

# Employing Representative Batch Normalization in Search Identification Network (SINet) for Polyp Image Segmentation

Adarsha Dhakal <sup>a</sup>, Sharad Kumar Ghimire <sup>b</sup>, Ganesh Gautam <sup>c</sup>

<sup>a, b, c</sup> Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal

✉ <sup>a</sup> adarshadhakal02@gmail.com, <sup>b</sup> skghimire@ioe.edu.np, <sup>c</sup> ganesh@pcampus.edu.np

## Abstract

Images of polyps obtained from colonoscopies can provide valuable insights into different polyp types and their associations with certain diseases. We are introducing an analysis that segments these polyp images. Oftentimes, polyps blend into the background due to their intensity being similar to it. Hence, it's vital to choose a model that can identify subtle details in images, especially when objects, like polyps, resemble their surroundings. Therefore, we have applied the SINet model, originally designed for concealed object detection, to our Polyp Image Segmentation (PIS) study. To enhance SINet's feature extraction, we adapted it with Representative Batch Normalization (RBN). Traditional Batch Normalization (BN) in SINet can sometimes add noise to the data. RBN addresses this by using methods like centering and scaling calibration, ensuring a cleaner and more consistent feature distribution. We've compared our adapted SINet-RBN with the original SINet and PraNet. For training, we used the clinicDB Dataset, and for testing and evaluation, we employed the LaribpolypDB, ColonDB, and CVC-300 datasets.

## Keywords

Concealed, Search Identification Network, Object Detection, Representative Batch Normalization, Feature Calibration, Polyp Image Segmentation

## 1. Introduction

Object detection [1], a computer vision technology, empowers us to identify and pinpoint specific elements or objects within an image or video. This technology not only allows for item counting within a given scene but also enables real-time tracking and accurate labeling through the utilization of localization and identification [2]. Object detection has shown to be extremely beneficial and applicable in a variety of fields, and the discipline has made incredible progress. Despite improvements in object detection, there are still obstacles to overcome, making it difficult to detect some objects, such as polyps that are somewhat camouflaged with the background and can provide or generate various types of significant data that can be helpful in knowledge discovery, analysis, and to improve computer vision. Objects that are camouflaged or concealed in images, like polyps, are challenging to find but can yield valuable information. Many researchers and developers are interested in this field because it is challenging yet valuable. Studies on them are still rare, despite recent increases in interest, primarily because there aren't enough good models or large enough datasets available.

Detecting things that are concealed or camouflaged [3] in visual settings is the goal of the computer vision task known as camouflaged/concealed object detection (COD). Comprehending the concept of COD possesses significance not only from a scientific standpoint but also holds practical implications for various foundational domains, including computer vision, agriculture, medicine, segmentation, security, forensic analysis, etc [4]. Robots that can detect, distinguish, and differ from anything in the scene will be amazing. That's where COD can be very beneficial and can enhance computer vision. One of the uses for COD is polyp

detection. For an early diagnosis of colorectal cancer, [5] polyp detection is crucial. Nevertheless, COD is difficult because of the wide range of object sizes, hazy object boundaries, and textural similarity between the objects and their surroundings.

Automated polyp detection[6] and segmentation can aid doctors in avoiding manual, time-consuming, and sometimes even promote accurate segmentation. A polyp is an abnormal growth of tissue that protrudes from the skin or mucous membrane of the body. The colon and rectum, ear canal, cervix, stomach, nose, uterus, throat, and bladder are just a few areas of the body where polyps can form. This paper discusses polyps, which develop on the inner lining of the colon or rectum and, when left unaddressed, can progress into cancer. Consequently, timely identification and removal of polyps play a vital role in averting colorectal cancer. Image segmentation techniques can help identify and locate polyps in medical images such as colonoscopy or computed tomography (CT) scans [7].

Polyp segmentation involves separating the polyp region from the background region in the image. This can be challenging due to factors such as noise, lighting variations, and complex shapes and textures of polyps. Accurate polyp segmentation can aid in the development of computer-aided diagnosis (CAD) [8] systems for detecting and diagnosing polyps. These tools can help medical professionals like radiologists spot problematic areas in images, which can help find polyps earlier and increase the likelihood that therapy will be effective. Polyp image segmentation(PIS) can also speed up colonoscopy processes by minimizing the time and effort needed for manual inspection. Identifying polyps at an early stage, may result in quicker and more accurate diagnosis and even save lives. Polyp image segmentation is far more difficult

to identify and discriminate due to the significant inherent similarity of intensity between the background and the target object than the traditional object detection task [4] [9]. Unlike other object detection techniques, the segmentation of polyp images can be thought of as a challenging example of generic objects.

In order to extract the specific attributes of hidden objects, COD requires fine-grained textures and features of images so better concealed object detection models [4] [10] is required to make applicable on PIS. Feature extraction is the first step in every object detection technique. The performance of any object recognition model depends on how well it can extract features and isolate noise from features. SINet has proven promising results in concealed object detection. For feature extraction, SINet employs res2net, which captures multi-scale features at a fine-grained level and expands the receptive field range for each layer within the network [11]. Res2net uses Batch Normalization as a normalizing technique. Batch Normalization does well on SINet, but it adds some noise to the data distribution, which can hurt the network's performance. Additionally, it may require more computing resources than other normalizing methods. In Batch normalization approach, unsuitable running variance results in an unstable feature distribution among channels, and mean statistics inconsistency adds extra noise or loses useful representation. So we will be employing a new normalization technique called Representative Batch Normalization (RBN) [12] to enhance the SINet model and study how this normalization technique performs on polyp images. After employing RBN on SINet we have named it SINet-RBN and have tested this new modified model with various polyp image datasets.

## 2. Previous Work

Numerous deep learning [13] based techniques for polyp segmentation have recently been developed.

ResUNet++ [14] is an architecture to address the need for more accurate segmentation of colorectal polyps found in colonoscopy examinations.

Extensive Experiments on novel architecture PraNet (Parallel reverse Attention Network)[15] to automatically segment polyps[16] from colonoscopy images demonstrated that PraNet consistently outperforms all state-of-art approaches by a large margin and also achieves very high accuracy without any pre-processing [17] or post-processing [18] [19].

## 3. Methodology

The SINet model is based on biological research on predator and prey[4]. SINet works for capturing concealed object detection like a predator searching for its prey and identifying it. We have introduced Representative Batch Normalization(RBN) in the SINet model for better model performance in polyp image segmentation. For this, we have used RBN in feature extraction with Res2net which we call this model SINet-RBN. So it has two phases.

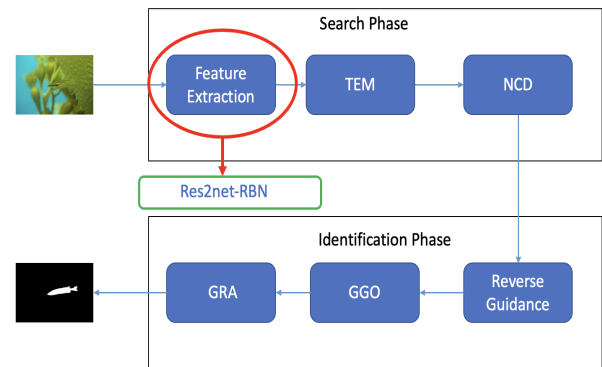


Figure 1: SINet-RBN

### 3.1 Search Phase

**Feature Extraction:** In Search search phase, our model is responsible for searching for a concealed object, which is a polyp in our case. Feature Extraction, Neighbor Connection Decoder, and Texture Enhanced module come under this stage. We will employ Representative Batch Normalization (RBN) [12] in the feature extraction phase in order to improve feature representation and performance. We have refined the SINet model in the feature extraction phase. The res2net represents multiscale features at a granular level and increases the range of receptive field of each network [11]. Representative Batch normalization(RBN) will be employed instead of Batch Normalization to improve stability and performance on feature extraction. Centering calibration feature of RBN moves features with instance-specific statistics, and scaling calibration reduces feature intensity to build a more stable feature distribution.RBN has proved itself a valuable tool for improving the performance of deep residual networks. Batch normalization can introduce some noises into the data distribution, which can reduce the performance of the network. RBN addresses this issue by adding two calibration steps to the standard Batch Normalization (BN) [20] [21] procedure. Instead of using BN, we will be modifying res2net by employing representative batch normalization.

**Reviewing Representative Batch Normalization:** In RBN there are two new steps or calibration compared with BN. Before RBN, BN consists of only two steps which are centering and scaling.

Centering:  $X_m = X - E(X)$ ,

Scaling:  $X_s = X_m / \sqrt{Var(X) + \epsilon'}$

The initial calibration phase, known as centering calibration, amplifies relevant features while minimizing irrelevant ones. The subsequent phase, termed scaling calibration, moderates feature strength to establish a consistent feature distribution. RBN provides a straightforward yet impactful method to enhance deep residual network efficiency. It offers the potential to boost neural network results across multiple tasks. Two new calibration steps that are added to standard Batch Normalization which are:

**Centering Calibration:** During the centering calibration process, features are adjusted around their average value. This action eliminates bias within the feature, ensuring it more

accurately mirrors the data's inherent distribution. This step is undertaken prior to the centering aspect of Batch Normalization.

Centering Calibration :  $X_{cm} = X + w_m \odot K_m$

Centering :  $X_m = X - E(X)$ ,

Centering calibration boosts the distinguishing capability of pertinent features by minimizing non-essential variations. Adjusting features according to the mean of a particular instance ensures they more accurately reflect that specific instance.

**Scaling Calibration:** During the scaling calibration phase, features are adjusted based on their standard deviation. This standardization ensures that features operate on a comparable scale. This procedure is executed following the scaling component of batch normalization.

Scaling:  $X_s = X_m / \sqrt{Var(X) + \epsilon}$

Scaling Calibration:  $X_{cs} = X_s \cdot R(w_v \odot K_s + w_b)$

Scaling calibration boosts the ability of significant features to differentiate by minimizing extraneous fluctuations. By adjusting features in relation to the standard deviation of a particular instance, they more accurately portray that specific instance.

**Texture Enhanced Module:** It incorporates feature representations that are highly discriminative[22] during the searching stage. The output of the feature extraction phase where we applied Representative Batch Normalization technique in res2net will be fed into TEM to obtain better feature representation.

**Neighbor Connection Decoder:** TEM cannot obtain the location information of the polyp so it is further fed into NCD. It is able to provide the location information of polyps which generates the coarse location map[23].

### 3.2 Identification Phase

The identification phase is same as original SINet model [4]. It consists of three components which are Reverse Guidance, Group Guidance Operation (GGO), and Group Reversal Attention(GRA).

**Reverse Guidance:** Methodical approach to mining discriminative hidden regions by employing sigmoid and reverse operations to erase objects.

**GGO:** It has the ability to clearly separate the candidate feature and guidance prior to the ensuing refinement process [24].

**GRA:** To increase performance, multistage refinement [25] is carried out with the aid of both reverse guidance and GGO.

## 4. Results and Discussion

The original SINet model employs the batch normalization approach, whereby unsuitable running variance results in an

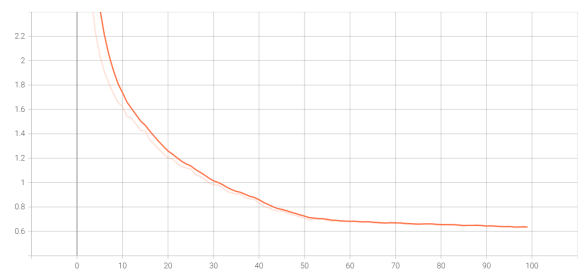
unstable feature distribution among channels, and mean statistics inconsistency adds extra noise or loses useful representation. Thus, rather than using batch normalization in the SINet model, we have employed the representative batch normalization technique. Since SINet showed promise in polyp identification, we have opted for Polyp Image Segmentation. For enhanced feature representations and better model performance, RBN is further incorporated into it.

### 4.1 Overview of Training

CVC-ClinicDB serves as a suitable dataset for training our model and examining the outcomes of polyp image segmentation. Given Parameters were used to train all models with clinicDB dataset.

Parameters:		
Epoch:	100	Learning Rate: 1e-4
Batch Size:	36	Train Size: 352
Clip:	0.5	Decay Rate: 0.1
Machine:	Google Colab	

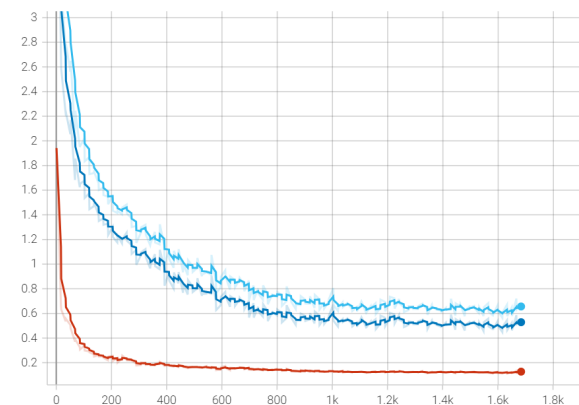
The following graphs show the analysis of Loss Statistics, Mean Average Error, and Loss per epoch when clinicDB was used to train both SINet and SINet-RBN models. These graphs are the output when 612 clinicDB datasets were used to train SINet and SINet-RBN.



(a) Loss per Epoch



(b) Mean Average Error



(c) Loss Statistics

Figure 2: Overview of training for SINet

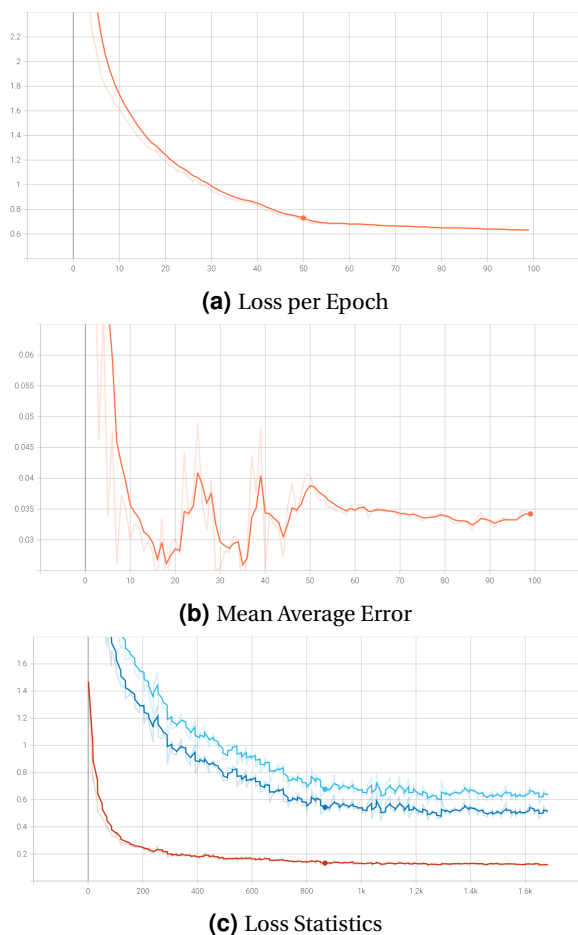


Figure 3: Overview of training for SINet-RBN

## 4.2 Results

The following qualitative outcomes were seen when various models were used to test the provided sample images. The second column from Table 1 shows the outcome of testing sample images into the SINet model using the three polyp image samples that were provided. The same images were tested using the SINet-RBN model in the third column, and the PraNet model's result is shown in the fourth column. These findings indicate that PraNet has not performed as well as SINet and SINet-RBN. In terms of visualization, SINet-RBN produces better results than SINet.

## 4.3 Evaluation

In this research, LaribPolypDB, ColonDB, and CVC-300 Dataset are employed for the purposes of testing and evaluation.

Table 1: Three Test Datasets and their symbol representation

DATASET	SYMBOL
LaribPolypDB	DS1
ColonDV	DS2
CVC-300	DS3

Structure measure( $S_\alpha$ ) [26] is a metric that is used to evaluate the segmentation accuracy and structural similarity between the ground truth segmentation of an object and the segmentation produced by the segmentation algorithm. The

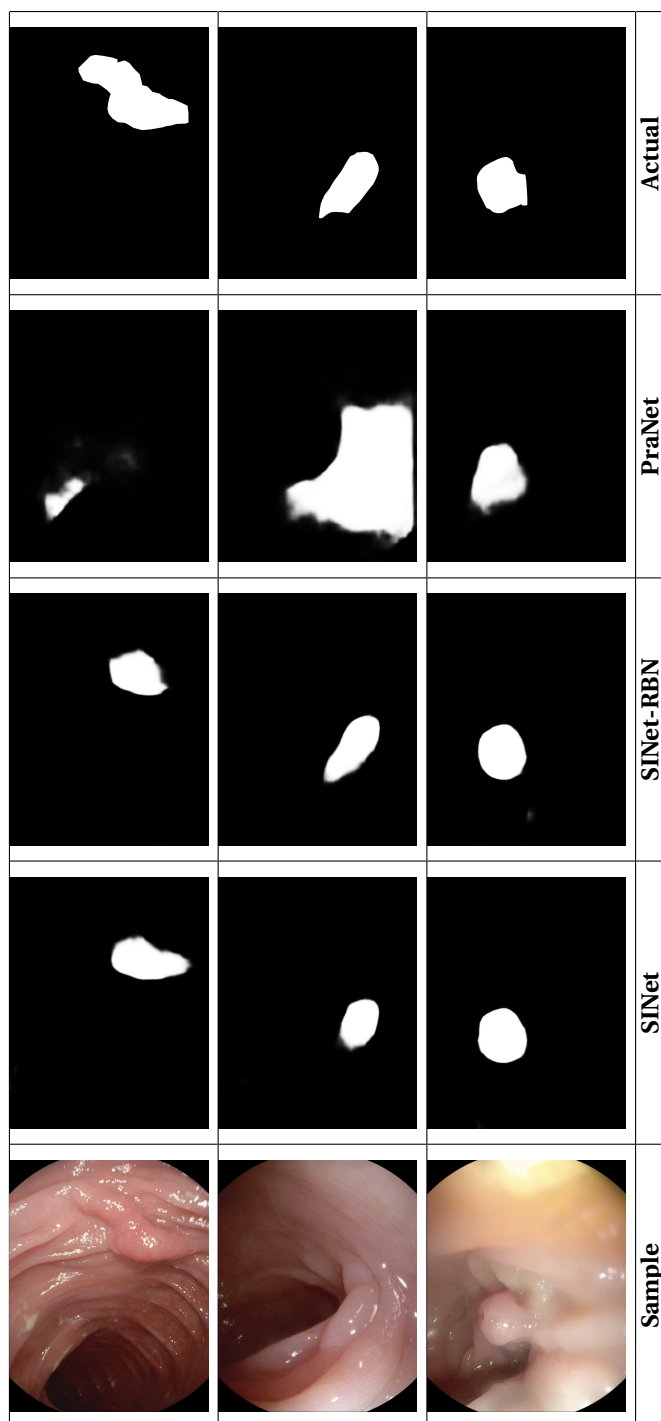


Figure 4: Qualitative outcomes from testing sample images in the SINet, SINet-RBN, and PraNet models

Smeasure ranges between 0 and 1, with higher values indicating better segmentation accuracy.

wFmeasure( $^w F_m$ ), [27] on the other hand, is a metric that is used to assess both the localization and classification accuracy of object detection algorithms [28].

Mean Absolute Error (MAE) is employed as a metric to assess the algorithm's overall accuracy in polyp detection.[29]

E-measure ( $E_\phi$ ) [30] integrates local pixel values with the mean value at the image level and concurrently assesses the similarity, both at the image and pixel levels, between a segmentation result and a ground truth (GT) result.

**Table 2:** Evaluation for each model on three test datasets

DATASET	$S_\alpha$	${}^w F_m$	MAE	$E_\phi$	MODEL
DS1	0.836	0.657	0.017	0.857	SINet
DS1	0.512	0.252	0.149	0.503	PraNet
DS1	0.832	0.649	0.019	0.859	SINet-RBN
DS2	0.888	0.819	0.011	0.947	SINet
DS2	0.660	0.389	0.055	0.696	PraNet
DS2	0.895	0.815	0.010	0.946	SINet-RBN
DS3	0.926	0.846	0.008	0.957	SINet
DS3	0.720	0.476	0.049	0.744	PraNet
DS3	0.940	0.878	0.007	0.970	SINet-RBN

Following successful training of SINet[4], SINet-RBN, and PraNet[15] using the ClinicDB Dataset, various datasets were examined. All of the models were tested and evaluated using LaribPolypDB, ColonDB, and CVC-300 datasets. In order to analyze the output provided by various models while using various datasets, predicted output was evaluated using metrics like  $S_\alpha$ ,  $E_\phi$ ,  ${}^w F_m$ , and MAE.

When the SINet, SINet-RBN, and PraNet models were tested using LaribPolypDB, SINet performed much better than PraNet while marginally outperforming SINet-RBN. When ColonDB was used, SINet-RBN outperformed PraNet in all metrics and outperformed SINet in terms of  $S_m$  and MAE. When CVC-300 was tested in every model, SINet-RBN outperformed both SINet and PraNet. When we used representative batch normalization, the performance of the model was improved overall.

## 5. Conclusion

We've conducted a thorough study on Polyp image segmentation. Utilizing the deep learning [13] model, Search Identification Network (SINet), we've recognized its capability to detect concealed or camouflaged objects. This model was initially trained using the ClinicDB dataset. To enhance its feature extraction, we incorporated Representative Batch Normalization, resulting in the SINet-RBN variant. This modified model was trained on the same datasets. In the feature extraction phase, it was further modified using Representative Batch Normalization, and the resulting modified model, called SINet-RBN, was trained using the same datasets. Those models along with PraNet were subsequently tested using different datasets, including LaribpolypDB, ColonDB, and CVC-300. Different evaluation metrics such as  $S_\alpha$ ,  $E_\phi$ ,  ${}^w F_m$  and MAE were used to compare the results, and SINet, which was initially developed for concealed object detection, outperformed the PraNet Model. The SINet-RBN model overall performed better than both models except for LaribPolypDB. In case of LaribPolypDB, SINet was slightly better than SINet-RBN. In the Future SINet and SINet-RBN models can benefit from the use of large datasets, and video-based polyp segmentation will be crucial and useful.

## Acknowledgments

The authors express deepest gratitude to Dr. Binod Tripathee, whose invaluable insights, guidance, and expertise played a

pivotal role in shaping this research. Additionally, heartfelt appreciation goes out to all those who lent their helping hands throughout this endeavor.

## References

- [1] Yali Amit, Pedro Felzenszwalb, and Ross Girshick. Object detection. In *Computer Vision*, pages 1–9. Springer International Publishing, 2020.
- [2] Rakhsith L. A, Anusha K. S, Karthik B. E, Arun Nithish D, and Kishore Kumar V. A survey on object detection methods in deep learning. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 1619–1626, 2021.
- [3] Joanna R. Hall, Innes C. Cuthill, Roland Baddeley, Adam J. Shohet, and Nicholas E. Scott-Samuel. Camouflage, detection and identification of moving targets. *Proceedings of the Royal Society B: Biological Sciences*, 280(1758):20130064, May 2013.
- [4] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.
- [5] Mojtaba Akbari, Majid Mohrekeh, Ebrahim Nasr-Esfahani, S. M. Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. Polyp segmentation in colonoscopy images using fully convolutional network, 2018.
- [6] Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Automatic polyp segmentation via multi-scale subtraction network, 2021.
- [7] J F O'Connor and J Cohen. Computerized tomography (cat scan, ct scan) in orthopaedic surgery. *The Journal of Bone & Joint Surgery*, 60(8):1096–1098, Dec 1978.
- [8] Maryellen L. Giger and Kenji Suzuki. Computer-aided diagnosis. In *Biomedical Information Technology*, pages 359–XXII. Elsevier, 2008.
- [9] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12997–13007, 2021.
- [10] Lu Cheng, Yicai Ji, Chao Li, Xiaojun Liu, and Guangyou Fang. Improved SSD network for fast concealed object detection and recognition in passive terahertz security images. *Scientific Reports*, 12(1), July 2022.
- [11] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xiao-Yang Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(2):652–662, Feb. 2021.
- [12] Shang-Hua Gao, Qi Han, Duo Li, Ming-Ming Cheng, and Pai Peng. Representative batch normalization with feature calibration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2021.
- [13] Iqbal H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), August 2021.
- [14] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pal Halvorsen, and Havard D. Johansen. ResUNet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*. IEEE, December 2019.

- [15] Cheng Li, Zhuo Zhang, and Li Tian. Parallel reverse attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 411–419. Springer, 2020.
- [16] Tongping Shen and Xueguang Li. Automatic polyp image segmentation and cancer prediction based on deep learning. *Front. Oncol.*, 12:1087438, 2022.
- [17] G. Mori. Guiding model search using segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. IEEE, 2005.
- [18] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011.
- [19] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.
- [20] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [21] Greg Yang, Jeffrey Pennington, Vinay Rao, Jascha Sohl-Dickstein, and Samuel S. Schoenholz. A mean field theory of batch normalization. *CoRR*, abs/1902.08129, 2019.
- [22] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Re-thinking the inception architecture for computer vision. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2818–2826, 2016.
- [23] Chenhongyi Yang, Zehao Huang, and Naiyan Wang. Querydet: Cascaded sparse query for accelerating high-resolution small object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13668–13677, June 2022.
- [24] Shuhan Chen and Ying Fu. Progressively guided alternate refinement network for rgb salient object detection (version 1), 2020.
- [25] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. Reverse attention for salient object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [26] Ming-Ming Cheng and Deng-Ping Fan. Structure-measure: A new way to evaluate foreground maps. *International Journal of Computer Vision*, 129(9):2622–2638, June 2021.
- [27] Ran Margolin, Lih Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2014.
- [28] Ruyong Ren, Shaozhang Niu, Hua Ren, Shubin Zhang, Tengyue Han, and Xiaohai Tong. ESRNet: Efficient search and recognition network for image manipulation detection. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(4):1–23, March 2022.
- [29] Fabio Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung. Contrast based filtering for salient region detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 733–740, 2012.
- [30] Xuebin Qin Dengping Fan, Ge Pengji and Mingming Cheng. Cognitive vision inspired object segmentation metric and loss function. *SCIENTIA SINICA Informationis*, 51(9):1475, September 2021.