frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 07 October 2022
DOI 10.3389/fbioe.2022.963590

OPEN ACCESS

EDITED BY
Zhihua Cui,
Taiyuan University of Science and Technology, China

REVIEWED BY
J. Blas Pagador,
Jesús Usón Minimally Invasive Surgery Center, Spain
Luisa F. Sánchez-Peralta,
Jesus Uson Minimally Invasive Surgery Centre, Spain

*CORRESPONDENCE
Feng Liu,
liuf@nankai.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Bionics and Biomimetics, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 07 June 2022
ACCEPTED 12 August 2022
PUBLISHED 07 October 2022

CITATION
Cui R, Yang R, Liu F and Cai C (2022), N-Net: Lesion region segmentations using the generalized hybrid dilated convolutions for polyps in colonoscopy images.
Front. Bioeng. Biotechnol. 10:963590.
doi: 10.3389/fbioe.2022.963590

N-Net: Lesion region segmentations using the generalized hybrid dilated convolutions for polyps in colonoscopy images

Rongsheng Cui[1†], Runzhuo Yang[1†], Feng Liu[1,2*] and Chunqian Cai[3,4]

[1]College of Electronic Information and Optical Engineering, Nankai University, Tianjin, China, [2]Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin, China, [3]First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, China, [4]National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin, China

Colorectal cancer is the cancer with the second highest and the third highest incidence rates for the female and the male, respectively. Colorectal polyps are potential prognostic indicators of colorectal cancer, and colonoscopy is the gold standard for the biopsy and the removal of colorectal polyps. In this scenario, one of the main concerns is to ensure the accuracy of lesion region identifications. However, the missing rate of polyps through manual observations in colonoscopy can reach 14%–30%. In this paper, we focus on the identifications of polyps in clinical colonoscopy images and propose a new N-shaped deep neural network (N-Net) structure to conduct the lesion region segmentations. The encoder-decoder framework is adopted in the N-Net structure and the DenseNet modules are implemented in the encoding path of the network. Moreover, we innovatively propose the strategy to design the generalized hybrid dilated convolution (GHDC), which enables flexible dilated rates and convolutional kernel sizes, to facilitate the transmission of the multi-scale information with the respective fields expanded. Based on the strategy of GHDC designing, we design four GHDC blocks to connect the encoding and the decoding paths. Through the experiments on two publicly available datasets on polyp segmentations of colonoscopy images: the Kvasir-SEG dataset and the CVC-ClinicDB dataset, the rationality and superiority of the proposed GHDC blocks and the proposed N-Net are verified. Through the comparative studies with the state-of-the-art methods, such as TransU-Net, DeepLabV3+ and CA-Net, we show that even with a small amount of network parameters, the N-Net outperforms with the Dice of 94.45%, the average symmetric surface distance (ASSD) of 0.38 pix and the mean intersection-over-union (mIoU) of 89.80% on the Kvasir-SEG dataset, and with the Dice of 97.03%, the ASSD of 0.16 pix and the mIoU of 94.35% on the CVC-ClinicDB dataset.

KEYWORDS

N-shape deep neural network, generalized hybrid dilated convolution, colonoscopy, colorectal polyp identification, lesion region segmentation, deep learning

1 Introduction

Colorectal cancer is the cancer with the second and the third highest incidence rates for the female and the male, respectively. Early diagnosis has a huge impact on the survival from colorectal cancer (Torre et al., 2015). Colorectal polyps are potential prognostic indicators of the colorectal cancer, and colonoscopy is the gold standard for the biopsy and the removal of colorectal polyps (van Toledo et al., 2022). Research shows that nearly half of the individuals taking the colonoscopy at the age of 50 are found to be suffered from colorectal polyps (Lima Pereira et al., 2020). This incidence rate even increases with the age (Rundle et al., 2008). The accurate identification of colorectal polyp lesion regions plays a preliminary role in the medical treatment of colorectal cancers (Ren et al., 2019; Qadir et al., 2020; Tan et al., 2020). However, several studies indicate that the missing rate of polyps through manual observations in colonoscopy can reach 14%–30%, depending on the types and the sizes of the polyps (Van Rijn et al., 2006). Thus, the development of accurate colorectal polyp segmentation methods is critical.

With the rapid development of computer and information techniques, computer-aided diagnosis methods have been used in polyp segmentation tasks. However, the techniques of computer-aided polyp segmentations is still immature, especially in the cases that some complex and uncontrolled environmental factors exist. For example, existing computer-aided diagnosis methods can not effectively deal with factors that can affect the accuracy of polyp segmentation, such as intraluminal folds and variations of the polyp textures and locations. In (Sasmal et al., 2018), the principal component pursuit (PCP) technique was used in colorectal polyp segmentations. However, the segmentation performance got worse in low-light situations. In (Ganz et al., 2012), the shape-UCM methods were used in colorectal polyp segmentations, but the polyps with heterogeneous shapes could not be extracted.

Later, some early machine learning methods were developed for colorectal polyp segmentations. However, these methods consumed a lot of computer memory and relied excessively on handcrafted features, thus were not robust enough. For example, the machine learning based methods in (Tajbakhsh et al., 2016; Yu et al., 2017) got constrained segmentation performance for polyps in the presence of high luminance or intestinal residues.

In the past few years, the deep learning technology has been adopted in the segmentations of medical images. Particularly, Olaf proposed the U-shape artificial neural network (U-Net) using an encoding-decoding structure (Ronneberger et al., 2015). In the U-Net, multiple encoding and decoding modules were included in a symmetrical framework. In addition, the skip connections were added between the encoding and decoding paths to enable multi-scale information transmissions. The encoder-decoder structure of the U-Net has now become the most commonly used network structure in medical image segmentations.

More recently, some network structures based on the U-Net were proposed for the medical image segmentations and showed outstanding performances (Zhou Z. et al., 2018; Chen et al., 2018; Oktay et al., 2018; Huang et al., 2020; Chen et al., 2021). In particular, the densely connected convolutional neural network (DenseNet) strengthened the feature propagation and alleviated the gradient-vanishing problem, with the correct training convergences and the good feature extraction performances ensured (Huang et al., 2017; Wang et al., 2021). Moreover, in (Alom et al., 2018; Gu et al., 2021; Zhang and Yang, 2021), the attention mechanisms were also proposed recently, greatly improving the precision of medical image segmentations. Additionally, in order to further reduce the training time of the accelerate the convergence of the network training, the transfer learning was adopted by scholars (Shao et al., 2015; He et al., 2020; Rozo et al., 2022).

In addition to the above works, researches showed that the fine-grained image features could be better captured by expanding the receptive fields in the multi-scale information transmissions. In (Zhou L. et al., 2018), L. Zhou *et al* connected the encoding and decoding paths with dilated convolutions, where the receptive fields were expanded. However, the gridding effects could be introduced in this case. In order to solve this problem, P. Wang et al proposed the hybrid dilated convolution (HDC) (Wang et al., 2018). The HDC was established with cascaded dilated convolutions and has been adopted for semantic segmentations (Fu et al., 2019; Cheng et al., 2020; Liu et al., 2020; Ma et al., 2022). For example, J. Liu et al implemented an HDC based algorithm in the detection of retinal pigment epithelium defective cells (Liu et al., 2020). However, the mathematical model of the strategy of HDC designing was not much described. In addition, the HDC architecture required all the convolutional kernel sizes to be equal, with its implementation flexibility constrained.

In order to overcome the shortcomings of the current methods and optimize the colorectal polyp lesion region segmentation performance using colonoscopy images, we propose a novel N-shaped artificial neural network (N-Net) structure. It should be specially noted that although this work focuses on the colorectal segmentation task, the N-Net structure can be used generally for image segmentations. The N-Net structure is briefly described in Figure 1. Compared with state-of-the-art methods through experiments on two public colonoscopy datasets for polyp segmentations, the proposed method achieves the best segmentation performance in the metrics of Dice, average symmetric surface distance (ASSD) and mean intersection-over-union (mIoU). The main contributions of this paper are as follows:

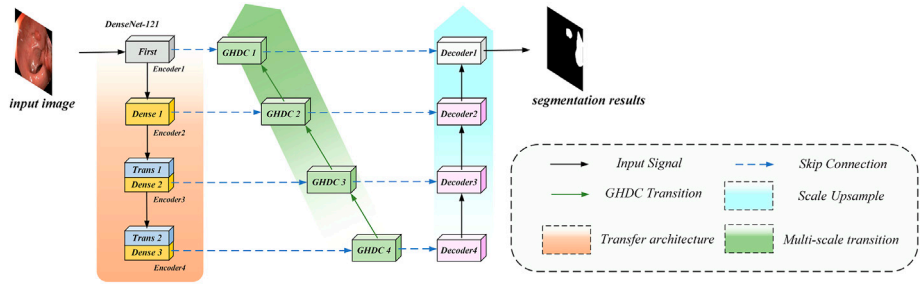

FIGURE 1
The overview of our proposed N-Net.

1) We propose a novel N-shaped artificial neural network (N-Net) structure to conduct the lesion region segmentations of polyps in colonoscopy images. The proposed N-Net is designed based on the encoding-decoding framework. Within the proposed structure, the multi-scale information can flow between the encoding and decoding paths. The pretrained DenseNet modules based on the ImageNet are implemented in the encoding path of the N-Net to ensure the fast training convergence and good feature extraction performance of the entire network structure.
2) To expand the receptive fields and facilitate the multi-scale information transmission between the encoding and decoding paths, we propose a strategy to design the generalized hybrid dilated convolution (GHDC). Compared with the existing works related to the dilated convolutions, the GHDC is established with a more flexible strategy to design cascaded dilated convolutional layers.
3) Based on the strategy of GHDC designing, four GHDC blocks are designed to connect the encoding path and the decoding path. With experiments on two public available datasets: the Kvasir-SEG dataset and the CVC-ClinicDB dataset, we show that the GHDC blocks outperform the HDC in (Wang et al., 2018). Moreover, comparative studies shows that the proposed N-Net, even with a small amount network parameters, outperforms the state-of-the-art methods including TransU-Net, DeepLabV3+ and CA-Net.

2 Methods

2.1 The structure of the N-Net

In this paper, motivated by the U-Net encoder-decoder framework and the DenseNet modules, we propose a novel N-shaped architecture. As shown in Figure 2, the proposed N-Net structure contains four stages both in the encoding path on the left side and the decoding path on the right side. In addition, the GHDC blocks are added to connect the encoding and the decoding paths.

In contrast to the U-Net structure, a "Conv (7 × 7)-BN-ReLU" operation (i.e. convolution with the kernel size of 7 × 7 followed by a batch normalization and a ReLU) and an "Maxpool" operation (maximum pooling with the kernel size of 7 × 7 and the stride of 2) are implemented as the first stage of the encoding path. With the idea of transfer learning, the three encoding modules in the encoding path are designed based on the three dense blocks in the pre-trained DenseNet-121 from the ImageNet, respectively. Each dense block is composed by certain dense layers and a transition layer. The pre-trained dense layers in the three dense blocks and the first two transition layers are directly transferred. The number of dense layers in the three dense blocks are 6, 12 and 24, respectively. The number of output channels from the four encoders are 64, 256, 512 and 1,024, respectively. The feature map size from the Encoder 1 is 224 × 224 and is halved after the processing of each encoder.

The decoding path in the proposed N-Net is composed with four decoding modules. Each of the decoding modules contains a "Conv (1 × 1)-BN-ReLU-TransposeConv (3 × 3)-BN-Conv (1 × 1)" procedure (i.e. 3 steps: the first and third steps consist of a convolution with 1 × 1 kernel and a batch normalization; the second step consists of a transpose convolution with 3 × 3 kernel and a batch normalization). The Decoder one contains two "Conv (3 × 3)" operations and a bilinear interpolation operation.

The encoding and decoding paths in the N-Net structure are connected with the proposed GHDC blocks, which are designed based on the novel strategy of GHDC designing in this paper. For $i = 1, 2, 3$, the outputs of the Encoder i and the GHDC block $i + 1$ are concatenated as the input of the GHDC block i. Besides, the outputs of the GHDC block i and the Decoder $i + 1$ are point-wise added as the input to the Decoder i.

With the connections between the encoding path and the decoding path using the GHDC blocks, the multi-scale features are deeply exploited with expanded respective fields. The detailed strategy of GHDC designing and the GHDC blocks are provided in the remainder of this section.

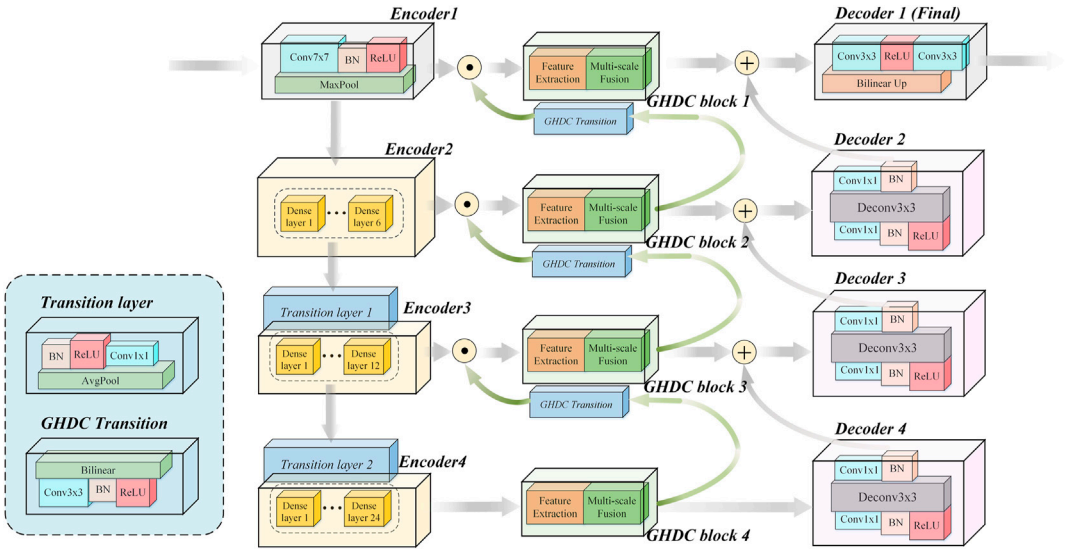

FIGURE 2
The detailed structure of our proposed N-Net.

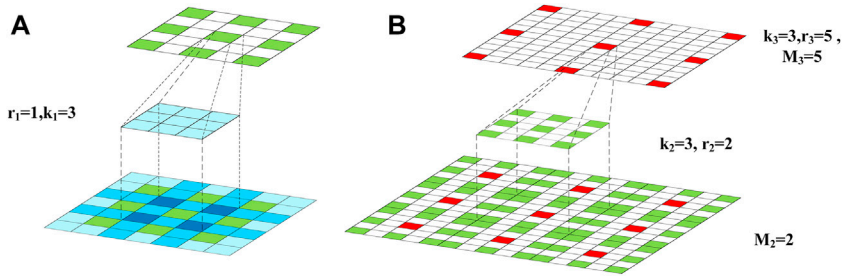

FIGURE 3
The cascaded dilated convolutional layers: **(A)** The operation of the first dilated convolutional layer; **(B)** the operation of the second dilated convolutional layer.

2.2 The strategy of generalized hybrid dilated convolution designing

In order to achieve sufficient multi-scale information transmissions, while expanding the respective fields without the gridding effects, we establish a more general and simplified strategy to design the cascaded dilated convolutions. In contrast to the HDC, flexible dilation rates and convolutional kernel sizes are enabled in the designed operation. Which is referred to as the generalized hybrid dilated convolution (GHDC) in this paper.

To explain the strategy of GHDC designing, we first provide a simple example to design the first three layers in a cascade of dilated convolutions in Figure 3. The convolutional kernel sizes for the three layers are denoted as k_1, k_2 and k_3, respectively. The dilation rates used in the three layers are denoted as r_1, r_2 and r_3, respectively.

In Figure 3A, the feature map into the first layer, the dilated convolutional kernel of the first layer and the feature map into the second layer are shown from down to up, respectively. In Figure 3B, the feature map into the second layer, the dilated convolutional kernel of the second layer and the feature map into the third layer are shown from down to up, respectively. In the feature map into the first layer, the darker color suggests that the corresponding pixels are used for more times during the convolution. In the feature maps into the second and third layers, the maximum distances between the neighbored pixels, which are sampled in the dilated convolutions, are denoted as M_2 and M_3, respectively. According to Figure 3, M_2 and k_1 can be used to observe the coverage of the convolutional kernel in the first layer.

In the proposed strategy of GHDC designing, a key point is to ensure that all of the pixels in the feature map into the first layer are utilized in the operations in order to avoid the gridding effect.

To meet this constraint, the parameters in the third and second layers are decided sequentially. For example, if M_3, k_2 and r_2 takes the value of 5, 3 and 2, respectively, then M_2 becomes 2. In this scenario, by taking $k_1 = 3$ and $r_1 = 1$, it can be guaranteed that no holes exist during the cascaded convolutions.

To be more general, we observe the dilated convolution operation on the arbitrary *lth* ($l > 0$) layer in a dilated convolution cascade. In this scenario, it can be found that the sampling positions in this operation are influenced by the parameters of both the *lth* and $(l + 1)^{th}$ layers. Let us define M_{l+1} as the maximum distance between neighbored sampled positions horizontally or vertically on the $(l + 1)^{th}$ ($l > 0$) layer, and define r_l and k_l to be the dilation rate and the convolutional kernel sizes on the *lth* layer, respectively. Then we can estimate M_l by the sliding of the dilated convolutional kernel on an *xOy* coordinate system. Due to the symmetry, there is no harm to simplify the problem by just observing the movement of the convolutional kernel on the *x*-axis.

If we consider k_{l+1} pixels separated by the distance of M_{l+1} on the feature map into the $(l + 1)^{th}$ layer, then the possible pixels covered by the convolutional kernel can be represented by: $\{(m_{l,i}r_l + n_{l,i}M_{l+1}, 0)\}$, where $m_{l,i} \in \{0, \pm 1, \ldots, \pm \frac{k_l-1}{2}\}$ and $n_{l,i} \in \{0, \pm 1, \ldots, \pm \frac{k_{l+1}-1}{2}\}$ stand for the coefficients to determine the location of *ith* pixel covered by the convolutional kernel. In other words, if M_{l+1}, k_{l+1}, r_l, k_l are determined, the parameter M_l can be derived.

Therefore, in the proposed GHDC model, the design of a dilated convolution cascade is done in a recursive manner. For the *lth* ($l > 0$) layer, we define $D_l = \{d_{l,i}\}$, $i = 1, \ldots, k_lk_{l+1}-1$ to be a non-decreasing sequence. Then given M_{l+1} and k_{l+1}, r_l and k_l ($l > 1$) are determined with the following constraints:

$$d_{l,i} = \|(m_{l,i+1}r_l + n_{l,i+1}M_{l+1}) - (m_{l,i}r_l + n_{l,i}M_{l+1})\|, \quad (1)$$

where $\|\cdot\|$ stands for the l-2 norm operation, $m_{l,i} \in \{0, \pm 1, \ldots, \pm \frac{k_l-1}{2}\}$ and $n_{l,i} \in \{0, \pm 1, \ldots, \pm \frac{k_{l+1}-1}{2}\}$. Then M_l, which is used to design the parameters of k_{l-1} and r_{l-1}, can be determined as follows:

$$M_l = Max\|d_{l,i} - d_{l,i-1}\| \text{ and } M_2 \le k_1. \quad (2)$$

In addition, it is preferred that the receptive field in the first dilated convolution layer, which can be calculated using:

$$F = 1 + \sum_l r_l(k_l - 1), \quad (3)$$

can cover the input feature map to the entire cascaded dilated convolutions.

For instance, if the kernel sizes of three are used in the cascaded dilated convolutions, we can get the following expression using Eqs 1, 2:

$$M_l = \begin{cases} r_l, 0 < M_{l+1} < r_l \\ \max\{r_l, M_{l+1} - 2r_l\}, M_{l+1} > 2r_l \end{cases} \quad (4)$$

By recursively solving Eq. 4, we can get a cascade of convolutional layers, which is consistent with the HDC.

2.3 The GHDC blocks

In the proposed N-Net, the GHDCs are conducted with four GHDC blocks between the encoding and decoding paths. As shown in Figure 4, a GHDC block generally consists of a feature extraction module and a multi-scale fusion module. The feature extraction module is established with parallel groups of cascaded dilated convolutional layers, which are designed according to the strategy of GHDC designing. In the multi-scale fusion module, the results from the groups of cascaded dilated convolutional layers are concatenated and then processed using a "Conv (3 × 3)-BN-ReLU" operation to keep the output size of each GHDC block consistent with its input. In addition, to make sure that the feature map dimensions from the GHDC block $i + 1$ are consistent with that from the Encoder i ($i = 1, 2, 3$), a GHDC transition layer, consisting of a bi-linear interpolation and a "Conv (3 × 3)-BN-ReLU" operation, is introduced.

In Table 1, the dilation rates and the convolutional kernel sizes of the dilated convolutional layers, as well as the receptive field sizes, are provided for the four GHDC blocks. The convolutional kernel sizes are odd. Thus, unique and consistent central pixel locations, as well as symmetric padding operations, can be guaranteed during the convolutions. Meanwhile, integer values of $m_{l,i}$ and $n_{l,i}$ in Eq. 1 are ensured. The detailed structures of the four GHDC blocks are shown in Figure 5.

3 Experiments and results

3.1 Dataset

In order to verify the proposed N-Net, two publicly available datasets on polyp segmentations of colonoscopy images were used in the experiments: 1) the Kvasir-SEG dataset (Jha et al., 2019); 2) the CVC-ClinicDB dataset (Bernal et al., 2015). Both the original images and corresponding masks for the ground truth of the lesion regions are included in the two datasets. The details of these two datasets are provided in Table 2. The images from the Kvasir-SEG dataset and the CVC-ClinicDB dataset were resized into 256 × 320 and 288 × 384, respectively, before used in the experiments. We randomly divided the datasets into the training sets, the test set and the validation set, according to the ratio of 7:2:1.

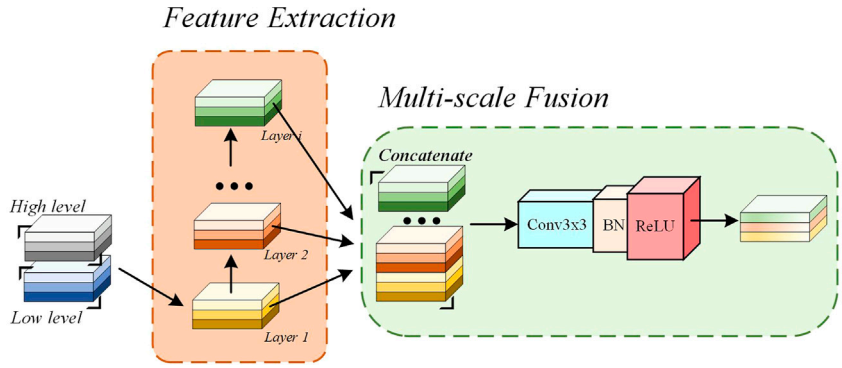

FIGURE 4
The Overview of the Proposed GHDC block.

TABLE 1 The arguments for the GHDC blocks in the proposed N-Net.

GHDC blocks	The number of Layers i $i = 1, 2, 3, 4, 5$	Dilation Rate $[r_1, r_2, \ldots, r_i]$	Kernel Size $[k_1, k_2 \ldots, k_i]$	Respective Fields
GHDC block4	$i = 2$	[1,2]	[3,3]	7
GHDC block3	$i = 3$	[1,2,5]	[3,3,3]	17
GHDC block2	$i = 4$	[1,2,5,7]	[3,3,3,5]	45
GHDC block1	$i = 5$	[1,2,5,7,9]	[3,3,3,5,5]	81

3.2 Implementation details

The proposed N-Net was trained based on the gradient descent method. The training process was performed with the Python package PyTorch 1.11.0 + GPU, using a computer with the Nvidia GTX 3090 GPU, and the RAM size of 32.00 GB. We used the Adaptive Moment Estimation (Adam) (Kingma and Ba, 2017) to control the learning rates in the training process, with the initial learning rate, the weight decay, the batch size and the number of epochs as 10^{-4}, 10^{-8}, 16 and 300, respectively. The sum of the binary cross-entropy and the soft dice coefficient (Sun et al., 2016) was used as the loss function in the training process of the N-Net. The binary cross-entropy and the soft dice coefficient are defined as:

$$L_{bce} = -\sum_{i=1}^{n}\left[\hat{y}_{i0}\ln y_{i0} + \left(1-\hat{y}_{i0}\right)\ln\left(1-y_{i0}\right)\right] \tag{5}$$

and

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{n} y_{i1}\hat{y}_{i1}}{\sum_{i=1}^{n} y_{i1} + \sum_{i=1}^{n}\hat{y}_{i1}}, \tag{6}$$

respectively.

In Eqs 5, 6, n denotes the number of samples in a training data batch. y_{im} and y_{im} represent the predicted and actual probabilities that the *ith* pixel belongs to the class m ($m \in \{0, 1\}$), respectively.

To verify the performance of the proposed method, the metrics of Dice, ASSD and mIoU were implemented on the lesion region segmentation results in the experiments.

The metric of Dice is defined as:

$$Dice = \frac{2|S \cap G|}{|S| + |G|}, \tag{7}$$

where S and G represent the region segmented by the N-Net and in the ground truth, respectively.

The metric of ASSD is defined as:

$$ASSD = \frac{\sum_{m\in P_a} d(m, P_b) + \sum_{n\in P_b} d(n, P_a)}{|P_a| + |P_b|}, \tag{8}$$

where P_a and P_b denote the set of boundary points segmented by the convolutional neural network (CNN) and in the ground truth, respectively. $d(u, P_b) = \min_{v\in P_b}\|u - v\|$ represents the minimum Euclidean distance from the point u to P_b.

The metric of mIoU is defined as:

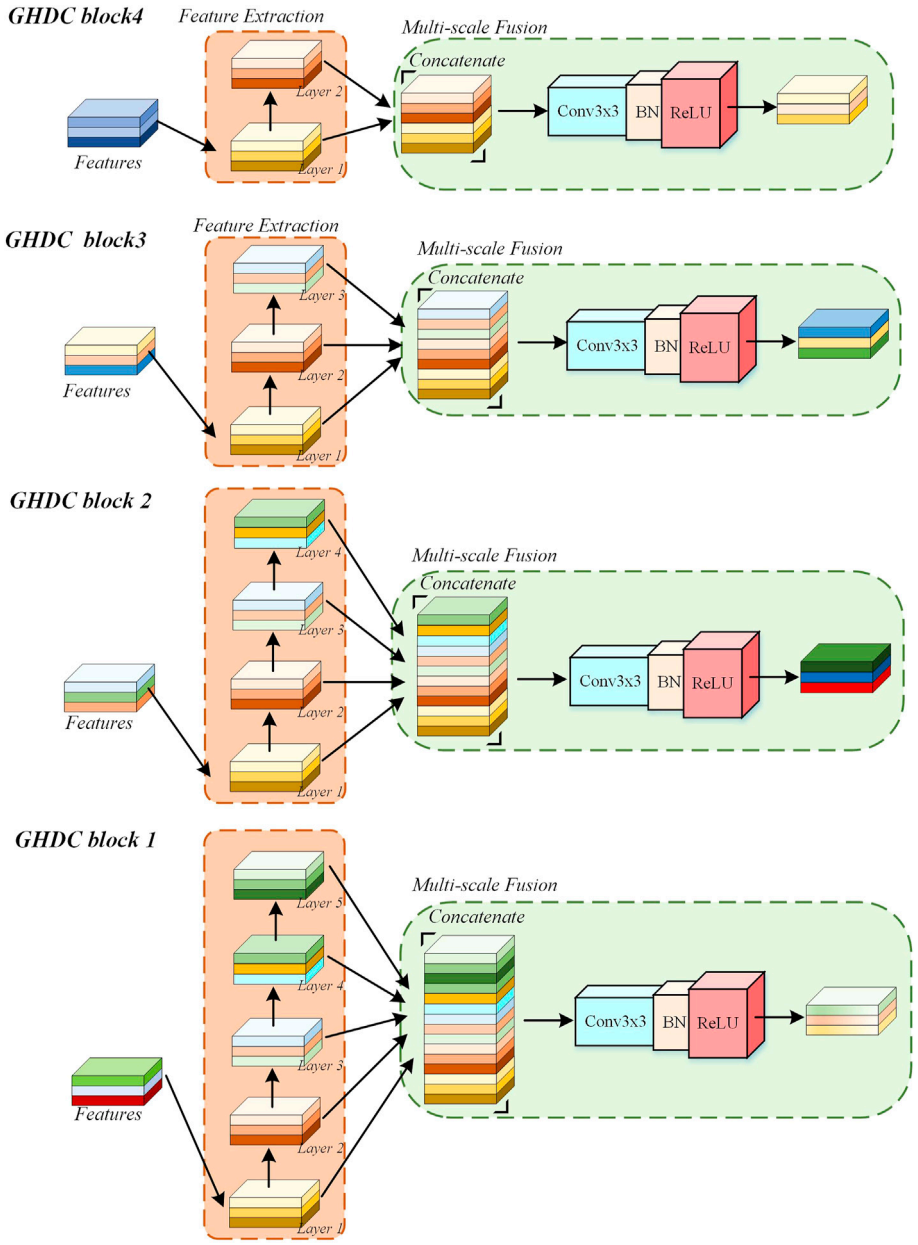

FIGURE 5
The detailed structure of the GHDC block.

$$mIoU = \frac{1}{2}\left(\frac{TP}{FN + FP + TP} + \frac{TN}{FP + FN + TN}\right), \quad (9)$$

where *TP*, FP, *TN* and *FN* represent the numbers of pixels with true positive, false positive, true negative and false negative decisions for the lesion regions, respectively.

3.3 Results

3.3.1 Ablation studies on the proposed GHDC blocks of the N-Net

In order to verify the effectiveness of the N-Net structure and analyze the contributions of the GHDC blocks to the network

TABLE 2 The two publicly available colorectal polyp segmentations datasets used in experiments.

Dataset	Images	Train	Validation	Test
Kvasir-SEG	1,000	715	104	189
CVC-ClinicDB	612	429	60	123

performance, we first performed the ablation experiments with the GHDC blocks on the two datasets. The results of the experiments are shown in Table 3. In that table, the U-Net structure with the encoding path transferred from the first four stages of the DenseNet-121 is denoted as the baseline. By observing the results from the networks with only a single GHDC block included, we can find that the GHDC blocks greatly help in the segmentation performance improvement in terms of the Dice, ASSD and mIoU values. In particularly, this contribution increases sequentially from the GHDC block four to the GHDC block 1. Moreover, it can also be observed that by integrating more GHDC blocks into the network structures, the segmentation performances of the resulting networks can be further improved, as the multi-scale information can be transmitted more sufficiently in these cases. At the end of Table 3, we can see that the N-Net, where all of the four GHDC blocks are included, achieve the best lesion region segmentation performance on both of the two datasets.

TABLE 3 Ablation Study Results of the GHDC Blocks on the Two Public Datasets of Colorectal Polyp Segmentations (The best results and the second best results are marked in red bold and blue bold fonts, respectively).

Results on the Kvasir-SEG Dataset.

Baseline	GHDC block *i*	Dice (%)	ASSD (pix)	mIoU (%)
√	×	92.06 (±4.74)	0.51 (±0.20)	85.92 (±4.63)
√	*i* = 4	92.23 (±5.02)	0.49 (±0.43)	86.08 (±7.14)
√	*i* = 3	92.26 (±3.64)	0.49 (±0.70)	86.19 (±3.63)
√	*i* = 2	92.43 (±4.31)	0.48 (±0.67)	86.34 (±4.69)
√	*i* = 1	92.97 (±3.86)	0.46 (±0.42)	86.60 (±5.53)
√	*i* = 4, 3	92.53 (±3.05)	0.46 (±0.36)	86.48 (±4.00)
√	*i* = 4, 2	92.70 (±3.66)	0.45 (±0.47)	86.61 ± 5.26)
√	*i* = 4, 1	93.15 (±4.28)	0.44 (±0.53)	86.83 (±4.37)
√	*i* = 3, 2	92.62 (±4.35)	0.47 (±0.49)	86.56 (±6.01)
√	*i* = 3, 1	93.24 (±7.68)	0.43 (±0.35)	87.07 (±4.49)
√	*i* = 2, 1	93.53 (±3.66)	0.42 (±0.17)	87.24 (±4.03)
√	*i* = 4, 3, 2	93.84 (±5.30)	0.40 (±0.20)	87.93 (±7.03)
√	*i* = 3, 2, 1	**94.22(±3.57)**	**0.41(±0.26)**	**89.42(±3.80)**
	N-Net	**94.45(±1.48)**	**0.38(±0.21)**	**89.80(±2.56)**

Results on the CVC-ClinicDB Dataset.

Baseline	GHDC block *i*	Dice (%)	ASSD (pix)	mIoU (%)
√	×	95.15 (±2.01)	0.32 (±0.40)	90.63 (±2.96)
√	*i* = 4	95.20 (±2.52)	0.30 (±0.31)	90.92 (±4.30)
√	*i* = 3	95.41 (±1.94)	0.29 (±0.29)	91.04 (±5.37)
√	*i* = 2	95.83 (±2.42)	0.27 (±0.33)	91.16 (±6.61)
√	*i* = 1	96.02 (±5.28)	0.26 (±0.27)	91.32 (±4.72)
√	*i* = 4, 3	95.49 (±5.45)	0.28 (±0.43)	91.20 (±6.20)
√	*i* = 4, 2	95.98 (±3.26)	0.26 (±0.28)	91.37 ± 3.65)
√	*i* = 4, 1	96.13 (±6.20)	0.24 (±0.19)	91.73 (±4.01)
√	*i* = 3, 2	96.05 (±6.07)	0.25 (±0.20)	92.09 (±3.84)
√	*i* = 3, 1	96.29 (±5.32)	0.23 (±0.39)	92.56 (±5.28)
√	*i* = 2, 1	96.46 (±5.00)	0.21 (±0.09)	92.89 (±6.25)
√	*i* = 4, 3, 2	96.62 (±7.08)	0.19 (±0.10)	92.01 (±2.02)
√	*i* = 3, 2, 1	**96.91(±5.34)**	**0.18(±0.21)**	**94.13(±4.60)**
	N-Net	**97.03(±0.82)**	**0.16(±0.06)**	**94.35(±1.49)**

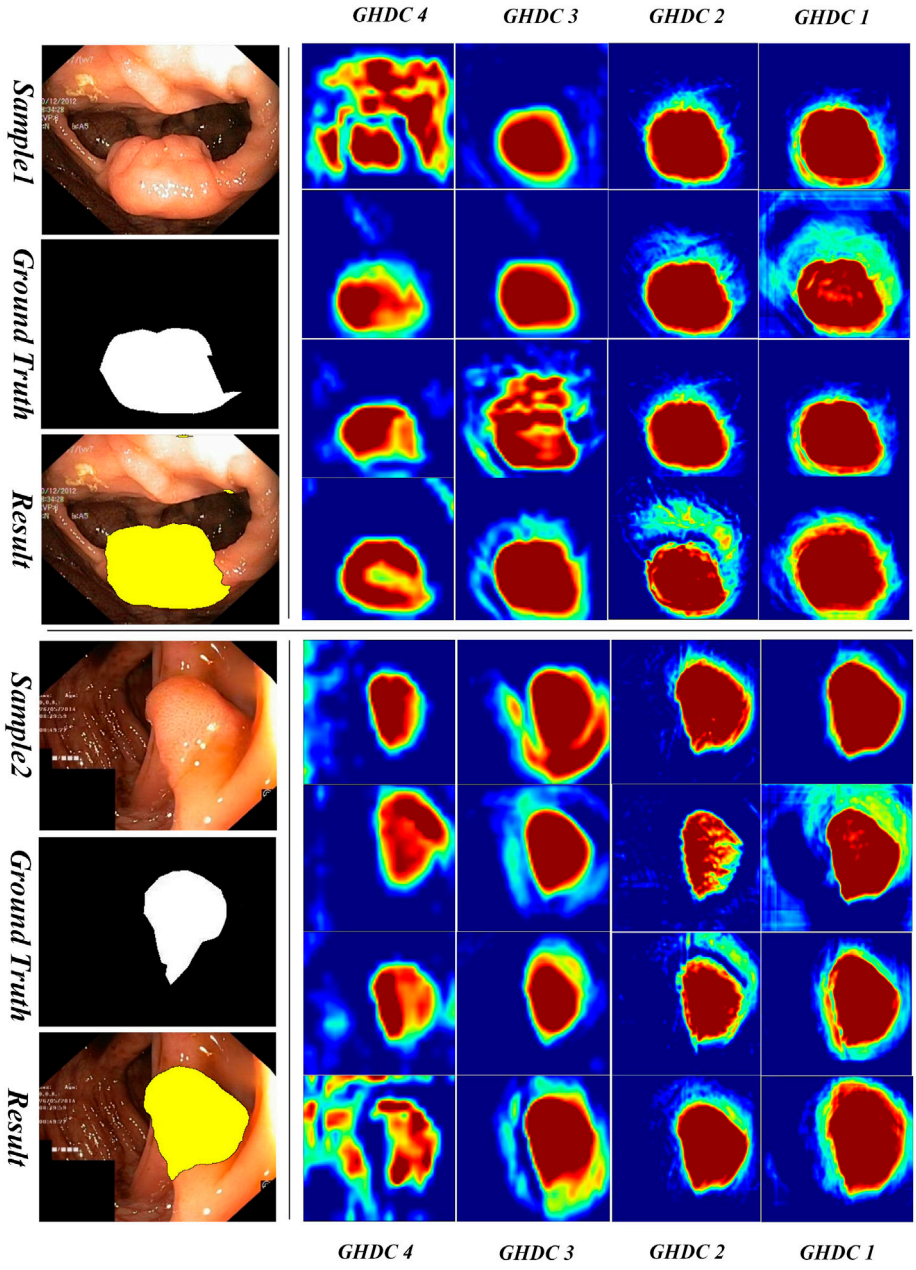

FIGURE 6
Visualization of the feature maps from the GHDC blocks on the Kvasir-SEG dataset.

Besides, we also conducted experiments to verify the advantage of the GHDC blocks compared the HDC on the two public colorectal polyp segmentation datasets. In particular, we compared the networks integrating the GHDC block 1 and/or the GHDC block 2 with the cases where the corresponding GHDC block(s) were/was replaced with the HDC(s). The HDCs in the experiments were implemented as proposed in (Wang et al., 2018) (i.e. the maximum number of cascaded convolutional layers was $n = 4$; the dilation rates and convolutional kernel sizes of the layers were taken as $R_1 = 1$, $R_2 = 2$, $R_3 = 5$, $R_4 = 9$ and $k_1 = k_2 = k_3 = k_4 = 3$, respectively). The experimental results are shown in Table 4.

From Table 4, we can observe that the networks integrating the GHDC blocks outperform the corresponding networks using the HDC on both of the two datasets, in terms of Dice, ASSD and mIoU values. As the receptive fields of both the GHDC block two and the GHDC block 1 are larger than that of the HDC, especially the receptive field generated by the GHDC block 1 is even more

TABLE 4 Experimental Results of GHDC and HDC Implementations Based on the Two Public Datasets of Colorectal Polyp Segmentations (The best results are marked in red bold font).

Results on the Kvasir-SEG Dataset

Network	Dilated Convolution Settings		RF	Dice (%)	ASSD (pix)	mIoU (%)
	Dilation Rate	Kernel Size				
Baseline	\	\	\	92.06 (±4.74)	0.51 (±0.20)	85.92 (±4.63)
Baseline + GHDC 2	**[1,2,5,7]**	**[3,3,3,5]**	**45**	**92.43(±4.31)**	**0.48(±0.67)**	**86.34(±4.69)**
Baseline + HDC	[1,2,5,9]	[3,3,3,3]	35	92.25 (±3.14)	0.49 (±0.50)	86.31 (±5.23)
Baseline + GHDC 1	**[1,2,5,7,9]**	**[3,3,3,5,5]**	**81**	**92.97(±3.86)**	**0.46(±0.42)**	**86.60(±5.53)**
Baseline + HDC	[1,2,5,9]	[3,3,3,3]	35	92.54 (±4.30)	0.48 (±0.22)	86.59 (±6.02)
Baseline + GHDC 2 + GHDC 1	–	–	–	**93.53(±3.66)**	**0.42(±0.17)**	**87.24(±4.03)**
Baseline + HDC + HDC	–	–	–	92.97 (±4.51)	0.45 (±0.29)	86.91 (±4.60)

Results on the CVC-ClinicDB Dataset

Network	Dilated Convolution Settings		RF	Dice (%)	ASSD (pix)	mIoU (%)
	Dilation Rate	Kernel Size				
Baseline	\	\	\	95.15 (±2.01)	0.32 (±0.40)	90.63 (±2.96)
Baseline + GHDC 2	**[1,2,5,7]**	**[3,3,3,5]**	**45**	**95.83(±2.42)**	**0.27(±0.33)**	**91.16(±6.61)**
Baseline + HDC	[1,2,5,9]	[3,3,3,3]	35	95.71 (±3.30)	0.28 (±0.44)	91.09 (±8.30)
Baseline + GHDC 1	**[1,2,5,7,9]**	**[3,3,3,5,5]**	**81**	**96.02(±5.28)**	**0.26(±0.27)**	**91.32(±4.72)**
Baseline + HDC	[1,2,5,9]	[3,3,3,3]	35	95.94 (±6.01)	0.27 (±0.45)	91.24 (±6.66)
Baseline + GHDC 2 + GHDC 1	–	–	–	**96.46(±5.00)**	**0.21(±0.09)**	**92.89(±6.25)**
Baseline + HDC + HDC	–	–	–	96.17 (±4.68)	0.22 (±0.34)	92.79 (±7.05)

than twice that of the HDC, the proposed GHDC blocks are able to exploit more features, resulting in more powerful networks.

3.3.2 Visual inspection of the feature maps obtained by GHDC blocks

In order to obtain a deeper understanding of the GHDC benefits, we analyzed the feature maps from the proposed GHDC blocks by visual inspection. In the proposed N-Net, the feature maps from the GHDC blocks 1, 2, 3 and 4 contain 64, 256, 512 and 1,024 channels, respectively. As The channels from the feature map of a GHDC block get similar features, four channels are randomly picked from the feature map from each GHDC block as representative channels. Results from two representative colonoscopy images in the testing set are shown in Figure 6, 7 for each of the two publicly available datasets, respectively.

From Figures 6, 7 we can observe that the feature map extracted by the GHDC blocks gets gradually richer details from the GHDC block 4 to the GHDC block 1. Compared to the GHDC blocks 3 and 4, the GHDC blocks 1 and 2 are able to capture more fine-grained features and generate feature maps with higher resolutions. Moreover, it is worth mentioning that the feature maps from all of the 4 GHDC blocks (not only the GHDC blocks 1 and 2) are highly correlated to the final segmentation results. This in turn illustrates that the semantic information exploited by each stage of the proposed N-Net is sufficiently utilized and that the transmission of multi-scale information is facilitated with the proposed GHDC blocks.

3.3.3 Comparative studies between the proposed N-Net and the state-of-the-art methods

Finally, we conducted comparative lesion region segmentation studies between the proposed N-Net and the state-of-the-art methods to verify the advantage of the proposed method. The experiments were conducted on the above publicly available polyp segmentation datasets on colonoscopy images. A variety of the state-of-the-art methods, including the U-Net (Ronneberger et al., 2015), the DeepLabV3+ (Chen et al., 2018), the U-Net++ (Zhou Z. et al., 2018), the U-Net+++ (Huang et al., 2020), Attention U-Net (Oktay et al., 2018), TransU-Net (Chen et al., 2021),OCR-Net (Wang et al., 2021) and CA-Net (Gu et al., 2021) were selected for comparison. The experimental results are shown in Table 5.

TABLE 5 Performance Comparison of the proposed N-Net to the State-of-the-art Methods on the Two Public Datasets of Colorectal Polyp Segmentations.

Results on the Kvasir-SEG Dataset.

Network	Parameter Size	Dice (%)	ASSD (pix)	mIoU (%)
U-Net	39.4 M	87.10 (±1.84)	0.82 (±0.15)	78.41 (±2.71)
U-Net++	47.2 M	88.16 (±3.72)	0.78 (±0.18)	80.12 (±5.40)
Attention U-Net	34.5 M	88.60 (±2.37)	0.81 (±0.31)	80.60 (±3.39)
U-Net+++	27.0 M	89.20 (±3.01)	0.73 (±0.30)	81.51 (±4.65)
CA-Net	44.4 M	89.58 (±3.40)	0.71 (±0.44)	82.17 (±5.19)
TransU-Net	133.4 M	91.84 (±3.44)	0.52 (±0.84)	85.71 (±5.41)
OCR-Net	70.4 M	92.24 (±2.85)	0.46 (±0.47)	86.26 (±4.52)
DeeplabV3+	39.6 M	**94.07(±1.73)**	**0.39(±0.22)**	**89.08(±2.85)**
N-Net	**20.6 M**	**94.45(±1.48)**	**0.38(±0.21)**	**89.80(±2.56)**

Results on the CVC-ClinicDB Dataset.

Network	Parameter Size	Dice (%)	ASSD (pix)	mIoU (%)
U-Net	39.4 M	91.58 (±1.99)	0.68 (±0.16)	85.29 (±3.04)
U-Net++	47.2 M	91.64 (±1.63)	0.71 (±0.25)	85.78 (±2.44)
Attention U-Net	34.5 M	93.53 (±1.78)	0.70 (±0.28)	88.39 (±2.88)
U-Net+++	27.0 M	94.70 (±1.31)	0.47 (±0.28)	90.30 (±2.16)
CA-Net	44.4 M	94.58 (±2.51)	0.47 (±0.43)	90.20 (±4.15)
TransU-Net	133.4 M	94.98 (±1.43)	0.33 (±0.15)	90.27 (±2.35)
OCR-Net	70.4 M	95.35 (±1.44)	0.39 (±0.19)	91.42 (±2.46)
DeeplabV3+	39.6 M	**96.87(±0.75)**	**0.19(±0.06)**	**94.14(±1.36)**
N-Net	**20.6 M**	**97.03(±0.82)**	**0.16(±0.06)**	**94.35(±1.49)**

Through the segmentation performance comparisons on the Kvasir-SEG dataset, it is important to mention that the N-Net can guarantee the best segmentation performance with the parameter amount minimized. The reduction of the parameters in turn greatly improves the efficiency of the algorithm. From Table 5, we can also observe that the performance of DeepLabV3+ is the closest to the N-Net with the above three metrics. However, the parameter size of 39.6 M is needed for the DeepLabV3+, while only 20.6 M is needed for the N-Net. Compared with the N-Net, the metrics of Dice, ASSD and mIoU for the U-Net+++ gets worse by 5.25%, 0.35 pix and 8.29%, respectively, though the size of its parameters is increased by 6.4 M.

From Table 5, we can also observe that the N-Net shows the best segmentation performance on the CVC-ClinicDB dataset with the metrics of Dice, ASSD and mIoU reaching 97.03%, 0.16 pix, and 94.35%, respectively, with the smallest computational parameter amount.

To be more intuitive, we also compared the lesion region segmentation results of the above networks. Some representative comparison results are shown in Figures 8, 9, respectively. The segmentation results from the colonoscopy images are marked in yellow. From those figures, we observe that the segmentation results with the proposed N-Net method is the closest to the ground truth. Moreover, the isolated false predictions by the proposed N-Net are also minimized. This is because the proposed N-Net is able to utilize the multi-scale features sufficiently, with the receptive fields ensured during the multi-scale information transmissions.

4 Discussion

In this paper, we propose a new N-shaped deep neural network (N-Net) structure to conduct the lesion region segmentations of the colorectal polyps from colonoscopy images. To facilitate the multi-scale information transmissions, we propose a strategy of generalized hybrid dilated convolution (GHDC) designing which enables flexible dilation rates and convolutional kernel sizes to facilitate transmission of the multi-scale information. Based on the proposed strategy of GHDC designing, we design four GHDC blocks to connect the encoding path and the decoding path of the N-Net.

The proposed method was evaluated on two publicly available colonoscopy image datasets for polyp segmentations: the Kvasir-SEG dataset and the CVC-ClinicDB dataset. The advantages of the proposed GHDC blocks were demonstrated through multiple sets of

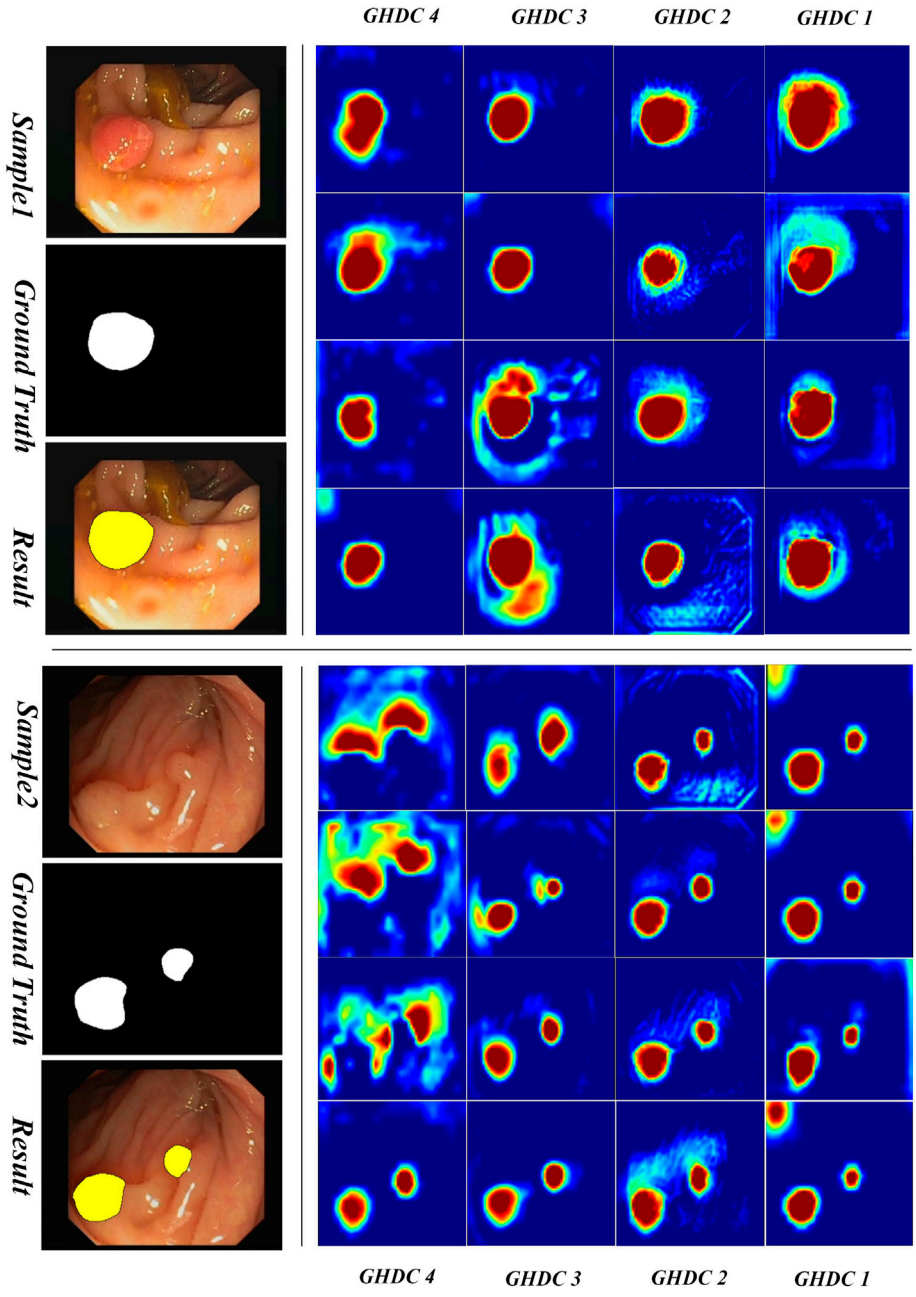

FIGURE 7
Visualization of the feature maps from the GHDC blocks on the CVC-ClinicDB dataset.

ablation experiments. In addition, the interpretability of the proposed GHDC blocks was analyzed through the visualization of the feature maps. Moreover, through comparative studies, the proposed N-Net was shown to outperform the state-of-the-art CNNs, including DeepLabV3+, TransU-Net and CA-Net, with the metrics of Dice, ASSD and mIoU as 94.45%, 0.38 pix and 89.80% on the Kvasir-SEG dataset and 97.03%, 0.16 pix and 94.35% on the CVC-ClinicDB dataset, respectively.

In this paper, the research was conducted on two publicly available datasets of polyp segmentations for colonoscopy images, the Kvasir-SEG dataset and the CVC-ClinicDB dataset, where consistency were shown in the results. As the patient amount included in the two datasets was still limited, additional datasets (Vázquez et al., 2017; Misawa et al., 2020; Sánchez-Peralta et al., 2020; Li et al., 2021; PIBAdb, 2022) can also be considered in the algorithm

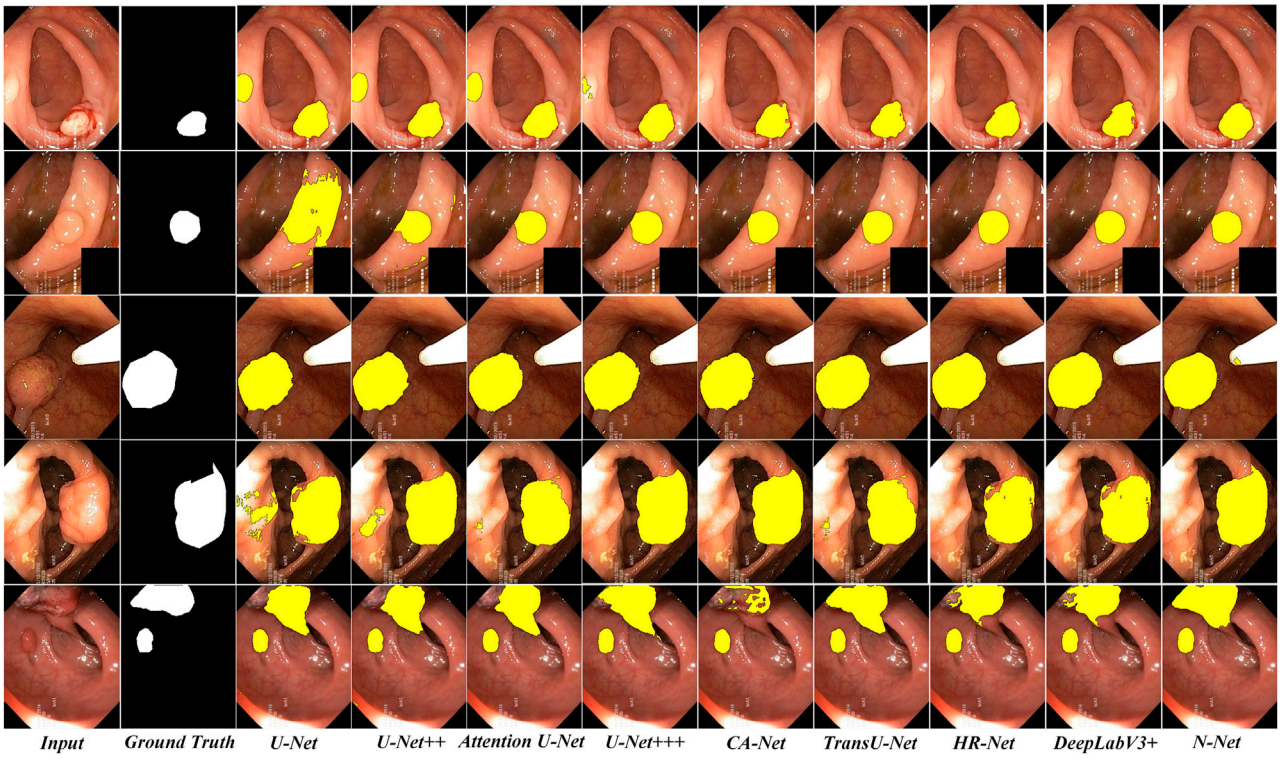

FIGURE 8
Segmentation results of the N-Net and the state-of-the-art methods on representative images of the Kvasir-SEG dataset.

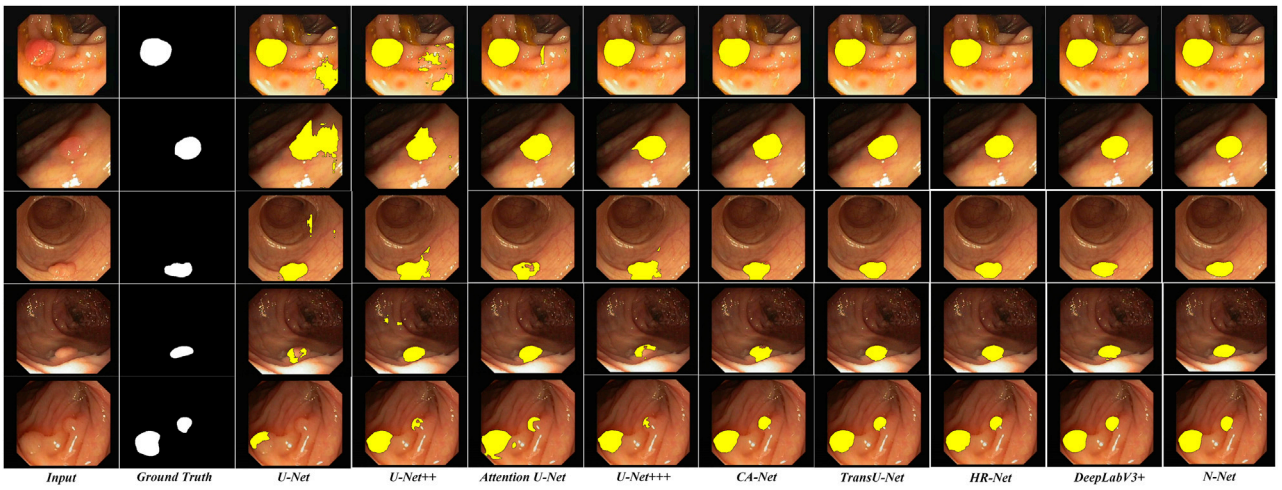

FIGURE 9
Segmentation results of the N-Net and the state-of-the-art methods on representative images of the CVC-ClinicDB dataset.

development and validations in our future work. In addition, we will also explore to implement the proposed GHDC theory and the designed blocks in other biomedical image segmentation tasks in our future work.

5 Conclusion

In this work, we proposed an N-Net structure based on the encoding-decoding structure to conduct the polyp lesion region segmentations of colonoscopy images. In the proposed N-Net, the pre-trained DenseNet module was transferred as the encoding path of the network. In particular, we proposed a strategy of generalized hybrid dilated convolution (GHDC) designing to facilitate transmission of the multi-scale information and expand the respective fields. Based on the strategy of GHDC designing, four GHDC blocks were designed to connect the encoding path and the decoding path of the N-Net. Experiments were performed on two publicly available colorectal polyp lesion region segmentation dataset: the Kvasir-SEG dataset and the CVC-ClinicDB dataset. The advantages of the GHDC blocks were verified. Moreover, experimental results also showed that the proposed N-Net outperforms with a small amount of parameters, compared with the state-of-the-art methods.

Data availability statement

Two publicly available polyp segmentation datasets of the colonoscopy images (the Kvasir-SEG dataset and the CVC-ClinicDB dataset) were used in the experiments of this study. The Kvasir-SEG dataset can be found at: https://datasets.simula.no/kvasir-seg/. The CVC-ClinicDB dataset can be found at: https://www.kaggle.com/datasets/balraj98/cvcclinicdb.

Author contributions

RC, RY, and FL proposed the main idea. RY implemented the experiments and analyze the result. RC wrote the manuscript and participated in the design of the experiments. FL supervised the project, wrote part of the manuscript and revised, read, and approved the final manuscript. CC conducted the biomedical research on the colonoscopy images, and participated in the manuscript revision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 61901233) and the Natural Science Foundation of Tianjin City of Peoples Republic of China (grant no. 19JCQNJC00900).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alom, Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *arXiv Preprint* , arXiv:1802.06955.

Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., and Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* 43, 99–111. doi:10.1016/j.compmedimag.2015.02.007

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv Preprint*, arXiv:2102.04306.

Chen, L., Zhu, Y., George, P., Florian, S., and Hartwig, A. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer vision – eccv 2018* (Cham: Springer), 833–851. doi:10.1007/978-3-030-01234-2_49

Cheng, W., Lu, J., Zhu, X., Hong, J., Liu, X., Li, M., et al. (2020). Dilated residual learning with skip connections for real-time denoising of laser speckle imaging of blood flow in a log-transformed domain. *IEEE Trans. Med. Imaging* 39, 1582–1593. doi:10.1109/TMI.2019.2953626

Fu, X., Cai, N., Huang, K., Wang, H., Wang, P., Liu, C., et al. (2019). M-Net: A novel U-net with multi-stream feature fusion and multi-scale dilated convolutions for bile ducts and hepatolith segmentation. *IEEE Access* 7, 148645–148657. doi:10.1109/ACCESS.2019.2946582

Ganz, M., Yang, X., and Slabaugh, G. (2012). Automatic segmentation of polyps in colonoscopic narrow-band imaging data. *IEEE Trans. Biomed. Eng.* 59, 2144–2151. doi:10.1109/TBME.2012.2195314

Gu, R., Wang, G., Song, T., Huang, R., Aertsen, M., Deprest, J., et al. (2021). CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Trans. Med. Imaging* 40, 699–711. doi:10.1109/TMI.2020.3035253

He, B., Zhang, Y., Zhou, Z., Wang, B., Liang, Y., Lang, J., et al. (2020). A neural network framework for predicting the tissue-of-origin of 15 common cancer types based on RNA-Seq data. *Front. Bioeng. Biotechnol.* 8, 737. doi:10.3389/fbioe.2020.00737

Huang, G., Liu, Z., Laurens, V. D. M., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017, 2261–2269. doi:10.1109/CVPR.2017.243

Huang, H., Lin, L., Tong, R., Hu, H., and Wu, J. (2020). UNet 3+: A full-scale connected UNet for medical image segmentation, 1055–1059. doi:10.1109/ICASSP40776.2020.9053405

Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., et al. (2019). "Kvasir-SEG: A segmented polyp dataset," in *International Conference on MultiMedia Modeling* (Springer, Cham), 451–462. doi:10.1007/978-3-030-37734-2_37

Kingma, D., and Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv Preprint*, arXiv:1412.6980.

Li, K., Fathan, M. I., Patel, K., Zhang, T., Zhong, C., Bansal, A., et al. (2021). Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations. *Plos One* 16, e0255809. doi:10.1371/journal.pone.0255809

Lima Pereira, A. A., Dos Santos Fernandes, G., Braga, G. T. P., Regina Marchetti, K., do Couto Mascarenhas, C., Gumz, B., et al. (2020). Differences in pathology and mutation status among colorectal cancer patients younger than, older than, and of screening age. *Clin. Colorectal Cancer* 19, 264–271. doi:10.1016/j.clcc.2020.06.004

Liu, J., Han, Y. J., Liu, T., Aguilera, N., and Tam, J. (2020). Spatially aware Dense-LinkNet based regression improves fluorescent cell detection in adaptive optics ophthalmic images. *IEEE J. Biomed. Health Inf.* 24, 3520–3528. doi:10.1109/JBHI.2020.3004271

Ma, Y., Shuai, H., and Cheng, W. (2022). Spatiotemporal dilated convolution with uncertain matching for video-based crowd estimation. *IEEE Trans. Multimed.* 24, 261–273. doi:10.1109/TMM.2021.3050059

Misawa, M., Kudo, S., Mori, Y., Hotta, K., Ohtsuka, K., Matsuda, T., et al. (2020). Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointest. Endosc.* 93, 960–967.e3. e3. doi:10.1016/j.gie.2020.07.060

Oktay, O., Schlemper, J., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., et al. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv Preprint*, arXiv:1804.03999.

PIBAdb (2022). Colorectal polyp image cohort. Available Online at: https://www.iisgaliciasur.es/home/biobanco/colorectal -polyp-image-cohort-pibadb/?lang=en (Accessed July, 2022).

Qadir, H. A., Balasingham, I., Solhusvik, J., Bergsland, J., Aabakken, L., and Shin, Y. (2020). Improving automatic polyp detection using CNN by exploiting temporal dependency in colonoscopy video. *IEEE J. Biomed. Health Inf.* 24, 180–193. doi:10.1109/JBHI.2019.2907434

Ren, Y., Ma, J., Xiong, J., Chen, Y., Lu, L., and Zhao, J. (2019). Improved false positive reduction by novel morphological features for computer-aided polyp detection in CT colonography. *IEEE J. Biomed. Health Inf.* 23, 324–333. doi:10.1109/JBHI.2018.2808199

Ronneberger, O., Fischer, P., Brox, T., and Guo, S. H. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention – miccai* (Cham: Springer), 234–241. doi:10.1007/978-3-319-24574-4_28

Rozo, A., Moeyersons, J., Morales, J., Garcia van der Westen, R., Lijnen, L., Smeets, C., et al. (2022). Data augmentation and transfer learning for data quality assessment in respiratory monitoring. *Front. Bioeng. Biotechnol.* 10, 806761. doi:10.3389/fbioe.2022.806761

Rundle, A. G., Lebwohl, B., Vogel, R., Levine, S., and Neugut, A. I. (2008). Colonoscopic screening in average-risk individuals ages 40 to 49 vs 50 to 59 years. *Gastroenterology* 134, 1311–1315. doi:10.1053/j.gastro.2008.02.032

Sánchez-Peralta, L. F., Pagador, J. B., Picón, A., Calderón, J., Polo, F., Andraka, N., et al. (2020). Piccolo white-light and narrow-band imaging colonoscopic dataset: A performance comparative of models and datasets. *Appl. Sci.* 10, 8501. doi:10.3390/app10238501

Sasmal, P., Iwahori, Y., Bhuyan, M. K., and Kasugai, K. (2018). "Active contour segmentation of polyps in capsule endoscopic images," in 2018 International Conference on Signals and Systems (ICSigSys), Bali, Indonesia, 01-03 May 2018, 201–204. doi:10.1109/ICSIGSYS.2018.8372666

Shao, L., Zhu, F., and Li, X. (2015). Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1019–1034. doi:10.1109/TNNLS.2014.2330900

Sun, Z., Xiao, Y., You, L., Yin, L., Yang, P., and Li, S. (2016). Cross-entropy-based antenna selection for spatial modulation. *IEEE Commun. Lett.* 20, 622–625. doi:10.1109/LCOMM.2016.2518662

Tajbakhsh, N., Gurudu, S. R., and Liang, J. (2016). Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans. Med. Imaging* 35, 630–644. doi:10.1109/TMI.2015.2487997

Tan, J., Gao, Y., Liang, Z., Cao, W., Pomeroy, M. J., Huo, Y., et al. (2020). 3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix-based cnn model for polyp classification via ct colonography. *IEEE Trans. Med. Imaging* 39, 2013–2024. doi:10.1109/TMI.2019.2963177

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.*, 65. 87–108. doi:10.3322/caac.21262

Van Rijn, J. C., Reitsma, J. B., Stoker, J., Bossuyt, P. M., Van Deventer, S. J., and Dekker, E. (2006). Polyp miss rate determined by tandem colonoscopy: A systematic review. *Am. J. Gastroenterol.* 101, 343–350. doi:10.1111/j.1572-0241.2006.00390.x

van Toledo, D. E. F. W. M., IJspeert, J. E. G., Bossuyt, P. M. M., Bleijenberg, A. G. C., van Leerdam, M. E., van der Vlugt, M., et al. (2022). Serrated polyp detection and risk of interval post-colonoscopy colorectal cancer: A population-based study. *Lancet Gastroenterology Hepatology* 7, 747–754. doi:10.1016/S2468-1253(22)00090-5

Vázquez, D., Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., López, A. M., Romero, A., et al. (2017). A benchmark for endoluminal scene segmentation of colonoscopy images. *J. Healthc. Eng.* 2017, 1–9. doi:10.1155/2017/4037190

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., et al. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3349–3364. doi:10.1109/TPAMI.2020.2983686

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., et al. (2018). "Understanding convolution for semantic segmentation," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12-15 March 2018, 1451–1460. doi:10.1109/WACV.2018.00163

Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P. A. (2017). Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE J. Biomed. Health Inf.* 21, 65–75. doi:10.1109/JBHI.2016.2637004

Zhang, Q., and Yang, Y. (2021). "SA-Net: Shuffle attention for deep convolutional neural networks," in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 06-11 June 2021, 2235–2239. doi:10.1109/ICASSP39728.2021.9414568

Zhou, L., Zhang, C., and Ming, W. (2018). "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 18-22 June 2018, Salt Lake City, UT, USA, 192–1924. doi:10.1109/CVPRW.2018.00034

Zhou, Z., Siddiquee, M., Tajbakhsh, N., and Liang, J. (2018). "UNet++: A nested U-net architecture for medical image segmentation," in *International workshop on deep learning in medical image analysis* (Springer, Cham), 3–11. doi:10.1007/978-3-030-00889-5_1