Liver Tumor Screening and Diagnosis in CT with Pixel-Lesion-Patient Network

Ke Yan[1,2](✉), Xiaoli Yin[3], Yingda Xia[1], Fakai Wang[1], Shu Wang[3], Yuan Gao[1,2], Jiawen Yao[1,2], Chunli Li[1,2,3], Xiaoyu Bai[1,2], Jingren Zhou[1,2], Ling Zhang[1], Le Lu[1], and Yu Shi[3]

[1] DAMO Academy, Alibaba Group, Hangzhou, China
yanke.yan@alibaba-inc.com
[2] Hupan Lab, 310023 Hangzhou, China
[3] Department of Radiology, Shengjing Hospital of China Medical University, Shenyang 110004, China

Abstract. Liver tumor segmentation and classification are important tasks in computer aided diagnosis. We aim to address three problems: liver tumor screening and preliminary diagnosis in non-contrast computed tomography (CT), and differential diagnosis in dynamic contrast-enhanced CT. A novel framework named Pixel-Lesion-pAtient Network (PLAN) is proposed. It uses a mask transformer to jointly segment and classify each lesion with improved anchor queries and a foreground-enhanced sampling loss. It also has an image-wise classifier to effectively aggregate global information and predict patient-level diagnosis. A large-scale multi-phase dataset is collected containing 939 tumor patients and 810 normal subjects. 4010 tumor instances of eight types are extensively annotated. On the non-contrast tumor screening task, PLAN achieves 95% and 96% in patient-level sensitivity and specificity. On contrast-enhanced CT, our lesion-level detection precision, recall, and classification accuracy are 92%, 89%, and 86%, outperforming widely used CNN and transformers for lesion segmentation. We also conduct a reader study on a holdout set of 250 cases. PLAN is on par with a senior human radiologist, showing the clinical significance of our results.

Keywords: Liver tumor · Lesion segmentation and classification · CT

1 Introduction

Liver cancer is the third leading cause of cancer death world-wide in 2020 [14]. Early detection and accurate diagnosis of liver tumors may improve overall

Partially supported by the National Natural Science Foundation of China (grant 82071885), Basic Research Projects of Liaoning Provincial Department of Education (LJKMZ20221160), the National Youth Talent Support Program of China, and Science and Technology Innovation Talent Project in Shenyang (RC210265).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_8.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 72–82, 2023.
https://doi.org/10.1007/978-3-031-43904-9_8

patient outcomes, in which imaging plays a key role [11]. Computed tomography (CT) is one of the most important imaging modalities for liver tumors. Dynamic contrast-enhanced (DCE) CT is widely used for diagnostics, but it requires iodine contrast injection which can cause reaction and potential risks in patients. Recently, non-contrast (NC) CT scans are gaining attention as they are cheaper and safer to acquire, thus can be potential tools for opportunistic tumor screening [18,20]. Meanwhile, finding and diagnosing tumors in NC CTs is also extremely challenging because of the poor contrast between tumors and normal tissues compared to those in DCE CTs. Prior works on pancreas [18] and esophagus [20] have shown that latest deep learning techniques can detect subtle texture and shape changes in NC CT that even human eyes may miss. Thus, we aim to investigate the performance of liver tumor segmentation and classification in NC CTs. Such an approach will be helpful to discover asymptomatic incidental tumors [12] from routine NC CT scans indicated for general diagnostic purposes at no additional cost and radiation exposure. After an incidental tumor is found, the patient may undergo further imaging examination such as a multi-phase DCE CT for differential diagnosis [11], which can provide useful discriminative information such as the vascularity of lesions and the pattern of contrast agent enhancement [19]. Liver is largest solid organ in body and is the site of many tumor types [11]. Therefore, accurate tumor type classification is important for the decision of treatment plans and prognosis.

Many researchers have developed algorithms to automatically segment [1,9,13,15,23] or classify [19,21,25] liver tumors in CT to help radiologists improve their accuracy and efficiency. For example, public datasets such as the Liver Tumor Segmentation Benchmark (LiTS) [1] fostered a series of works aiming to segment liver tumors with improved convolutional neural network (CNN) backbones [9,13] and lesion edge information [15]. LiTS only has single-phase CTs (venous phase). Several studies investigated methods to exploit multi-phase CT by methods such as hetero-phase fusion [5] and modality-aware mutual learning [23]. There are few work discussing liver tumor analysis in NC CT [5]. Besides lesion segmentation, CNN-based lesion classification algorithms have been studied to distinguish common lesion types [19,21,25].

In this paper, we build a comprehensive framework to address both tumor screening and diagnosis. (1) Tumor screening involves finding tumor patients in a large pool of healthy subjects and patients. Most existing works in tumor segmentation and detection did not explicitly consider it since their training and testing images are all tumor patients. Such models may generate false positives in real-world screening scenario when facing diverse tumor-free images. We collect a large-scale dataset with both tumor and non-tumor subjects, where the non-tumor subjects includes not only healthy ones, but also patients with various diffuse liver diseases such as steatosis and hepatitis to improve the robustness of the algorithm. (2) Most works studied liver tumor segmentation alone without differentiating tumor types, while a few works classify liver tumors on cropped tumor patches [19,21,25]. Meanwhile, we learn tumor segmentation and classification with one network using an instance segmentation framework [3]. We train two networks for NC and multi-phase DCE CTs, respectively. (3) For evaluation, previous segmentation works typically use *pixel-level* metrics such as Dice coefficient. Such metrics cannot reflect the *lesion-level* accuracy (how many lesion instances are correctly detected and classified) and may bias to large lesions when a patient has multiple tumors. *Patient-level* metrics (e.g. classifying whether a subject has malignant tumors) are also useful for treatment recommendation in clinical practice [18,20]. Therefore, we assess our algorithm thoroughly with pixel, lesion, and patient-level metrics.

Algorithms for liver tumor segmentation have focused on improving the feature extraction backbone of a fully-convolutional CNN [9,13,15,23]. The pixel-wise segmentation architectures may not be optimal for lesion and patient-level evaluation metrics since they cannot consider a lesion or an image holistically. Recently, a series of mask transformer algorithms [3,4,17] have emerged in the computer vision community and achieved the state-of-the-art performance in instance segmentation tasks. In brief, they use object queries to interact with image feature maps and with each other to produce mask and class predictions for each instance. Inspired by them, we propose a novel end-to-end framework named Pixel-Lesion-pAtient Network (PLAN) for lesion segmentation and classification, as well as patient classification. It contains three branches with bottom-up cooperation: The segmentation map from the *pixel branch* helps to initialize the *lesion branch*, which is an improved mask transformer aiming to segment and classify each lesion; The *patient branch* aggregates information from the whole image and predicts image-level labels of each lesion type, with regularization terms to encourage consistency with the lesion branch.

We collected a large-scale multi-phase dataset containing 810 non-tumor subjects and 939 tumor patients. 4010 tumor instances of eight types are extensively annotated based on pathological reports. On the non-contrast tumor screening and diagnosis task, PLAN achieves 95.0%, 96.4%, and 0.965 in patient-level sensitivity, specificity, and average AUC for malignant and benign patients, in contrast to 94.4%, 93.7%, and 0.889 for the widely-used nnU-Net [8]. On multi-phase DCE CT, our lesion-level detection precision, recall, and classification accuracy are 92.2%, 89.0%, 85.9%, outperforming nnU-Net [8] and Mask2Former [3]. We further conduct a reader study on a holdout set of 250 cases. Our algorithm is on par with a senior radiologist (16 yrs experience), showing the clinical significance of our results. Our codes will be made public upon institutional approval.

2 Method

2.1 Preliminary on Mask Transformer

Mask transformers are a series of latest works achieving superior accuracy on various segmentation tasks [3,4,17,22]. Different from traditional fully-convolutional segmentators [8] that predict a class label for each pixel, mask transformers predict a class label and a binary mask for each object. Take Mask2Former [3] as an example. It includes a pixel encoder and a pixel decoder that extract a high-resolution pixel embedding tensor $\mathbf{P} \in \mathbb{R}^{M \times D \times H \times W}$ from

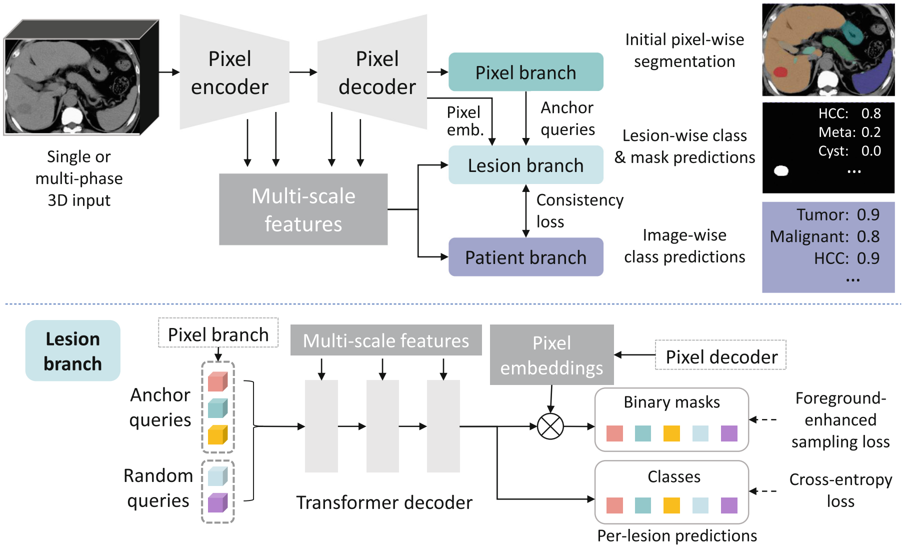

Fig. 1. Framework of the Pixel-Lesion-pAtient Network (PLAN).

the image, where M is the embedding dimension, $D \times H \times W$ is the shape of the 3D image. A group of Q learnable feature vectors $\{\mathbf{q}_i \in \mathbb{R}^M\}_{i=1}^Q$ are randomly initialized as object queries. They are processed by a transformer decoder to interact with multi-scale image features and each other using cross and self-attention operations. After processing, each query is supposed to contain information of one object, which can be used to predict the class probability $\mathbf{c} \in \mathbb{R}^{C+1}$ of the object. Here C is the number of object classes, and we add 1 to indicate an additional "no-object" class if the query does not match with any object. In training, Mask2Former uses bipartite matching [2] to assign each query to a ground-truth object (or "no-object"). Multiplying $\mathbf{q}_i$ with $\mathbf{P}$ gives the binary mask $\mathbf{m}_i \in \mathbb{R}^{D\times H\times W}$ of object i. During inference, the class and mask predictions of all queries can be merged by matrix multiplication to obtain the final semantic segmentation result $\hat{\mathbf{Y}} \in \mathbb{R}^{C\times D\times H\times W}$. We refer readers to [3] for more details.

Mask transformers have various advantages when applied to our task. They can classify a lesion as a whole instead of classifying each pixel, thus can view each lesion holistically. Cross-attention is used to aggregate global features for each lesion. Inter-lesion relation can also be exploited by self-attention operations. In liver CT, inter-lesion relation is diagnostically useful, e.g., metastases and cysts are often multiple. Therefore, We pioneer mask transformers' adaptation for lesion segmentation and classification in 3D medical images. Given a ground-truth or a predicted lesion mask image, we perform connected component (CC) analysis and treat each CC as a lesion instance for training and evaluation.

2.2 Pixel-Lesion-Patient Network (PLAN)

Our goal is to segment the mask and classify the type of each tumor in a liver CT. We also hope to make patient-level diagnoses for each CT scan. PLAN is inspired by Mask2Former [3] with three key improvements: (1) A pixel branch is added to provide anchor queries to the lesion branch. (2) The lesion branch is composed of the transformer decoder in Mask2Former, and we improve its segmentation loss to enhance recall of small lesions. (3) A patient branch is attached to make dedicated image-level predictions with a proposed lesion-patient consistency loss. Our framework is shown in Fig. 1.

Pixel Branch and Anchor Queries. The pixel branch is a convolutional layer after the pixel decoder and learns to predict pixel-wise segmentation maps similar to traditional segmentators. We do CC analysis to the predicted mask to extract lesion instances, and then average the pixel embeddings inside each predicted lesion to obtain a feature vector. The feature vectors are regarded as anchor queries and work the same way as the randomly initialized queries in the lesion branch. Compared to the random queries in the original Mask2Former, the anchor queries contain prior information of the lesions to be segmented, helping the lesion branch to match with the lesion targets more easily [10].

Lesion Branch and Foreground-Enhanced Sampling Loss. Similar to Mask2Former, the lesion branch predicts a binary mask and a class label for each query, see Fig. 1. Mask2Former calculates its segmentation loss on K sampled pixels instead of on the whole image, which is shown to both improve accuracy and reduce GPU memory usage [3]. However, in lesion segmentation, some tumors are very small compared to the whole 3D image. The importance sampling strategy [3] can hardly select any foreground pixels in such cases, so the loss only contains background pixels, degrading the segmentation recall of small lesions. We propose a simple approach to remedy this issue by sampling an extra n foreground pixels for each lesion.

Patient Branch. A patient-level diagnosis is useful for triage. For example, diagnosing the subject as normal, benign, or malignant will result in completely different treatments [24]. Intuitively, we can also infer patient-level labels from segmentation results by checking if there is any lesion in the predicted mask. However, certain tumors are often related to signs outside the tumor, e.g. hepatocellular carcinoma and cirrhosis, cholangiocarcinoma and bile duct dilatation, etc. We equip PLAN with a dedicated patient branch to aggregate such global information to make better patient-level prediction. Since one patient can have multiple liver tumors of different types, in our problem, we give each image several hierarchical binary labels. The first label classifies normal and tumor subjects (whether the image contains any tumor); The second and third labels indicate the existence of respectively benign and malignant tumors; The rest C labels suggest the existence of C fine-grained types of tumors. We employ the dual-path transformer block [17] to fuse multi-scale features from the pixel encoder and decoder to generate a feature map, followed by global average pooling and a linear classification layer to predict the $C + 3$ labels.

A **lesion-patient consistency loss** is further proposed to encourage coherence of the lesion and patient-level predictions. Inspired by multi-instance learning [6], we compute a pseudo patient-level prediction $\tilde{\mathbf{c}} \in \mathbb{R}^C$ from the lesion-level predictions by max-pooling the class probability of each class across all lesion queries (discarding the no-object class). We also have the probability vector from the patient branch $\tilde{\mathbf{p}} \in \mathbb{R}^C$ corresponding to the C fine-grained classes. Then, we compute the L2 loss between them: $\mathcal{L}_{\text{consist}} = \|\tilde{\mathbf{p}} - \tilde{\mathbf{c}}\|^2$.

The overall loss of PLAN is listed in Eq. 1, where $\mathcal{L}_{\text{pixel}}$ is the combined cross-entropy (CE) and Dice loss for the pixel branch as in nnU-Net [8]; $\mathcal{L}_{\text{lesion-class}}$ is the CE loss [3] for lesion classification in the lesion branch; $\mathcal{L}_{\text{lesion-mask}}$ is the combined CE and Dice loss [3] for binary lesion segmentation in the lesion branch with the foreground-enhanced sampling strategy; $\mathcal{L}_{\text{patient}}$ is the binary CE loss for the multi-label classification task in the patient branch.

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{pixel}} + \lambda_{2\text{c}} \mathcal{L}_{\text{lesion-class}} + \lambda_{2\text{m}} \mathcal{L}_{\text{lesion-mask}} + \lambda_3 \mathcal{L}_{\text{patient}} + \lambda_4 \mathcal{L}_{\text{consist}}. \quad (1)$$

3 Experiments

Data. Our dataset contains 810 normal subjects and 939 patients with liver tumors. Each normal subject has a non-contrast (NC) CT, while each patient has a dynamic contrast-enhanced (DCE) CT scan with NC, arterial, and venous phases. We use DEEDS [7] to register NC and arterial phases to the venous phase, and then invite a senior radiologist with 10 years of experience to annotate on the multi-phase CTs using CT Labeler [16]. The 3D mask and the type of all liver tumors are annotated based on pathological reports and magnetic resonance scans if necessary. Eight tumor types are considered in our study: hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), metastasis (meta), hepatoblastoma (hepato), hemangioma (heman), focal nodular hyperplasia (FNH), cyst, and others (all other tumor types). If a lesion's type cannot be determined according to image signs [11] and pathology, it will be marked as "unknown" and ignored in training and evaluation. In total, 4010 tumor instances are annotated, whose volumes range from 11 to 3.7×10^6 mm^3. Detailed statistics and examples of the lesions are shown in the supplementary material. We train two separate networks for NC and DCE CTs. In the former setting, both normal and patient data are used and randomly split into 1149 training, 100 validation, and 500 testing. In the latter one, only patient data are used with 641 training, 100 validation, and 200 testing. Another hold-out set of 150 patients and 100 normal CTs are used for reader study to compare our accuracy with two radiologists.

Implementation Details. Each CT is resampled to $0.7 \times 0.7 \times 5$mm in spacing. We first train an nnU-Net on public datasets to segment liver and surrounding organs (gallbladder, hepatic vein, spleen, stomach, and pancreas), and then crop the liver region to train PLAN. To help PLAN differentiate liver tumors and other organs, we train the network to segment both tumors and organs

Table 1. Patient-level performance on the test set of 500 cases. Spec. 1: specificity on the 202 completely normal cases; Spec. 2: specificity on the 100 hard non-tumor cases.

	NC tumor screening (%)			NC diagnosis AUC		DCE diagnosis AUC
	Sens.	Spec. 1	Spec. 2	Malignant	Benign	8-class Average
nnU-Net [8]	94.4	95.1	91.0	0.948	0.829	0.863
Mask2Former [3]	93.9	97.0	**94.0**	0.924	0.828	0.873
PLAN (ours)	**95.0**	**97.5**	**94.0**	**0.961**	**0.968**	**0.898**

using the predicted organ labels. PLAN is built on top of the nnU-Net framework [8]. Its pixel encoder is a U-Net encoder, whereas its pixel decoder is a light-weight feature pyramid network [3]. The lesion branch incorporates three transformer decoder blocks with masked attention [3] which use feature maps of strides 16, 8, 4 from the pixel decoder. The number of random queries is $Q = 20$; the embedding dimension is $M = 64$; the number of sampled pixels is $K = 12544$ [3], foreground pixels $n = 3$; the loss weight is 0.1 for the no-object class while 1 for other classes in the lesion branch [3]. The weights in Eq. 1 are $\lambda_1 = \lambda_{2\mathrm{c}} = 2, \lambda_{2\mathrm{m}} = 5, \lambda_3 = 1, \lambda_4 = 0.1$. We use the RAdam optimizer with an initial learning rate of 0.0001. Each training batch contains two patches of size $256 \times 256 \times 24$. For DCE CT, the three phases form a 3-channel image as the network input. Extensive data augmentation is applied including random cropping, scaling, flipping, elastic deformation, and brightness adjustment [8]. During training, we first pretrain the backbone and the pixel branch for 500 epochs, and then train the whole network for another 500 epochs.

Patient-Level Results. This paper has three major goals: tumor screening in NC CT (classifying a subject as normal or tumor), preliminary diagnosis in NC CT (predicting the existence of malignant and benign tumors), and fine-grained diagnosis in DCE CT (predicting the existence of 8 tumor types). Among the 8 tumor types, HCC, ICC, meta, and hepato are malignant; heman, FNH, and cyst are benign. "Others" can be either malignant or benign, thus are excluded in the preliminary diagnosis task. The NC test set contains 198 tumor cases, 202 completely normal cases, and 100 "hard" non-tumor cases which may have larger image noise, artifact, ascites, diffuse liver diseases such as hepatitis and steatosis. These cases are used to test the robustness of the model in real-world screening scenario with diverse tumor-free images. We compare PLAN with a widely-used strong baseline, nnU-Net [8]. The recent mask transformer, Mask2Former [3], is also adapted to 3D for comparison. For the baselines, patient-level labels are inferred from their predicted masks by counting lesion pixels. As displayed in Table 1, PLAN achieves the best accuracy on all tasks, especially in NC preliminary diagnosis tasks, which demonstrates the effectiveness of its dedicated patient branch that can explicitly aggregate features from the whole image.

Lesion and Pixel-Level Results. In lesion-level evaluation, we treat a prediction as a true positive if its overlap with a ground-truth lesion is >0.2 in Dice.

Table 2. Lesion-level performance (precision, recall, recall of lesions with different radius, classification accuracy of 8 tumor types), and pixel-level performance (Dice per case). Precision, recall, and Dice are computed without considering the tumor types.

		Prec.	Recall	R<5 mm	5∼10	10∼20	>20 mm	Acc.	Dice
NC	nnU-Net	78.8	77.3	19.7	63.6	90.1	96.5	75.7	**78.3**
	Mask2Former	**85.7**	74.0	10.0	60.5	**91.9**	97.4	77.9	76.4
	PLAN	80.1	**81.9**	**21.9**	**64.6**	90.1	**98.3**	**78.5**	77.2
DCE	nnU-Net	88.1	88.3	22.5	**76.4**	93.7	**98.3**	83.1	**84.2**
	Mask2Former	90.3	83.5	11.7	74.4	**94.6**	97.4	84.8	82.9
	PLAN	**92.2**	**89.0**	**25.6**	74.9	**94.6**	**98.3**	**85.9**	**84.2**

Table 3. Reader study results on 150 tumor cases and 100 normal cases. 3-class acc. means classification accuracy of normal vs. benign vs. malignant.

	NC			DCE
	Sens.	Spec.	3-class Acc.	8-class Acc.
Radiologist 1	94.1	**99.0**	90.8	**75.6**
Radiologist 2	85.5	**99.0**	72.0	40.5
PLAN	**96.7**	98.0	**91.3**	**75.6**

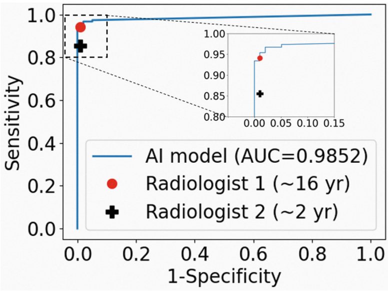

Fig. 2. ROC curve of our method versus 2 radiologists' performance.

Lesions smaller than 3 mm in radius are ignored. As shown in Table 2, the pixel-level accuracy of nnU-Net and PLAN are comparable, but PLAN's lesion-level accuracy is consistently higher than nnU-Net. In this work, we focus more on patient and lesion-level metrics. Although NC images have low contrast, they can still be used to segment and classify lesions with $\sim 80\%$ precision, recall, and classification accuracy. It implies the potential of NC CT, which has been understudied in previous works. Mask2Former has higher precision but lower recall in NC CT, especially for small lesions, while PLAN achieves the best recall using the foreground-enhanced sampling loss. Both PLAN and Mask2Former achieve better classification accuracy, which illustrates the mask transformer architecture is good at lesion-level classification.

Comparison with Radiologists. In the reader study, we invited a senior radiologist with 16 years of experience in liver imaging, and a junior radiologist with 2 years of experience. They first read the NC CT of all subjects and provided a diagnosis of normal, benign, or malignant. Then, they read the DCE scans and provided a diagnosis of the 8 tumor types. We consider patients with only one tumor type in this study. Their reading process is without time constraint. In Table 3 and Fig. 2, all methods get good specificity probably because the normal subjects are completely healthy. Our model achieves comparable accuracy with the senior radiologist but outperforms the junior one by a large margin in sensitivity and classification accuracy.

An ablation study for our method is shown in Table 4. It can be seen that our proposed anchor queries produced by the pixel branch, FES loss, and lesion-patient consistency loss are useful for the final performance. The efficacy of the lesion and patient branches has been analyzed above based on the lesion and patient-level results. Due to space limit, we will show the accuracy for each tumor type and more qualitative examples in the supplementary material.

Comparison with Literature. In the pixel level, we obtain Dice scores of 77.2% and 84.2% using NC and DCE CTs, respectively. The current state of the art (SOTA) of LiTS [1] achieved 82.2% in Dice using CTs in venous phase; [23] achieved 81.3% in Dice using DCE CT of two phases. In the lesion level, our precision and recall are 80.1% and 81.9% for NC CT, 92.2% and 89.0% for DCE CT, at 20% overlap. [25] achieved 83% and 93% for DCE CT. SOTA of LiTS achieved 49.7% and 46.3% at 50% overlap. [21] classified lesions into 5 classes, achieving 84% accuracy for DCE and 49% for NC CT. We classify lesions into 8 classes with 85.9% accuracy for DCE and 78.5% for NC CT. In the patient level, [5] achieved AUC=0.75 in NC CT tumor screening, while our AUC is 0.985. In summary, our results are superior or comparable to existing works.

Table 4. Ablation study on NC data. FES loss: foreground enhanced sampling loss.

	Tumor screening (%)		Prelim. diagnosis AUC		Lesion and pixel-level (%)			
	Sens.	Spec.	Malignant	Benign	Precision	Recall	Acc.	Dice
PLAN (proposed)	**95.0**	96.4	**96.1**	**96.8**	80.1	**81.9**	**78.5**	**77.2**
w/o anchor queries	94.4	95.4	94.9	93.5	78.9	78.1	77.1	75.0
w/o FES loss	93.4	96.0	94.0	96.4	**86.6**	75.1	77.7	**77.2**
w/o consistency loss	93.9	**96.7**	95.4	96.3	79.1	80.7	78.2	76.6

4 Conclusion

Three tasks are investigated in this paper: liver tumor screening and preliminary diagnosis in NC CT, and the diagnosis of 8 tumor types in DCE CT. The pixel-lesion-patient network is proposed that can accomplish lesion-level segmentation and classification, and patient-level classification. Comprehensive evaluation on a large-scale dataset confirms the effectiveness and clinical significance of our method. It can serve as a powerful tool for automated screening and diagnosis of various liver tumors. Our future work includes further improving the specificity of hard non-tumor cases and sensitivity of small lesions.

References

1. The Liver Tumor Segmentation Benchmark (LiTS). Med. Image Anal. **84** (2023)
2. 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

3. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: CVPR, pp. 1280–1289 (2022)
4. Cheng, B., Schwing, A.G., Kirillov, A.: Per-Pixel classification is not all you need for semantic segmentation. In: NeurIPS, vol. 22, pp. 17864–17875 (2021)
5. Cheng, C.T., Cai, J., Teng, W., Zheng, Y., Huang, Y.T.: A flexible three-dimensional hetero-phase computed tomography hepatocellular carcinoma (HCC) detection algorithm for generalizable and practical HCC screening. Hepatol. Commun. (2022)
6. Cheplygina, V., de Bruijne, M., Pluim, J.P.: Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med. Image Anal. **54**, 280–296 (2019)
7. 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)
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. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans. Med. Imaging **37**(12), 2663–2674 (2018)
10. Liu, S.: DAB-DETR : dynamic anchor boxes are better queries for DETR. In: ICLR, pp. 1–19 (2022)
11. Marrero, J.A., Ahn, J., Rajender Reddy, K.: Americal college of gastroenterology: ACG clinical guideline: the diagnosis and management of focal liver lesions. Am. J. Gastroenterol. **109**(9), 1328–1347 (2014)
12. Semaan, A., et al.: Incidentally detected focal liver lesions-a common clinical management dilemma revisited. Anticancer Res. **36**(6), 2923–2932 (2016)
13. Seo, H., Huang, C., Bassenne, M., Xiao, R., Xing, L.: Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. IEEE Trans. Med. Imaging **39**(5), 1316–1325 (2020)
14. Sung, H., et al.: Global cancer statistics 2020 : GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clin. **71**(3), 209–249 (2021)
15. Tang, Y., Tang, Y., Zhu, Y., Xiao, J., Summers, R.M.: E^2Net: an edge enhanced network for accurate liver and tumor segmentation on CT scans. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 512–522. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_50
16. Wang, F., et al.: A Cascaded Approach for Ultraly High Performance Lesion Detection and False Positive Removal in Liver CT Scans (2023). http://arxiv.org/abs/2306.16036
17. Wang, H., Adam, H., Yuille, A., Chen, L.c.: MaX-DeepLab : end-to-end panoptic segmentation with mask transformers. In: CVPR, pp. 5463–5474 (2021)
18. 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
19. Xu, X., Zhu, Q., Ying, H., Li, J., Cai, X., Li, S.: A knowledge-guided framework for fine-grained classification of liver lesions based on multi-phase CT images. IEEE J. Biomed. Health Inf. **27**(1), 386–396 (2023)

20. 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.) MICCAI 2022. LNCS, vol. 13433, pp. 344–354. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_33
21. Yasaka, K., Akai, H., Abe, O., Kiryu, S.: Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology **286**(3), 887–896 (2018)
22. Yu, Q., et al.: K-means mask transformer. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13689, pp. 288–307. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19818-2_17
23. Zhang, Y., Yang, J., Tian, J., Shi, Z., Zhong, C., He, Z.: Modality-aware mutual learning for multi-modal medical image segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 589–599. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_56
24. Zhao, T., et al.: 3D graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management. In: CVPR, pp. 13738–13747 (2021)
25. Zhou, J., et al.: Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. Front. Oncol. **10**, 1 (2021)