

CONTENTS

ENDORSEMENT PAGE	ii
PAGE OF DEDICATION	iii
STATEMENT	iv
PREFACE	vi
NOMENCLATURE AND ABBREVIATION	viii
ABSTRACT	ix
INTISARI	x
CONTENTS	xi
LIST OF FIGURES	xiii
LIST OF TABLES	xviii
CHAPTER I Introduction	1
1.1 Background	1
1.2 Motivation	2
1.3 Contributions	4
1.4 Demarcations	6
1.5 Organization	8
CHAPTER II Preliminaries and Related Works	9
2.1 Fully Connected Neural Networks	9
2.1.1 Forward Propagation	10
2.1.2 Feedforward Neural Networks	11
2.1.3 Backward Propagation	12
2.2 Convolutional Neural Networks	14
2.2.1 Forward Propagation	15
2.2.2 Backward Propagation	19
2.3 Involution	21
2.4 Normalization Layers	23
2.4.1 Batch Normalization	23
2.4.2 Instance Normalization	25
2.5 Residual Connection	26
2.6 Lightweight CNN Architectures	28
2.7 Lightweight Gastrointestinal Polyp Segmentation Models	31
CHAPTER III Method	35
3.1 Architecture	35
3.1.1 ConvInCEDNet Efficient Convolution Block	35
3.1.2 ConvInCEDNet Involution Block	41
3.1.3 Encoder	43
3.1.4 Bridge	44
3.1.5 Decoder	45
3.1.6 Replacement With Instance Normalization	46
3.2 Assessment Approach	46
3.2.1 Datasets	46
3.2.2 Evaluation Metrics	52

3.2.3	Computational Performance Frameworks	55
3.2.4	Statistical Analysis	55
CHAPTER IV	Experiments	57
4.1	Experimental Setup	57
4.1.1	Training Setup	57
4.1.2	Testing Setup	59
4.2	Comparison With Prior Models	59
4.2.1	Segmentation Performance	60
4.2.1.1	CVC-ClinicDB	60
4.2.1.2	CVC-ColonDB (extended)	64
4.2.1.3	Endotect	65
4.2.1.4	ETIS-LaribPolypDB	69
4.2.1.5	KvasirCapsule-SEG	71
4.2.1.6	Kvasir-SEG	72
4.2.2	Computational Performance	74
4.3	Ablation Study	77
4.3.1	Downsampling-Upsampling Choices	78
4.3.2	Stem Layer	81
4.3.3	Generalization Capacity Enhancement	83
4.3.4	CInvo Spatial-Channel Attention Effectiveness	85
4.3.5	Involution Non-Linearity	87
4.3.6	Involution Kernel Size	87
4.3.7	ECA Module Kernel Size	87
4.3.8	CEConv Module Cheap Operation Kernel Size	91
CHAPTER V	Conclusion and Future Work	94
5.1	Conclusion	94
5.2	Future Work	95
REFERENCES	97

LIST OF FIGURES

Figure 1.1	Segmentation region (greenish) of the gastrointestinal polyps.	1
Figure 2.1	An illustration of an output neuron from a fully connected neural network. It receives the aggregation of signals from all independent input neurons (combined with respective weights) and a bias. This aggregation is then activated via a certain function, which can be linear or non-linear, producing a numeric value for the output.	10
Figure 2.2	The visualization of (a) a fully connected neural network and (b) a feedforward neural network. The main difference between these two is the hidden layer in the middle of the network. The bias term is omitted for convenience.	11
Figure 2.3	CNN layers (first part).	16
Figure 2.4	CNN layers (second part).	17
Figure 2.5	The graphical abstraction of involution in the 2D case. Here ϕ denotes a non-linear mapping pixel-to-kernel via a feedforward neural network, \otimes refers to element-wise multiplication or Hadamard product, and \oplus is a spatial aggregation.	21
Figure 2.6	(a) Batch normalization for 2D data normalizes the entire constituent within the height, width, and mini-batch of the feature at a particular channel; (b) Instance normalization for 2D data normalizes the whole element across the height and width at a certain channel and instance of the mini-batch.	25
Figure 2.7	A visualization of a residual network block consisting of the three-layered neural network at the main path (a) and a shortcut path (b).	27
Figure 2.8	The visualization of four lightweight blocks: (a) the MobileNetV3 block, (b) the ShuffleNetV2 unit, (c) the IGCV3 block, and (d) the Ghost module. Each square represents a channel of the feature map. Here, blue connections are the pointwise convolution, green connections are the depthwise convolution (typically 3×3 kernel), yellow connections are the element-wise multiplication, black connections are the identity operation, cyan connections are the channel shuffle operation, red connections are 1×1 grouped convolution, and pink connections are the depthwise convolution followed by the channel shuffle operation. The normalization layer, activation function, and skip connection are omitted for simplicity.	29

Figure 2.9	Segmentation results denoted by the greenish mask from pre-trained U ² -Net [1] (a deep learning architecture specifically designed for salient object detection) on several samples of the ETIS-LaribPolypDB dataset. The samples contain salient polyps (a) and (c) and camouflaged polyps (b), (d), and (e). The ground truth is indicated by the red bounding box. (a) The model successfully recognizes the salient polyp. (b) The segmentation region includes the camouflaged polyp, yet it does not accurately indicate the actual polyp location. (c) The salient polyp is detected, but it still incorporates the background. (d) The model fails to segment the camouflaged polyp completely. (e) The model does not detect the camouflaged polyp, but it detects the background.	31
Figure 3.1	The blueprint of ConvInCEDNet. Here, CEConv and CInvo correspond to the ConvInCEDNet efficient convolution block and the ConvInCEDNet involution block, respectively. Width multiplier G determines the channel expansion size of ConvInCEDNet. The downsampling stages of ConvInCEDNet require the stride S of two. The abbreviations <i>in</i> , <i>mid</i> , and <i>out</i> refer to the number of <i>input</i> channels, the number of <i>middle</i> channels, and the number of <i>output</i> channels of a certain module, respectively.	36
Figure 3.2	Four forms of ConvInCEDNet efficient convolution (CEConv) block: (a) the standard CEConv, (b) the downsampling CEConv, (c) the CEConv with attention, and (d) the residual CEConv with attention.	37
Figure 3.3	(a) The CEConv module with ReLU; (b) The CEConv module without ReLU.	39
Figure 3.4	The schematic of ConvInCEDNet Involution (CInvo) block. ...	41
Figure 3.5	The distribution of the seven polyp datasets, i.e., PolypGen, CVC-ClinicDB, CVC-ColonDB (extended), Endotect, ETIS-LaribPolypDB, KvasirCapsule-SEG, and Kvasir-SEG, based on the 2D latent space of VQ-VAE [2].	47
Figure 3.6	Some image-mask pair specimens from PolypGen.	48
Figure 3.7	Several samples from CVC-ClinicDB.	49
Figure 3.8	Some exemplars of images (top) and the related masks (bottom) of CVC-ColonDB (extended).	49
Figure 3.9	The examples of the image-mask pairs of Endotect.	50
Figure 3.10	The polyp in ETIS-LaribPolypDB can be big (left), tiny (right), or multi-presence (middle).	50
Figure 3.11	Typical image-mask pairs of KvasirCapsule-SEG: barely any room for the background, just the polyp.	51

Figure 3.12	Several selected samples of image-mask pairs of Kvasir-SEG. The image may contain a close view of a polyp (right), an occluded polyp due to the imaging artifact (middle), and multiple polyps in a scene.	51
Figure 3.13	Red, blue, green, and white regions denote false positives, false negatives, true positives, and true negatives, respectively. Eight possible cases of image segmentation evaluation are: (a) the predicted foreground matches perfectly with the ground-truth foreground; (b) the predicted foreground overwhelms the image and successfully captures the ground-truth foreground; (c) the predicted foreground does not wrongfully claim the non-polyp region as the polyp, yet it does not fully populate the correct foreground; (d) the most common occurrence case, in which the foreground populates the image yet it does not fully capture the valid foreground and considers the non-polyp region as the polyp; (e) no ground-truth foreground is available, but the predicted foreground exists anyway; (f) no predicted foreground exists; (g) no presence of the predicted foreground or ground-truth foreground; and (h) both ground truth and predicted foregrounds exist, yet there is no overlapping region.	53
Figure 4.1	Some of the best-performing (near-perfect segmentation) results of the proposed models on the CVC-ClinicDB dataset. The red, blue, and green regions denote false positive, false negative, and true positive, respectively.	61
Figure 4.2	Some of the worst-performing results (polyps are overlooked) of the proposed models on the CVC-ClinicDB dataset. The red, blue, and green regions denote false positive, false negative, and true positive, respectively.	62
Figure 4.3	Qualitative intra-comparisons of several segmentation results of the proposed models on the CVC-ClinicDB dataset. The red, blue, and green regions denote false positive, false negative, and true positive, respectively.	63
Figure 4.4	Qualitative comparisons of several samples of the best-performing results of the proposed models and the prior models on the CVC-ColonDB (extended) dataset. The reddish region indicates a false positive. The bluish region indicates a false negative. The greenish region indicates a true positive.	66
Figure 4.5	Qualitative comparisons of several samples of the worst-performing results of the proposed models and the prior models on the CVC-ColonDB (extended) dataset. The reddish region indicates a false positive. The bluish region indicates a false negative. The greenish region indicates a true positive.	67

Figure 4.6	Qualitative intra-comparisons of some segmentation results of the proposed models on the CVC-ColonDB (extended) dataset. The reddish region indicates a false positive. The bluish region indicates a false negative. The greenish region indicates a true positive.	68
Figure 4.7	Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the big-sized polyp of the proposed models and the prior models on the Endotect dataset. The reddish region denotes a false positive. The bluish region denotes a false negative. The greenish region denotes a true positive.	69
Figure 4.8	Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the small-sized polyp of the proposed models and the prior models on the Endotect dataset. The reddish region denotes a false positive. The bluish region denotes a false negative. The greenish region denotes a true positive.	70
Figure 4.9	Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the big-sized polyp of the proposed models and the prior models on the ETIS-LaribPolypDB dataset. The reddish region represents a false positive. The bluish region represents a false negative. The greenish region represents a true positive.	72
Figure 4.10	Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the small-sized polyp of the proposed models and the prior models on the ETIS-LaribPolypDB dataset. The reddish region indicates a false positive. The bluish region indicates a false negative. The greenish region indicates a true positive.	73
Figure 4.11	Qualitative intra-comparisons of some segmentation results of the proposed models on the ETIS-LaribPolypDB dataset. The reddish region denotes a false positive. The bluish region indicates a false negative. The greenish region represents a true positive.	74
Figure 4.12	Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the proposed models and the prior models on the KvasirCapsule-SEG dataset. The reddish region represents a false positive. The bluish region denotes a false negative. The greenish region marks a true positive.	76

- Figure 4.13 Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the big-sized polyp of the proposed models and the prior models on the Kvasir-SEG dataset. The reddish region represents a false positive. The bluish region denotes a false negative. The greenish region indicates a true positive..... 77
- Figure 4.14 Qualitative comparisons of samples of the worst-performing (top) and the best-performing (bottom) results of the small-sized polyp of the proposed models and the prior models on the Kvasir-SEG dataset. The reddish region denotes a false positive. The bluish region indicates a false negative. The greenish region represents a true positive. 78

LIST OF TABLES

Table 2.1	Summary of prior models and the proposed model for more efficient gastrointestinal polyp segmentation.	34
Table 3.1	The summary of datasets employed in the experiments.	47
Table 4.1	Quantitative results of the models on the CVC-ClinicDB dataset. The best outcomes are in boldface. The first-best, the second-best, and the third-best are denoted with ¶, §, and ◇, respectively.	60
Table 4.2	Quantitative results of the models on the CVC-ColonDB (extended) dataset. Bold texts indicate the best results. The first-best, the second-best, and the third-best are denoted with ¶, §, and ◇, respectively.	64
Table 4.3	The quantitative comparison of the segmentation performance of the models on the Endotect dataset. Bold texts indicate the best results. The first-best, the second-best, and the third-best are denoted with ¶, §, and ◇, respectively.	65
Table 4.4	The quantitative comparison of the segmentation performance of the models on the ETIS-LaribPolypDB dataset. The bold texts with ¶ indicate the first-best results. The second-best results are denoted by the bold texts with §. The results with ◇ typed in boldface are the third-best.	71
Table 4.5	The quantitative comparison of the segmentation performance of the models on the KvasirCapsule-SEG dataset. The bold texts with ¶ indicate the first-best results. The second-best results are denoted by the bold texts with §. The results with ◇ typed in boldface are the third-best.	75
Table 4.6	Quantitative results of the segmentation performance of the models on the Kvasir-SEG dataset. The bold texts with ¶ indicate the first-best results. The second-best results are denoted by the bold texts with §. The results with ◇ typed in boldface are the third-best.	75
Table 4.7	Computational performance of the models. The best outcome is typed in bold.	76
Table 4.8	Comparison between alternative combinations of downsampling and upsampling methods. MP, SC, TC, BI, and MU stand for max pooling, strided convolution, transposed convolution, bilinear interpolation, and max unpooling. REC, PRC, and SPC indicate recall, precision, and specificity. Boldface text denotes the best outcome.	79

Table 4.9	Comparison between ConvInCEDNet with strided convolution as the stem layer and ConvInCEDNet with the downsampling CEConv block as the stem layer. REC, PRC, and SPC indicate recall, precision, and specificity. The text written in bold denotes the best outcome.	82
Table 4.10	Comparison between ConvInCEDNet with instance normalization and ConvInCEDNet with batch normalization. REC, PRC, and SPC denote recall, precision, and specificity. The text written in bold indicates the best result.	84
Table 4.11	Comparison between the baseline of the proposed model, the proposed model without the modified involution, the proposed model without the ECA module in the CInvo block, and the proposed model without both the modified involution and the ECA module in the CInvo block. REC, PRC, and SPC indicate recall, precision, and specificity. Boldface text denotes the best outcome.	86
Table 4.12	Comparison between the different employments of non-linearity after involution kernel generation in ConvInCEDNet. REC, PRC, and SPC refer to recall, precision, and specificity. The text in bold indicates the best result.	88
Table 4.13	The results of the ablation study on the kernel size of the modified involution. REC, PRC, and SPC denote recall, precision, and specificity. The text written in boldface represents the best outcome.	89
Table 4.14	The yields of the ablation study on the kernel size of the ECA module. REC, PRC, and SPC denote recall, precision, and specificity. The text typed in boldface represents the best outcome.	90
Table 4.15	The outcomes of the ablation study on the kernel size of the cheap operation of the CEConv module. REC, PRC, and SPC represent recall, precision, and specificity. Bold text indicates the best yield.	92