



Cluster-Induced Mask Transformers for Effective Opportunistic Gastric Cancer Screening on Non-contrast CT Scans

Mingze Yuan^{1,2,3}, Yingda Xia^{1(✉)}, Xin Chen^{4(✉)}, Jiawen Yao^{1,3}, Junli Wang⁵,
Mingyan Qiu^{1,3}, Hexin Dong^{1,2,3}, Jingren Zhou¹, Bin Dong^{2,6}, Le Lu¹,
Li Zhang², Zaiyi Liu^{4(✉)}, and Ling Zhang¹

¹ DAMO Academy, Alibaba Group, Hangzhou, China
yingda.xia@alibaba-inc.com

² Peking University, Beijing, China

³ Hupan Lab, 310023 Hangzhou, China

⁴ Guangdong Province People's Hospital, Guangzhou, China
wolfchenxin@163.com, zyliu@163.com

⁵ The First Affiliated Hospital of Zhejiang University, Hangzhou, China

⁶ Peking University Changsha Institute for Computing and Digital Economy,
Changsha, China

Abstract. Gastric cancer is the third leading cause of cancer-related mortality worldwide, but no guideline-recommended screening test exists. Existing methods can be invasive, expensive, and lack sensitivity to identify early-stage gastric cancer. In this study, we explore the feasibility of using a deep learning approach on non-contrast CT scans for gastric cancer detection. We propose a novel cluster-induced Mask Transformer that jointly segments the tumor and classifies abnormality in a multi-task manner. Our model incorporates learnable clusters that encode the texture and shape prototypes of gastric cancer, utilizing self- and cross-attention to interact with convolutional features. In our experiments, the proposed method achieves a sensitivity of 85.0% and specificity of 92.6% for detecting gastric tumors on a hold-out test set consisting of 100 patients with cancer and 148 normal. In comparison, two radiologists have an average sensitivity of 73.5% and specificity of 84.3%. We also obtain a specificity of 97.7% on an external test set with 903 normal cases. Our approach performs comparably to established state-of-the-art gastric cancer screening tools like blood testing and endoscopy, while also being more sensitive in detecting early-stage cancer. This demonstrates the potential of our approach as a novel, non-invasive, low-cost, and accurate method for opportunistic gastric cancer screening.

Keywords: Gastric cancer · Large-scale cancer screening · Mask Transformers · Non-contrast CT

M.Yuan—Work was done during an internship at DAMO Academy, Alibaba Group.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 146–156, 2023.

https://doi.org/10.1007/978-3-031-43904-9_15

1 Introduction

Gastric cancer (GC) is the third leading cause of cancer-related deaths worldwide [19]. The five-year survival rate for GC is approximately 33% [16], which is mainly attributed to patients being diagnosed with advanced-stage disease harboring unresectable tumors. This is often due to the latent and nonspecific signs and symptoms of early-stage GC. However, patients with early-stage disease have a substantially higher five-year survival rate of around 72% [16]. Therefore, early detection of resectable/curable gastric cancers, preferably before the onset of symptoms, presents a promising strategy to reduce associated mortality. Unfortunately, current guidelines do not recommend any screening tests for GC [22]. While several screening tools have been developed, such as Barium-meal gastric photofluorography [5], upper endoscopy [4, 7, 9], and serum pepsinogen levels [15], they are challenging to apply to the general population due to their invasiveness, moderate sensitivity/specificity, high cost, or side effects. Therefore, there is an urgent need for novel screening methods that are noninvasive, highly accurate, low-cost, and ready to distribute.

Non-contrast CT is a commonly used imaging protocol for various clinical purposes. It is a non-invasive, relatively low-cost, and safe procedure that exposes patients to less radiation dose and does not require the use of contrast injection that may cause serious side effects (compared to multi-phase contrast-enhanced CT). With recent advances in AI, opportunistic screening of diseases using non-contrast CT during routine clinical care performed for other clinical indications, such as lung and colorectal cancer screening, presents an attractive approach to early detect treatable and preventable diseases [17]. However, whether early detection of gastric cancer using non-contrast CT scans is possible remains unknown. This is because early-stage gastric tumors may only invade the mucosal and muscularis layers, which are difficult to identify without the help of stomach preparation and contrast injection. Additionally, the poor contrast between the tumor and normal stomach wall/tissues on non-contrast CT scans and various shape alterations of gastric cancer, further exacerbates this challenge.

In this paper, we propose a novel approach for detecting gastric cancer on non-contrast CT scans. Unlike the conventional “segmentation for classification” methods that directly employ segmentation networks, we developed a cluster-induced Mask Transformer that performs segmentation and global classification simultaneously. Given the high variability in shape and texture of gastric cancer, we encode these features into learnable clusters and utilize cluster analysis during inference. By incorporating self-attention layers for global context modeling, our model can leverage both local and global cues for accurate detection. In our experiments, the proposed approach outperforms nnUNet [8] by 0.032 in AUC, 5.0% in sensitivity, and 4.1% in specificity. These results demonstrate the potential of our approach for opportunistic screening of gastric cancer in asymptomatic patients using non-contrast CT scans.

2 Related Work

Automated Cancer Detection. Researchers have explored automated tumor detection techniques on endoscopic [13, 14], pathological images [20], and the prediction of cancer prognosis [12]. Recent developments in deep learning have significantly improved the segmentation of gastric tumors [11], which is critical for their detection. However, our framework is specifically designed for non-contrast CT scans, which is beneficial for asymptomatic patients. While previous studies have successfully detected pancreatic [25] and esophageal [26] cancers on non-contrast CT, identifying gastric cancer presents a unique challenge due to its subtle texture changes, various shape alterations, and complex background, *e.g.*, irregular gastric wall; liquid and contents in the stomach.

Mask Transformers. Recent studies have used Transformers for natural and medical image segmentation [21]. Mask Transformers [3, 24, 29] further enhance CNN-based backbones by incorporating stand-alone Transformer blocks, treating object queries in DETR [1] as memory-encoded queries for segmentation. CMT-Deeplab [27] and KMaX-Deeplab [28] have recently proposed interpreting the queries as clustering centers and adding regulatory constraints for learning the cluster representations of the queries. Mask Transformers are locally sensitive to image textures for precise segmentation and globally aware of organ-tumor morphology for recognition. Their cluster representations demonstrate a remarkable balance of intra-cluster similarity and inter-class discrepancy. Therefore, Mask Transformers are an ideal choice for an end-to-end joint segmentation and classification system for detecting gastric cancer.

3 Methods

Problem Formulation. Given a non-contrast CT scan, cancer screening is a binary classification with two classes as $\mathcal{L} = \{0, 1\}$, where 0 stands for “normal” and 1 for “GC” (gastric cancer). The entire dataset is denoted by $\mathcal{S} = \{(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{P}_i) | i = 1, 2, \dots, N\}$, where \mathbf{X}_i is the i -th non-contrast CT volume, with \mathbf{Y}_i being the voxel-wise label map of the same size as \mathbf{X}_i and K channels. Here, $K = 3$ represents the background, stomach, and GC tumor. $\mathbf{P}_i \in \mathcal{L}$ is the class label of the image, confirmed by pathology, radiology, or clinical records. In the testing phase, only \mathbf{X}_i is given, and our goal is to predict a class label for \mathbf{X}_i .

Knowledge Transfer from Contrast-Enhanced to Non-contrast CT. To address difficulties with tumor annotation on non-contrast CTs, the radiologists start by annotating a voxel-wise tumor mask on the contrast-enhanced CT, referring to clinical and endoscopy reports as needed. DEEDs [6] registration is then performed to align the contrast-enhanced CT with the non-contrast CT and the resulting deformation field is applied to the annotated mask. Any misaligned ones are revised manually. In this manner (Fig. 1d), a relatively coarse yet highly reliable tumor mask can be obtained for the non-contrast CT image.

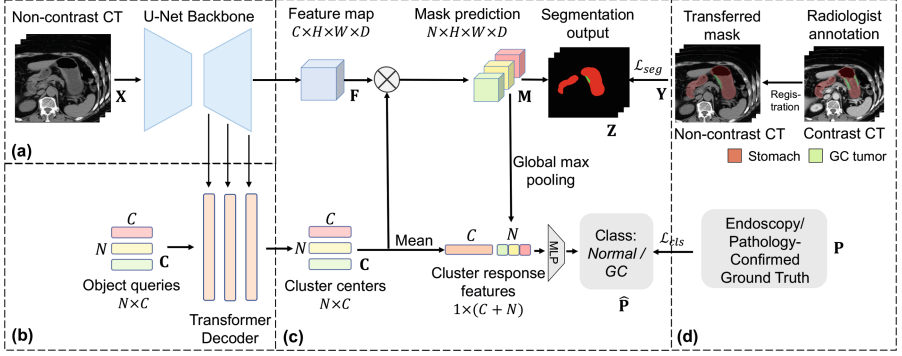


Fig. 1. Method overview. (a) The non-contrast CT image is first forwarded into a U-Net [8, 18] to extract a feature map. (b) Learnable object queries interact with the multi-level U-Net features through a Transformer Decoder and produce learned cluster centers. (c) All the pixels are assigned to cluster centers by matrix multiplication. The cluster assignment (*a.k.a.* mask prediction) is further used to generate the final segmentation output and the classification probability. (d) The entire network is supervised by transferred masks from radiologists’ annotation on contrast-enhanced CT and endoscopy or pathology-confirmed ground truth.

Cluster-Induced Classification with Mask Transformers. Segmentation for classification is widely used in tumor detection [25, 26, 32]. We first train a UNet [8, 18] to segment the stomach and tumor regions using the masks from the previous step. This UNet considers local information and can only extract stomach ROIs well during testing. However, local textures are inadequate for accurate gastric tumor detection on non-contrast CTs, so we need a network of both local sensitivity to textures and global awareness of the organ-tumor morphology. Mask transformer [3, 24] is a well-suited approach to boost the CNN backbone with stand-alone transformer blocks. Recent studies [27, 28] suggest interpreting object queries as cluster centers, which naturally exhibit intra-cluster similarity and inter-class discrepancy. Inspired by this, we further develop a deep classification model on top of learnable cluster representations.

Specifically, given image $\mathbf{X} \in \mathbb{R}^{H \times W \times D}$, annotation $\mathbf{Y} \in \mathbb{R}^{K \times H \times W \times D}$, and patient class $\mathbf{P} \in \mathcal{L}$, our model consists of three components: 1) a CNN backbone to extract its pixel-wise features $\mathbf{F} \in \mathbb{R}^{C \times H \times W \times D}$ (Fig. 1a), 2) a transformer module (Fig. 1b), and 3) a multi-task cluster inference module (Fig. 1c). The transformer module gradually updates a set of randomly initialized object queries $\mathbf{C} \in \mathbb{R}^{N \times C}$, *i.e.*, to meaningful mask embedding vectors through cross-attention between object queries and multi-scale pixel features,

$$\mathbf{C} \leftarrow \mathbf{C} + \arg \max_N (\mathbf{Q}^c (\mathbf{K}^p)^T) \mathbf{V}^p, \quad (1)$$

where c and p stand for query and pixel features, $\mathbf{Q}^c, \mathbf{K}^p, \mathbf{V}^p$ represent linearly projected query, key, and value. We adopt cluster-wise argmax from KMax-DeepLab [28] to substitute spatial-wise softmax in the original settings.

We further interpret the object queries as cluster centers from a cluster analysis perspective. All the pixels in the convolutional feature map are assigned to different clusters based on these centers. The assignment of clusters (*a.k.a.* mask prediction) $\mathbf{M} \in \mathbb{R}^{N \times HWD}$ is computed as the cluster-wise softmax function over the matrix product between the cluster centers \mathbf{C} and pixel-wise feature matrix \mathbf{F} , *i.e.*,

$$\mathbf{M} = \text{Softmax}_N(\mathbf{R}) = \text{Softmax}_N(\mathbf{C}\mathbf{F}). \quad (2)$$

The final segmentation logits $\mathbf{Z} \in \mathbb{R}^{K \times HWD}$ are obtained by aggregating the pixels within each cluster according to cluster-wise classification, which treats pixels within a cluster as a whole. The aggregation of pixels is achieved by $\mathbf{Z} = \mathbf{C}_K \mathbf{M}$, where the cluster-wise classification \mathbf{C}_K is represented by an MLP that projects the cluster centers \mathbf{C} to K channels (the number of segmentation classes).

The learned cluster centers possess high-level semantics with both inter-cluster discrepancy and intra-cluster similarity for effective classification. Rather than directly classifying the final feature map, we first generate the cluster-path feature vector by taking the channel-wise average of cluster centers $\bar{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_i \in \mathbb{R}^C$. Additionally, to enhance the consistency between the segmentation and classification outputs, we apply global max pooling to cluster assignments \mathbf{R} to obtain the pixel-path feature vector $\bar{\mathbf{R}} \in \mathbb{R}^N$. This establishes a direct connection between classification features and segmentation predictions. Finally, we concatenate these two feature vectors to obtain the final feature and project it onto the classification prediction $\hat{\mathbf{P}} \in \mathbb{R}^2$ via a two-layer MLP.

The overall training objective is formulated as,

$$\mathcal{L} = \mathcal{L}_{seg}(\mathbf{Z}, \mathbf{Y}) + \mathcal{L}_{cls}(\hat{\mathbf{P}}, \mathbf{P}), \quad (3)$$

where the segmentation loss $\mathcal{L}_{seg}(\cdot, \cdot)$ is a combination of Dice and cross entropy losses, and the classification loss $\mathcal{L}_{cls}(\cdot, \cdot)$ is cross entropy loss.

4 Experiments

4.1 Experimental Setup

Dataset and Ground Truth. Our study analyzed a dataset of CT scans collected from Guangdong Province People’s Hospital between years 2018 and 2020, with 2,139 patients consisting of 787 gastric cancer and 1,352 normal cases. We used the latest patients in the second half of 2020 as a hold-out test set, resulting in a training set of 687 gastric cancer and 1,204 normal cases, and a test set of 100 gastric cancer and 148 normal cases. We randomly selected 20% of the training data as an internal validation set. To further evaluate specificity in a larger population, we collected an external test set of 903 normal cases from Shengjing Hospital. Cancer cases were confirmed through endoscopy (and pathology) reports, while normal cases were confirmed by radiology reports and a two-year follow-up. All patients underwent multi-phase CTs with a median spacing of $0.75 \times 0.75 \times 5.0$ mm and an average size of (512, 512, 108) voxel. Tumors

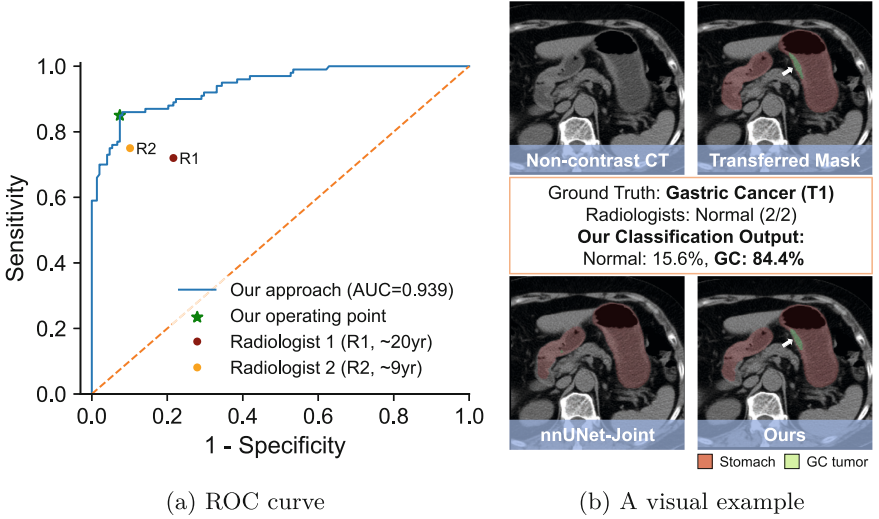


Fig. 2. (a) ROC curve for our model versus two experts on the hold-out test set of $n = 248$ patients for binary classification. (b) A visual example in the test set. This early-stage GC case is miss-detected by both radiologists and nnUNet [8] but our model succeeds to locate the tumor.

were annotated on the venous phase by an experienced radiologist specializing in gastric imaging using CTLabeler [23], while the stomach was automatically annotated using a self-learning model [31].

Implementation Details. We resampled each CT volume to the median spacing while normalizing it to have zero mean and unit variance. During training, we cropped the 3D bounding box of the stomach and added a small margin of (32, 32, 4). We used nnUNet [8] as the backbone, with four transformer decoders, each taking pixel features with output strides of 32, 16, 8, and 4. We set the number of object queries N to 8, with each having a dimension of 128, and included an eight-head self-attention layer in each block. The patch size used during training and inference is (192, 224, 40) voxel. We followed [8] to augment data. We trained the model with RAdam using a learning rate of 10^{-4} and a (backbone) learning rate multiplier of 0.1 for 1000 epochs, with a frozen backbone of the pre-trained nnUNet [8] for the first 50 epochs. To enhance performance, we added deep supervision by aligning the cross-attention map with the final segmentation map, as per KMax-Deeplab [27]. The hidden layer dimension in the two-layer MLP is 128. We also trained a standard UNet [8, 18] to localize the stomach region in the entire image in the testing phase.

Evaluation Metrics and Reader Study. For the binary classification, model performance is evaluated using area under ROC curve (AUC), sensitivity (Sens.), and specificity (Spec.). And successful localization of the tumors is considered when the overlap between the segmentation mask generated by the model and

Table 1. Results on binary classification: gastric cancer vs. normal. The 95% confidence interval of each metric is listed. †: $p < 0.05$ for DeLong test (ours vs. nnUNet-S4C). *: $p < 0.05$ for permutation test (ours vs. nnUNet-S4C and radiologist experts). Sens.: Sensitivity. Spec.: Specificity.

	Internal Hold-out ($n = 248$)			External ($n = 903$)
Method	AUC	Sens.(%)	Spec.(%)	Spec.(%)
Mean of radiologists	-	73.5	84.1	-
nnUNet-S4C [8]	0.907 (0.862, 0.942)	80.0 (72.0, 87.5)	88.5 (83.3, 93.5)	96.6 (95.2, 97.8)
TransUNet-S4C [2]	0.916 (0.876, 0.952)	82.0 (74.7, 89.5)	90.5 (86.1, 94.8)	96.0 (94.8, 97.2)
nnUNet-Joint [8]	0.924 (0.885, 0.959)	81.0 (73.0, 87.9)	90.5 (85.1, 95.0)	97.6 (96.5, 98.6)
Ours	0.939 [†] (0.910, 0.964)	85.0 * (78.1, 91.1)	92.6 * (88.0, 96.5)	97.7 (96.7, 98.7)

Table 2. Patient-level detection and tumor-level localization results (%) over gastric cancer across different T-stages. Tumor-level localization evaluates how segmented masks overlap with the ground-truth cancer (Dice > 0.01 for correct detection). Miss-T: Missing of T stage information.

Method	Criteria	T1	T2	T3	T4	Miss-T
nnUNet-Joint [8]	Patient	30.0(3/10)	66.7(6/9)	94.1(32/34)	100.0(9/9)	86.1(31/36)
	Tumor	20.0(2/10)	55.6(5/9)	94.1(32/34)	100.0(9/9)	80.6(29/36)
Ours	Patient	60.0(6/10)	77.8(7/9)	94.1(32/34)	100.0(9/9)	86.1(31/36)
	Tumor	30.0(3/10)	66.7(6/9)	94.1(32/34)	100.0(9/9)	80.6(30/36)
Radiologist 1	Patient	50.0(5/10)	55.6(5/9)	76.5(26/34)	88.9(8/9)	77.8(28/36)
Radiologist 2	Patient	30.0(3/10)	55.6(5/9)	85.3(29/34)	100.0(9/9)	80.6(29/36)

the ground truth is greater than 0.01, measured by the Dice score. A reader study was conducted with two experienced radiologists, one from Guangdong Province People’s Hospital with 20 years of experience and the other from The First Affiliated Hospital of Zhejiang University with 9 years of experience in gastric imaging. The readers were given 248 non-contrast CT scans from the test set and asked to provide a binary decision for each scan, indicating whether the scan showed gastric cancer. No patient information or records were provided to the readers. Readers were informed that the dataset might contain more tumor cases than the standard prevalence observed in screening, but the proportion of case types was not disclosed. Readers used ITK-SNAP [30] to interpret the CT scans without any time constraints.

Table 3. Comparison with a state-of-the-art blood test on gastric cancer detection [10], UGIS and endoscopy screening performance in large population [4], and early stage gastric cancer detection rate of senior radiologists on narrow-band imaging with magnifying endoscopy (ME-NBI) [7]. (*: We leave out two tumors *in situ* within the test set in accordance with the setting in [10]. †: We also merely consider early-stage gastric cancer cases, including Tumor *in situ*, T1, and T2 stages, among whom we successfully detect 17 of 19 cases.)

Method	Spec.(%)	Sens.(%)	Our sensitivity(%) at the same specificity
Blood Test [10]	99.5	66.7	69.4*
Upper-gastrointestinal series [4]	96.1	36.7	85.0
Endoscopy screening [4]	96.0	69.0	85.0
ME-NBI (early-stage) [7]	74.2	76.7	89.5†

Compared Baselines. Table 1 presents a comparative analysis of our proposed method with three baselines. The first two approaches belong to “Segmentation for classification” (S4C) [26,32], using nnUNet [8] and TransUNet [2]. A case is classified as positive if the segmented tumor volume exceeds a threshold that maximizes the sum of sensitivity and specificity on the validation set. The third baseline (denoted as “nnUNet-Joint”) integrates a CNN classification head into UNet [8] and trained end-to-end. We obtain the 95% confidence interval of AUC, sensitivity, and specificity values from 1000 bootstrap replicas of the test dataset for statistical analysis. For statistical significance, we conduct a DeLong test between two AUCs (ours vs. compared method) and a permutation test between two sensitivities or specificities (ours vs. compared method and radiologists).

4.2 Results

Our method Outperforms Baselines. Our method outperforms three baselines (Table 1) in all metrics, particularly in AUC and sensitivity. The advantage of our approach is that it captures the local and global information simultaneously in virtue of the unique architecture of mask transformer. It also extracts high-level semantics from cluster representations, making it suitable for classification and facilitating a holistic decision-making process. Moreover, our method reaches a considerable specificity of 97.7% on the external test set, which is crucial in opportunistic screening for less false positives and unnecessary human workload.

AI Models Surpass Experienced Radiologists on Non-contrast CT Scans. As shown in Fig. 2a, our AI model’s ROC curve is superior to that of two experienced radiologists. The model achieves a sensitivity of 85.0% in detecting gastric cancer, which significantly exceeds the mean performance of doctors (73.5%) and also surpasses the best performing doctor (R2: 75.0%), while maintaining a high specificity. A visual example is presented in Fig. 2b. This early-stage cancer (T1) is miss-detected by both radiologists, whereas classified and localized precisely by our model.

Subgroup Analysis. In Table 2, we report the performance of patient-level detection and tumor-level localization stratified by tumor (T) stage. We compare our model’s performance with that of both radiologists. The results show that our model performs better in detecting early stage tumors (T1, T2) and provides more precise tumor localization. Specifically, our model detects 60.0% (6/10) T1 cancers, and 77.8% (7/9) T2 cancers, surpassing the best performing expert (50% T1, 55.6% T2). Meanwhile, our model maintains a reliable detection rate and credible localization accuracy for T3 and T4 tumors (2 of 34 T3 tumors missed).

Comparison with Established Screening Tools. Our method surpasses or performs on par with established screening tools [4, 7, 10] in terms of sensitivity for gastric cancer detection at a similar specificity level with a relatively large testing patient size ($n = 1151$ by integrating the internal and external test sets), as shown in Table 3. This finding sheds light on the opportunity to employ automated AI systems to screen gastric cancer using non-contrast CT scans.

5 Conclusion

We propose a novel Cluster-induced Mask Transformer for gastric cancer detection on non-contrast CT scans. Our approach outperforms strong baselines and experienced radiologists. Compared to other screening methods, such as blood tests, endoscopy, upper-gastrointestinal series, and ME-NBI, our approach is non-invasive, cost-effective, safe, and more accurate for detecting early-stage tumors. The robust performance of our approach demonstrates its potential for opportunistic screening of gastric cancer in the general population.

Acknowledgement. This work was supported by Alibaba Group through Alibaba Research Intern Program. Bin Dong and Li Zhang was partly supported by NSFC 12090022 and 11831002, and Clinical Medicine Plus X-Young Scholars Project of Peking University PKU2023LCXQ041.

References

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 213–229. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_13
2. Chen, J., et al.: TransuNet: transformers make strong encoders for medical image segmentation. arXiv preprint [arXiv:2102.04306](https://arxiv.org/abs/2102.04306) (2021)
3. Cheng, B., Schwing, A., Kirillov, A.: Per-pixel classification is not all you need for semantic segmentation. In: NeurIPS, vol. 34, pp. 17864–17875 (2021)
4. Choi, K.S., et al.: Performance of different gastric cancer screening methods in Korea: a population-based study. PLoS One **7**(11), e50041 (2012)
5. Hamashima, C., et al.: The Japanese guidelines for gastric cancer screening. Jpn. J. Clin. Oncol. **38**(4), 259–267 (2008)

6. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans. Med. Imaging* **32**(7), 1239–1248 (2013)
7. Hu, H., et al.: Identifying early gastric cancer under magnifying narrow-band images with deep learning: a multicenter study. *Gastrointest. Endosc.* **93**(6), 1333–1341 (2021)
8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: NNU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)
9. Jun, J.K., et al.: Effectiveness of the Korean national cancer screening program in reducing gastric cancer mortality. *Gastroenterology* **152**(6), 1319–1328 (2017)
10. Klein, E., et al.: Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**(9), 1167–1177 (2021)
11. Li, H., et al.: 3d IFPN: improved feature pyramid network for automatic segmentation of gastric tumor. *Front. Oncol.* **11**, 618496 (2021)
12. Li, J., et al.: CT-based delta radiomics in predicting the prognosis of stage iv gastric cancer to immune checkpoint inhibitors. *Front. Oncol.* **12**, 1059874 (2022)
13. Li, L., et al.: Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* **23**, 126–132 (2020)
14. Luo, H., et al.: Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *The Lancet Oncol.* **20**(12), 1645–1654 (2019)
15. Miki, K.: Gastric cancer screening using the serum pepsinogen test method. *Gastric Cancer* **9**, 245–253 (2006)
16. National Cancer Institute, S.P.: Cancer stat facts: Stomach cancer. <https://seer.cancer.gov/statfacts/html/stomach.html> (2023)
17. Pickhardt, P.J.: Value-added opportunistic CT screening: state of the art. *Radiology* **303**(2), 241–254 (2022)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Smyth, E.C., Nilsson, M., Grabsch, H.I., van Grieken, N.C., Lordick, F.: Gastric cancer. *The Lancet* **396**(10251), 635–648 (2020)
20. Song, Z., et al.: Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nat. Commun.* **11**(1), 4294 (2020)
21. Tang, Y., et al.: Self-supervised pre-training of Swin transformers for 3d medical image analysis. In: *CVPR*, pp. 20730–20740 (2022)
22. USPSTF: U.S. Preventive Services Task Force, Recommendations. https://www.uspreventiveservicestaskforce.org/uspstf/topic_search_results?topic.status=P (2023)
23. Wang, F., et al.: A cascaded approach for ultraly high performance lesion detection and false positive removal in liver CT scans. *arXiv preprint arXiv:2306.16036* (2023)
24. Wang, H., Zhu, Y., Adam, H., Yuille, A., Chen, L.C.: Max-deeplab: end-to-end panoptic segmentation with mask transformers. In: *CVPR*, pp. 5463–5474 (2021)
25. Xia, Y., et al.: Effective pancreatic cancer screening on non-contrast CT scans via anatomy-aware transformers. In: de Bruijne, M., et al. (eds.) *MICCAI 2021. LNCS*, vol. 12905, pp. 259–269. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_25

26. Yao, J., et al.: Effective opportunistic esophageal cancer screening using noncontrast CT imaging. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *Medical Image Computing and Computer Assisted Intervention. MICCAI 2022*. LNCS, vol. 13433, pp. 344–354. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_33
27. Yu, Q., et al.: CMT-DeepLab: clustering mask transformers for panoptic segmentation. In: *CVPR*, pp. 2560–2570 (2022)
28. Yu, Q., et al.: k-means mask transformer. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) *Computer Vision. ECCV 2022*. LNCS, vol. 13689, pp. 288–307. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19818-2_17
29. Yuan, M., et al.: Devil is in the queries: advancing mask transformers for real-world medical image segmentation and out-of-distribution localization. In: *CVPR*, pp. 23879–23889 (2023)
30. Yushkevich, P.A., et al.: User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**(3), 1116–1128 (2006)
31. Zhang, L., Gopalakrishnan, V., Lu, L., Summers, R.M., Moss, J., Yao, J.: Self-learning to detect and segment cysts in lung CT images without manual annotation. In: *ISBI*, pp. 1100–1103 (2018)
32. Zhu, Z., Xia, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11769, pp. 3–12. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_1