RESEARCH ARTICLE

An Optimized Multi-Organ Cancer Cells Segmentation for Histopathological Images Based on CBAM-Residual U-Net

HASNAIN ALI SHAH AND JAE-MO KANG, (Member, IEEE)
Department of Artificial Intelligence, Kyungpook National University, Daegu 41566, South Korea

Corresponding author: Jae-Mo Kang (jmkang@knu.ac.kr)

ABSTRACT In digital pathology, the accurate segmentation of cell nuclei in histopathology images is essential for medical image analysis. Histopathologists visually evaluate the patterns of cellular architecture and tissue patterns in histopathology image analysis for cancer detection to determine the malignant tissue portions and assess the severity of malignancy. However, manually analyzing scans using a high-resolution microscope requires significant effort and time. A computer-assisted diagnosis system utilizing deep learning (DL) algorithms rapidly, reliably, and automatically segments cell nuclei. However, the existing research studies have limited accuracy, high computational costs, and a lack of robustness and generalizability on diverse datasets. To address these issues, this paper proposes a novel and improved DL architecture based on the U-Net, namely, the *CBAM-Residual U-Net* for improving accuracy, robustness, and generalized segmentation algorithm that can be applied to various staining techniques and tissue structures. The proposed architecture utilizes a ResConv and convolution block attention modules (CBAM). These modules help the proposed architecture learn the image's shallow and deep features. The CBAM module uses an attention mechanism concentrating on essential features such as cell nuclei's shape, texture, and intensity to accurately segment the raw input patterns. The proposed CBAM-Residual U-Net involves fewer trainable parameters, reducing the computational and time cost s compared to state-of-the-art techniques. Extensive experiments and comprehensive evaluations are conducted to demonstrate the performance of the proposed scheme on publicly available datasets: i) Data Science Bowl (DSB) 2018, ii) The GlaS, iii) Triple-Negative Breast Cancer (TNBC). The experimental results show that our proposed model considerably outperforms the state-of-the-art techniques and detects cellular boundaries well, providing fine-grained segmentation results.

INDEX TERMS Digital pathology, deep learning, medical image segmentation, cell nuclei, cancer detection.

I. INTRODUCTION

Pathology is a branch of medicine that treats and prevents various diseases. It is essential in many medical diagnostic applications and is the basis for many medical regimens. Pathologists use biopsies to make predictive and diagnostic decisions based on cell arrangement and shape [1]. Detecting and segmenting cell nuclei is critical for pathologists and pathology researchers. Histopathological image analysis is

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

performed to investigate tissue growth patterns and cell shapes. Pathologists use a microscope to examine tissue slides at different magnifications. They examine the entire tissue sample at modest magnifications ($4\times$, $10\times$) to determine the locations of malignant areas, particularly the architectural patterns of cancer. Slides are then carefully examined at greater magnifications ($40\times$, $60\times$, and $100\times$) to assess their appearance at the cellular level. [2].

Digital pathology (DP), which benefits from enhanced microscope imaging technologies, has become prevalent in clinical applications. The DP also facilitates the remote

study of multispectral whole-slide images (WSIs) containing detailed features from the tissue to cellular levels. Whole-slide histopathology scans are commonly used to determine malignancy grades and diagnoses. One can extract strong features for nuclear morphological characteristics and other factors through nuclear segmentation based on these images. Early diagnosis is crucial for preserving a patient's life, and the accurate examination of histopathology scans is critical for diagnosing a patient [3]. Cell nuclei are separated from cells using either manual or image processing approaches. However, manually segmenting nuclei from cells is laborious and time-consuming. Furthermore, board-certified doctors and pathologists often find it difficult to travel globally, resulting in a time-consuming and continuous effort to assess large numbers of WSIs. Such constraints limit the widespread implementation of effective evaluation and diagnosis in developing nations and regions with a shortage of pathologists and high patient populations.

Integrating artificial intelligence (AI) and DP facilitates precise, rapid, and complete diagnosis, leading to more accurate targeted therapy [4]. As technology changes continuously, it is vital to understand the current status of AI technologies in DP. Therefore, studying AI in DP is necessary for evaluating its implementation and research development. AI, particularly a deep neural network architecture, has recently dominated ML and computer vision studies using a wide range of well-annotated databases. Neural architectures can effectively address fundamental issues such as detection, segmentation, and classification, allowing them to outperform human specialists [5]. Based on the rapid improvement of DL, CAD systems have been intensively explored and implemented for medical imaging tasks. Several datasets and challenges for DP have been released for numerous diagnostic and exploratory research tasks, including detecting malignancies using WSIs, categorizing malignancies as cancerous/noncancerous or dividing them into specific subcategories, segmenting cells/nuclei, and identifying cells. As a result, many CAD approaches based on deep neural architectures have been developed, including completely supervised, weakly supervised, semi-supervised, and unsupervised methods [6], [7]. These approaches can help pathologists diagnose and screen for malignancies in the lungs, lymph nodes, breasts, blood, and other organs. Few pathologists are available to evaluate malignancies, but numerous cases require evaluation, particularly in underdeveloped nations and regions [8].

However, to classify cancer types accurately, extensive and accurate datasets are required to assist in the development of robust and reliable DL algorithms. Therefore, databases and methodologies for studying various tumors are urgently needed. The shapes of nuclei may vary depending on factors such as cellular life span, cell type, and disease severity. The context of a nucleus may contain a variety of cellular morphologies that can affect segmentation, as shown in Fig. 1. Additionally, nuclei may be close to each other, making segmentation challenging. To overcome these

obstacles, this paper aims to develop a CAD system for overlapping/clustered nuclei segmentation in histopathology images. Our work stands out due to the unique integration of the Convolutional Block Attention Module (CBAM) with the Residual U-Net architecture. This combination is novel and has not been explored extensively in the existing literature. The CBAM block, which efficiently combines attention and convolutional mechanisms, is designed to focus on informative features in the input data, thereby improving the model's performance.

Our work uniquely combines the CBAM with the Residual U-Net architecture. This integration has yet to be extensively explored. The CBAM block is innovatively designed to concentrate on valuable features in the input data by effectively merging attention and convolutional mechanisms. This design enhances the performance of the model. Our model has been fine-tuned explicitly for segmenting multi-organ cancer cells in histopathological images. By capitalizing on the power of both the CBAM and Residual U-Net, our model successfully extracts pertinent features, thus enhancing segmentation accuracy. The model has proven superior in terms of Dice scores, sensitivity, and specificity compared to existing models. This significant improvement can refine diagnostic accuracy and improve patient outcomes.

- We propose an automated method called the convolution block attention module (CBAM) residual U-net architecture for histopathological cell nuclei image segmentation. This method combines attention and convolutional information across shallow and deep layers. The CBAM Residual U-Net can minimize segmentation gaps and improve features by combining a learning algorithm with residual connections.
- Our proposed model utilizes the residual convolution (ResConv) module for extracting more comprehensive features. It compensates for the loss of local features during convolution operations and enables the proposed model to learn diverse information from previous inputs by utilizing skip connections. Spatial and channel attention modules suppress redundant details, concentrate on rich features, and enhance the model's resilience to noise.
- We conduct a thorough ablation study to guide the selection and optimization of various data augmentation techniques for the proposed model, through which we propose to use various data augmentation transformations for the proposed model to achieve better results using limited datasets while reducing overfitting.
- The extensive experimentation conducted in this study demonstrates the pivotal role of improved feature selection and optimization in achieving the proposed model's superior performance, surpassing the previous state-of-the-art models.

This paper is divided into five sections. Section I overview nuclear segmentation and its importance for cancer detection. Section II discusses the existing literature. Section III describes the proposed method and Section IV explains

our detailed experimentation and ablation analysis, and Section VI concludes the paper with future directions.

II. LITERATURE REVIEW

This section reviews related works that have used traditional image processing, ML, and DL methods to identify and segment abnormal cell nuclei in histopathology scans. Wang et al. [9] employed an image preprocessing method to identify and segment cell nuclei. To find regions of interest (ROIs), they introduced a wavelet transform and multi-scale region growth. Additionally, an adaptable morphological operation combined with the curvature-scale-space corner detection technique was used to separate overlapping cells. This approach yielded excellent and consistent outcomes in terms of specificity and sensitivity. Feng et al. [10] developed a DL network for colon pathological WSI evaluation, which includes tissue identification and lesion segmentation. It includes an enhanced U-Net with a VGG net as the backbone and two strategies for inference and learning, respectively. Researchers specifically suggested a class-wise Dice loss function to balance the proportions of benign and cancerous samples. The experiments showed that their approach produced excellent results on the benchmark dataset.

Albusayli et al. [11] suggests utilizing digital scores of stroma cancer lymphocytes and cancer stroma to predict disease-specific lifespan in TNBC patients. The researchers developed a deep learning system on a sizable dataset of labeled breast cancer tissue samples to locate and measure disease. They subsequently predicted disease-specific survival in TNBC patients using the computerized ratings that were produced. The findings show that in TNBC patients, the AI-based digital scores significantly indicate disease-specific survival. Makem et al. [12] proposed an arithmetic operational approach led by a control parameter to detect and extract white blood cell nuclei from blood smear images. Images with a resolution greater than 300×300 were first cropped to recover white blood cell nuclei. The resulting color features were fed into a Fourier transform, followed by an average circular filter and mean-shift clustering operation. This approach was evaluated on five distinct image datasets, yielding Dice coefficient (DC) scores ranging from 90 to 95%. Sharma et al. [13] suggests a novel technique for segmenting colon polyps in medical images using a scalable deep-learning framework. The research emphasizes the significance of early colorectal polyp diagnosis, which can aid in preventing colorectal cancer. The proposed Li-SegPNet performs well in segmenting the polyps after being trained using the Kvasir-SEG and CVC-ClinicDB datasets of annotated clinical data.

Tomar et al. [14] proposed a novel FAnet architecture based on a feedback attention mechanism. When utilized with current feature maps to strengthen feature representations, the feedback mechanism included in the FANet architecture enhanced intricate operational attention. Zidan et al. [15] proposed the SwinCup model, which consists of two stages of segmentation: the first stage uses a Swin Transformer

encoder-decoder architecture to perform coarse segmentation of the entire image, followed by refinement of the segmentation results using a second Swin Transformer network with attention mechanisms. They also provide a unique function, the ''hard negative mining loss,'' intended to alleviate class imbalance and increase the model's capacity to recognize tiny objects. The suggested technique delivers cutting-edge results in accuracy, speed, and efficiency. Xu et al. [16] proposed a novel convolutional neural network (CNN)-based structure called DCSAU-Net for various medical image segmentation applications. This framework uses multi-resolution mixed features and a broad receptive field in its CSA and DC layers. DCSAU-Net can retrieve optimal information from input images. Experiments demonstrated that DCSAU-Net outperformed other SOTA networks regarding F1 score and mean intersection over union (mIoU). Liu et al. [17] proposed an augmenting strategy for transferring microscopy nucleus images from one modality to another using a multi-modality style-transfer generative adversarial network. This augmentation technique supports training a masked region convolution neural network (R-CNN) for nucleus segmentation by increasing training image variety, thereby making the trained model more resilient to testing data and leading to higher accuracy.

Jian and Kamata [18] proposed a two-step DL architecture for cell-nucleation segmentation. In the first phase, their architecture performed coarse segmentation. The second phase then performed sophisticated segmentation. They also included a unique weighted loss function to improve the resistance of their model to blurred borders, ambient noise, and staining strength inhomogeneity. Dahamija et al. [19] presented complete DL architectures for segmenting different medical images, including cell nuclei, brain tumors, lungs, and skin lesions. Their model can assist experts and accelerate the medical diagnosis procedure. Additionally, they combined a transformer encoder with a fully connected neural encoder. The results demonstrated significant improvement compared to previous methods. Alom et al. [20] proposed the implementation of a recurrent residual U-net (R2UNet) model for nuclear segmentation. They validated the R2UNet model using the DSB 210 Grand Challenge dataset. They obtained an impressive DC of nearly 92.15%. Additionally, experimental findings revealed good qualitative ability for multiple modalities of medical image segmentation (retinal blood vessels, skin cancer, and lung segmentation).

Priyal et al. [21] developed a DL framework (modified UNet) with fewer trainable parameters to reduce computing costs and training time. They analyzed the performance of the modified U-Net and SegNet architectures. They reported that the U-Net model outperformed SegNet while using fewer parameters. Their study proved that a modified UNet can perform excellently on medical imaging applications. Jha et al. [22] proposed a novel architecture consisting of an encoder-decoder U-net. The proposed double U-net utilized VGG19 as a backbone network for the encoder, squeeze excitation blocks, and atrous spatial
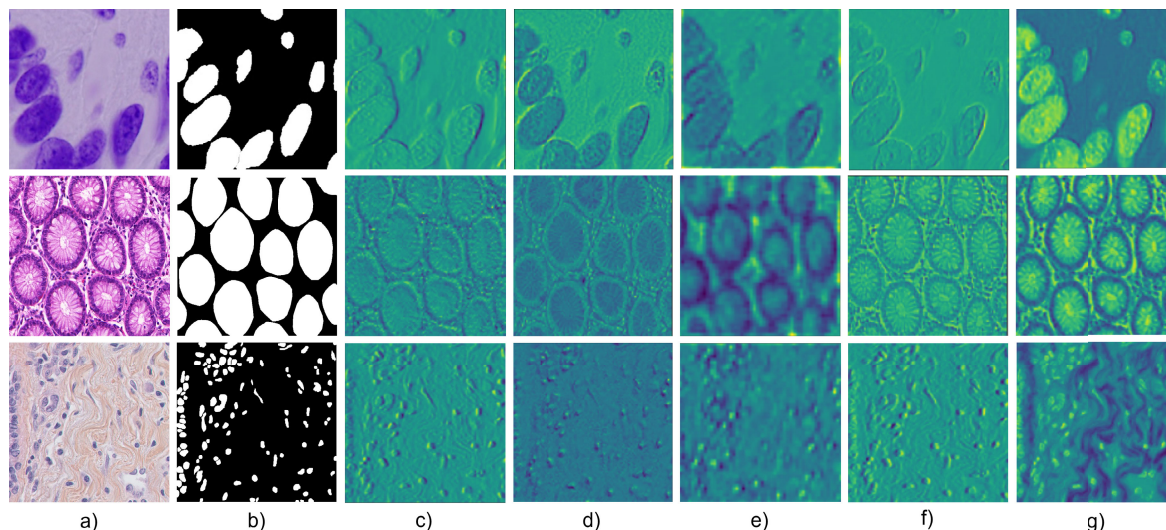
FIGURE 1. Illustration of activation maps received from proposed CBAM Residual U-Net model using various experimental combinations of several datasets organized as (a) Original image of tissues, (b) Ground truth masked images Here, from (c) to (g) denote the model feature extraction capability as the convolutional layers get deeper.

pyramid pooling. The authors evaluated their model using different medical imaging datasets and achieved SOTA performance. Chanchal et al. [23] proposed a convolution neural architecture called the deep structured residual encoder-decoder network that addressed fundamental issues in automated nucleus segmentation. The main challenge is identifying nuclei in histopathology images with a broad spectrum and many defects. This challenge was solved by developing an encoder-decoder architecture with additional discriminative pathways capable of retrieving relevant and concise textural information.

Previous research shows that significant efforts have been made to develop different CNN-based architectures for naturalistic and pathological imaging segmentation. However, the CNNs usually suffer from losing essential features, due to which residual blocks are frequently used to address the issue. Recently, additional studies have been put into creating generalized algorithms, resulting in most experts evaluating the techniques using multiple databases [12], [19], [21]. The availability of validation data constitutes one of the significant issues in the healthcare field. Table 1 shows the overall results of the previous results. There are several complex appearances in medical visualizations that are frequently overlooked throughout the inspection and, therefore, can progress to malignancy, provided the initial diagnosis is not carried out immediately. Thus, a precise segmentation of medical images method is required to handle such complex data. Therefore, in the next section, we propose the CBAM Residual U-Net architecture to meet this demand, which generates effective, accurate segmentation masks from complex images.

III. PROPOSED MODEL: CBAM RESIDUAL U-NET
This paper presents an improved segmentation model called the CBAM residual U-Net, which extends the capabilities of

the base U-Net model initially developed in [24] in medical image segmentation. The encoder and decoder are the two basic components of the U-Net. The encoder is composed of convolution layers that are maximally pooled to extract the characteristics of an image. The decoder uses transposed convolution layers with upsampling layers to enable localization. In our proposed architecture, in addition to utilizing the regular convolution blocks proposed in the original U-Net, we incorporate a variety of customized ResConv blocks based on residual connections, improved channel attention, and spatial attention modules. In this section, we first outline the architecture of the proposed CBAM residual U-Net by presenting a schematic of its general structure. We then introduce the construction processes and functions of each attention module. Finally, the loss function for training our proposed model is presented.

A. ARCHITECTURE OF THE PROPOSED CBAM RESIDUAL U-NET
In the U-Net, spatial and regional information loss during convolution operations is unavoidable. We propose the CBAM residual U-Net to address this drawback, which utilizes a unique CNN encoder and decoder model for cell nucleus segmentation. Our proposed model uses detailed spatial context information for image segmentation and provides more accurate results than baseline networks. Fig. 2 presents the entire architecture of the CBAM residual U-Net. In the CBAM residual U-Net, the twin convolutional block in the original U-Net is replaced by two fully connected ResConv 3×3 modules such that the convolution kernel can lose the loss of valid information can be compensated by obtaining deeper semantic information. An efficient CBAM is inserted after each fully connected ResConv block. In other words, the ResConv blocks and relevant CBAM modules appear in pairs. The decoder is similar to the encoder.

TABLE 1. Detailed analyses of the proposed and state-of-the-art studies.

Reference	Model	Advancements	Limiting factors
[10]	U-Net (VGG-16)	- Designed class-wise loss function	- Cannot generalize well to other datasets or imaging modalities.
[11]	ResNet	- Improved training and inference time - innovatively combines deep CNNs and graph theory to create digital sTME markers for predicting TNBC outcomes	- More diverse datasets is needed. More data and validation needed to confirm findings
[12]	Fourier Transform, K-means	- Strong generalization ability through prioritizing reliable and informative samples - Exceptional performance achieved with multiple datasets	- Postprocessing overhead and requires many learn-able parameters - Complex method
[13]	Li-SegPNet	- Skip connections help in training deeper networks - Big impact augmentation process	- Relies on Large Dataset - Structure Complexity
[14]	FANet	- Efficient for biomedical image segmentation - Good at capturing context and localization	- Struggles with variable lighting conditions - Requires large amounts of data for training
[15]	SwinCup Transformer	- Good for jointly segmented nuclei - Utilizes hierarchical representations	- Complex architecture - Has a long runtime
[16]	Deep CSA U-Net	- CSA module captures detailed information - performs better on the multi-class segmentation	- relies on large data
[18]	Style GAN, Mask R-CNN	- High accuracy on nuclear segmentation with different modalities - Good performance for the prediction	- Computationally expensive due to two stages networks - Long inference time
[19]	USegTransformer-P	- Achieved better performance - Low cost	- preprocessing required - Has a long runtime
Ours	CBAM Residual U-Net	- Improved Segmentation than prior methods - Robust to noise and better generalization capability	- Extensive training is required

The arrow connecting the decoder and encoder represents a skip connection, and a CBAM block is inserted after each skip connection layer. These CBAM blocks are utilized to enhance the effective extraction of optimized image features while suppressing erroneous image features through convolution. The deep features of input cell nuclei images are retrieved using the CBAM residual U-Net encoder, and the decoder restores enhanced characteristics. Finally, all nuclei are segmented and a targeted structure-based image is generated.

B. CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)
The CBAM comprises two consecutive submodules, the channel attention module (CAM) and spatial attention module (SAM), used in a specific order. A CBAM is used in every convolutional block in a deep network to generate successively refined feature maps from initial input feature maps. The CBAM is utilized to enhance the main properties of the spatial and channel dimensions. Such an attention module can improve the expressiveness of a network by focusing on essential aspects while suppressing irrelevant features. As shown in Fig. 3 a CBAM combines CAM and SAM. The attention is performed separately for the channels and spatial dimensions and then combined with input features to achieve adaptive feature improvement. The CBAM involves one-dimensional channel attention mapping M_c of size $C \times 1 \times 1$ and a two-dimensional spatial attention

mapping M_s of size $1 \times H \times W$ as follows:

$$F' = M_c(F) \otimes F,$$
$$F'' = M_s(F') \otimes F'. \tag{1}$$

Here, $\otimes$ represents an element-wise multiplication, $\mathbf{F}'$ represents the one-dimensional channel attention map developed by the CAM module. and $\mathbf{F}''$ is the ultimate refined output produced by SAM. Through the experiments, we found that a sequential order produces better results than a parallel configuration. Also, the experimental results indicate that the channel-first-order is significantly better than a spatial design for the sequential process arrangement.

C. CHANNEL ATTENTION MODULE (CAM)
The CAM initially uses average and maximum pooling methods to combine spatial information from feature components before generating two potential contextual descriptive features $F_{avg}^s \in R^{1 \times H \times W}$ and $F_{max}^s \in R^{1 \times H \times W}$, which are calculated from the average pooling layer and obtained using the max pooling layer, respectively. Then, by using an activation function, these two defining features are sent into a multilayer perceptron (MLP) and added element by element to generate a channel attention map:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)))$$
$$= \sigma\left(W_1\left(W_0\left(F_{\text{Avg}}^c\right)\right) + W_1\left(W_0\left(F_{\text{Max}}^c\right)\right)\right). \tag{2}$$
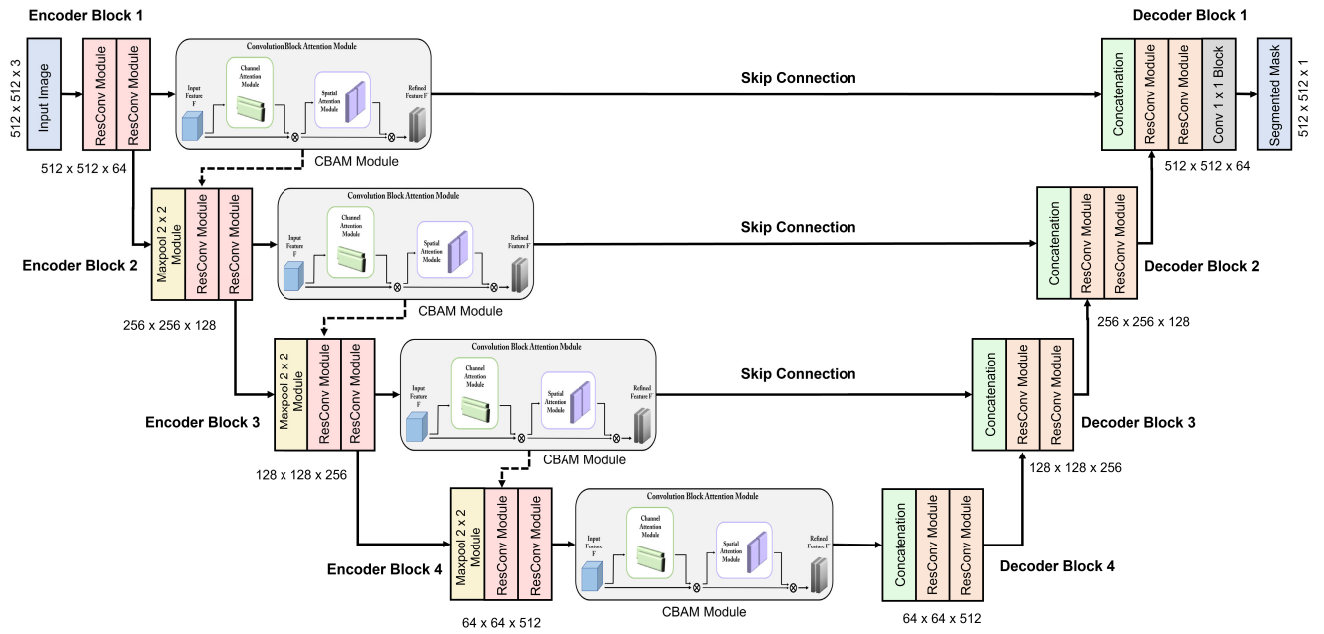
FIGURE 2. Block diagram of the CBAM Residual U-Net architecture. The ResConv modules are followed by CBAM Module and Max pool (2 × 2) block in the Encoder. The characteristics retrieved from these are concatenated with Up-sampling ResConv Module in the Decoder.
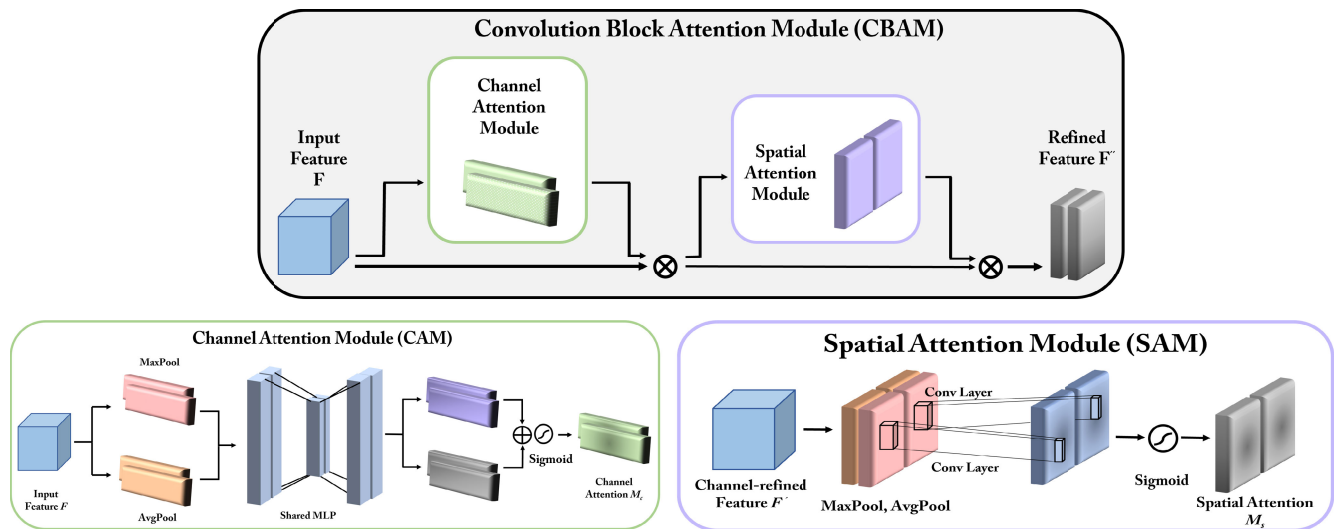

FIGURE 3. Architecture of CBAM Module utilized in the proposed CBAM-Residual-UNet architecture. The input features are passed to (CAM) module for amplifying the most relevant channels through 1 × 1 Maxpool, the Average pool with shared MLPs. The refined features are passed to (SAM) module to highlight the most informative regions of the feature maps using Conv 7 × 7.

Here, F_{avg}^c and F_{max}^c represent the average- and max-pooled characteristics, respectively. Subsequently, a common layer receives these characteristics as inputs and generates an M_c-channel attention map. Shared MLP systems with hidden units form a unified connection. The hidden activation size is fixed at $R/C = r \times 1 \times 1$, where r is the reduction ratio, which decreases parameterization complexity. Following the application of the shared network based on every descriptor, all of the extracted features are combined using an element-wise sum. Also, $W_0 \in R^{C/r \times C}$, $W_1 \in R^{C \times C/r}$ are the weights of the MLP. The two inputs F_{avg}^c and F_{max}^c are shared based on these two weights. $M_c(F)$ is multiplied by (F) in element-by-element fashion to generate (F'). Finally, the channel

attention map is obtained through the sigmoid activation function..

D. SPATIAL ATTENTION MODULE (SAM)

SAM mainly concentrates on the information components that synchronize with channel attention. The input characteristics of the spatial attention module are processed simultaneously through max and average pooling. In the SAM, the channel dimensions are pooled, and two descriptive characteristics are acquired after pooling. $F_{max}^s \in R^{1 \times H \times W}$ is obtained from the max pooling layer and $F_{avg}^s \in R^{1 \times H \times W}$ is obtained from the average pooling layer. These two

descriptive characteristics are patched together based on the channel dimension. The two feature mappings are patched simultaneously and a two-dimensional spatial attention-weighted map $M_s(F')$ is produced by applying the sigmoid activation function and $f^{7 \times 7}$ convolution layer, as follows:

$$
\begin{aligned}
M_s\left(F'\right) &= \sigma\left(f^{7 \times 7}\left(\left[\text{AvgPool}\left(F'\right); \text{MaxPool}\left(F'\right)\right]\right)\right) \\
&= \sigma\left(f^{7 \times 7}\left(\left[F_{\text{avg}}^s\, F_{\text{max}}^s\right]\right)\right).
\end{aligned}
\tag{3}
$$

Here, $f^{7 \times 7}$ refers to the convolution process using a 7×7 convolution filter and $\sigma(x) = \frac{1}{1+e^{-x}}$ denotes the sigmoid activation function. After feature screening, a final high-level refined feature map F'' is produced at the end by combining the channel-attention-extracted features $M_c(F')$ and spatial attention weighted map $M_s(F')$.

E. ResConv MODULE

Instead of using traditional double convolutional layers, we introduce a different approach for completely connected residual blocks called ResConv. Fig. 4 presents the architecture of the ResConv module. In ResConv, features are obtained through a sequence of convolution procedures, each accompanied by rectified linear unit activation and batch normalization. The results are summed sequentially according to the weights w_1, w_2, and w_3 to achieve a superior level of integration between shallow and deep traits while retaining intermediate layer characteristics. In this case, w_1, w_2, and w_3 are all adjusted to one, implying that the feature maps include both deep and shallow feature knowledge and intermediate details. These feature weights have the same ratio and are equally important because they compensate for the loss of semantic information and aid in retrieving more extensive and adaptable features. Consequently, the ResConv module can be considered as three separate convolution layers, resulting in three different receptive domain regions, which boosts the flexibility and generalization potential of the network. The proposed ResConv module adds the information retrieved by many convolutional filters rather than concatenating it, which can significantly reduce memory usage. Additionally, ResConv is relatively simple to implement.

F. LOSS FUNCTION

The effectiveness of a DL model is highly dependent on its loss function. The binary cross-entropy (BCE) loss function is a frequently used loss function in classification, detection, and segmentation applications, of which the fundamental benefit is that it produces smooth loss curves, leading to fast model training [25]. The BCE loss is given by

$$
BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log\left(\hat{y}_i\right) + (1-y_i) \cdot \log\left(1-\hat{y}_i\right).
\tag{4}
$$

Here, segmented output and true images are represented by y_i and $\hat{y}_i$, respectively, where N denotes the number of training sets for the ith slice ($I \in N$). Dice loss is a loss function
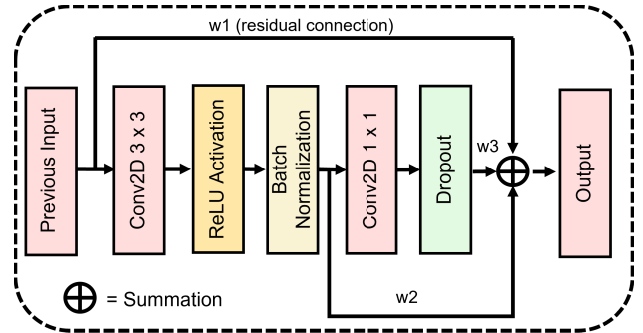

FIGURE 4. ResConv module architecture where w_1, w_2, and w_3 are three different residual skip connections followed by ReLU activation and batch normalization layers to include both shallow and deep characteristics while retaining intermediate layer features.

that allows an algorithm to obtain results closer to the ground truth. It is gradually gaining popularity based on its intrinsic optimization of the DC and positive effect on class imbalance. The Dice loss is computed mathematically as [26]

$$
\text{Dice Loss} = 1 - \sum_{i=0}^{i=N} \frac{2 * \left|y_i \cap \hat{y}_i\right|}{|y_i| + |\hat{y}_i|}.
\tag{5}
$$

Note that the Dice and BCE losses can be used to optimize the DC and manage imbalanced data, respectively. The proposed CBAM Residual U-Net is trained to minimize both the BCE loss in (4) and the Dice loss in (5), considering the mutual connection between these two loss functions. The combined loss (CL) function is given by:

$$
\begin{aligned}
CL = {}&-\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log\left(\hat{y}i\right) + (1-yi) \cdot \log\left(1-\hat{y}i\right) \\
&+ 1 - \sum_{i=0}^{i=N} \frac{2 * \left|y_i \cap \hat{y}_i\right|}{|y_i| + |\hat{y}_i|}.
\end{aligned}
\tag{6}
$$

Here, N is the number of training sets, and y_i and $\hat{y}_i$ represent the actual image and segmentation image predicted for the ith slice ($I \in N$), respectively.

IV. RESULTS AND ANALYSIS

This section presents the training and validation results for the proposed CBAM residual U-Net trained on benchmark DSB 2018, Glas, and TNBC datasets images taken from Kaggle. Several data augmentation strategies are used to increase the size of the training data. We employ a range of hyperparameter settings to optimize the proposed algorithm for improved learning. Fig. 5 shows the training and validation process of the proposed CBAM Residual U-Net for semantic segmentation. Fig. 6 shows the loss curves graph for the proposed CBAM residual U-Net model on DSB 2018, GlaS, and TNBC datasets. The graph reveals that as the number of epochs increases, the loss decreases steadily over a shorter period using the optimized hyperparameters until it reaches a point of stability. For loss curves, one can see that the proposed model converges within 25 to 35 epochs,

TABLE 2. Layer wise details of the proposed CBAM-Residual U-Net.

| Block	Name/Size	Filters	Dimensions	Parameters	Activation Function	Operation Type
Encoder Block 1	ResConv Block	64	$512 \times 512 \times 64$	3,840,51	ReLU	Convolution
	ResConv Block	64	$512 \times 512 \times 64$	36,864	ReLU	Convolution
	CBAM Block	-	$512 \times 512 \times 64$	-	-	Attention
	Maxpool 2x2	-	$256 \times 256 \times 64$	-	-	Pooling
Encoder Block 2	ResConv Block	128	$256 \times 256 \times 128$	73,728	ReLU	Convolution
	ResConv Block	128	$256 \times 256 \times 128$	147,456	ReLU	Convolution
	CBAM Block	-	$256 \times 256 \times 128$	-	-	Attention
	Maxpool 2x2	-	$128 \times 128 \times 128$	-	-	Pooling
Encoder Block 3	ResConv Block	256	$128 \times 128 \times 256$	294,912	ReLU	Convolution
	ResConv Block	256	$128 \times 128 \times 256$	589,824	ReLU	Convolution
	CBAM Block	-	$128 \times 128 \times 256$	-	-	Attention
	Maxpool 2x2	-	$64 \times 64 \times 256$	-	-	Pooling
Encoder Block 4	ResConv Block	512	$64 \times 64 \times 512$	1,179,648	ReLU	Convolution
	ResConv Block	512	$64 \times 64 \times 512$	2,359,296	ReLU	Convolution
	CBAM Block	-	$64 \times 64 \times 512$	-	-	Attention
Decoder Block 4	ResConv Block	512	$64 \times 64 \times 512$	2,359,296	ReLU	Convolution
	ResConv Block	512	$64 \times 64 \times 512$	2,359,296	ReLU	Convolution
	Skip Connection	-	$64 \times 64 \times 512$	-	-	Skip Connection
	UpSample 2x2	-	$128 \times 128 \times 512$	-	-	Pooling
Decoder Block 3	ResConv Block	256	$128 \times 128 \times 256$	1,179,648	ReLU	Convolution
	ResConv Block	256	$128 \times 128 \times 256$	1,179,648	ReLU	Convolution
	Skip Connection	-	$128 \times 128 \times 256$	-	-	Skip Connection
	UpSample 2x2	-	$256 \times 256 \times 256$	-	-	Pooling
Decoder Block 2	ResConv Block	128	$256 \times 256 \times 128$	295,040	ReLU	Convolution
	ResConv Block	128	$256 \times 256 \times 128$	147,584	ReLU	Convolution
	Skip Connection	-	$256 \times 256 \times 128$	-	-	Skip Connection
	UpSample 2x2	-	$512 \times 512 \times 128$	-	-	Pooling
Decoder Block 1	ResConv Block	64	$512 \times 512 \times 64$	73,792	ReLU	Convolution
	ResConv Block	64	$512 \times 512 \times 64$	36,928	ReLU	Convolution
	Skip Connection	-	$512 \times 512 \times 64$	-	-	Skip Connection
Output	Conv 1×1	1	$512 \times 512 \times 1$	3,328	Sigmoid	Convolution
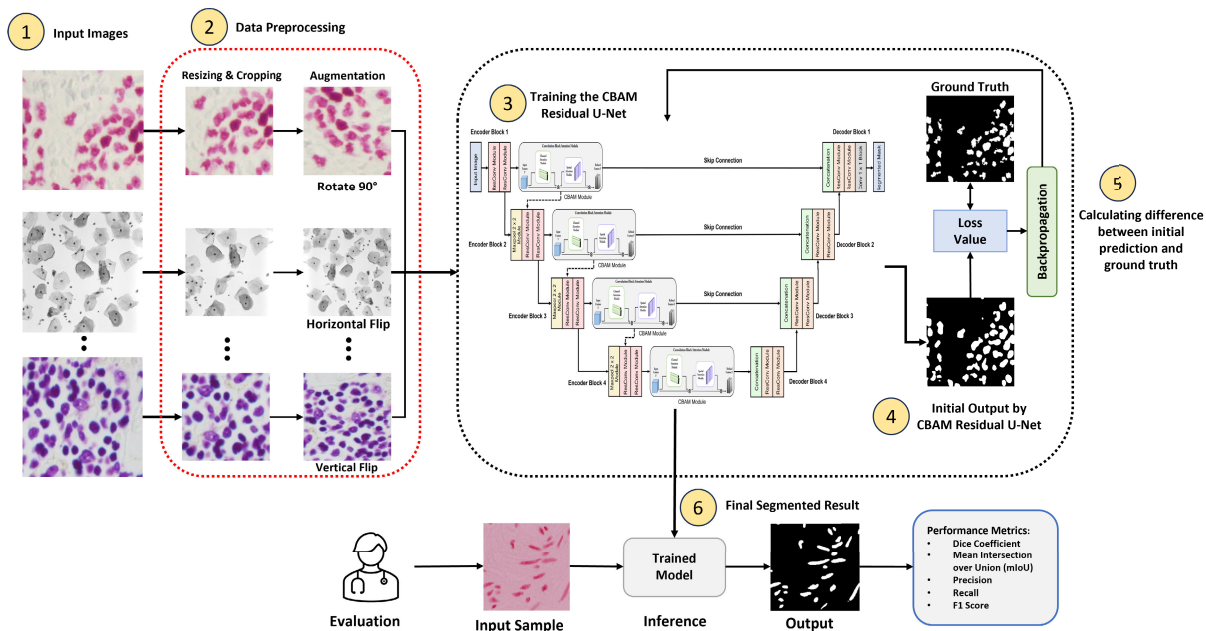
FIGURE 5. Flowchart depicting the data preprocessing and training procedures. In the first stage, the input data is preprocessed and augmented. Following that, data is processed by the proposed CBAM Residual U-Net for semantic segmentation.

showing that the training is performed successfully and very quickly.

A. DATASETS DESCRIPTION

1) DATA SCIENCE BOWL 2018

This dataset publicly available on Kaggle [27], in which the data was obtained under diverse circumstances and varied in terms of cell type, staining type, cell brightness, cell magnification, and cell imaging modality (fluorescence versus bright field). A combination of 841 images with diverse biological and experimental heterogeneities from 30 biological experiments was provided by multiple laboratories. Each image was assigned a different number, and the images and masks for each nucleus in the images were

saved in a specific folder based on the image numbers. Images and masks of various sizes were scaled to the same size of 256 × 256 pixels.

2) GlaS

The GlaS colon dataset contains digital histopathology images of stained colon tissue samples with hematoxylin and eosin. Researchers at the University of Warwick developed the dataset and is openly available for use in research [28]. The dataset includes 165 scans, each accompanied by annotations of specific glands inside the colon tissue. Expert pathologists generated gland annotations, which provide a foundation for training and assessing algorithms for automated identification and categorization of colon tissue anomalies such as colorectal cancer.

3) TROPLE NEGATIVE BREAST CANCER

The TNBC cell histopathological dataset is a publicly accessible collection of digitized histopathology imaging of breast cancer tissue samples [29]. The dataset is produced in order to help in the development of algorithms for the automated identification and categorization of breast cancer subtypes. TNBC includes 81 digitized images of hematoxylin and eosin-stained breast cancer tissue samples. These samples came from individuals with triple-negative breast cancer, a form of breast cancer that lacks the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2. Each image in the dataset is labeled with pixel-level annotations of the breast cancer cells.

B. HYPERPARAMETER SETTINGS

Now, we elaborate on the hyperparameter settings selected for the target task to obtain optimal results. We exploit Adam as an optimizer for training to achieve the maximum loss reduction [30]. This optimization strategy employs an adjustable gradient descent algorithm to adjust weights as close to their local minima. We particularly select Adam rather than other optimization approaches such as stochastic gradient descent [31] or RMSProp [32] thanks to its easy implementation, faster learning experience, and efficient memory consumption. Adam has recently achieved outstanding performance in various DL applications, especially for the segmentation of medical images. The settings of the hyperparameters with a low learning rate (LR) that were modified to operate alongside the other hyperparameters are listed in Table 3. The Adam optimization method is utilized with an initial LR of $10e^{-3}$ because it is relatively stable for hyperparameter optimization to tolerate sparse gradients for a complex problem such as a nucleus or cell segmentation. The LR concerning validation loss is reduced to a minimum learning rate of $10e-7$ with a batch size of eight and a combined loss function using the ReduceLROnPlateau callback. Our U-Net architecture is trained using the Keras API with the TensorFlow backend. 90% of the dataset is used to train the proposed model, and 10% is used for evaluation.

TABLE 3. Hyperparameter settings for training process.

No.	Hyperparameters	Settings
1	Optimizer	Adam
2	Callbacks	ReduceLROnPlateau
3	Initial LR	$10e^{-3}$
4	Reduced LR	$10e^{-7}$
5	Batch size	8
6	Epochs	100
7	Loss function	Combined Loss

TABLE 4. System specifications used for our experiments.

No.	System	Specification
1	CPU	Intel Core i5- 11400 CPU
2	System type	Windows 10, 64 bit OS
3	RAM	16 GB
4	GPU	Nvidia RTX A5000
5	Library	TensorFlow
6	Development tool	Python 3.7

C. IMPLEMENTATION SETUP

The proposed CBAM residual U-Net is implemented using Python with the Keras and TensorFlow frameworks. The DSB 2018, Glas, and TNBC datasets validate the proposed model. The entire model was trained on a server with the specifications listed in Table 4. To validate our model, we compare its performance to that of the existing U-Net models. The validation processes adopted for the specific dataset utilized in this study will be discussed in detail later.

D. EVALUATION METRICS

Nucleus segmentation algorithms must be evaluated using criteria that penalize both pixel-level and object-level inaccuracy. Therefore, we employ two distinct assessment measures. First, the F1 measure is used for object-level evaluations. The Dice Coefficient (DC) are also employed for pixel-level assessment. The harmonic average of precision and recall defines the F1 measure. If ground-truth images are represented by X_i and segmented objects are represented by Y_j, then the precision, recall, and F1 measure, are assessed using false positives (FP), false negatives (FN), and true positives (TP). TP is the total number of ground-truth objects X_i that have been accurately segmented among Y_j. FP is the number of wrongly segmented images Y_j that are not ground-truth objects X_i. The count of erroneously un-segmented items Y_j that are ground-truth objects X_i defines FN. The F1 measure, precision, recall, and Dice Coefficient are given by Equations (7) to (10).

$$F1 - \text{Score} = \frac{2TP}{2TP + FP + FN}, \quad (7)$$

$$Precision = \frac{TP}{TP + FP}, \quad (8)$$

$$Recall = \frac{TP}{TP + FN}. \quad (9)$$

The F1 measure is less effective and sufficient to account for pixel level. The DC assesses segmentation quality at the pixel level. If the pixels of a ground-truth nucleus are represented by X_i and the pixels of its associated segmented

nucleus are represented by Y_j, then the DC is defined as:

$$DC = 2 \cdot \frac{|X_i \cap Y_j|}{|X_i| + |Y_j|}. \tag{10}$$

The predicted nucleus X_i maximizes the Dice score index of the ground-truth nucleus Y_j. The remaining cells are a collection of X_i with no matching. DC is the proportion of segmented results among the confined area of the matched elements. Each segmentation error (both under and over-segmentation) reduces the DC.

E. EXPIREMENTAL RESULTS

Now, to demonstrate the superiority of our proposed CBAM Residual U-Net algorithm for nuclei segmentation, we compare the performance of the CBAM residual U-Net to the other SOTA algorithms, including U-Net (several backbones), DeepLabv3, and Attention U-Net, and a pixel-wise segmentation network (SegNet). For a fair comparison, the training and validation data and assessment standards are set to be the same for all compared schemes.

1) RESULTS ON DSB 2018 DATASET

Table 8 presents the results of experimental comparisons on the DSB 2018 dataset. We investigate the performance of a U-Net architecture using various neural networks as a backbone. We report all of the outcomes. The proposed technique's F1 score, DC, precision, and recall are 0.963, 0.945, 0.966, and 0.953, respectively. The proposed model has greater recall and F1 scores than the other models, demonstrating that our approach is clinically advantageous. Also, the proposed model improves the F1 score by 4.13%, DC by 5.35%, recall by 4.72%, and precision by 4.83% compared to the conventional U-Net. Additionally, our approach provides more balanced performance for the relevant parameters, indicating outstanding capabilities across the board for nucleus segmentation.

2) RESULTS ON GlaS DATASET

The GlaS colon dataset intends to create an autonomous AI system for segmenting colon structures to overcome challenges brought on by manual labeling. The quantitative findings from the GlaS dataset are displayed in Table 8. The F1, Dice, precision, and recall values for the proposed CBAM Residual U-Net are 0.82, 0.77, 0.86, and 0.85. The results show that the proposed network generates SOTA outcomes outperform other more current approaches in terms of the metrics and other methods utilized for evaluation.

3) RESULTS ON TNBC DATASET

The segmentation of TNBC cell images can be an important step in analyzing the histopathological features of the cancer cells and aiding in diagnosis and treatment planning. The quantitative findings from the GlaS dataset are shown in Table 8. We have used the K-fold method for these experiments to train and test every image available for
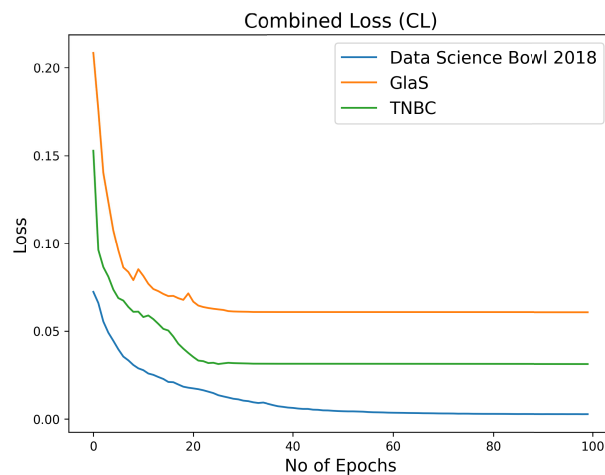

FIGURE 6. Training loss curves of the proposed CBAM Residual U-Net with DSB 2018, GlaS and TNBC dataset.

training and evaluation. The proposed model's F1, Dice, and recall values are 0.90, 0.86, 0.91, and 0.90. The results demonstrate that the proposed network yields state-of-the-art outcomes, surpassing other contemporary approaches in the metrics adopted by those methods for assessment.

Fig. 7 shows the qualitative segmenting performance of our proposed architecture with heat maps. We present findings on cell segmentation in Table 7 - 8 to illustrate the efficiency of our suggested architecture on cell nuclei and colon histology semantic segmentation. CBAM Residual U-Net competed favorably in testing against various CNN and hybrid models. Moreover, we tested the effectiveness of our method on the DSB 2018, GlaS, and TNBC datasets, showing generally superior performance compared to previous approaches and highlighting our method's significant generalizability across many datasets. The channel and spatial attention mechanism loops over the image to capture the image's long-range patterns, engaging with several windows at each level. This method helps attention-based networks, which need fewer layers to achieve a global frame of reference. The CBAM block can help selectively emphasize the most instructive characteristics in the feature maps derived from the encoder in the Residual U-Net framework for cancer cell segmentation. It is especially helpful for locating small, delicate structures of significance, such as cancerous cells, which can be hard to distinguish from surrounding tissues.

V. ABLATION STUDY

A. PERFORMANCE ANALYSIS

We investigated the function and effects of several components inside our CBAM Residual U-Net model. These elements included the CBAM, residual connections, and skip connections. The CBAM module, known for its ability to focus on significant features, showed a critical role in enhancing the performance of our model. Its elimination potentially decreases the model's accuracy due to the inability

TABLE 5. Total Number of images produced in data augmentation.

Dataset	Original Images	Random Cropping	Horizontal Flipping	Vertical Flipping	Random Rotation	Total Images
DSB	841	841	1682	3364	10092	16820
GlaS	165	330	495	990	495	5445
TNBC	81	162	243	486	1458	3888
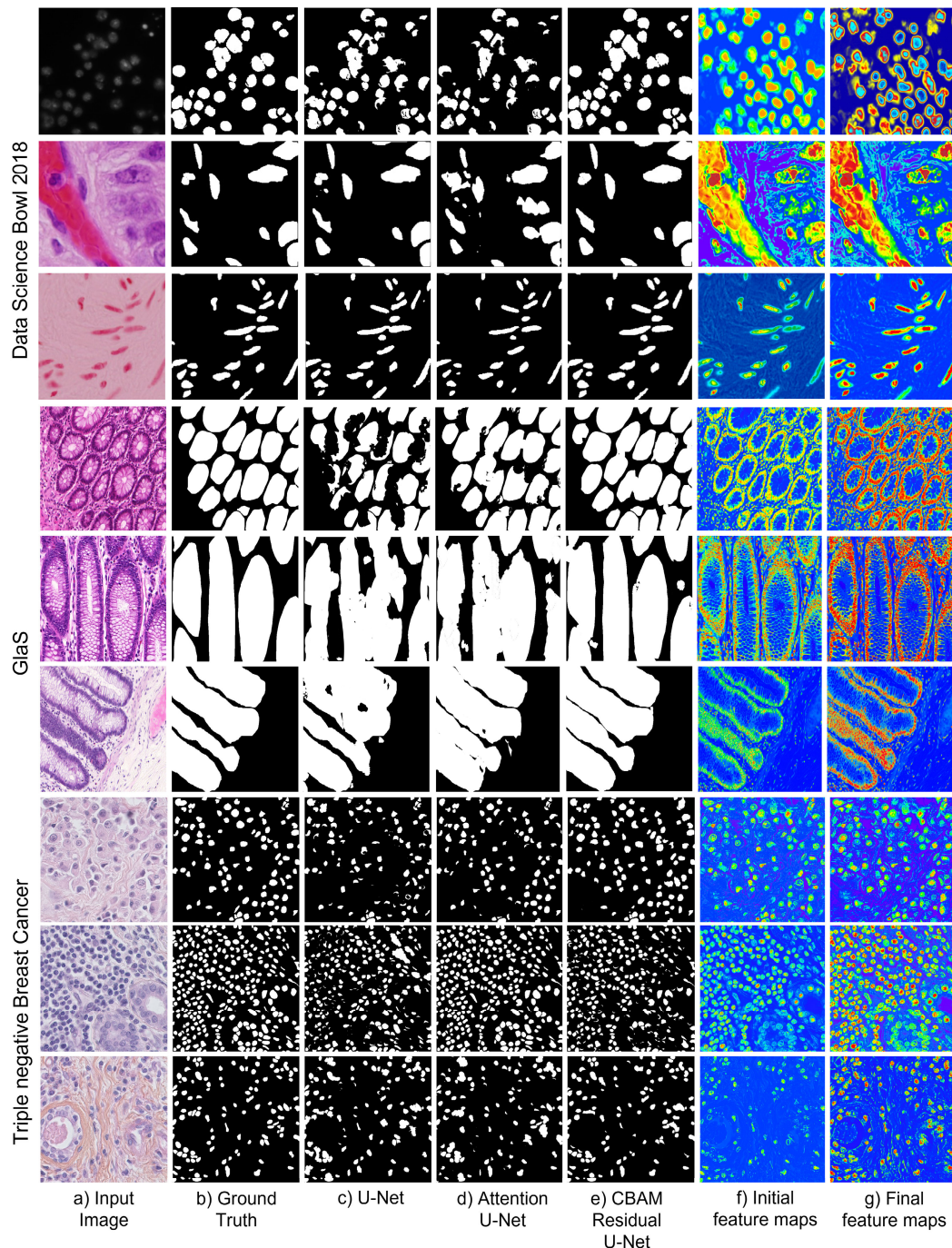
FIGURE 7. Qualitative outcomes of the various approaches evaluated on the DSB 2018, GlaS, and TNBC datasets. CNN networks and attention-based U-Net networks exhibit similar distinctions. Homogeneous forms can be captured quickly, while more complicated shapes need a global understanding of the limits. a) Input Image b) Ground Truth masked image c) Standard U-Net output d) Attention U-Net e) Our proposed CBAM Residual U-Net segmentation results, while f) shows initial feature maps and g) shows the final feature maps of the proposed architecture.

to adequately highlight crucial areas within the images, which are pivotal for cancerous nuclei segmentation. Moreover, the absence of CBAM makes the model more susceptible to noise, reducing its performance, as shown in Table 6.

TABLE 6. Performance analysis of the proposed technique with and without CBAM and ResConv on the DSB 2018, GlaS, and TNBC dataset.

Methods	DSB 2018 Dataset				GlaS Dataset				TNBC Dataset			
	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall
U-Net (Standard)	89.7	85.3	91.4	88.2	75.7	69.3	68.4	72.2	83.7	75.3	81.4	84.2
Proposed Architecture — (No CBAM Module)	90.3	89.3	90.6	93.3	74.8	64.3	67.4	71.8	83.1	73.9	80.3	82.5
Proposed Architecture — (No skip connections)	88.2	84.6	91.1	85.8	73.9	68.8	68.1	71.7	83.1	74.1	81.2	81.3
Proposed Architecture — (Single ResConv in Encoder)	93.3	92.3	**94.1**	93.1	81.5	76.3	80.1	80.6	**90.4**	85.3	91.1	90.2
Proposed Architecture (CBAM + ResConV + Skip Connection)	**94.5**	**93.7**	93.6	**94.3**	**82.5**	**77.7**	**86.6**	**85.3**	90.1	**86.6**	**91.2**	**90.5**

TABLE 7. Effect of data augmentation, loss functions, and resolution on the proposed CBAM residual U-Net.

Methods	DSB 2018 Dataset				GlaS Dataset				TNBC Dataset			
	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall
Augmented	**94.5**	**93.7**	**95.6**	**95.3**	**82.5**	62.7	**86.6**	**85.3**	86.5	83.7	82.8	83.9
Non-Augmented	93.3	86.4	92.1	93.8	76.3	57.6	77.5	79.5	82.3	80.1	82.6	81.8
Dice Loss	91.7	87.9	94.9	93.1	79.7	74.9	78.9	79.1	79.7	80.9	78.9	79.1
BCE Loss	91.3	87.1	93.5	92.5	76.3	77.1	77.5	79.5	78.3	77.1	79.5	69.5
Combined Loss	**94.5**	**93.7**	93.6	94.3	**82.5**	**77.7**	**86.6**	**85.3**	86.5	83.7	82.8	83.9
500×500 pixels	91.3	89.2	90.3	90.5	**82.5**	**77.7**	**86.6**	85.3	86.5	83.7	82.8	83.9
256×256 pixels	94.5	93.7	93.6	94.3	80.3	75.2	81.3	75.5	**89.3**	83.9	90.3	**90.5**
128×128 pixels	92.1	90.6	91.2	93.3	81.1	77.6	85.2	**88.3**	94.5	90.1	86.6	88.3

We studied the value of residual connections, which are crucial in DL models for overcoming the vanishing gradient issue. According to the experimental findings, removing these connections could render it more challenging to train the model since backpropagation would be more complex and result in worse model performance. Similarly, skip connections provide accurate segmentation by enabling the network to use data from high-level semantic characteristics and low-level geographical details. Their removal results in a loss of spatial detail, resulting in a drop in performance and a less successful model at segmenting malignant nuclei.

B. EFFECT OF DATA AUGMENTATION

A systematic analysis is conducted to determine the effects of the addition or removal of different elements or modules on the efficiency of the proposed model. We investigate the overall impact of data augmentation, other loss functions, and model robustness and stability. The quality, quantity, and significance of training data affect the performance of most ML and DL algorithms. A shortage of information is among the most common issues in deploying ML in enterprise settings. Acquiring relevant data can be time-consuming and costly in many cases. Data augmentation allows us to increase the quantity of existing data by performing various alterations to available data. We generated new training data and applied apply several image transformations, random cropping, horizontal and vertical flipping, and random rotation (90°, 180°, 270°). The experimental findings show that data augmentation is critical for improving the training efficacy of the proposed model. Table 5 presents the number of images produced in data augmentation, and Table 7 reveals that training the proposed model using augmented data increases the accuracy compared to training it without augmented data.

The effectiveness of segmentation models can be significantly improved by data augmentation. Since there are fewer training data in the DSB 2018, GlaS, and TNBC datasets, data augmentation can generate more training data by applying various alterations to the original images. It can assist the model in learning from a more varied collection of training data by creating new images from current data. As a result, the model could better manage changes in tissue texture and form and diverse tissue types and disease patterns. Also, by increasing the variety of the training data and strengthening the model's resistance to noise, data augmentation can assist in preventing overfitting.

C. EFFECT OF LOSS FUNCTIONS

In the following, through the experiments, we study the effects of using several other loss functions with the same hyperparameter settings defined in Table 7. A loss function informs the algorithm of its proximity to the optimal solution. It leads a model to seek a perfect prediction that transfers input information into output data (images into masks in the case of image segmentation). We consider binary cross-entropy, Dice loss, and a combination of cross-entropy and Dice loss called combined loss. The experimental results reveal that the combined loss function performs better than the other two.

The performance can be improved for cancer cell detection by combining the Dice loss and binary cross-entropy (BCE) loss functions. Table 7 shows the performance of our proposed CBAM Residual U-Net model on Dice loss, BCE loss, and Combined loss. Although the BCE loss function evaluates the discrepancy between the anticipated likelihood of each pixel belonging to a particular class and the ground truth label, the Dice algorithm measures the intersection between the predicted segmentation map and the ground truth segmentation pattern. The two algorithms can be combined to provide the model advantages from both methods. The BCE loss function can help imbalance the positive and negative

TABLE 8. Comparison of proposed CBAM Residual U-Net with SOTA architectures on DSB 2018, GlaS, and TNBC datasets.

Models / Backbone	DSB 2018 Dataset				GlaS Dataset				TNBC Dataset			
	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall	F1	Dice	Precision	Recall
U-Net	89.7	85.3	91.4	88.2	75.7	69.3	68.4	72.2	83.7	75.3	81.4	84.2
U-Net (DenseNet)	90.6	88.5	92.8	88.5	70.6	61.5	74.8	70.5	84.6	75.5	85.8	82.5
U-Net (Inception)	91.0	88.8	92.7	89.5	69.0	60.8	72.7	70.5	81.0	73.8	80.7	82.5
U-Net (ResNet)	88.8	86.8	90.3	88.1	71.8	65.8	69.3	74.1	85.8	80.8	81.3	88.1
U-Net (EfficientNet)	90.7	87.9	91.2	90.3	80.7	76.2	85.2	80.3	86.7	85.9	93.2	80.3
U-Net (VGG16)	71.1	67.8	72.2	70.2	71.1	67.8	72.2	70.2	82.1	79.8	92.2	70.2
DeepLabV3 (ResNet)	91.4	88.2	92.4	89.5	79.6	72.2	70.4	75.7	83.4	80.2	81.4	84.5
Attention U-Net	88.5	83.6	87.8	90.3	81.5	75.6	75.8	**86.3**	89.2	84.6	**93.8**	85.3
CBAM-Residual U-Net (Ours)	**94.5**	**93.7**	**93.6**	**94.3**	**82.5**	**77.7**	**86.6**	85.3	**90.1**	**86.6**	91.2	**90.5**

TABLE 9. Performance comparison with SOTA algorithms.

Models	No. of Parameters	Complexity	Time Cost
U-Net	31M	Medium	Medium
PSPNet	46.8M (ResNet101)	High	High
DeepLabv3	58M (including ASPP)	High	High
Mask R-CNN	44M (ResNet50)	High	High
FCN	134M (FCN-8s, VGG16)	Very High	Very High
ResNeXt	88M (ResNeXt-101)	Very High	Very High
Ours	22M (CBAM)	Low	Low

classes, while the Dice loss function can assist with the model learning the intricate boundaries between tissue sections.

D. EFFECTS OF RESOLUTIONS
In the next set of experiments, we resize the images to 500×500, 256×256, and 128×128 pixels to study the robustness of the proposed model for different image sizes. When we scale the images up to 500×500 pixels, the image noise increases, which degrades performance. Despite the decrease in size, the model performance on images of 256×256 pixels is better than that of images of 500×500 pixels as a result of noise reduction. When we further reduce the size to 128×128 pixels, the model performance declines based on the reduction in nucleus size. Table 7 further reveals that the model trained on 256×256 pixel images performs best regarding the F1-score, Dice, precision, and recall values. For the GlaS dataset, the proposed CBAM residual U-Net performed better on 500×500 pixels by achieving the best precision and recall values.

The performance findings show that our suggested design performs better than traditional U-Net and backbone architectures. Table 9 shows the performance comparison of the proposed model with other SOTA algorithms. The proposed CBAM residual U-net model has a size of only 296.8 MB with 22M distinct parameters, making it one of the most effective architectures for segmentation.

VI. CONCLUSION AND FUTURE WORK
We introduced the CBAM residual U-Net, a novel DL model for identifying and segmenting cell nuclei. This model employed more precise spatial context data and delivered more exact findings than baseline image segmentation networks by convolution. The CBAM consisted of two successive sub-modules, the CAM and the SAM, utilized in

every convolution block to construct refined feature maps from input initial feature maps. Also, we introduced a novel type of completely connected residual block (ResConv) to obtain initial features from input images. The residual skip connection weights w_1, w_2, and w_3 ensured that both shallow and deep feature maps were received during the convolution process. At last, we compared our proposed framework to various SOTA DL models in terms of DC, F1-score, weight size, and other factors.

Overall, the proposed model outperformed the other SOTA models in all aspects. Therefore, the proposed model is suitable for real-world applications because of its high performance, stability, and minimal computational complexity. Future work will include researching and implementing more powerful, robust, and precise CNN, transformer, and diffusion algorithms for medical and clinical purposes. Input image with feature maps extracted from the intermediate layers of the proposed CBAM residual U-Net model with different staining intensities.

REFERENCES
[1] V. Kumar, A. K. Abbas, and J. Aster, *Robbins Basic Pathology E-Book*. Amsterdam, The Netherlands: Elsevier, 2017.

[2] C. Liu, F. Shang, J. A. Ozolek, and G. K. Rohde, "Detecting and segmenting cell nuclei in two-dimensional microscopy images," *J. Pathol. Informat.*, vol. 7, no. 1, p. 42, Jan. 2016.

[3] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.

[4] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology," *Nature Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019.

[5] H. A. Shah, F. Saeed, S. Yun, J. Park, A. Paul, and J. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022.

[6] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, and J. Kini, "NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104075.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[8] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Med. Image Anal.*, vol. 48, pp. 107–116, Aug. 2018.

[9] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Process.*, vol. 122, pp. 1–13, May 2016.

[10] R. Feng, X. Liu, J. Chen, D. Z. Chen, H. Gao, and J. Wu, "A deep learning approach for colonoscopy pathology WSI analysis: Accurate segmentation and classification," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3700–3708, Oct. 2021.

[11] R. Albusayli, J. D. Graham, N. Pathmanathan, M. Shaban, S. E. A. Raza, F. Minhas, J. E. Armes, and N. Rajpoot, "Artificial intelligence-based digital scores of stromal tumour-infiltrating lymphocytes and tumour-associated stroma predict disease-specific survival in triple-negative breast cancer," *J. Pathol.*, vol. 260, no. 1, pp. 32–42, May 2023.

[12] M. Makem, A. Tiedeu, G. Kom, and Y. P. K. Nkandeu, "A robust algorithm for white blood cell nuclei segmentation," *Multimedia Tools Appl.*, vol. 81, no. 13, pp. 17849–17874, May 2022.

[13] P. Sharma, A. Gautam, P. Maji, R. B. Pachori, and B. K. Balabantaray, "Li-SegPNet: Encoder–decoder mode lightweight segmentation network for colorectal polyps analysis," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 4, pp. 1330–1339, Apr. 2023.

[14] N. K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, and S. Ali, "FANet: A feedback attention network for improved biomedical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 25, 2022, doi: 10.1109/TNNLS.2022.3159394.

[15] U. Zidan, M. M. Gaber, and M. M. Abdelsamea, "SwinCup: Cascaded Swin transformer for histopathological structures segmentation in colorectal cancer," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119452.

[16] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation," 2022, *arXiv:2202.00972*.

[17] Y. Liu, S. J. Wagner, and T. Peng, "Multi-modality microscopy image style augmentation for nuclei segmentation," *J. Imag.*, vol. 8, no. 3, p. 71, Mar. 2022.

[18] P. Jian and S.-I. Kamata, "A two-stage refinement network for nuclei segmentation in histopathology images," in *Proc. 4th Int. Conf. Image, Video Signal Process.*, Mar. 2022, pp. 8–13.

[19] T. Dhamija, A. Gupta, S. Gupta, R. Katarya, and G. Singh, "Semantic segmentation in medical images through transfused convolution and transformer networks," *Appl. Intell.*, vol. 53, no. 1, pp. 1–17, 2022.

[20] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R^2U-Net)," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, Jul. 2018, pp. 228–233.

[21] S. Vishnu Priyal, M. T. Vyshnav, V. Sowmya, and K. P. Soman, "Modified UNet architecture with less number of learnable parameters for nuclei segmentation," in *Soft Computing and Signal Processing*, V. S. Reddy, V. K. Prasad, J. Wang, K. T. V. Reddy, Eds. Singapore: Springer, 2022, pp. 101–111.

[22] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 558–564.

[23] A. K. Chanchal, S. Lal, and J. Kini, "Deep structured residual encoder–decoder network with a novel loss function for nuclei segmentation of kidney and breast histopathology images," *Multimedia Tools Appl.*, vol. 81, no. 7, pp. 9201–9224, Mar. 2022.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[25] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–16.

[26] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7.

[27] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, and A. E. Carpenter, "Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, Dec. 2019.

[28] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The GlaS challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, Jan. 2017.

[29] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[31] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.

[32] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, "A sufficient condition for convergences of Adam and RMSProp," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11119–11127.

HASNAIN ALI SHAH received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2019. He is currently pursuing the M.Sc. degree in intelligent ICT with the Computer Vision and Reinforcement Learning Laboratory, Department of Artificial Intelligence, Kyungpook National University, Daegu, South Korea. He is doing the master's program with a fully funded scholarship. His current research interests include computer vision, machine learning, deep learning, medical image processing, and pattern recognition.

JAE-MO KANG (Member, IEEE) received the Ph.D. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2017. He was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada, and an Assistant Professor with the School of Intelligent Mechatronics Engineering, Sejong University, Seoul, South Korea. He is currently an Assistant Professor with the Department of Artificial Intelligent, Kyungpook National University, Daegu, South Korea. His current research interests include the IoT, LoRa, 6G, machine/deep learning, reinforcement learning, federated Learning, and edge AI.

• • •