A Novel Three-way fusion image segmentation for early esophageal cancer detection

Xintong Ren1, Zhen Zhang2, Junchao Jiang, Guodong Li2, Jiahua Jiang3, Wuwei Ren4, Xinyong Jia2

1Lily AI

2The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital

3University of Birmingham

4ShanghaiTech University

XR, ZZ contributed equally.

Correspondence to

Xinyong Jia, Junchao Jiang and Guodong Li

jiaxinyong0505@126.com, jiangjunchao0830@yahoo.com, lgd252340@163.com.

Abstract

Background/aims

Esophageal cancer (EC) is a prevalent malignancy worldwide and is ranked as the sixth most common cause of cancer-related deaths, where nearly half of the new cases are reported in China annually. The detection of esophageal cancer (EC) is crucial for improving patient survival. However, EEC is a highly aggressive form of cancer with a poor prognosis, even among experienced endoscopists. Therefore, computer-aided diagnosis (CAD) has emerged as a valuable tool to assist endoscopists in the accurate and rapid detection of EEC.

Methods

We have developed a novel three-way fusion CAD system for detecting EEC, which allows physicians to take a comprehensive approach by investigating both local and overall boundaries in the clinical setting.

Results

In terms of achieving an accurate treatment boundary in clinical practice, we obtained a top performance of 0.9236 mAP on ME-NBI from the local perspective using a novel dual-annotation method that we innovated. The treatment boundary mAP should be even higher than this by definition. From the overall perspective, we also achieved a top performance of 0.862 mAP on LCE.

Conclusions

Our three-way fusion CAD system has demonstrated excellent diagnostic performance for EEC detection using deep learning, providing accurate treatment boundaries on both image and patient level. This shows great potential for clinical application. Moving forward, we plan to collect and analyze patient data from a broader area to achieve real-time performance in clinical settings.

Introduction

Esophageal cancer (EC) is a significant global health challenge, ranking as the sixth highest cause of cancer-related deaths. In particular, China experiences a high incidence of EC, with nearly half of all new cases reported. Therefore, early detection and intervention are crucial for improving patient outcomes. The five-year survival rate for EC exceeds 80% [1, 2] when diagnosed at early stages. However, it drops to just 20% with late detection [3]. To improve survival rates, there is a need for effective techniques that enable earlier diagnosis and curative treatment, as well as proper screening and early detection practices.

Accurately defining the boundaries of esophageal cancer is a critical challenge in clinical practice to improve the survival rate of early-stage patients. The treatment boundary is a fundamental reference for endoscopic submucosal dissection (ESD) surgery, which is essential for subsequent ESD treatment. However, even experienced endoscopists may struggle to identify early-stage esophageal cancer, particularly when distinguishing intramucosal neoplasias that appear as small

erosions, flat mucous membranes, or normal mucosa in biopsy specimens. Incorrect boundary definition can lead to adverse outcomes, such as scar stenosis and difficulty in eating after surgery if the boundary is larger than the actual lesion. On the other hand, if the boundary is smaller than the actual lesion, it can lead to residual lesions prone to recurrence with a poor prognosis. To address these challenges, computer-aided diagnosis (CAD) is gaining importance as a tool to improve accuracy, save time, reduce burden, and standardize diagnosis. CAD has shown promising developments and has the potential to enhance the diagnostic process in EC.[4]

To address this clinical challenge, it is imperative for physicians to take a comprehensive approach that investigates both local and overall boundaries, in addition to pathological comparison. In this article, we present our pioneering use of a three-way fusion CAD technology. By optimizing the local boundary, we utilize CAD technology to assist in the diagnosis of ME-NBI images. To optimize the overall boundary, we also employ CAD technology to assist in the diagnosis of Lugol's chromoendoscopy (LCE) images, providing an innovative and effective approach to the precise diagnosis of gastrointestinal lesions.

ME-NBI imaging magnified lesions to 80-110 times, presents several advantages in terms of enhancing diagnostic accuracy, lesion visibility, and image quality consistency. Additionally, it also provides better visualization of microvascular patterns and tissue structures, which aids physicians in making more precise assessments of lesion boundaries.[5] However, traditional ME-NBI CAD technology has not been effective in providing accurate treatment boundaries in clinical settings. This is primarily due to the unevenness of lesions, resulting in distinct information on the main plane, while indistinct on side planes made the boundary ambiguous. Moreover, the focus of the image may vary, creating shadows in the image, and image reconstruction may result in blurriness due to motion artifacts, image processing, scatter radiation, and equipment issues. The final treatment boundary required in clinical practice comprises the distinct as well as indistinct area of the lesion. The indistinct area lies between the distinct lesion area and the non-lesion area, with ambiguous features, making the treatment boundary hard to determine. Additionally, shadows and blurriness can directly interfere with the treatment boundary determination.

We propose two annotation categories to make enhancement. 1) distinct boundary category (DBC): includes only the distinct portions of the lesion and excludes areas of indistinctness, shadowing, and blurriness. 2) comprehensive boundary category (CBC): includes both distinct and indistinct areas of the lesion, while excluding focus and shadow interference areas. Physicians determine this boundary after reviewing all ME-NBI images of a patient in a comprehensive manner.

As a single ME-NBI image may not cover all areas of a lesion due to its magnification, physicians must review multiple images to diagnose the overall lesion region. Different images may capture the same region from different angles, resulting in indistinct areas being displayed as distinct in some images. Our dual annotation categories offer advantages over traditional CAD methods that only predict the CBC. Physicians can compare the prediction results between DBC and CBC across all related ME-NBI images, thus alleviating the problem of low reliability caused by

indistinct areas' unclear features. This greatly improves the accuracy of treatment boundary prediction in clinical practice.

We also introduced a novel technique, CAD-assisted LCE imaging, for precise treatment boundary visualization from a macroscopic perspective. Compared to ME-NBI imaging, LCE imaging has distinct advantages, including non-magnification, complete lesion boundary display, low equipment requirements, low medical costs, and a wide range of applications. Endoscopists have found LCE imaging to enhance the visualization of certain lesions, making it easier to identify areas of abnormal tissue. Additionally, LCE imaging can aid in real-time diagnosis of neoplastic changes during endoscopy, potentially reducing the need for biopsy sampling. Notably, LCE imaging has demonstrated high sensitivity (>95%) in detecting EEC. [6]

Methods

2.1 Experimental dataset

We sourced data from patients diagnosed with EEC between January 2010 and January 2020 at the First Affiliated Hospital of Shandong First Medical University. Our focus was on endoscopic images of patients diagnosed with high-grade dysplasia or T1 stage adenocarcinoma, with endoscopies performed using Olympus series upper endoscopes (260 HQ and 290 HQ; Olympus, Tokyo, Japan). We collected a total of 3500 images from 800 patients, including images of abnormal hyperplasia confirmed by EEC histology. The histological evaluations were performed by four experienced gastrointestinal pathologists, and the endoscopy images were annotated by two endoscopists with over fifteen years of experience in surgical endoscopy. The criteria for defining non-abnormal EC images and annotation methodology were described in the appendix.

2.2 CAD system Results

In this research endeavor, we propose eight frameworks that were built by combining several recently published advanced modules [21]. We selected some critical modules implemented and introduced in Appendix. The model name abbreviations and descriptions are delineated in (Table 1). Our experimentation involved applying all eight frameworks to two ME-NBI categories and one LCE category. In order to measure the overall performance, we used Mean Average Precision (mAP). Additionally, to evaluate the lesion detection at the pixel level, we provided average IoU (avgIoU), average Dice (avgDice), Average Recall Rate (ARR), and Average Precision Rate (APR). Furthermore, we used True Positive (TP), False Positive (FP), False Negative (FN), Precision, Recall, and F1 Score to evaluate the performance of the lesion detection at the object level. The results of all 24 experiments, along with the measurements at 0.5 IoU level, are presented in the Appendix.

The obtained results demonstrate that the mAP for CBC is nearly 0.91 and the average IoU is 0.88, surpassing the mAP of 0.84 and average IoU of 0.80 for DBC. This is consistent with our theoretical expectations, as CBC corresponds to the actual treatment boundary, while DBC only encompasses the distinct lesion area and serves to improve the clinical accuracy of CBC. This suggests that in practical clinical settings, the accuracy of the treatment boundary is better than the performance of CBC. From a deep learning perspective, the superior performance of CBC over

DBC implies that differentiating between non-lesion areas and indistinct lesion areas is easier than distinguishing indistinct and distinct lesion areas.

The utilization of LCE category prediction has demonstrated its effectiveness in improving the precision of the clinical treatment boundary by enabling a comprehensive view of the entire lesion. The evaluation of mAP and average IoU for LCE resulted in values of 0.86 and 0.84, respectively. It should be noted that these metrics are lower than those achieved by CBC due to the fact that ME-NBI magnifies the image 80-110 times, providing a more detailed and intricate depiction of the lesion. However, since LCE covers the entire lesion boundary, which is typically a small proportion of the image, it poses a more challenging task than CBC.

Below, we present one predicted example for each category. The images in each row, from left to right, show the original image, the image with the annotated lesion region, and the image with the predicted lesion region.

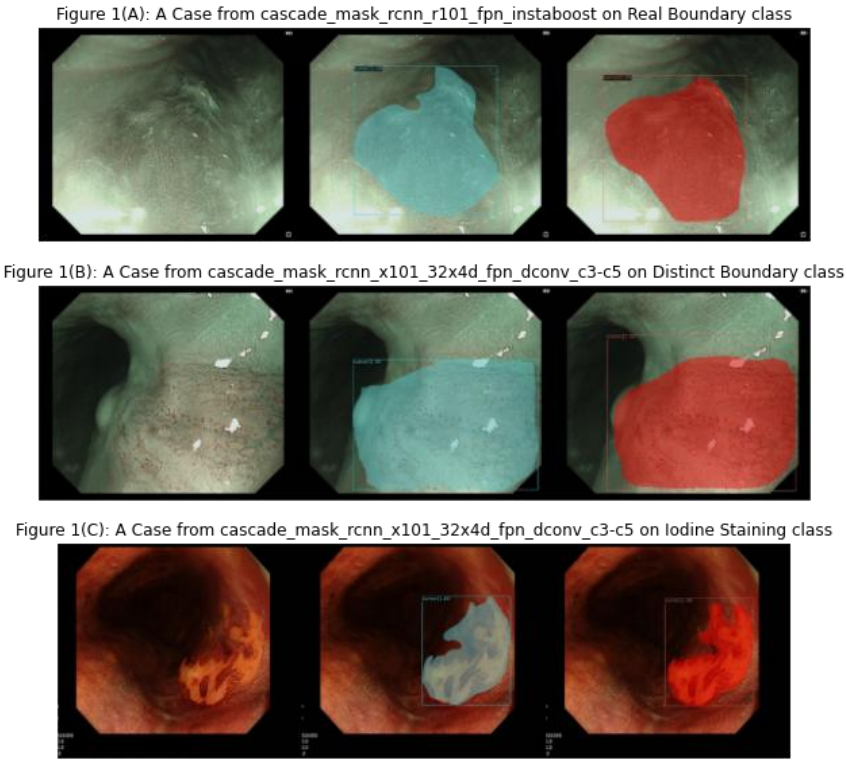

Figure 1 *Predicted example for each category*

Discussion

3.1 Image level

ME-NBI imaging can be classified into two categories: Whole Lesion Coverage Imaging (WLCI) and Incomplete Lesion Coverage Imaging (ILCI). In WLCI, the entire image is encompassed by CBC, while in ILCI, only a portion of the image is covered by CBC.

3.1.1 Whole Lesion Coverage Imaging

In our study, we found that 36% of all ME-NBI images fall under the category of WLCI. In such cases, the image only captures a partial area of the lesion, and the clinical treatment boundary lies outside the image. Our analysis of the annotations showed that the area of indistinct regions in WLCI are not insignificant, and in some cases, the edges mainly consist of indistinct regions, as shown in (Figure 2a). Physicians differentiate indistinct regions as either lesions or non-lesions only after careful consideration, comparison, and discussion during the image annotation process. However, in actual clinical practice, physicians may face challenges in quickly distinguishing between indistinct lesion areas, non-lesion areas, and eliminating shadows and blurriness. Our CAD system's performance in WLCI on CBC shows a nearly 1.00 mAP and an average IoU greater than 0.95 (Figure 2b). In other words, our CAD system can assist physicians in making judgments on indistinct regions in clinical practice for WLCI promptly.

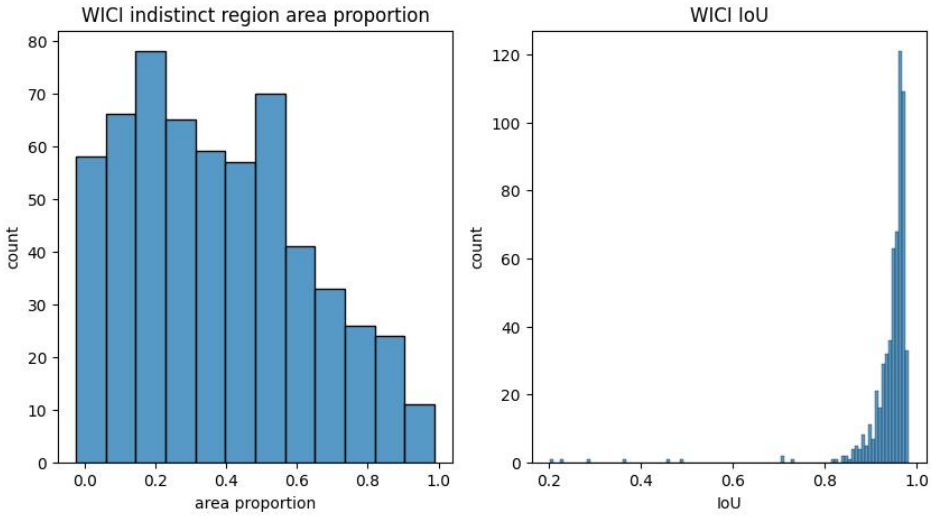

Figure 2(a)(b) *WLCI indistinct region area proportion and IoU histogram*

3.1.2 Incomplete Lesion Coverage Imaging

ILCI can be further subdivided into two groups based on the proportion of indistinct areas in the image. The first group comprises images that do not feature any indistinct regions, leading to the CBC region being the same as the DBC region. The second group, however, contains indistinct regions, resulting in the CBC region being more extensive than the DBC region as depicted in (Figure 3a). In both categories, physicians must determine the clinical treatment boundary.

Our experimental findings reveal that ILCI achieves an mAP of 0.93 in CBC, and the distribution of IoU values is illustrated in Figure 3(b). Despite ILCI achieving a high mAP score, indicating excellent performance, it still performs slightly worse than WLCI in CBC detection. This observation suggests that differentiating between the indistinct lesion area and the non-lesion area is a challenging task, and ILCI tends to predict a larger region than the ground truth region. Specifically, ILCI exhibits higher sensitivity but lower specificity, making it more prone to falsely predicting non-lesion indistinct areas as lesions and unable to resist shadow and blur interference.

To address these issues, our CAD system incorporates DBC prediction to mitigate the problems. For instance, region A in image F1 is an indistinct non-lesion area, but ILCI incorrectly predicted it as a CBC. However, in image F2, region A is distinctly visible from a different angle, and the DBC prediction in F2 excludes region A. This assists physicians in determining that region A is, in fact, a non-lesion area. Our study emphasizes the importance of developing effective dual annotation methods to aid physicians in accurately diagnosing and treating CBC lesions.

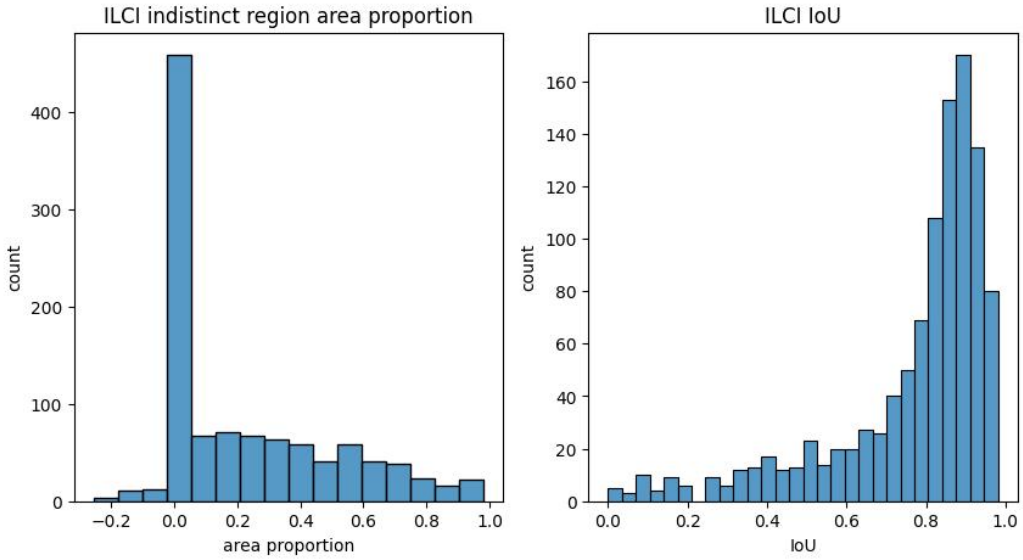

Figure 3(a)(b) *ILCI indistinct region area proportion and IoU histogram*

In addition, the correlation between the proportion of indistinct areas and the IoU is only 0.1, which can be considered irrelevant. Furthermore, the t-test results for the proportion of indistinct areas between the IoU>0.5 and IoU<0.5 buckets are also insignificant. This indicates that the IoU is not directly related to the proportion of indistinct areas, but rather to the distinctiveness of the image itself at the boundary between the indistinct lesion area and the non-lesion area.

Our results show that DBC outperforms CBC, with an average IoU of 0.79 compared to 0.74 for the 424 images that do not contain any indistinct regions. We also found that smaller lesions are more challenging to predict than larger ones for both CBC and DBC. However, DBC performs significantly better than CBC for lesions with an area smaller than 400000, with an average IoU of 0.74 compared to 0.65 for CBC. Our t-test results indicate that the difference in IoU between DBC and CBC is statistically significant. Overall, our findings suggest that DBC can better assist physicians in determining the treatment boundary, especially when all lesions are in a distinct state, particularly when the lesion area is small, compared to traditional CBC annotation.

Figure 4 depicts a small lesion with all distinct regions. The figure includes (a), (b), and (c), representing the annotated area, the predicted area by CBC, and the predicted area by DBC, respectively. Our analysis reveals that CBC's prediction result is interfered with due to the presence of indistinct regions caused by the side plane and factors such as focus-induced glare in the upper part of the image. As a result, CBC's prediction falsely predicted extra areas as lesions. In contrast, the DBC prediction result in (c) is highly accurate, with an IoU of up to 98%. These

results demonstrate the effectiveness of DBC in accurately predicting lesion boundaries, especially in cases where all lesions are in a distinct state.

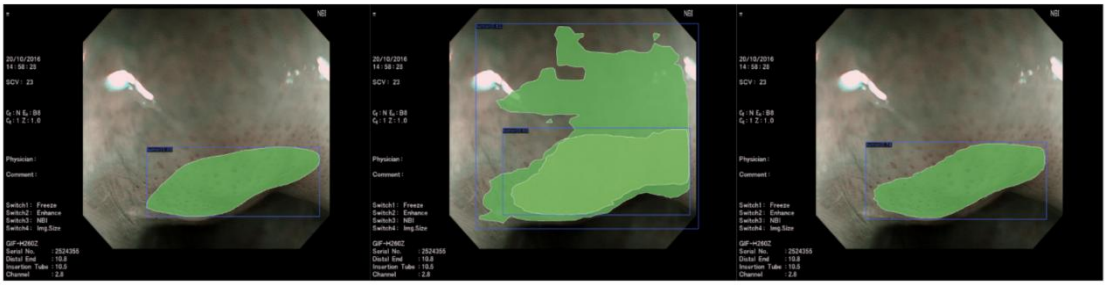

Figure 4(a)(b)(c) *An example with Annotated region, CBC predicted region, and DBC predicted region*

3.2 Patient level

On average, a patient may have 4-5 ME-NBI images, with some patients as many as 40. At the patient level, the mAP in CBC can reach up to 0.95. For patients below 0.95, after our CAD system assists in the DBC category, there are at least 43% of patients have mAP higher than 0.95, indicating that our CAD system's high performance is not limited to the high diagnostic rate of CBC but also extends to assisting patients who do not meet the standard line through DBC. As a result, the performance of the clinical treatment boundary at the patient level is also superior to that of CBC.

3.3 Lugol's chromoendoscopy

In comparison to ME-NBI images, LCE images exhibit higher sensitivity in lesion boundaries but lower specificity. Furthermore, certain non-tumorous regions with inflammation can also display shallow staining areas, leading to potential overestimation of the predicted area compared to the ground truth. However, in our CAD system, the APR and ARR are generally equal, indicating that the predicted area is similar to the ground truth area. Additionally, the IoU performance is not directly influenced by the staining chromatics of the lesion or the color difference between the lesion and non-lesion regions. This is because we conducted color normalization on LCE images during model training, providing the model with strong robustness in distinguishing between lesion and non-lesion regions under different staining effects and identifying diverse chromatic differences caused by staining. Furthermore, akin to ME-NBI, the IoU performance of LCE is not related to the size of the lesion itself but rather to the disparity in the image at the junction between the lesion and non-lesion regions.

3.4 Causes of FN, FP

In most cases where ME-NBI images are falsely predicted, the model erroneously identifies non-lesion areas as lesions, particularly in CBC. Two common factors that contribute to such outcomes are indistinct areas on the side plane that make it challenging for the model to distinguish accurately, as seen in (Figure 4b), and interference caused by imaging flashes and shadows, as depicted in (Figures 5), which respectively display the CBC labeled area and predicted area. The glare point in the upper left corner and the shadow area in the upper right

corner both cause interference to the model's prediction. Additionally, the DBC IoU in this example is higher than 0.90, which can assist physicians.

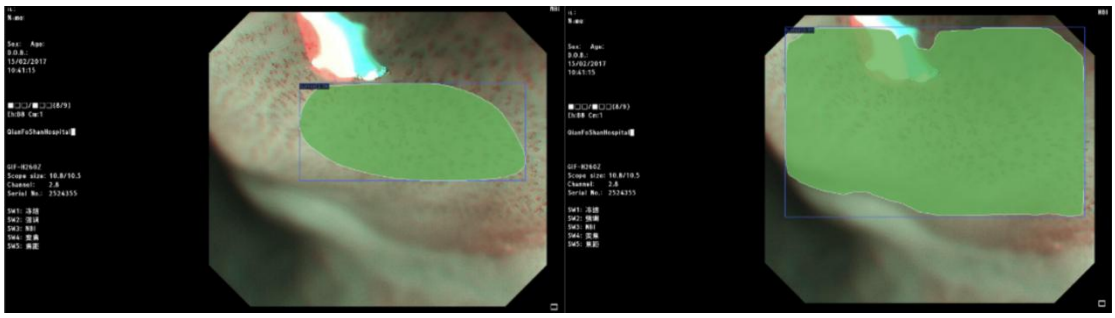

Figure 5(a)(b) *Interference caused by imaging flashes and shadows*

Conclusion

This paper presents our novel three-way fusion CAD system for detecting EEC. We conducted 24 experiments on two ME-NBI categories and one LCE category. By using DBC results assisting, our dual annotation categories offer advantages over traditional CAD methods that only predict the CBC. Our best results for CBC category were achieved using the hybrid task cascade ResNeXt101 with deformable convolutional networks, with 0.9236 mAP, 0.907 APR, and 0.9173 ARR at a 0.5 threshold IoU level. We also achieved promising results of 93.98% Precision and 93.05% Recall. The LCE category showed the best results using the hybrid task cascade framework, with 0.8620 mAP, 0.9386 APR, and 0.9433 ARR, along with 82.89% Precision and 88.32% Recall. Our three-way fusion method demonstrated excellent diagnostic performance with accurate treatment boundaries on both image and patient level. These results confirm the feasibility of deep learning for detecting EEC in both ME-NBI and LCE. We plan to collect and analyze patient data from a wider area to achieve real-time performance for clinical applications.

Reference

1. Wang GQ,Jiao GG,Chang FB,et al.Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening.Ann Thorac Surg.2004 May;77(5):1740-4.PMID: 15111177.

2. Yang S,Wu S,Huang Y,et al.Screening for oesophageal cancer. Cochrane Database Syst Rev.2012 Dec12;12:CD007883.PMID: 23235651.

3. Zeng H,Zheng R,Guo Y,et al.Cancer survival in China,2003-2005: a population-based study.Int J Cancer.2015 Apr 15;136(8):1921-30.PMID: 25242378.

4. S.Van Riel,F.Van Der Sommen,S.Zinger,et al."Automatic Detection of Early Esophageal Cancer with CNNS Using Transfer Learning," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 1383-1387.

5. Lecleire S,Antonietti M,Iwanicki-Caron I,et al.Lugol chromo-endoscopy versus narrow band imaging for endoscopic screening of esophageal squamous-cell carcinoma in patients with a history of cured esophageal cancer: a feasibility study. Dis Esophagus. 2011 Aug;24(6):418-22.PMID: 21309914.

6. Protano MA, Xu H, Wang G,et al.Low-Cost High-Resolution Microendoscopy for the Detection of Esophageal Squamous Cell Neoplasia:An International Trial.Gastroenterology.2015 Aug;149(2):321-329.PMID: 25980753.

7. Setio,A.A.A,Sommen,van der,F. Zinger,et al.Evaluation and comparison of textural feature representation for the detection of early stage cancer in endoscopy.In Proceedings of the 8th International Conference on Computer Vision Theory and Applications,2013Feb;https://doi.org/10.5220/0004204502380243

8. Markus HA,Sommen FVD,Zinger S,et al."Early esophageal cancer detection using RF classifiers",Medical Imaging 2016: Computer-Aided Diagnosis, 2016 Mar; doi:10.1117.

9. Ghatwary N, Zolgharni M, Ye X. Early esophageal adenocarcinoma detection using deep learning methods. Int J Comput Assist Radiol Surg. 2019 Apr;14(4):611-621.PMID: 30666547.

10. Ghatwary N,Ye X,Zolgharni M,"Esophageal Abnormality Detection Using DenseNet Based Faster R-CNN With Gabor Features," 2019 IEEE Access, vol. 7, pp. 84374-84385.

11. Hashimoto R, Requa J, Dao T,et al.Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). Gastrointest Endosc. 2020 Jun;91(6):1264-1271.e1.PMID: 31930967.

12. Wang YK,Syu HY,Chen YH,et al.Endoscopic Images by a Single-Shot Multibox Detector for the Identification of Early Cancerous Lesions in the Esophagus: A Pilot Study. Cancers (Basel). 2021 Jan 17;13(2):321.PMCID: PMC7830509.

13. Fang H,Sun J, Wang R,et al."InstaBoost: Boosting Instance Segmentation via Probability Map Guided Copy-Pasting."2019 IEEE/CVF International Conference on Computer Vision (ICCV), 682-691.

14. J. Dai et al."Deformable Convolutional Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 764-773, doi: 10.1109.

15. Cai Z,Vasconcelos N.Cascade R-CNN: "Delving Into High Quality Object Detection."2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6154-6162.

16. He K,Gkioxari G,Dollár P,et al."Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV),2017,pp. 2980-2988, doi: 10.1109.

17. Chen K,Pang J,Wang J,et al."Hybrid Task Cascade for Instance Segmentation."2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),4969-4978.

18. He K,Zhang X,Ren S,et al."Deep Residual Learning for Image Recognition."2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),770-778.

19. Huang Z,Huang L,Gong Y,et al. "Mask Scoring R-CNN," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2019,pp.6402-6411,doi: 10.1109.

20. Cao Y,Xu J,Lin S,et al."GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW),2019,pp.1971-1980,doi: 10.1109.

21. Chen KW,Jiaqi P,Jiangmiao C,et al."Open MMLab Detection Toolbox and Benchmark,"arXiv preprint arXiv:1906.07155 2019.

Appendix

Annotation Criteria

In the course of processing the experimental dataset, we meticulously followed a set of plans and procedures that encompassed defining the non-abnormity of EEC images and performing annotations. The corresponding details are presented below.

To classify an image as non-abnormity, it had to meet all of the following three criteria: 1) The patient had no prior history of esophageal lesions detected during examination. 2) Confirmation was obtained from endoscopists with over a decade of experience in endoscopy. 3) Unanimous agreement was reached on the non-abnormity status.

For the annotation of abnormity images, we proposed an annotation methodology and adhered strictly to the following steps to ensure consistency and professionalism: 1) The first endoscopist reviewed the gastroscopy image and marked the lesion as a polygon. 2) The second endoscopist examined the annotation results and added notes as necessary. 3) If there were no disagreements, the annotation was considered complete. If a disagreement arose, a third endoscopist was consulted, and the three endoscopists collaborated to arrive at the final decision.

As a single ME-NBI image may not cover all areas of a lesion due to its magnification, physicians must review multiple images to diagnose the overall lesion region. Commonly, there are 4 to 5, up to 40 ME-NBI images per patient.

Deep learning modules

In this research, modules implemented in our CAD like Deformable Convolutional Networks (DCN) [14] introduced a flexible sampling grid through variable convolution and variable RoI pooling; Random InstaBoost [13] achieved significant improvements in mean average precision through image randomization inspired by stochastic grammar of images; Cascade RCNN [15] addressed overfitting in multistage target detection architecture by utilizing a series of detectors with equal positive training set sizes; And other advanced methods such as the Hybrid Task Cascade series [17], Mask Scoring R-CNN [19], and GCNet [20] also have been applied in our system.

Model performance overview

We proposed eight frameworks for each category by combining multiple advanced modules published recently [21]. The model name abbreviations and descriptions are shown in Table 1. We experimented with all eight frameworks on two ME-NBI categories and one LCE category, and the results of the total 24 experiments are presented in Table 2. The optimal EEC detection framework was obtained by comparing all performance results, the hybrid task cascade ResNeXt101 with deformable convolutional networks stands out from the other frameworks among all categories.

As shown in Table 2, the bounding box level outperforms the polygon level, which is as expected. Predicting polygons involves fine boundaries, making it more challenging to predict accurately. However, in terms of reliability, the performance of pixel-level polygons is more convincing since it considers the actual area of the tumor lesion object and excludes any background noise interference.

Table 3 displays the Precision, Recall, and F1 Scores for all frameworks across all categories, along with the number of True Positive (TP), False Positive (FP), and False Negative (FN) objects detected on the test set. The performance metrics are calculated for both Bounding boxes level and Polygon level.

From Table 3, we observe that all eight frameworks performed well at the bounding box level, achieving an F1 score of more than 80%. Notably, the use of Cascade RoI Head to select the region of interest (RoI) resulted in better performance due to reduced FP. This approach prevents the model from overfitting on the incremental IoU, making it more robust, especially for objects with large sizes.

Additionally, we found that all frameworks had higher Recall scores than Precision scores. This is desirable as we intend to use these applications in a CAD system, where the goal is to present prediction results to doctors before arriving at a final diagnosis of tumor regions. A better Recall model would ensure a lower missing detection rate, minimizing the omission and negligence of EEC cases, which is critical for the intended application.

$$\mathrm{IoU} = \frac{B^g \cap B^p}{B^g \cap B^p}$$

$$\mathrm{avgIoU} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IoU}$$

$$\text{avgDice} = \frac{1}{N}\sum_{i=1}^{N}\frac{2 * \text{IoU}}{\text{IoU} + 1}$$

$$\text{APR} = \frac{1}{N}\sum_{i=1}^{N}\frac{B^g \cap B^p}{B^p}$$

$$\text{ARR} = \frac{1}{N}\sum_{i=1}^{N}\frac{B^g \cap B^p}{B^g}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Metrics	**Explanation**
N	total number of images
B^g	the ground truth area (bounding box/polygon) of the tumor region
B^p	the predicted area (bounding box/polygon) of the tumor region
mAP	the average of Average Precision (AP), area under the precision-recall curve above
TP	count of True Positive tumor object, with a 0.5 IoU threshold
FP	count of False Positive tumor object, with a 0.5 IoU threshold
FN	count of False Negative tumor object, with a 0.5 IoU threshold

Table 1: Explanation of the metrics and the corresponding description

Real Boundary on Bounding Box					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.8180	0.8945	0.8955	0.9119	0.8414
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.8087	0.8890	0.8933	0.9041	0.8356
cascade_mask_rcnn+ResNet101+FPN	0.8044	0.8858	0.8718	0.9207	0.7863
mask_scoring_rcnn+ResNeXt101+FPN	0.7926	0.8793	0.8780	0.9018	0.8273
hybrid_task_cascade+ResNeXt101+FPN	0.8141	0.8921	0.8874	0.9149	0.8127
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.8233	0.8984	0.9060	0.9081	0.8511
mask_rcnn+ResNet101+FPN	0.7965	0.8820	0.8798	0.9020	0.8487
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.8202	0.8964	0.8987	0.9124	0.8455

Real Boundary on Polygon Mask					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.7704	0.8654	0.8679	0.8845	0.7947
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.7634	0.8608	0.8796	0.8639	0.7829
cascade_mask_rcnn+ResNet101+FPN	0.7429	0.8473	0.8544	0.8638	0.7165
mask_scoring_rcnn+ResNeXt101+FPN	0.7440	0.8483	0.8712	0.8489	0.7709
hybrid_task_cascade+ResNeXt101+FPN	0.7570	0.8570	0.8681	0.8676	0.7606
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.7778	0.8705	0.8868	0.8747	0.8277
mask_rcnn+ResNet101+FPN	0.7666	0.8636	0.8778	0.8673	0.7871
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.7775	0.8703	0.8768	0.8849	0.7915

Distinct Boundary on Bounding Box					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.8866	0.9354	0.9256	0.9570	0.9191
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.8773	0.9302	0.9285	0.9432	0.9158
cascade_mask_rcnn+ResNet101+FPN	0.8795	0.9315	0.9238	0.9518	0.9211
mask_scoring_rcnn+ResNeXt101+FPN	0.8684	0.9257	0.9276	0.9334	0.9117
hybrid_task_cascade+ResNeXt101+FPN	0.8835	0.9342	0.9270	0.9514	0.9059
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.8865	0.9363	0.9386	0.9433	0.9236
mask_rcnn+ResNet101+FPN	0.8643	0.9224	0.9192	0.9385	0.9189
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.8877	0.9367	0.9371	0.9463	0.9364

Distinct Boundary on Polygon Mask					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.8619	0.9215	0.9136	0.9406	0.8753
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.8483	0.9134	0.9177	0.9219	0.8926
cascade_mask_rcnn+ResNet101+FPN	0.8537	0.9168	0.9141	0.9319	0.8935
mask_scoring_rcnn+ResNeXt101+FPN	0.8445	0.9110	0.9208	0.9128	0.8792
hybrid_task_cascade+ResNeXt101+FPN	0.8559	0.9183	0.9174	0.9310	0.8795
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.8511	0.9154	0.9245	0.9181	0.9079
mask_rcnn+ResNet101+FPN	0.8439	0.9106	0.9163	0.9172	0.8848
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.8562	0.9182	0.9246	0.9243	0.9087

Iodine Staining on Bounding Box					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.8417	0.9104	0.9083	0.9250	0.8315
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.8170	0.8958	0.9068	0.8988	0.8042
cascade_mask_rcnn+ResNet101+FPN	0.8390	0.9090	0.9144	0.9167	0.8332
mask_scoring_rcnn+ResNeXt101+FPN	0.8035	0.8881	0.9026	0.8883	0.8122
hybrid_task_cascade+ResNeXt101+FPN	0.8305	0.9037	0.9061	0.9155	0.8319
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.8320	0.9050	0.9070	0.9173	0.8620
mask_rcnn+ResNet101+FPN	0.7997	0.8849	0.8938	0.8920	0.8434
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.8332	0.9055	0.8999	0.9237	0.8408

Iodine Staining on Polygon Mask					
Model	avgIoU	avgDice	APR	ARR	mAP
cascade_mask_rcnn+ResNet101+FPN+instaboost	0.7899	0.8791	0.8892	0.8800	0.8423
cascade_mask_rcnn+ResNeXt101+FPN+DCN	0.7753	0.8701	0.8866	0.8657	0.7960
cascade_mask_rcnn+ResNet101+FPN	0.7867	0.8771	0.8990	0.8679	0.8445
mask_scoring_rcnn+ResNeXt101+FPN	0.7753	0.8702	0.8933	0.8602	0.8325
hybrid_task_cascade+ResNeXt101+FPN	0.7897	0.8793	0.8967	0.8738	0.8403
hybrid_task_cascade+ResNeXt101+FPN+DCN	0.7933	0.8817	0.8984	0.8756	0.8501
mask_rcnn+ResNet101+FPN	0.7744	0.8695	0.8944	0.8571	0.8545
hybrid_task_cascade+ResNeXt101+FPN+instaboost	0.7914	0.8805	0.8889	0.8832	0.8295

Table 2 *Image level performance based on Bounding Box and Polygon Masks instance segmentation*

Real Boundary on Bounding Box						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	325	58	47	0.8486	0.8737	0.8609
cascade_mask_rcnn+ResNeXt101+FPN+DCN	328	59	44	0.8475	0.8817	0.8643
cascade_mask_rcnn+ResNet101+FPN	312	62	60	0.8342	0.8387	0.8365
mask_scoring_rcnn+ResNeXt101+FPN	323	68	49	0.8261	0.8683	0.8467
hybrid_task_cascade+ResNeXt101+FPN	320	54	52	0.8556	0.8602	0.8579
hybrid_task_cascade+ResNeXt101+FPN+DCN	331	48	41	0.8734	0.8898	0.8815
mask_rcnn+ResNet101+FPN	331	75	41	0.8153	0.8898	0.8509
hybrid_task_cascade+ResNeXt101+FPN+instaboost	330	68	42	0.8291	0.8871	0.8571

Real Boundary on Polygon Mask						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	310	73	62	0.8094	0.8333	0.8212
cascade_mask_rcnn+ResNeXt101+FPN+DCN	312	75	60	0.8062	0.8387	0.8221
cascade_mask_rcnn+ResNet101+FPN	290	83	82	0.7775	0.7796	0.7785
mask_scoring_rcnn+ResNeXt101+FPN	309	81	63	0.7923	0.8306	0.8110
hybrid_task_cascade+ResNeXt101+FPN	306	69	66	0.8160	0.8226	0.8193
hybrid_task_cascade+ResNeXt101+FPN+DCN	323	56	49	0.8522	0.8683	0.8602
mask_rcnn+ResNet101+FPN	315	85	57	0.7875	0.8468	0.8161
hybrid_task_cascade+ResNeXt101+FPN+instaboost	313	85	59	0.7864	0.8414	0.8130

Distinct Boundary on Bounding Box						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	280	24	22	0.9211	0.9272	0.9241
cascade_mask_rcnn+ResNeXt101+FPN+DCN	279	24	23	0.9208	0.9238	0.9223
cascade_mask_rcnn+ResNet101+FPN	280	22	22	0.9272	0.9272	0.9272
mask_scoring_rcnn+ResNeXt101+FPN	278	34	24	0.8910	0.9205	0.9055
hybrid_task_cascade+ResNeXt101+FPN	276	25	26	0.9169	0.9139	0.9154
hybrid_task_cascade+ResNeXt101+FPN+DCN	281	18	21	0.9398	0.9305	0.9351
mask_rcnn+ResNet101+FPN	281	33	21	0.8949	0.9305	0.9123
hybrid_task_cascade+ResNeXt101+FPN+instaboost	286	19	16	0.9377	0.9470	0.9423

Distinct Boundary on Polygon Mask						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	270	34	32	0.8882	0.8940	0.8911
cascade_mask_rcnn+ResNeXt101+FPN+DCN	275	28	27	0.9076	0.9106	0.9091
cascade_mask_rcnn+ResNet101+FPN	274	27	28	0.9103	0.9073	0.9088
mask_scoring_rcnn+ResNeXt101+FPN	271	37	31	0.8799	0.8974	0.8885
hybrid_task_cascade+ResNeXt101+FPN	270	31	32	0.8970	0.8940	0.8955
hybrid_task_cascade+ResNeXt101+FPN+DCN	278	22	24	0.9267	0.9205	0.9236
mask_rcnn+ResNet101+FPN	273	40	29	0.8722	0.9040	0.8878
hybrid_task_cascade+ResNeXt101+FPN+instaboost	281	25	21	0.9183	0.9305	0.9243

Iodine Staining on Bounding Box						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	181	39	33	0.8227	0.8458	0.8341
cascade_mask_rcnn+ResNeXt101+FPN+DCN	178	35	36	0.8357	0.8318	0.8337
cascade_mask_rcnn+ResNet101+FPN	183	29	31	0.8632	0.8551	0.8592
mask_scoring_rcnn+ResNeXt101+FPN	185	62	29	0.7490	0.8645	0.8026
hybrid_task_cascade+ResNeXt101+FPN	183	33	31	0.8472	0.8551	0.8512
hybrid_task_cascade+ResNeXt101+FPN+DCN	189	39	25	0.8289	0.8832	0.8552
mask_rcnn+ResNet101+FPN	190	69	24	0.7336	0.8879	0.8034
hybrid_task_cascade+ResNeXt101+FPN+instaboost	184	38	30	0.8288	0.8598	0.8440

Iodine Staining on Polygon Mask						
Model	TP	FP	FN	Precision	Recall	F1
cascade_mask_rcnn+ResNet101+FPN+instaboost	183	35	31	0.8394	0.8551	0.8472
cascade_mask_rcnn+ResNeXt101+FPN+DCN	177	36	37	0.8310	0.8271	0.8290
cascade_mask_rcnn+ResNet101+FPN	185	25	29	0.8810	0.8645	0.8726
mask_scoring_rcnn+ResNeXt101+FPN	189	50	25	0.7908	0.8832	0.8344
hybrid_task_cascade+ResNeXt101+FPN	184	31	30	0.8558	0.8598	0.8578
hybrid_task_cascade+ResNeXt101+FPN+DCN	187	41	27	0.8202	0.8738	0.8462
mask_rcnn+ResNet101+FPN	192	60	22	0.7619	0.8972	0.8240
hybrid_task_cascade+ResNeXt101+FPN+instaboost	182	34	32	0.8426	0.8505	0.8465

Table 3 *Object level performance based on Bounding Box and Polygon Masks instance segmentation*