S., Aslam, S.M.: Automated detection of retinopathy of prematurity using quantum machine learning and deep learning techniques. IEEE Access **11**, 94306–94321 (2023)

[33] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 36–46 (2021). Springer

[34] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision, pp. 205–218 (2022). Springer

[35] Chen, A., Yin, H.-L., Chen, Z.-B., Wu, S.: Hybrid quantum-inspired resnet and densenet for pattern recognition. arXiv preprint arXiv:2403.05754 (2024)

[36] Chen, C.-S., Kuo, E.-J.: Quantum adaptive self-attention for quantum transformer models. arXiv preprint arXiv:2504.05336 (2025)