LIDIA: Precise Liver Tumor Diagnosis on Multi-Phase Contrast-Enhanced CT via Iterative Fusion and Asymmetric Contrastive Learning

Wei Huang[1,2,3], Wei Liu[2,3], Xiaoming Zhang[2,3]✉, Xiaoli Yin[4], Xu Han[5], Chunli Li[2,3,4], Yuan Gao[2,3], Yu Shi[4], Le Lu[2], Ling Zhang[2], Lei Zhang[1]✉, and Ke Yan[2,3]

[1] College of Computer Science, Sichuan University, 610065, Chengdu, China.
[2] DAMO Academy, Alibaba Group
[3] Hupan Lab, 310023, Hangzhou, China
[4] Department of Radiology, Shengjing Hospital of China Medical University, 110004, Shenyang, China
[5] Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital of Zhejiang University, 310006, Hangzhou, China
`zxiaoming360@gmail.com`; `leizhang@scu.edu.cn`;

Abstract. The early detection and precise diagnosis of liver tumors are tasks of critical clinical value, yet they pose significant challenges due to the high heterogeneity and variability of liver tumors. In this work, a precise LIver tumor DIAgnosis network on multi-phase contrast-enhanced CT, named LIDIA, is proposed for real-world scenario. To fully utilize all available phases in contrast-enhanced CT, LIDIA first employs the iterative fusion module to aggregate variable numbers of image phases, thereby capturing the features of lesions at different phases for better tumor diagnosis. To effectively mitigate the high heterogeneity problem of liver tumors, LIDIA incorporates asymmetric contrastive learning to enhance the discriminability between different classes. To evaluate our method, we constructed a large-scale dataset comprising 1,921 patients and 8,138 lesions. LIDIA has achieved an average AUC of 93.6% across eight different types of lesions, demonstrating its effectiveness. Besides, LIDIA also demonstrated strong generalizability with an average AUC of 89.3% when tested on an external cohort of 828 patients.

Keywords: Liver tumor · Lesion segmentation · Multi-phase fusion.

1 Introduction

Liver is the largest solid organ in human body and plays a crucial role in various physiological functions. Meanwhile, it can be a common site for many malignant and benign tumors. According to global cancer statistics, liver cancer became

---

✉ Corresponding authors. The work was done during W. Huang's internship at Alibaba DAMO Academy.

the third leading cause of cancer death worldwide in 2020 [1]. This high mortality rate is partially attributed to the late diagnosis of liver cancer, since patients with late-stage liver cancer discovered have limited treatment options and often a poor prognosis as well. Therefore, early detection and accurate diagnosis of liver tumors have become an urgent clinical task. Dynamic contrast-enhanced computed tomography (DCE-CT) is a widely utilized imaging technology for the diagnosis of liver tumors. To obtain DCE-CTs, multiple images are scanned at consecutive time points after intravenous injection of contrast agents. These multi-phase images provide valuable diagnostic information about the characteristics (e.g. vascularity) of lesions via the pattern of contrast agent enhancement [12]. However, these characteristics may be difficult to interpret due to the high degree of diversity and heterogeneity of liver tumors, especially for rare tumor types and atypical imaging signs. Additionally, manual analysis of CT images is time-consuming and often influenced by personal experience and biases of radiologists [20].

Previous studies [2,16,18,20] have demonstrated the capability of deep learning technology in identifying subtle textural details and shape variations of tumors that are imperceptible to human observation. In recent years, considerable efforts have been made on the segmentation, detection, and classification of liver tumors. A large proportion of works focus on the segmentation [1,10,13,15,21,8] or detection [5] of tumors without differentiating their types. These works proposed improved convolutional neural network (CNN) backbones [10], novel losses using lesion edge information [13,15], weakly-supervised teacher-student network [21], or synthetic training data [8]. Some studies investigated methods to classify tumor types with manually drawn region-of-interests (ROIs) [14,19]. Two-stage methods [23,20] first detect tumor ROIs with Faster R-CNN, and then classify each ROI with 3D CNN. Recently, a 3D instance segmentation framework is proposed in [16] to jointly segment and classify liver tumors. Besides, multi-phase image fusion has been studied using early fusion [16], hetero-modal image fusion [5], ConvLSTM [17], spatial and channel attention [20], etc.

Despite progresses have been made, there are still two issues that need to be addressed in real-world scenario. First, the clinical guideline for liver tumor diagnosis [12] recommends that a liver DCE-CT includes arterial, venous, and delayed phases. However, in practice, the delayed phase is not always scanned. This is partially because delayed phase prolongs scan duration, and many liver tumors are found in abdominal DCE-CTs not specially designed for liver, thus do not include delayed phases. Most existing algorithms neglected the delayed phase [20,16,23,14], resulting in the omission of valuable information for diagnosis. Second, most existing algorithms only considered common lesion types. Meanwhile, in practice there are numerous less common tumor types that hold significant clinical importance as well and need to be differentiated from the common ones. These rare lesions may present with imaging features similar to other types of liver tumors, or they might occur infrequently, resulting in a lack of ample data for effective training and recognition.

To address these problems, we propose a precise LIver tumor DIAgnosis network for multi-phase contrast-enhanced CTs, named LIDIA. LIDIA iteratively fuses all available CT phases to comprehensively capture and analyze the characteristics of lesions at different time points. Additionally, we introduce an asymmetric contrastive learning approach to address the heterogeneity of indeterminate categories and rare lesions in real-world scenarios. To verify the effectiveness of LIDIA, we collected a large-scale contrast-enhanced CT dataset with 1921 patients, 2/3 of them with the delayed phase. 8,138 lesions of 8 classes were comprehensively annotated. Besides 7 common tumor classes, there is an "others" class including more than 20 relatively rare tumor types. LIDIA can not only effectively fuse multi-phase CT under incomplete phase conditions, but also accurately differentiate rare lesion types from common types. It achieves a mean classification AUC of 93.6%, outperforming the widely used baseline models. We also test LIDIA on an external cohort of 828 patients and achieve a mean AUC of 89.3%, showing good generalization ability.

2   Method

2.1   Preliminaries

Problem Definition. We define our liver tumor diagnosis task as follows. Let $P = \{\text{NC}, \text{A}, \text{V}, \text{D}\}$ denote images of the non-contrast, arterial, venous, and delayed phases, respectively. Consider a dataset $\mathcal{D}$ composed of N patient cases, with each case containing multi-phase CT images $\boldsymbol{X} = \{\boldsymbol{x}^{\text{NC}}, \boldsymbol{x}^{\text{A}}, \boldsymbol{x}^{\text{V}}, [\boldsymbol{x}^{\text{D}}]\}$. Here, the delayed phase image $\boldsymbol{x}^{\text{D}}$ is optionally included, as denoted by the brackets, to reflect its potential unavailability. Each case is paired with K instance-level lesion masks $\{\boldsymbol{S}_j\}_{j=1}^{K}$ and the corresponding lesion classifications $\{\boldsymbol{C}_j\}_{j=1}^{K}$. Note that K is the number of tumors of each case and may be different among cases. The objective is to develop a model $f : \boldsymbol{X} \to (\{\boldsymbol{S}_j\}_{j=1}^{K}, \{\boldsymbol{C}_j\}_{j=1}^{K})$ that utilizes the multi-phase images $\boldsymbol{X}$ to accurately predict the lesion masks and their associated classes. This model must be designed to effectively utilize the all available phases to accurately identify the lesions and predict the correct class of each lesion, accommodating the scenario in which the delayed phase may not be available for all cases.

Mask Transformer. Recently, mask transformers have been proposed for various segmentation tasks. Rather than performing per-pixel classification as in traditional semantic segmentation methods, they predict a set of binary masks and assigns a class label to each mask, enabling instance segmentation [4,3,22]. An example is Mask2Former [3], which uses a pixel encoder-decoder to generate multi-scale features and employs learnable embeddings as object queries. These queries interact with image features as well as themselves through a transformer decoder to segment objects and also identify their classes. Benefiting from the masked attention mechanism, this approach allows the transformer block to concentrate on specific local regions where tumors are situated, making Mask2Former particularly effective for the segmentation and diagnosis of lesions, which are often small in size.

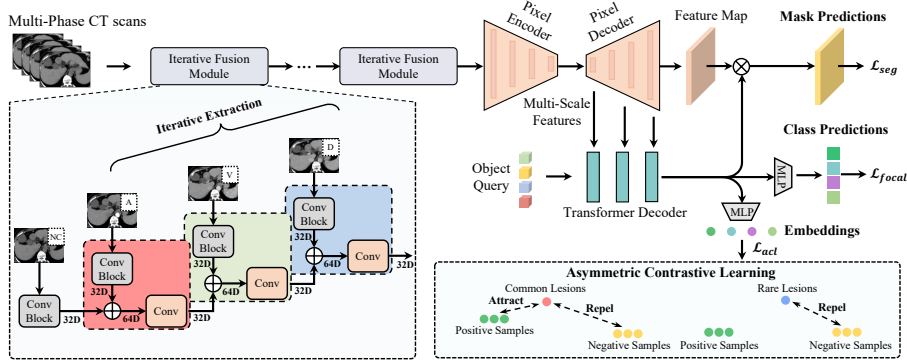## 2.2   Liver tumor diagnosis network (LIDIA)

Fig. 1. Illustration of the overall framework of LIDIA.

We build LIDIA based on Mask2Former and introduce two key enhancements. First, we propose a multi-phase iterative fusion module, designed to handle multi-phase input and potential phase incompleteness. Second, we introduce an asymmetric contrastive learning loss to promote the discrimination between tumor types, especially for rare types. We also make a series of improvements in the training and inference procedure to enhance the robustness and accuracy of LIDIA. Our framework is shown in Fig. 1.

Iterative Fusion Module (IFM). In clinical workflows, physicians often determine the type of lesions based on the differences in features of lesion regions of interest (ROIs) across various phases. Inspired by this practice, to effectively utilize the complementary information between different phases, we first propose an iterative multi-phase fusion module. Formally, the specific information extraction for each phase can be defined as a function mapping from the phase image to feature map:

$$\boldsymbol{h}^p = \mathcal{F}_p(\boldsymbol{x}^p) \in \mathbf{R}^C, \quad p \in P, \tag{1}$$

where $\mathcal{F}_p$ represents the context-specific block for phase p, and $\boldsymbol{h}^p$ is the resultant feature map for that phase with C channels. The feature extraction layer is comprised of two convolutional blocks, each employing 3D convolutions, instance normalization, and LeakyReLU activations to progressively extract information. Then, to effectively aggregate multi-phase information, we perform the fusion of information following the temporal sequence of the phases. Specifically, we fuse them in the order of non-contrast, arterial, venous, and delayed phases. The fusion process iteratively incorporates feature maps from each phase p into a fused feature $\boldsymbol{h}^{fuse}$. The fusion operation is performed by concatenating the current fused $\boldsymbol{h}^{fuse}$ with the feature of the next phase $\boldsymbol{h}^i$ and applying the convolutional

block $\mathcal{F}_{conv}$ to extract new features. This is mathematically represented by the following recursive equation:

$$\boldsymbol{h}_{k+1}^{fuse} = \mathcal{F}_{conv}(\text{concat}(\boldsymbol{h}_k^{fuse}, \boldsymbol{h})) \in \mathbf{R}^C, \quad \boldsymbol{h} \in \{\boldsymbol{h}^p\}_{p \in P}, \quad k \in \{1, 2, 3\}, \quad (2)$$

where $\boldsymbol{h}_1^{fuse} = \boldsymbol{h}^{\text{NC}}$ and $\boldsymbol{h}_k^{fuse}$ represents the fused feature after incorporating the k-th subsequent phase's feature. The final fused feature is $\boldsymbol{h}_4^{fuse}$ if delayed phase is present and $\boldsymbol{h}_3^{fuse}$ if not. This process provides a way to adaptively incorporate variable phase numbers. It is able to capture dynamic contrast changes among phases similar to ConvLSTM [17], while being more lightweight.

Asymmetric Contrastive Learning (ACL). To increase intra-class compactness and inter-class discriminability for liver tumors, we propose an asymmetric contrastive loss for the embedding of lesions, denoted as $\mathcal{L}_{acl}$. The liver contains numerous rare lesion types that are underrepresented (typically less than 10 samples in our dataset). It is impractical to assign each rare type as a separate class because the network could not effectively learn given such few samples, so we collectively assign as them as the "others" class. The ordinary supervised contrastive loss applies a uniform attraction within each class and repulsion between any two distinct classes, treating all classes equally. Due to the inherent heterogeneity of "others" class, clustering them using ordinary supervised contrastive loss would limit the model's flexibility when dealing with unseen lesions or those with similar appearances. On the other hand, we differentiates between common and rare lesions: $\mathcal{L}_{acl}$ only performs attraction within each common class and repulsion among common classes; while for samples within the "others" class, we do not try to cluster them together, but instead only keep them distant from the common classes. The original supervised contrastive loss can be represented by the following equation:

$$\mathcal{L} = \mathbb{E}_{x \in \mathcal{I}}[\mathcal{L}(x)], \quad (3)$$

where

$$\mathcal{L}(x) = \sum_{p \in \mathcal{P}(x)} -\frac{1}{|\mathcal{P}(x)|} \log \frac{\exp(z(x)^T z(p)/\tau)}{\sum\limits_{a \in \mathcal{A}(x)} \exp(z(x)^T z(a)/\tau)}. \quad (4)$$

Here, $z(x)$ is a non-linear projection function. $\mathcal{P}(x)$ is the set of positive samples with the same label as x. $\mathcal{A}(x)$ is the set of all contrastive samples with respect to x, and $\mathcal{I}$ is the set of training samples.

For common lesions, we perform intra-class attraction and inter-class repulsion using Eq. (4). For rare lesions, we simply push them away from the common classes; hence, there is no attraction amongst positive samples. Therefore, the contrastive loss becomes:

$$\mathcal{L}(x_r) = \log \sum_{a \in \mathcal{A}_c(x_r)} \exp(z(x_r)^T z(a)/\tau), \quad (5)$$

where $\mathcal{A}_c(x_r)$ represents the set of common lesion samples with respect to x_r. Specifically, we project the object queries into an additionally embedding space

to conduct above process. Meanwhile, to increase the sample size for contrastive learning, we perform cross-batch contrast via maintaining a memory bank.

Training and Inference. Our task includes binary mask prediction and lesion type classification. For mask prediction, we employ a hybrid loss function, $\mathcal{L}_{seg}$, which combines Cross-Entropy (CE) loss with Dice loss. Similar to [16], we adopt a *foreground-enhanced sampling strategy* when computing $\mathcal{L}_{seg}$, which increases the ratio of foreground pixels in the loss calculation, improving the recall of small lesions.For lesion type classification, due to the class imbalance issue present among liver lesions, we employ *focal loss* $\mathcal{L}_{focal}$ [11] instead of the CE loss in [3] to enhance the learning focus on the underrepresented classes. Therefore, the final loss function is expressed as a combination of above loss, mathematically represented as

$$\mathcal{L}_{final} = \lambda_1 \mathcal{L}_{seg} + \lambda_2 \mathcal{L}_{focal} + \lambda_3 L_{acl}. \tag{6}$$

Liver lesion classification faces the challenge of different tumors can present similar features while small lesions in segmentation tasks are easily missed due to their size. Therefore, when updating the model, we employ the *sharpness aware minimization* strategy [6], which seeks weights that demonstrate low sensitivity to input noises. This leads to more robust predictions, especially for subtle or ambiguous features, thereby improving the model's performance on both classification and segmentation of lesions.

In the inference stage, our goal extends beyond achieving lesion-wise detection and pixel-wise segmentation. We also aim to obtain patient-wise diagnostic results, i.e. the overall probability of the patient having each type of tumor. To achieve this goal, Zhu *et al.* [24] suggested utilizing the size of the predicted lesion mask to infer the patient-level diagnosis. However, this method fails to consider the confidence of the mask prediction and tends to neglect small lesions. Yan *et al.* [16] adopted an additional network branch for patient-level classification, which is prone to overfitting compared to pixel-wise predictions. In this work, we propose a simple yet effective approach called *LiverMax* to obtain patient-wise diagnosis probabilities from pixel-wise segmentation probabilities. We take the softmax output of the semantic segmentation result of LIDIA, compute the maximum value of each channel (tumor type) from all voxels inside the liver, and get the probability of the patient having each type of tumor. This strategy outperforms the previous strategies in our experiments.

3   Experiments

Dataset. We established a CT dataset comprising 1921 patients with 8,138 liver tumors annotated. Each patient underwent dynamic contrast-enhanced CT scans, with all patients having non-contrast (NC), arterial, and venous phases. 2/3 of them have delayed phase images. We registered all phases to the venous phase using DEEDS [7]. Then, we invited a senior radiologist with 10 years of experience to delineate the tumors and annotate the types of all liver tumors based on pathological reports, imaging signs of CT and MRI, and follow-up information. Our study encompasses seven common types of liver lesions: hepatocellular

carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), metastases (meta), hemangioma (heman), focal nodular hyperplasia (FNH), calcification (calc) and cyst. Additionally, we created a separate category for other rare lesions, which includes over 20 uncommon lesion types with each type typically fewer than 10 samples. We split the dataset into three subsets: 1305 samples for training, 298 for validation, and 318 for testing.

Implementation Details. LIDIA is built upon of the nnU-Net [9] framework. The pixel-encoder is based on the encoder from the U-Net architecture, but the first layer have been replaced with IFM. For the decoder, LIDIA employs a Feature Pyramid Network. LIDIA is configured with 50 learnable queries. As for the loss weights, $\lambda_1 = 5$, $\lambda_2 = 5$, and $\lambda_3 = 0.01$. Additionally, the temperature is set as 0.1, and the memory bank size is set as 1024. During training, we use a batch size of 2, a learning rate of 1e-5, and train the model for 1000 epochs.

3.1 Experimental results

Comparisons with other methods. We compare our proposed method with the widely-used robust baseline, nnU-Net [9]. Mask2Former [3] achieved outstanding accuracy in instance segmentation of natural objects, thus we adapted it for 3D data and included it in the comparison. PLAN [16] is a latest instance segmentation framework specially designed for liver tumor diagnosis. For nnU-Net and Mask2Former, patient-level results are inferred by counting the number of lesion pixels in their predicted masks as described in [24]. We report the patient-wise diagnosis (AUC-8: mean AUC of 8 tumor types; AUC-2: mean AUC of malignant and benign classification), lesion-wise detection (precision, sensitivity, and lesion classification accuracy), and pixel-wise lesion segmentation metrics in Table 1.
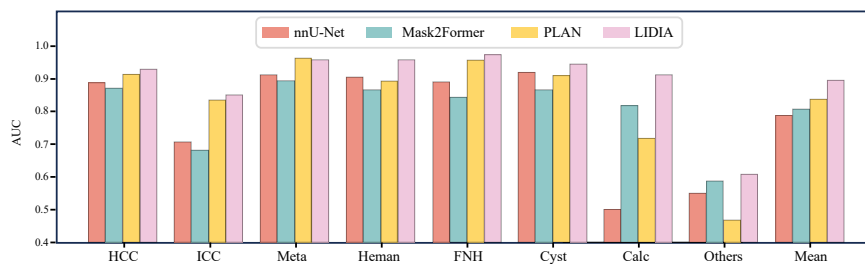

Fig. 2. Illustration of AUC for all classes on an external cohort.

Firstly, LIDIA achieves the highest patient-wise AUC, which are the primary metric of focus in this work that are important for clinical diagnosis. For lesion-wise classification, LIDIA achieves the highest accuracy. The segmentation accuracy of all methods are comparable. To evaluate the generalizability of LIDIA, we test its performance on an independent external cohort. As shown in Fig. 2,

Table 1. Comparisons with several state-of-the-art methods on the internal cohort.

Method	Patient-wise		Lesion-wise			Pixel-wise
	AUC-8	AUC-2	Prec.	Sens.	Acc.	Dice
nn-UNet [9]	0.822	0.915	**0.909**	0.854	0.8575	0.873
Mask2Former [3]	0.841	0.895	0.832	0.861	0.8658	**0.875**
PLAN[16]	0.905	0.878	0.907	**0.886**	0.8699	0.869
LIDIA	**0.936**	**0.946**	0.886	0.866	**0.8812**	0.869

Table 2. Comparison of the performance of various multi-phase fusion methods.

Method	HCC	ICC	Meta	Heman	FNH	cyst	calc	others	Mean
Baseline nnU-Net	0.907	0.728	0.884	**0.973**	0.866	0.955	0.500	0.762	0.822
HEMIS [5]	0.900	0.749	**0.899**	0.945	0.832	0.948	0.500	0.808	0.823
ConvLSTM [17]	0.890	0.715	0.890	0.972	0.900	0.935	0.500	0.781	0.823
IFM	**0.913**	**0.757**	**0.899**	0.961	**0.902**	**0.956**	0.500	**0.815**	**0.838**

Table 3. Ablation study. IFM: iterative fusion module, ACL: asymmetric contrastive learning, SAM: sharpness-aware minimization

Method	HCC	ICC	Meta	Heman	FNH	cyst	calc	others	Mean
Baseline	0.861	0.711	0.857	0.935	0.797	0.880	0.790	0.753	0.823
+Focal Loss	0.880	0.696	0.882	0.909	0.799	0.926	0.897	0.736	0.841
+LiverMax	0.934	**0.853**	0.947	0.974	0.968	0.942	0.912	0.826	0.919
+IFM	0.941	0.846	0.940	0.975	0.953	0.943	0.958	0.846	0.925
+ACL	**0.954**	0.845	0.948	0.973	0.966	0.946	0.937	0.856	0.928
+SAM(LIDIA)	0.951	0.850	**0.951**	**0.979**	**0.970**	**0.948**	**0.963**	**0.874**	**0.936**

our method achieved optimal performance in almost all lesions, indicating good generalization capability. In summary, LIDIA achieves the highest diagnosis accuracy. We will display qualitative examples and lesion-wise confusion matrix in the Appendix.

Effect of iterative fusion module. We used nn-UNet with three input phases (discarding the delayed phase) as the baseline method and compared the performance of various phase fusion approaches, where the AUC was calculated using the method described in [24]. The advantages of IFM are particularly evident across various lesions, including HCC, ICC, cysts, and "others". All methods except baseline nnU-Net can accept variable number of input phases and thus take advantage of the complete phase information (including delayed phase), yet IFM achieves better performance.

Ablation study. As shown in Table 3, focal loss is beneficial for improving the underrepresented calc class samples. LiverMax effectively confines the foreground within the liver region, significantly reducing false positives and thereby increasing accuracy. Moreover, IFM efficiently exploits multi-phase information, substantially enhancing performance in heterogeneous categories such as HCC and others. Lastly, with the regularization and optimization methods, ACL and

SAM, LIDIA achieves the highest average AUC of 93.6%. The findings from the ablation study clearly demonstrate the efficacy of our individual modules.

4   Conclusion

In this work, we introduce an effective approach to fuse multi-phase liver CT images to address the incomplete phase issue. Moreover, to minimize the impact of liver tumor heterogeneity on the model's classification performance, we propose an asymmetric contrastive loss. Our comprehensive evaluation on a large-scale dataset and external test set confirms the efficacy, generalizability, and clinical significance of our method.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., et al.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023)
2. Cao, K., Xia, Y., Yao, J., Han, X., Jiang, H., Fang, X., Nogues, I., Hou, Y., Kovarnik, T., Vocka, M.: Large-scale pancreatic cancer detection via non-contrast CT and deep learning. Nature Medicine (2023)
3. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1280–1289 (2022)
4. Cheng, B., Schwing, A., Kirillov, A.: Per-pixel classification is not all you need for semantic segmentation. Advances in Neural Information Processing Systems **34**, 17864–17875 (2021)
5. Cheng, C.T., Cai, J., Teng, W., Zheng, Y., Huang, Y.T.: A flexible three-dimensional heterophase computed tomography hepatocellular carcinoma detection algorithm for generalizable and practical screening. Hepatology Communications (2022)
6. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization. arXiv preprint arXiv:2010.01412 (2020)
7. Heinrich, M.P., Jenkinson, M., Brady, M., et al.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transactions on Medical Imaging **32**(7), 1239–1248 (2013)

8. Hu, Q., Chen, Y., Xiao, J., Sun, S., Chen, J., Yuille, A.L., Zhou, Z.: Label-free liver tumor segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7422–7432 (2023)

9. 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. Nature Methods **18**(2), 203–211 (2021)

10. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-DenseUNet: hybrid densely connected unet for liver and tumor degmentation from CT volumes. IEEE Transactions on Medical Imaging **37**(12), 2663–2674 (2018)

11. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2980–2988 (2017)

12. Marrero, J.A., Ahn, J., Reddy, R.K., of the American College of Gastroenterology, P.P.C., et al.: Acg clinical guideline: the diagnosis and management of focal liver lesions. Official Journal of the American College of Gastroenterology **109**(9), 1328–1347 (2014)

13. Tang, Y., Tang, Y., Zhu, Y., Xiao, J., Summers, R.M.: E2Net: An edge enhanced network for accurate liver and tumor segmentation on CT scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 512–522. Springer (2020)

14. Xu, X., Zhu, Q., Ying, H., Li, J., Cai, X., Li, S., Liu, X., Yu, Y.: A knowledge-guided framework for fine-grained classification of liver lesions based on multi-phase ct images. IEEE Journal of Biomedical and Health Informatics **27**(1), 386–396 (2022)

15. Xu, Y., Cai, M., Lin, L., Zhang, Y., Hu, H., Peng, Z., Zhang, Q., Chen, Q., Mao, X., Iwamoto, Y., Han, X.H., Chen, Y.W., Tong, R.: PA-ResSeg:A phase attention residual network for liver tumor segmentation from multiphase CT images. Medical Physics **48**(7), 3752–3766 (2021)

16. Yan, K., Yin, X., Xia, Y., Wang, F., Wang, S., Gao, Y., Yao, J., Li, C., Bai, X., Zhou, J., et al.: Liver tumor screening and diagnosis in ct with pixel-lesion-patient network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 72–82. Springer (2023)

17. Yao, J., Cao, K., Hou, Y., Zhou, J., Xia, Y., Nogues, I., Song, Q., Jiang, H., Ye, X., Lu, J., Jin, G., Lu, H., Xie, C., Zhang, R., Xiao, J., Liu, Z., Gao, F., Qi, Y., Li, X., Zheng, Y., Lu, L., Shi, Y., Zhang, L., Hospital, C.: Deep learning for fully automated prediction of overall survival in patients undergoing resection for pancreatic cancer: a retrospective multicenter study. Annals of Surgery (2022)

18. Yao, J., Ye, X., Xia, Y., Zhou, J., Shi, Y., Yan, K., Wang, F., Lin, L., Yu, H., Hua, X.S., et al.: Effective opportunistic esophageal cancer screening using non-contrast ct imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 344–354. Springer (2022)

19. 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)

20. Ying, H., Liu, X., Zhang, M., Ren, Y., Zhen, S., Wang, X., Liu, B., Hu, P., Duan, L., Cai, M., Jiang, M., Cheng, X., Gong, X., Jiang, H., Jiang, J., Zheng, J., Zhu, K., Zhou, W., Lu, B., Zhou, H., Shen, Y., Du, J., Ying, M., Hong, Q., Mo, J., Li, J., Ye, G., Zhang, S., Hu, H., Sun, J., Liu, H., Li, Y., Xu, X., Bai, H., Wang, S., Cheng, X., Xu, X., Jiao, L., Yu, R., Lau, W.Y., Yu, Y., Cai, X.: A multicenter clinical AI system study for detection and diagnosis of focal liver lesions. Nature Communications **15**(1), 1131 (feb 2024)

21. Zhang, D., Chen, B., Chong, J., Li, S.: Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images. Medical Image Analysis **70** (2021)
22. Zhang, H., Li, F., Xu, H., Huang, S., Liu, S., Ni, L.M., Zhang, L.: Mp-former: Mask-piloted transformer for image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18074–18083 (2023)
23. Zhou, J., Wang, W., Lei, B., Ge, W., Huang, Y., Zhang, L., Yan, Y., Zhou, D., Ding, Y., Wu, J., Wang, W.: Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. Frontiers in Oncology **10**, 1 (2021)
24. Zhu, Z., Xia, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12. Springer (2019)