Oncology
and Radiotherapy©
18(6) 2024: 001-009 • RESEARCH ARTICLE

Analyzing the evaluation metrics of detecting gastrointestinal tumor using segmentation techniques in endoscopic images

Pooja K, Jerritta S

Department of Biomedical Engineering, Vels Institute of Science Technology and Advanced Studies, Tamil Nadu, India

ABSTRACT

Convolutional Neural Networks (CNNs) are gaining popularity for analyzing endoscopic images due to their many benefits. Since certain gastric polyps can lead to stomach cancer, it's crucial to detect and remove them accurately and promptly. CNN-based semantic segmentation helps by precisely outlining polyp areas, aiding endoscopists in identifying and treating them effectively. Despite the potential benefits, there is a scarcity of studies employing CNN for automated gastric polyp identification, particularly in the realm of semantic segmentation. Thus, we present groundbreaking research focused on segmenting gastric polyps in endoscopic images using CNNs. Various traditional semantic segmentation models, such as U-Net, DeepNet, SegNet, FuNet, and CustomNet (referred to as GISTNet), employing encoders like U-Net, ResNet50, MobileNetV2, or EfficientNet-B1, were constructed and scrutinized using a comprehensive dataset. Given the complexity of the problem and the multitude of criteria, selecting the most suitable CNN model poses a challenge. To address this, we propose an integrated evaluation approach that combines subjective considerations with objective data to identify the optimal CNN model. Our proposed network, CustomNet (GIST-Net), employing ResNet as the encoder, emerged as the top performer according to our integrated evaluation method and was selected to construct the automated polyp segmentation system. This investigation underscores the clinical significance of semantic segmentation models in gastric polyp diagnosis and highlights the efficacy of the integrated evaluation approach in impartially selecting suitable models. Additionally, our research has the capacity to progress the identification techniques of gastric cancer, and the proposed evaluation methodology has implications for selecting diagnostic techniques based on mathematical models.

Keywords: endoscopic image, Gastric Cancer (GC), denoising, gastric polyps, semantic segmentation, Convolutional Neural Networks (CNNs)

Address for correspondence:

Pooja K.,

Department of Biomedical Engineering, Vels Institute of Science Technology and Advanced Studies, Tamil Nadu, India

Email: poojapreethi26711@gmail.com

Word count: 5222 **Tables:** 03 **Figures:** 07 **References:** 23

Received: 06 May, 2024, Manuscript No. OAR-24-134195

Editor Assigned: 20 May, 2024, Pre-QC No. OAR-24-134195(PQ)

Reviewed: 23 May, 2024, QC No. OAR-24-134195(Q)

Revised: 27 May, 2024, Manuscript No. OAR-24-134195(R)

Published: 31 May, 2024, Invoice No. J-134195

INTRODUCTION

Gastric cancer ranks among the most prevalent malignant tumors worldwide, with approximately 1 million new cases diagnosed annually. Particularly in China, it stands as one of the top three cancers, boasting a daunting 12.4% mortality rate.

Given its significant impact on morbidity and mortality, gastric cancer is recognized as a formidable health threat. Current diagnostic methods primarily rely on gastroscopy, a procedure heavily dependent on the expertise of skilled physicians. Studies indicate a modest accuracy rate of 69%-79% for manual gastroscopy [1]. With the advent of deep learning techniques in medical imaging, Convolutional Neural Networks (CNNs) have emerged as a promising tool for segmenting stomach cancer images. Hirasawa et al. utilized CNNs to detect stomach cancer in endoscopic images, albeit encountering limitations due to complex lesion characteristics. PAN et al. improved the SSD model for distinguishing between early-stage stomach cancer and non-cancerous images. They introduced the DSF module to enable better feature fusion across various levels. Zhang et al. introduced SSD-GPNet, which enhances the SSD architecture by including cross-layer interaction to improve the network's receptive field and feature extraction abilities. While CNNs have improved identification accuracy, their outputs may not always meet supplementary medical diagnostic criteria. Consequently, there's a growing need for customized network architectures to enhance segmentation performance. In 2022, Ronneberger et al. introduced U-Net, which employs skip connections to incorporate richer low-level feature information into the final recovered feature map, demonstrating wide applicability in medical image segmentation. Several studies have improved the U-Net for segmenting gastric cancer lesions. It refined the U-Net model by incorporating a pyramidal structure to precisely identify lesion locations in gastric cancer. Additionally, Zhang et al. introduced SERES and DAGC modules are integrated into a modified U-Net network to substitute pooling operations, thereby improving the fusion of high-level and low-level feature information. Despite the advancements achieved with upgraded U-Net techniques, inherent limitations hinder their ability to capture explicit long-range relationships [2]. Due to the complex folds of the stomach mucosa, models need to capture global information effectively in order to differentiate lesion characteristics from background noise. Therefore, the accurate and efficient gathering of global contextual information remains a crucial challenge that requires focus.

Zhang et al. introduced the Transfuse model, which integrates both Transformers and CNNs to collaborate. Similarly, Chen et al. proposed Transnet, which employs U-Net to retain local spatial information while incorporating the Transformer as an encoder for medical image segmentation. However, directly feeding deep features from convolutional layers into the Transformer leads to insufficient global information and underutilizes the Transformer's potential. To balance global context modeling efficiency with the retention of low-level features, Zhang et al. introduced the TransFuse model, which combines Transformers and CNNs to work together synergistically. Similarly, suggested TranUnet, which utilizes U-Net to maintain local spatial details while incorporating the Transformer as an encoder for medical image segmentation. However, directly inputting deep features from convolutional layers into the Transformer results in insufficient capture of global information. Utilization, limiting the Transformer's effectiveness. Thus, a recommendation is made to independently pass original images through Transformers and CNNs are employed to comprehensively extract unique features from each, although this leads to heightened model complexity [3]. To address the computational burden imposed by deep-level features, a decoder structure is suggested to replace the original, thereby isolating the influence of low-level features. Additionally, Transformers utilize computationally demanding global self-attention mechanisms. Liu et al. tackled this issue by introducing Windows Multi-Head Self-Attention (W-MSA) in the Swin-Transformer, thereby notably decreasing computational complexity.

In medical image analysis, Hirasawa et al. reported a sensitivity of 89.2% and a positive predictive value of 29.06% for automated stomach cancer detection using a Single-Shot Multi-Box Detector (SSD) for object detection. However, the high false-positive rate and inability to delineate invasive areas remain unresolved challenges. Hence, the focus shifts to segmentation for precise identification and extraction of objects from images. The U-Net architecture consists of an encoder layer for down sampling images while extracting features and a decoder layer for up sampling images. A bypass link between these layers allows for object segmentation without compromising image resolution [4]. However, traditional semantic segmentation methods process the entire image, leading to time constraints and an inability to identify individual items within an image.

To address these limitations, like researchers, introduced Mask R-CNN, which combines object segmentation with individual object identification in an image. This approach, presented Mask R-CNN, which was recognized with the Best Paper Award at the 16th International Conference on Computer Vision (ICCV) in 2019. Identifies object regions through bounding boxes and marks the corresponding real regions in a mask layer [5]. By providing class names and probabilities along with bounding boxes, detection reliability is improved.

Expanding on this study, we suggest a Dual-Branch Hybrid Network for segmenting stomach cancer images, which integrates both the Swin-Transformer and U-Net architectures. The combination of the Swin-Transformer and U-Net forms a decoder structure aimed at capturing intricate feature details to precisely localize lesions [6]. Our investigation focuses on developing an efficient feature fusion approach to merge features extracted from both the U-Net and Transformer components. This multi-modal fusion process enhances correlation information extraction across different scales, facilitated by the linear Hadamard product. During network training, we calculate losses from the Transformer, U-Net, and fused outputs with ground truth labels to produce high-quality segmentation results.

This study focuses on several key aspects: proposing a novel approach for detecting stomach cancer using Deep Learning techniques, showcasing cutting-edge technology, addressing the problem at hand, presenting the basic proposal, and demonstrating the effectiveness of the suggested strategy through various trials and results [7]. We also provide a summary and outline future efforts at the end of the publication.

Our Dual-Branch Hybrid Network simultaneously employs the Swin-Transformer and U-Net to segment lesions in stomach cancer images. We introduce the Deep Feature Aggregation Decoder (DFA) to replace the original decoders, which reduces model complexity while retaining information about lesion regions. Furthermore, the Feature Fusion (FF) module facilitates multi-modal fusion mechanisms for independent feature interaction. It employs the linear Hadamard product for feature fusion. Experimental results demonstrate the model's accurate segmentation of gastric cancer lesion regions, surpassing cutting-edge techniques [8]. The rest of this document is structured as follows: "Related Works" examines the utilization of enhanced U-Net and Transformer architectures in medical image segmentation, "Method" outlines our proposed framework, encompassing the Swin-Transformer branch, U-Net branch, DFA, FF, and Decoder modules, and "Experiments" presents experimental results on multiple datasets. Finally, we visualize the results and draw conclusions.

Related work

The objective of image categorization is to assign labels to photographs based on predefined categories. Deep learning methodologies, notably those leveraging Convolutional Neural Networks (CNNs), yield optimal results for tasks such as object recognition and segmentation by creating hidden feature representations [9]. These technologies offer faster analysis and computation compared to traditional methods. In the medical field, systems are being developed to aid clinicians in early cancer detection, particularly in stomach cancer. More research and applications for classification, detection, and segmentation in stomach cancer have advanced, with an increase in annotated datasets that are freely accessible or available with limited access restrictions.

One approach involves the creation of artificial networks capable of learning characteristics from publicly available datasets. However, much of the research focuses on the progression of stomach cancer rather than on the identification and localization of tumors, which is crucial for diagnostic accuracy. Models like YOLO-V2 are employed for multi-object detection within the same image. In a similar vein, researchers have developed detectors for small objects by augmenting feature information from fundamental network models [10]. Additionally, techniques like label smoothing weighted loss have been utilized to improve feature representation, as seen in Retina Net, an enhanced version of Faster R-CNN. Lesion-based CNNs, a type of deep learning method, utilize endoscopic images to detect complete lesions. While endoscopy aids in the identification of Gastro Intestinal (GI) tract issues, its diagnostic accuracy is limited by practitioner experience and environmental factors. For instance, the false-negative rate for gastric cancer detection using Esopha Gogastro Duodenoscopy (EGD) ranges from 3.2% to 30.8%. These models are typically built using the Cafe deep learning framework, a widely used and versatile platform. Recent studies have shown promising results, with CNNs achieving a sensitivity of 88.2% and a positive prediction value of 28.6% in identifying gastric cancer lesions. Further advancements include improving the network topology of models like Mask R-CNN for medical image detection and utilizing additional data augmentation techniques. Researchers have explored various approaches for stomach cancer detection using convolutional neural networks. For instance, Sakai et al. achieved an accuracy of 81.8% using 764 endoscopic images to train their CNN model. Ishihara et al. employed a patch-based CNN on X-ray scans of the stomach, though limited by single-angle imaging [11]. Researchers trained their CNN with 11,372 endoscopic images, accurately detecting 70 of 71 lesions. Similarly, Ishioka et al. utilized a CNN to identify stomach cancer from video images, achieving success rates of 92.2% for smaller lesions and 98.6% for larger ones.

Xu Zhang et al. utilize a straight forward Convolutional Neural Network (CNN) model to classify gastric precancerous conditions, which, if misdiagnosed, may progress to cancer. Despite achieving an accuracy of 72.90%, this GPDNet model exhibits lower accuracy compared to other detection methods. The focus of this article is on leveraging CNNs for the segmentation of endoscopic images to detect gastrointestinal disorders. Segmentation, the process of partitioning an image into relevant sections, plays a crucial role in identifying and characterizing specific abnormalities within the gastrointestinal system.

This study aims to provide insights into the strengths, limitations, and potential future directions of CNN-based segmentation approaches in gastrointestinal disease detection by reviewing the current state of research and breakthroughs in this area [12]. The utilization of CNNs in endoscopic image segmentation holds promise for revolutionizing gastrointestinal disease detection by enhancing accuracy, reducing false positives, and expediting the analysis process.

Our investigation will delve into the methodologies, datasets, and performance metrics employed in research focused on leveraging CNNs for gastrointestinal disease identification through endoscopic image segmentation [13]. Through a comprehensive examination, we aspire to contribute to the evolving landscape of medical imaging technologies, ultimately fostering breakthroughs that enhance clinical outcomes and patient care in the realm of gastrointestinal health.

MATERIALS AND METHODS

Data collection

Data collection is a pivotal stage in any research endeavor. For this study, datasets were sourced from various online platforms and reputable hospitals. These datasets consist of endoscopic scan images captured over a specific timeframe, with each patient contributing a sequence of images is acquired at different pixel densities over a specific timeframe.

This dataset, valuable, requires careful handling due to variations in endoscopic results. Fiber-optic endoscopes utilize bundles of optical fibers to transmit images, with each fiber conveying images to the top of the bundles. Subsequently, these images are viewed through an eyepiece after passing through a focusing lens. Endoscopies, as minimally invasive procedures, utilize natural orifices like the mouth or anus, employing optical sensors for imaging.

The aim of this study is to provide easily accessible information that can ultimately improve medical care quality, particularly in the context of malignancy diagnosis in images of the Gastro Intestinal (GI) tract, these devices have the potential to assist in image processing and the identification of endoscopic discoveries within the GI tract, including the intestines, bowel, and bladder. Significant benefits include improved identification accuracy, reduced workload for healthcare professionals, lower average costs, decreased patient discomfort, and potentially increased patient participation in procedures. Enhanced screening is expected to reduce mortality rates and instances of basal GI disease.

This dataset is versatile and can be applied in various applications aimed at developing and testing image analysis algorithms [14]. By utilizing the same dataset, researchers can compare methodologies and experimental outcomes more efficiently, facilitating result replication and further analysis.

Pre-processing techniques

The endoscopic imaging technique typically generates images characterized by unique contrasts and artifact coloring, which may vary based on the particular type of endoscopy employed. These images are crafted to enhance sensitivity in dimly lit regions, preserve image sharpness, and mitigate interference from noise. To achieve this objective, we have devised a dependable method for augmenting endoscopic contrast and mitigating noise [15]. This method entails the utilization of various techniques, including Median Filtering, Gaussian Filtering, Weiner Filtering, and Wavelet Filtering. Additionally, we leverage texture analysis metrics such as mean, standard deviation, variance, kurtosis, and skewness as part of this process. By experimenting with various combinations of these techniques, we aim to identify active areas within the stomach in regular endoscopy imaging.

Problematic

This research endeavors to identify abnormal tissue areas, potentially indicative of tumors, within the gastric organ. To accomplish this objective, we have developed a methodology utilizing the U-Net architecture, which we've named GAS-Net. U-Net is a widely used architecture in medical image segmentation, originally designed for understanding and segmenting medical images, and has been pivotal in medical imaging research. The U-Net architecture consists of two main pathways: a contracting path and an expansive path. In the contracting path, a variety of operations including convolutional, normalization, max-pooling, activation, and concatenation are employed to extract crucial features from input images, resulting in a feature vector of predetermined length. Subsequently, the expansive path utilizes the information obtained from the contracting path to perform up-convolutions, progressively generating an output segmentation map [16]. The underlying concept of the GAS-Net architecture revolves around constructing a network that accepts a series of input images along with their corresponding ground truth masks. Using probabilistic and generalized functions, the network learns to identify tumor areas from these ground truth masks based on the input images. Subsequently, a correspondence is created between the input images and their corresponding ground truth masks to facilitate accurate segmentation.

Existing semantic segmentation methods

Numerous semantic segmentation algorithms have shown impressive efficacy by leveraging Convolutional Neural Networks (CNNs) in natural images. Typically, CNNs are trained to learn a mapping from input images to ground-truth masks through a series of operations, including convolution, pooling, and up-sampling. In this research, different classical semantic segmentation models including U-Net, DeepNet, SegNet, FuNet, and CustomNet (GISTNet), are utilized with different encoders for accomplishing end-to-end semantic segmentation of gastric polyps. To prevent redundancy, a generic U-Net architecture is employed to illustrate the underlying concept shared by most semantic segmentation models. The encoder phase typically diminishes spatial resolution and extracts image features using convolution and max-pooling to aid feature extraction. Conversely, the symmetric decoder phase upscales the extracted features to the original input resolution, generating low-dimensional predictions [17]. Within the symmetric decoder, feature maps are combined from the up-sampling path and the skip connection. The skip connections in U-Net are notable and innovative because they supplement and enhance semantic information from the up-sampling process. These connections play a crucial role in retaining detailed information and improving the performance of semantic segmentation models.

Performances of the network

The network implementations were performed on a workstation featuring an Intel i7-6800K CPU with 64 GB of RAM. Throughout the training process, a batch size of 64 was employed, and a uniform learning rate of 0.004 was set for all layers. The learning rate was subsequently reduced to 0.0004 every 15 epochs starting from the 100^{th} epoch. The primary objective across the literature is to enhance efficiency in both training and testing phases, aiming for a more user-friendly system. In our study, the execution time during both training and testing phases varied depending on the number of regions of interest processed per image, ranging from 400 milliseconds to 930 milliseconds for images sized 256 pixels × 256 pixels. Additionally, networks heavily rely on hyperparameters [18]. Hence, we conducted numerous experiments to determine the optimal values for these hyperparameters and validate the effectiveness of our approach.

Performance metrics

This stage of the study is crucial for assessing the model's performance and determining whether it fulfills its objectives. Various performance metrics are commonly employed in the literature for tasks such as in table 1.

Tab. 1. Equational metrics

Accuracy	TP+TN/TP+TN+FP+FN
Precision	TP/TP+FP
Recall (Sensitivity)	TP/TP+FN
F-measure	2 × (Precision × Recall)/Precision+Recall
Dice coefficient	2 × TP/(TP+FN+FP)

These metrics provide valuable insights into different aspects of the model's performance, helping to evaluate its effectiveness in various tasks [19].

The following note that (TP) refers to True Positives, (TN) denotes True Negatives, (FP) represents False Positives, and (FN) indicates False Negatives [20]. These parameters are calculated using the confusion matrix, which provides insights into the accurate and inaccurate classification of images across all categories. Additionally, sensitivity played a crucial role when there was an overlap observed between the identified tumor area and the ground truth generated by the team mentioned earlier.

RESULTS AND DISCUSSION

In semantic segmentation, the role of feature extraction networks is pivotal in categorizing pixels accurately. Therefore, the encoders employed in this study comprise three distinct advanced CNNs: ResNet50, MobileNetV2, and EfficientNet-B1. Re Net50, a member of the ResNet series algorithms, stands out for its speed and accuracy, making it a popular choice it offers faster process-ing while still achieving impressive results. Moreover, ResNet50 outperforms shallower networks such as ResNet18 and ResNet34 in terms of both training outcomes and resource efficiency, thanks to its superior feature extraction capabilities. Hence, based on the criteria of speed and precision for the polyp segmentation task, ResNet50 was selected as the encoder for the proposed Custom-Net (GIST-Net) class.

The newly proposed network, CustomNet (GIST-Net), adopts the U-Net architecture with ResNet as its encoder, achieving the highest segmentation accuracy among the conducted experiments. It attains an Intersection Over Union (IOU) of 91.22%, accuracy of 98.12%, recall of 96.18%, precision of 98.12%, and F1-score of 98.14%. However, its detection speed is the slowest at 25 frames per second (FPS) compared to other U-Net experiments. When ResNet50 is utilized as the encoder for U-Net, the model exhibits a significantly higher number of parameters and Multiply Accumulate Operations (MACs) compared to others, requiring greater storage and processing resources.

Additionally, its segmentation accuracy is comparatively lower. While $UNet^{++}$ models boast more parameters than U-Net, they effectively enhance segmentation accuracy compared to the designed network. Among the $UNet^{++}$ models, $UNet^{++}$ with the Efficient Net-B1 encoder achieves the highest segmentation accuracy but has the slowest detection speed at 20 FPS. $UNet^{++}$ with the MobileNetV2 encoder strikes a balance between segmentation accuracy and computational efficiency, outperforming other $U\text{-}Net^{++}$ models and obtaining the highest score among all semantic segmentation models tested in the trial.

Efficient Net-B1 encoded with ResNet50 demonstrates the highest segmentation accuracy within the U-Net class. It achieves an Intersection over Union (IOU) of 90.18%, accuracy of 95.95%, recall of 95.99%, precision of 96.73%, and F1-score of 97.63%. Despite its complexity, this model achieves exceptional detection speed, reaching 29 Frames Per Second (FPS).

In table 1, FuNet exhibits the best performance in terms of model complexity when MobileNet-v2 serves as the encoder. Although it delivers somewhat satisfactory segmentation accuracy, with an IoU of 91.83%, accuracy of 92.09%, recall of 98.86%, precision of 92.71%, and F1-score of 95.52%, FuNet with the ResNet50 encoder shows significant improvements in segmentation accuracy and detection speed.

When EfficientNet-B1 serves as the encoder, Seg-Net attains the highest segmentation accuracy and model complexity. Specifically, it achieves an Intersection over Union (IOU) of 96.17%, accuracy of 93.12%, recall of 95.16%, precision of 92.14%, and F1-score of 93.14%. Nonetheless, its notably slow detection speed leads to an underwhelming overall score. Conversely, SegNet using the ResNet50 encoder and with the EfficientNet-B1 encoder yield satisfactory segmentation accuracy results.

When ResNet50 is utilized as the encoder, DeepNet exhibits below-average performance concerning model complexity. Furthermore, its segmentation accuracy is notably deficient, marked by an Intersection over Union (IOU) of 69.53%, accuracy of 78.19%, recall of 81.96%, precision of 83.21%, and F1-score of 85.12%. Despite this, DeepNet, when equipped with the ResNet50 encoder, does not realize substantial improvements in segmentation accuracy or detection speed.

Each model exhibits distinct characteristics as per quantitative metrics. Usually, achieving high segmentation accuracy entails increased complexity, which may compromise computational efficiency. Conversely, models prioritizing computational efficiency often sacrifice segmentation accuracy. The suggested integrated assessment approach provides a measurable score for comparison. As per the findings, the proposed network, CustomNet (GIST-Net), employing U-Net with the ResNet encoder, garners the highest scores across subjective and overall categories. This indicates the model's superior overall performance, regardless of the factors evaluated, including quality, quantity, or a blend of both.

The reflecting patches and stomach mucosal folds in endoscopic images are often mistaken as gastric polyps. Even with the naked eye, these areas are often misinterpreted due to lighting or angles. It demonstrates that the system fails to detect certain polyps. These missing polyps are little polyps that appear at the borders of the photographs. Due to technological restrictions, endoscopic pictures are reduced to 256 × 256, affecting the accuracy of tiny polyp segmentation process, while having room for enhancement, still achieves a commendable level of accuracy with the chosen model (Figure 1).

Overall, the automated polyp-segmentation approach showcases notable clinical benefits owing to its consistently high performance [21]. The system can act as a secondary observer by instantly segmenting identified polyps on a separate display next to the primary monitor. This feature holds promise for to address skill disparities among endoscopists and enhance the quality of routine Esopha Gogastro Duodenoscopy (EGD). However, it's important to note that the final decision remains with the endoscopist. Nevertheless, it may encourage endoscopists to perform additional procedures, thus decreasing the chances of missing particular polyps through visual inspection thus decreasing the chances of missing particular polyps through visual inspection.

From the patients' perspective, the proposed approach can streamline physical examinations, reducing waiting times and examination durations. Moreover, it has the capability to provide expert-level diagnostics, thereby boosting patients' confidence, fostering collaboration, and enhancing overall satisfaction [22]. Additionally, the system is expected to become open-source and freely available, contrasting with commercial software, which will make it highly accessible for medical students and young doctors to acquire and improve their clinical abilities (Table 2).

However, this system is subject to several limitations. Firstly, it is limited to segmenting gastric polyps and cannot simultaneously identify other gastric diseases during Esopha Gogastro Duodenoscopy (EGD). Secondly, it can be difficult to differentiate between benign and malignant lesions using traditional white-light imaging endoscopy [23]. However, incorporating histopathological evaluation alongside other advanced endoscopic techniques can offer more clinically relevant insights after identification. Thirdly, the system's robustness in managing diverse clinical scenarios has not been thoroughly evaluated through multicenter clinical trials with extensive datasets. Below, a comparison of neural networks is presented in a table format to illustrate the differences (Figure 1-6 and Table 3).

The output images generated through Python utilizing deep learning algorithms convey a compelling narrative of segmentation analysis, revealing insights into the accuracy levels attained. These images depict the intricate delineation of objects within the input data, showcasing the efficacy of the employed algorithms in identifying and segmenting distinct features. Accompanying the visual representations, the associated accuracy levels provide quantitative validation of the segmentation performance. Such precision in segmentation analysis is paramount across various domains, including medical imaging, autonomous driving, and industrial quality control, where accurate delineation of objects is critical for decision-making processes. Through this amalgamation of transformative potential of deep learning methodologies in advancing image processing tasks with unparalleled efficacy and precision.

In reviewing the table detailing the strengths and weaknesses of various CNN models, several notable patterns emerge. The table underscores the robustness of models like UNet and DeepNet, CustomNet, FuNet, SegNet in handling deep architectures, excelling in feature extraction and classification tasks. Conversely, while effective, exhibits a weakness in computational efficiency due to its substantial parameter count.Overall, this comprehensive comparison highlights the diverse landscape of CNN architectures, each with its distinct strengths and weaknesses, catering to different needs and constraints in machine learning applications (Figure 7).

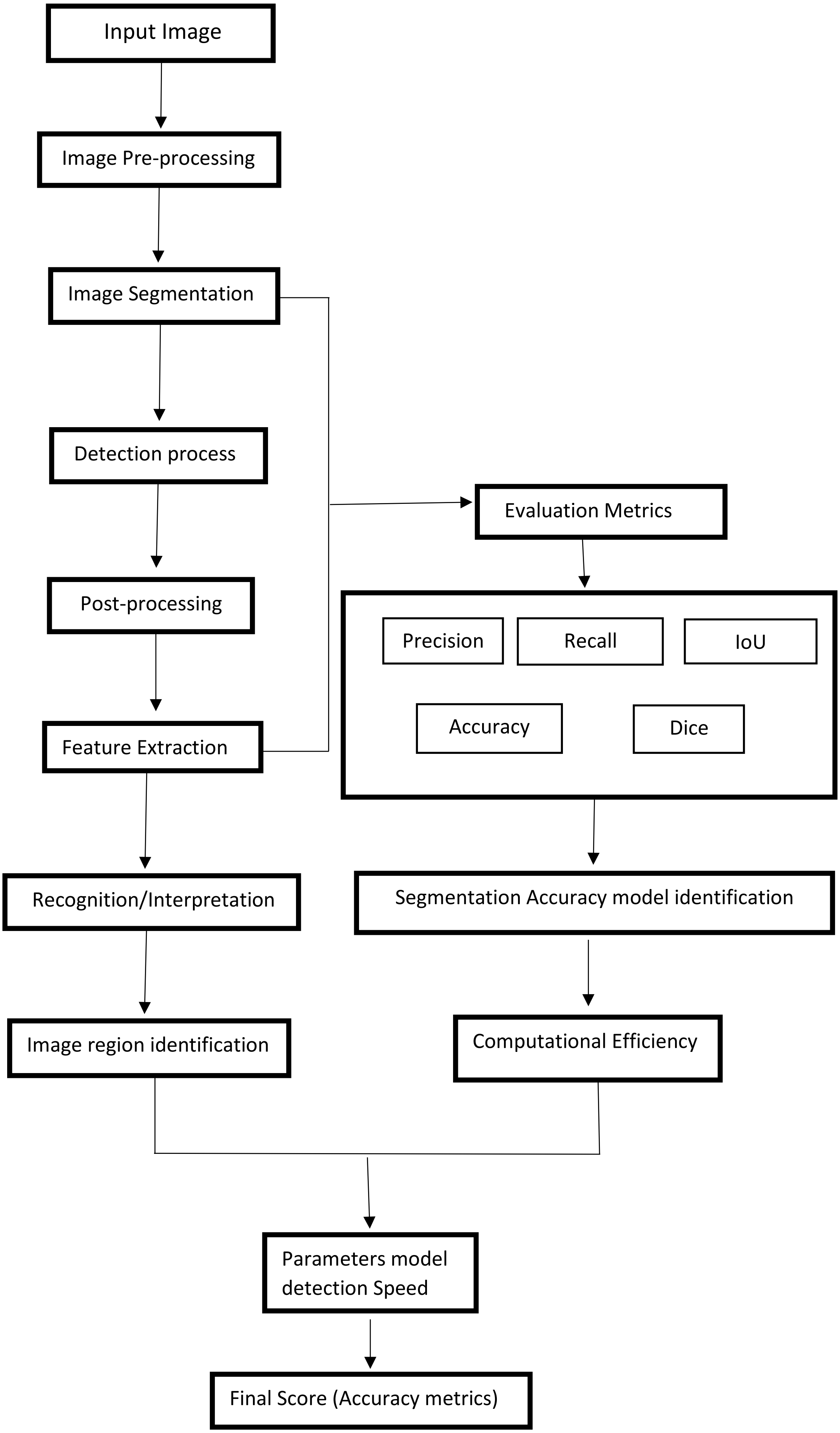

Fig. 1. Flow Chart of segmentations techniques

Tab. 2. Configuration details

Configuration	Version
Operating system	Windows 10
Programming language	Python 3.9
Frame	Pytorch-1.10.0

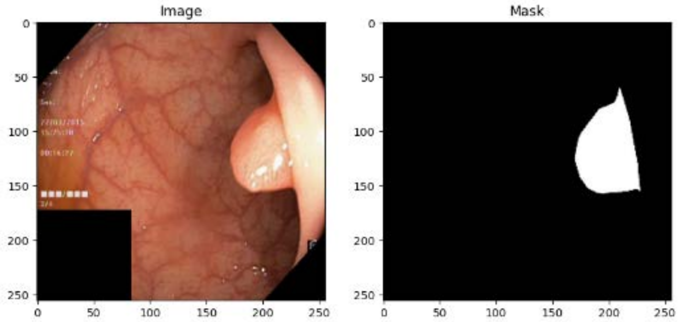

Fig. 2. Seg-Net

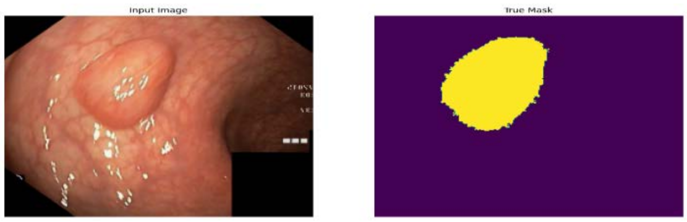

Fig. 3. U-Net

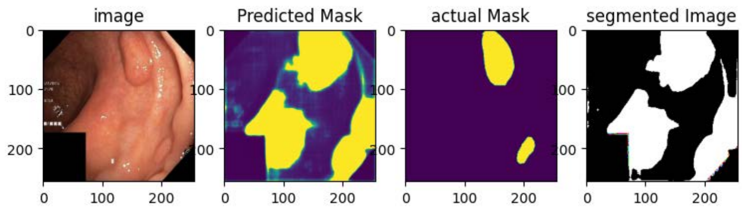

Fig. 4. Deep-Net

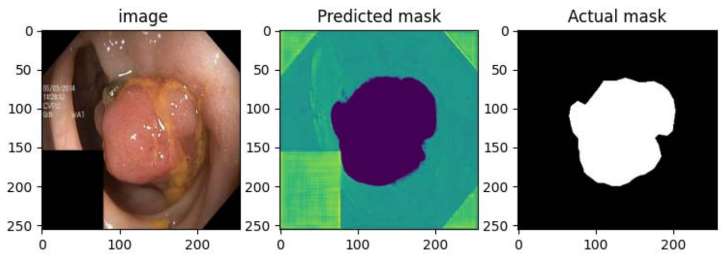

Fig. 5. Custom-Net (GIST-Net)

Fig. 6. Fu-Net

Tab. 3. Comparison of Networks Model visual outputs and accuracy metrics, the Python-driven segmentation analysis

Model	Strength	Weakness
U-Net	Minimalistic architecture	The optimal architecture's depth remains unknown
	Requires fewer training samples	The fusion scheme for feature maps at the same scale is needlessly restrictive
DeepNet	A broader framework	Issue of gridding effect
	It effectively segments objects across multiple scales	Long-range information may not be pertinent
Segnet	It effectively utilizes global contextual information.	A substantial volume of training data is necessary
	It requires lightweight and minimal computational resources	Precisely pinpointing small objects poses a significant challenge
FuNet	Focused on real-time operation	A substantial volume of training data is necessary
	Features a straightforward structure and minimal computational requirements	Limited precision or accuracy is observed
CustomNet	It has the capability to capture intricate contextual relationships	Demanding and costly in terms of computational resources
	It exhibits commendable performance	Requires a significant amount of time

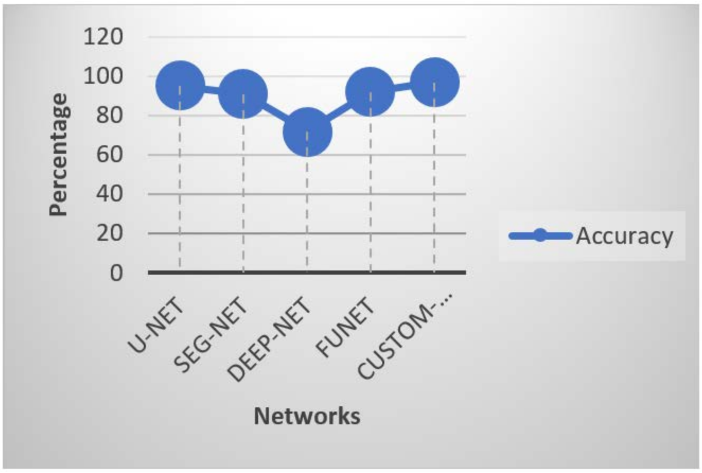

Fig. 7. Representation of graph on networks accuracy level

RESULT

- U-Net and GIST-Net Networks provide good accuracy results for segmentation and Analysis.
- Seg-Net also provide results but not that satisfactory
- So, a Custom Network has been Designed from U-Net and Res-Net called GIST-Net (Gastrointestinal stromal tumors) is improves the level of accuracy in analysis for segmentation.

DISCUSSION

To address class imbalance and variance distribution, we employed two loss functions during training. Our dataset comprised images from a private clinic and a secondary dataset to validate our approach against ground truth annotations. Experimental results demonstrate the proposal's ability to accurately distinguish lesions in input images and achieve precise tumor segmentation.

Future efforts will focus on extending the adopted backbone to enhance instance segmentation of tumor regions. Moreover, our aim is to create a model that can distinguish between healthy and diseased data and identify specific diseased areas like cancerous lesions. We'll achieve this by employing different training sets or classifier networks. This advancement will pave the way for designing an automated system that can detect and categorize abnormalities in endoscopic stomach images, ultimately enhancing disease management and treatment decision-making.

CONCLUSION AND FUTURE SCOPE

In this research, we presented an innovative method utilizing state-of-the-art deep learning techniques to tackle the difficulties associated with automatically detecting, identifying, and categorizing different types of tumors in gastric medical images. We designed and assessed various architectural models to accomplish this task. It includes U-Net, DeepNet, SegNet, FuNet, and CustomNet (GIST-Net), with the latter showing the most promising results in terms of accuracy performance. This proposed network, GIST-Net, effectively identifies abnormal tissue within the stomach. By generating binary masks for the five types of tumors across different grades and also achieved segmentation with minimized training time for improved generalization.

REFERENCES

1. Hatami S, Shamsaee DR, Olyaei MH. Detection and classification of gastric precancerous diseases using deep learning. 2020:28-30;1-5.
2. Zhang Q, Wang F, Chen ZY, Wang Z, Zhi FC, et al. Comparison of the diagnostic efficacy of white light endoscopy and magnifying endoscopy with narrow band imaging for early gastric cancer: a meta-analysis. Gastric Cancer. 2016;19:543-552.
3. Sugimoto M, Kawai Y, Morino Y, Hamada M, Iwata E, Niikura R, et al. Efficacy of high-vision transnasal endoscopy using texture and colour enhancement imaging and narrow-band imaging to evaluate gastritis: a randomized controlled trial. Ann Med. 2022;54:1004-1013.
4. Ang TL, Li JW, Wong YJ, Tan YLJ, Fock KM, et al. A prospective randomized study of colonoscopy using blue laser imaging and white light imaging in detection and differentiation of colonic polyps. Endosc Int Open. 2019;7:1207-1213.
5. Liu L, Liu H, Feng Z. A narrative review of postoperative bleeding in patients with gastric cancer treated with endoscopic submucosal dissection. J Gastrointest Oncol. 2022;13:413.
6. Martin DR, Hanson JA, Gullapalli RR, Schultz FA, Sethi A, et al. A deep learning convolutional neural network can recognize common patterns of injury in gastric pathology. Arch Pathol Lab Med. 2020;144:370-378.
7. Yoon HJ, Kim JH. Lesion-based convolutional neural network in diagnosis of early gastric cancer. Clin Endosc. 2020;53:127-131.
8. Kuchkorov TA, Sabitova NQ, Ochilov TD. Detection of gastric ulcers and lesions applying CNN architecture. Int J Contemp Sci Tech Res. 2022;200-204.
9. Pooja K, Kishore KR. Diagnosis of gastric cancer in role of endoscopic imaging techniques in artificial intelligence and machine learning applications: An overview. E3S Web Conf. 2024;491:03016.
10. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. Trans Pattern Anal Mach Intell. 2016;39:1-1.
11. Lee S-A, Cho HC, Cho H-C. A novel approach for increased convolutional neural network performance in gastric-cancer classification using endoscopic images. IEEE Access. 2021;9:51847-51854.
12. Oukdach Y, Kerkaou Z, El Ansari M, Koutti L, El Ouafdi AF. Gastrointestinal diseases classification based on deep learning and transfer learning mechanism. Rabat, Morocco 2022:26-28;1-6.
13. Pang X, Zhao Z, Weng Y. The role and impact of deep learning methods in computer-aided diagnosis using gastrointestinal endoscopy. Diagnostics. 2021;11:694.
14. Lee JY, Jeong J, Song EM, Ha C, Lee HJ, et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. Sci Rep. 2020;10:1-9.
15. Pooja K, Jerritta S. Denoising technique and analysis of statistical parameters for endoscopic images of gastric cancer. IEEE;2023.1-7.
16. Mushtaq D, Madni TM, Janjua UI, Anwar F, Kakakhail A. An automatic gastric polyp detection technique using deep learning. Int J Imaging Syst Technol. 2023.
17. Doniyorjon M, Madinakhon R, Shakhnoza M, Cho YI. An improved method of polyp detection using custom YOLOv4-Tiny. Appl Sci. 2022;12:10856.
18. Zeng Q, Li H, Zhu Y, Feng Z, Shu X, Wu A, et al. Development and validation of a predictive model combining clinical, radiomics, and deep transfer learning features for lymph node metastasis in early gastric cancer. Front Med. 2022;9:998416.
19. Ma L, Su X, Ma L, Gao X, Sun M. Deep learning for classification and localization of early gastric cancer in endoscopic images. Biomed Signal Process Control. 2023;79:104200.
20. Zhang J, Wen T, He T, Wang X, Hao R, Liu J, et al. Human stools classification for gastrointestinal health based on an improved ResNet18 model with dual attention mechanism. 2022.2096-2103.
21. Sakai Y, Takemoto S, Hori K, Nishimura M, Ikematsu H, et al. Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. 2018.4138-4141.
22. Pooja K, Kishore KR. A systematic review on detection of gastric cancer in endoscopic imaging system in artificial intelligence applications. 2023 16; Springer Nat Singap.337-346.
23. Ikenoyama Y, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, et al. Detecting early gastric cancer: comparison between the diagnostic ability of convolutional neural networks and endoscopists. Dig Endosc. 2022;33:141-150.