PLUS: Plug-and-Play Enhanced Liver Lesion Diagnosis Model on Non-Contrast CT Scans

Jiacheng Hao[1,2,3(*)], Xiaoming Zhang[2,3(✉)], Wei Liu[2,3], Xiaoli Yin[4], Yuan Gao[2,3], Chunli Li[4], Ling Zhang[2], Le Lu[2], Yu Shi[4], Xu Han[5], Ke Yan[2,3]

[1] School of Biomedical Engineering, Tsinghua University, 100084, Beijing, 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

Abstract. Focal liver lesions (FLL) are common clinical findings during physical examination. Early diagnosis and intervention of liver malignancies are crucial to improving patient survival. Although the current 3D segmentation paradigm can accurately detect lesions, it faces limitations in distinguishing between malignant and benign liver lesions, primarily due to its inability to differentiate subtle variations between different lesions. Furthermore, existing methods predominantly rely on specialized imaging modalities such as multi-phase contrast-enhanced CT and magnetic resonance imaging, whereas non-contrast CT (NCCT) is more prevalent in routine abdominal imaging. To address these limitations, we propose PLUS, a plug-and-play framework that enhances FLL analysis on NCCT images for arbitrary 3D segmentation models. In extensive experiments involving 8,651 patients, PLUS demonstrated a significant improvement with existing methods, improving the lesion-level F1 score by 5.66%, the malignant patient-level F1 score by 6.26%, and the benign patient-level F1 score by 4.03%. Our results demonstrate the potential of PLUS to improve malignant FLL screening using widely available NCCT imaging substantially. Code is availabel at https://github.com/alibaba-damo-academy/plug-and-play-diagnosis.

Keywords: Focal liver lesions · Lesion diagnosis · Non-Contrast CT

1 Introduction

The incidental detection of focal liver lesions (FLLs) during routine physical examination has become increasingly prevalent [15]. However, accurate differentiation between malignant and benign lesions remains a significant challenge

[*] The work was done during J. Hao's internship at Alibaba DAMO Academy.
[✉] Corresponding author.

for radiologists due to the subtle heterogeneity present in images. Benign lesions are usually harmless or require regular monitoring, whereas malignant tumors require immediate intervention as delayed diagnosis can lead to rapid disease progression and poor outcomes [2,15]. In this scenario, early detection and differentiation of the malignant lesions plays a crucial role in patient survival, with five-year survival rates exceeding 70% for early-stage diagnoses, underscoring the critical importance of effective screening programs, particularly for high-risk populations with chronic liver conditions [5].

Contrast-enhanced CT (CECT) and magnetic resonance imaging (MRI) are primary modalities for FLL diagnosis due to their high sensitivity [9]. However, both face significant limitations in the context of large-scale opportunistic screening. CECT requires contrast agents that carry risks of nephrotoxicity and allergic reactions, restrict follow-up examination frequency, and substantially increase healthcare costs. MRI, a radiation-free imaging modality, is unsuitable for large-scale screening due to limited availability, lengthy protocols, and device contraindications. Meanwhile, non-contrast CT (NCCT) presents advantages for large-scale malignancy screening [27], combining widespread availability, cost-effectiveness, rapid acquisition, and absence of contrast-related risks, making it ideal for broad FLL screening.

Despite NCCT's advantages, its fundamental reliance on tissue density differences makes it challenging to detect FLLs, especially early-stage or subtle lesions with minimal distinction against surrounding tissues, inherently hindering detection and complicating classification due to subtle variations among tumor types. Chronic liver conditions also further complicate tumor identification, e.g., fatty liver, cirrhosis, and biliary dilation. While radiologists' expertise is crucial, the subtle imaging characteristics make diagnosis highly experience-dependent and prone to errors, highlighting the need for computer-aided methods.

Recently, deep learning methods have shown promise in enhancing automatic FLL detection by accurately capturing the lesion patterns on CT images [25]. Although well-configured CNN-based architectures remain competitive for general medical image segmentation [11,23,24], FLL classification presents unique challenges beyond detection. Despite various architectural innovations including Transformers [7], and Mamba-based designs [18], the accurate subtype classification remains a key challenge. MaskFormer-based approaches [4,3], despite their promising results through dense cross-attention, lack explicit mechanisms to model crucial anatomical relationships between lesions and liver tissue. Existing classification methods [29,22,31] typically focus on isolated lesion-level or patient-level predictions, overlooking the complex interplay between global liver morphology and local lesion characteristics. While recent approaches [17,20,26] attempted global-local feature modeling, they failed to capture clinically relevant and multi-scale liver pathology information. Drawing insights from multi-organ cohesion analysis [13], we propose a classification-focused framework that emphasizes feature discrimination within pre-detected regions of interest (ROIs), addressing the limitations of detection-based approaches.

To address these limitations, we propose a **PLU**g-and-play enhanced liver lesion diagnosi**S** model (PLUS), a framework compatible with arbitrary 3D segmentation models for improved FLL analysis on NCCT. PLUS introduces three key components: (1) a hierarchical dual attention (HDA) mechanism facilitating bidirectional exchange between global liver and local lesion features; (2) a graph-based prior reasoning (GPR) module selectively leveraging prior pretrained segmentation knowledge for FLL type classification; and (3) a combined optimization strategy for lesion-level detection with patient-level diagnosis.

We curated the largest NCCT FLL dataset containing 28,853 annotated lesions from 8,651 patients and healthy subjects. PLUS achieves F1 scores of 65.11% for lesion-level detection, 90.12% for malignant, and 73.10% for benign patient-level diagnosis, surpassing other methods. Through extensive ablations and visualizations, we demonstrate PLUS's effectiveness and potential in enhancing large-scale FLL screening and differentiation via widely available NCCT.

2 Method

2.1 Problem Definition

Considering a 3D NCCT volume $I \in \mathbb{R}^{H\times W\times D}$, the PLUS framework builds on an arbitrary pre-trained segmentation model $\phi(\cdot)$ that processes the input volume and generates segmentation masks $M = \phi_{\text{seg}}(I) \in \{0,1\}^{H\times W\times D}$ and preliminary lesion classification logits $P = \phi_{\text{cls}}(I) \in \mathbb{R}^{C}$, where C denotes the number of lesion categories. The segmentation mask M inherently contains two anatomical regions: the liver region M_H and the set of lesion instances $\mathcal{S} = \{(M_{s_i}, P_i)\}_{i=1}^{N}$, which serve as spatial attention guides for subsequent feature extraction, where N is the number of detected lesions, the objective of the plug-in framework $f_\theta(\cdot)$ is to leverage these segmentation-derived priors and enhance the final classification prediction through $f_\theta(I, M_H, \mathcal{S})$.

2.2 Plug-and-Play Enhanced Liver Lesion Diagnosis

For the input NCCT volume I and its segmentation masks of liver M_{H} and pre-segmented lesion instances $\mathcal{S} = \{(M_{s_i}, P_i)\}_{i=1}^{N}$ from model $\phi(\cdot)$, the framework extracts region-specific features using two encoders: $F_H = E_\theta^H(I \odot M_H)$ for liver tissue and $F_{s_i} = E_\theta^S(I \odot M_{s_i})$ for each lesion instance i, see Fig. 1. The resulting feature tensors F_H and $\{F_{s_i}\}_{i=1}^{N}$ encode contextual information from the liver and detailed characteristics of individual lesions, respectively.

Hierarchical Dual Attention. The initial representations lack cross-scale feature interaction between anatomical context and local lesion patterns. To enable effective multi-scale feature interaction, we propose HDA, a module enabling bidirectional intra-/inter-scale semantic fusion with liver and lesions.

Given lesion features F_s and liver features F_H, the HDA mechanism can be formulated as concatenated features $\mathbf{F} = \mathcal{F}_{\theta 1}([\text{DCA}(\text{Pool}_i(F_H); F_S)]_{i=1}^{K})$, where the dual cross-attention (DCA) combines bidirectional attention flows:

$$F'_x = \mathcal{F}_{\theta 2}(\mathcal{A}(F_x, F_y)), \quad \text{DCA}(F_x, F_y) = \left[F'_x; \mathcal{F}_{\theta 3}(\mathcal{A}(F_y, F'_x))\right], \tag{1}$$

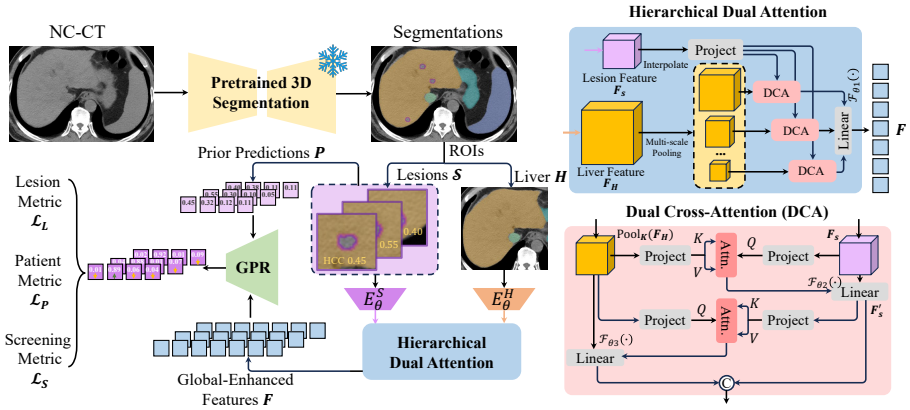


Fig. 1: Illustration of the overall pipeline of PLUS.

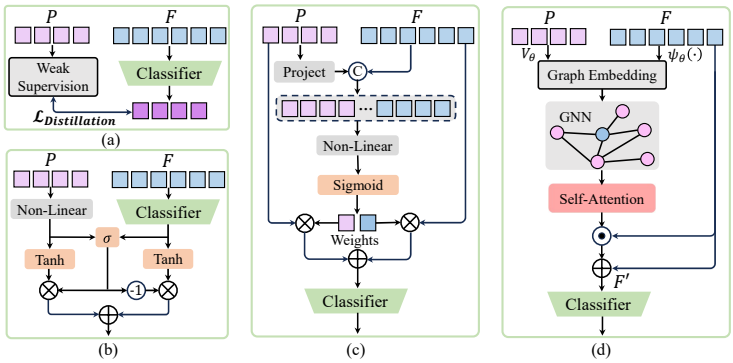


Fig. 2: Comparison of classifier enhancement strategies: (a) Knowledge Distillation, (b) Gated Fusion, (c) Weighted Fusion, (d) Graph-Based Prior Reasoning.

with attention operation $\mathcal{A}(F_a, F_b) = \text{Attn.}(\mathcal{T}_Q(F_a), \mathcal{T}_K(F_b), \mathcal{T}_V(F_b))$, $\mathcal{T}_{\{Q,K,V\}}$ are query/key/value projection operations, $\text{Pool}_i(\cdot)$ represents multi-scale pooling at the i-th scale, $\mathcal{F}_\theta$ are learnable linear projections, and the standard attention operation is defined as $\text{Attn.}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}$.

The design of HDA is motivated by two key clinical observations in liver lesion diagnosis. First, radiologists zoom in and out the image when examining lesions, as different lesions exhibit distinctive patterns on various scales. HDA mimics this diagnostic process through efficient multi-scale feature interaction with progressive pooling operations. Furthermore, the interpretation of the lesion depends heavily on the surrounding liver conditions, where bidirectional interaction is crucial. For example, a suspicious lesion of a cirrhosis patient tends to be a hepatocellular carcinoma, while a low-density region on the liver of a steatosis patient is probably a false positive tumor.

Algorithm 1 Graph-based Prior-Aware Reasoning

Require: Features $\mathbf{F} \in \mathbb{R}^{B\times D}$, Prior Prediction $\mathbf{P} \in \mathbb{R}^{B\times C}$, Prototypes $\mathbf{V}_\theta \in \mathbb{R}^{C\times D}$
Ensure: Prior-aware Features $\mathbf{F}' \in \mathbb{R}^{B\times D}$
1: $\mathbf{F}_\psi = \psi_\theta(\mathbf{F}), \mathbf{V}_w = \mathbf{P}_\psi \times \mathbf{V}_\theta$ ▷ Feature projection and weighted prototypes
2: $\mathbf{G} = [\mathbf{F}_\phi; \mathbf{V}_\theta]$ ▷ Node set construction: B original nodes + B prototypes nodes
3: $\mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathcal{W}_q(\mathbf{G}), \mathcal{W}_k(\mathbf{G}), \mathcal{W}_v(\mathbf{G})$; $\alpha_{ij} = \text{softmax}(\frac{\mathbf{q}_i^T \mathbf{k}_j}{\sqrt{d}})$, $\mathbf{m}_i = \sum_{j=1}^{2B} \alpha_{ij}\mathbf{v}_j$
4: **return** $\mathbf{F}' = \mathbf{F} + \mathcal{F}_\theta([\mathbf{F}_\phi \parallel \mathbf{m}_{1:B}]) \odot \mathbf{V}_w$ ▷ Updated feature of central node

Graph-based Prior-aware Reasoning. We propose a graph-based prior-aware reasoning (GPR) module to refine potentially inaccurate prior lesion predictions through graph neural networks [12]. As outlined in Algorithm 1 and (Fig. 2(d)), GPR constructs a homogeneous graph $\mathbf{G}$ where enhanced features $\mathbf{F}$ serve as center nodes, connected with class-specific prototypes V_θ as surrounding nodes. Instead of directly using the initial prior prediction P, which may be unreliable [16], we leverage it as a conditional message to guide prototype-based reasoning. Through message passing and self-attention [21] between features and prototypes, the graph-based structure automatically learns to enhance relevant prior knowledge while suppressing unreliable predictions, enabling more robust feature enhancement compared to conventional fusion methods.

Training Paradigm. Clinically, the diagnosis of FLL through medical imaging follows a systematic process: radiologists first analyze individual lesion characteristics, and then compile these findings into patient-level diagnoses. Both lesion-level and patient-level analyses are crucial: the former provides detailed lesion-level insights while the latter guides overall treatment planning. This is particularly challenging in cases with multiple lesions, where effective aggregation of lesion-level information becomes essential for accurate patient-level assessment. Inspired by this clinical workflow, we propose a comprehensive optimization strategy that aligns with radiologists' decision-making process.

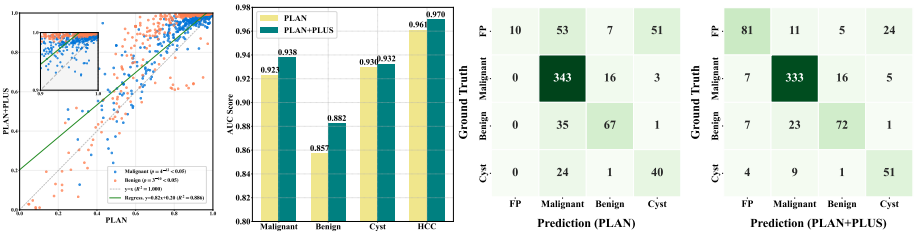
The lesion level loss $\mathcal{L}_L = -\frac{1}{N}\sum_i^N \sum_c^C \mathbf{z_c}(1-p_{i,c})^\gamma y_{i,c} \log p_{i,c}$ addresses class imbalance through lesion-specific weights z_c, and focal parameter γ, with N lesions across C categories. The overall lesion-level loss is the average of the individual losses for each patient's lesions. Taking malignancy prediction as example, the patient-level loss $\mathcal{L}_P = -\frac{1}{M}\sum_j^M [Y_j \log(\max_{i\in\mathcal{S}_j} p^j_{i,malig}) + (1-Y_j)\log(1 - \max_{i\in\mathcal{S}_j} p^j_{i,malig})]$ aggregates malignancy probabilities over M patients, prioritizing clinically critical findings. Screening loss $\mathcal{L}_S = -\frac{1}{M}\sum_j^M [y_j \log q_j + (1-y_j)\log(1-q_j)]$ uses $q_j = \max_{i\in\mathcal{S}_j}(p_{i,malig} \parallel p_{i,beni.})$ to detect whether each subject has any tumor. The overall loss function integrates three hierarchically structured components $\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_L + \beta\mathcal{L}_P + (1-\alpha-\beta)\mathcal{L}_S$. This tripartite design validates PLUS through the diagnostic cascade: screening → lesion analysis → diagnosis, ensuring consistency across clinical workflow stages.

Table 1: Performance comparison on test set. (●: malignant, ○: benign.)

Method	Lesion-level			Patient-level Diagnosis						Screening	
	F1	Prec.	Recall	F1●	Prec.●	Recall●	F1○	Prec.○	Recall○	F1	Acc.
nnUNet [10]	56.19	46.43	**71.16**	80.91	74.93	**87.92**	70.34	75.27	66.02	87.45	88.02
+PLUS	**62.97**	**57.62**	69.42	**83.74**	**81.56**	86.04	**72.68**	**77.53**	**68.41**	**87.83**	**88.45**
Mask2Former [3]	58.19	49.56	**70.48**	82.80	80.32	**85.45**	67.51	72.02	63.54	87.78	88.14
+PLUS	**63.46**	**58.85**	68.87	**86.02**	**89.32**	82.97	**72.18**	**76.54**	**68.30**	**87.96**	**88.84**
PLAN [26]	59.45	48.27	**77.39**	83.86	75.38	**94.49**	69.07	73.63	65.05	88.89	89.00
+PLUS(distill.) [6]	61.08	58.48	63.94	88.47	87.58	89.38	71.39	79.67	64.67	88.39	89.08
+PLUS(gated) [1]	57.92	51.47	66.24	88.53	86.49	90.68	66.39	**84.93**	54.50	88.82	88.96
+PLUS(weighted) [28]	64.47	56.38	75.26	87.84	86.12	89.65	64.41	76.75	55.50	**89.04**	89.42
+PLUS(GPR)	**65.11**	**60.28**	74.56	**90.12**	**88.56**	91.74	**73.10**	76.60	**69.90**	88.97	**89.73**

Table 2: Ablation study on proposed components. (●: malignant, ○: benign.)

Method	Lesion-level			Patient-level Diagnosis					
	F1	Prec.	Recall	F1●	Prec.●	Recall●	F1○	Prec.○	Recall○
PLAN	59.45	48.27	**77.39**	83.86	75.38	**94.49**	69.07	73.63	65.05
+HDA	64.67	58.72	71.96	89.25	87.53	91.05	71.56	75.33	68.16
+GPR	**65.87**	**60.51**	72.23	89.54	88.24	90.89	71.84	76.12	68.02
+Comb. Loss	65.11	60.28	74.56	**90.12**	**88.56**	91.74	**73.10**	**76.60**	**69.90**



(a) (b) (c)

Fig. 3: Quantitative analysis. (a) Binary classification significance test. (b) Lesion-wise AUC comparisons. (C) Patient-wise confusion matrix.

3 Experiments

Dataset. We constructed a large-scale NCCT dataset comprising 8,651 patients with or without liver tumors from two medical centers. To efficiently ensure high-quality annotations, we adopted a human-in-the-loop approach where two senior radiologists (>10 yrs experience) first manually annotated a subset of cases to train a preliminary model, which then generated results on remaining samples for iterative refinement. Ground-truth lesion diagnosis comes from pathology reports, imaging features, and follow-up data. The dataset includes 28,853 FLLs in nine types, including malignant lesions : hepatocellular carcinoma

Ethics approval numbers by institutional review board: 2024PS954K, IIT20220356B-R2.

(HCC; 3,321), cholangiocarcinoma (CCA;1,866), metastasis (Meta;8,299), other malignant (34); and benign lesions: hemangioma (Heman;1,381), focal nodular hyperplasia (128), calcification (755), cyst (13,049), other benign (23). We stratified the dataset into training (6,500), validation (951), and testing (1,200) sets.

Implementation Details. The experiments were conducted on a single NVIDIA A800 GPU. All pre-segmentation models $\phi(\cdot)$ were trained using 5-fold cross-validation on the training set for 1,000 epochs. The network architecture consists of two 3D ResNet-18 [8] encoders, i.e, E_θ^S and E_θ^H, to capture lesion-specific and global context liver features. For each detected lesion from $\phi(\cdot)$, we extracted volumetric ROIs of size $64 \times 64 \times 16$ voxels centered at the lesion. Input volumes were pre-processed using the nnUNet [10] pipeline, including resampling to isotropic voxel spacing and normalization with an abdominal window setting. The network was trained for 100 epochs using the AdamW optimizer [14] (learning rate=10^{-4}, weight decay=0.05, batch size=2) with cosine annealing scheduling. The loss weights are $\alpha = 0.5$ and $\beta = 0.3$. For HDA, we set $K = 4$.

Main Experimental Results. We evaluate the proposed PLUS framework by integrating it with three off-the-shelf liver lesion segmentation baselines with high FLL detection sensitivity: nnUNet [10] is widely recognized as the de facto gold standard in medical image segmentation with its robust generalizability [11]; Mask2Former [3] demonstrates superior performance in instance segmentation with dense cross-attention architecture; and PLAN [26] extends Mask2Former with improved anchor queries and foreground-enhanced sampling loss, achieving state-of-the-art performance in joint liver tumor segmentation and diagnosis.

As summarized in Table 1, PLAN+PLUS achieves the best performance across most metrics, with F1 scores of 65.11%, 90.12%, and 73.10% for lesion-wise detection, malignant patient-level, and benign patient-level assessment, respectively. Notably, the plugin PLUS demonstrates consistent performance improvements when integrated with existing architectures. Compared to vanilla PLAN, PLAN+PLUS shows significant gains of 5.66% in lesion-wise F1 score, 6.26% in malignant patient-level F1 score, and 4.03% in benign patient-level F1 score. Similar improvement patterns are also observed when applying the plugin module to nnUNet and Mask2Former, validating the effectiveness and architecture-agnostic robustness of PLUS.

Fig. 3 presents a comprehensive quantitative comparison between PLAN and PLAN+PLUS. The probability correlation analysis in Fig. 3(a) demonstrates that PLAN+PLUS increases the correct confidence score of each class than the baseline PLAN, particularly for cases in the mid-probability range where classification is typically more challenging. Fig. 3(b) reveals substantial improvements in both malignant vs. benign classification and the diagnosis of common lesion types. The confusion matrices in Fig. 3(c) highlight PLAN+PLUS's superior performance through reduced false positives for malignant lesions, improved benign lesion classification, and better identification of cystic lesions, suggesting that PLUS effectively further enhances patient-wise classification accuracy.

Effect of GPR. To validate the effectiveness of GPR, we compare it with several alternative fusion strategies as shown in Fig. 2, including distillation [6], gated-

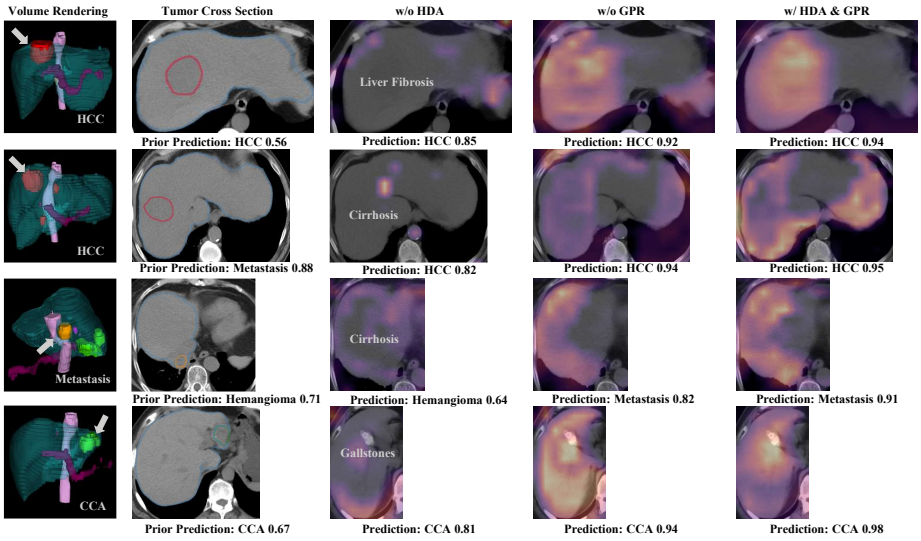


Fig. 4: Qualitative results showing ablation study of model components. Each case is annotated with the predicted lesion category and its probability.

fusion [1], and weighted fusion methods [28]. Given the potential inaccuracies in prior P and its dimensional imbalance with $\mathbf{F}$, existing approaches show limitations: knowledge distillation restricts priors to supervisory signals, gated fusion lacks adaptability to varying prototype relevance, and weighted fusion fails to capture complex feature-prior relationships. As shown in Table 1, GPR (last row) addresses these challenges and outperforms them across different metrics through GNN-based relationship modeling and adaptive feature aggregation.

Ablation Study. We conducted ablation studies to validate each component of the PLUS framework. Starting with the PLAN baseline, adding HDA significantly enhances lesion-level F1 by 4.40% and malignant patient-level F1 by 3.53%. GPR further improves performance, and after adding combined loss, despite causing slight deterioration at the lesion level, it achieves optimal results at the patient level with substantial gains in malignant and benign F1.

Saliency method visualizations (e.g., Grad-CAM [19,30]) show the baseline exhibits diffuse attention, often missing critical regions. HDA focuses on diagnostically pivotal areas (e.g., fibrotic patterns, uneven liver border that indicates cirrhosis, and gallstones that correlate with CCA) and resolves ambiguities between challenging subtypes. GPR refines attention to pathology-specific regions by fusing segmentation priors with liver features. Combined, HDA+GPR produces anatomically coherent attention maps that capture both localized lesion characteristics and structural context, confirming they enhance the model's focus on clinically meaningful regions.

4 Conclusion

In this work, we present PLUS, a plugin framework that enhances existing segmentation models for NCCT FLL analysis. PLUS further improves the state of the art while maintaining architectural flexibility. Future work will extend this paradigm to multi-modal imaging integration and real-world clinical trials to support clinical decision-making and intervention.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 82471971); General Program of the Liaoning Provincial Department of Education (LJKMZ20221160); Liaoning Province Science and Technology Joint Plan (2023JH2/101700127); the Leading Young Talent Program of Xingliao Yingcai in Liaoning Province (XLYC2203037).

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

References

1. Arevalo, J., Solorio, T., y Gómez, M.M., González, F.A.: Gated multimodal units for information fusion. In: International Conference on Learning Representations. IEEE (2017)
2. 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)
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. De Franchis, R., Bosch, J., Garcia-Tsao, G., Reiberger, T., Ripoll, C., Abraldes, J.G., Albillos, A., Baiges, A., Bajaj, J., Bañares, R., et al.: Baveno vii–renewing consensus in portal hypertension. Journal of hepatology **76**(4), 959–974 (2022)
6. Gou, J., Yu, B., Maybank, S.J., Tao, D.: Knowledge distillation: A survey. International Journal of Computer Vision **129**(6), 1789–1819 (2021)
7. Hatamizadeh, A., Yang, D., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision pp. 1748–1758 (2021)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition pp. 770–778 (2015)
9. Huang, W., Liu, W., Zhang, X., Yin, X., Han, X., Li, C., Gao, Y., Shi, Y., Lu, L., Zhang, L., et al.: Lidia: Precise liver tumor diagnosis on multi-phase contrast-enhanced ct via iterative fusion and asymmetric contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 394–404. Springer (2024)

10. 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)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
13. Li, C., Zhang, X., Gao, Y., Yin, X., Lu, L., Zhang, L., Yan, K., Shi, Y.: Improved esophageal varices assessment from non-contrast ct scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 349–359. Springer (2024)
14. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
15. 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)
16. Meng, Y., Zhang, H., Zhao, Y., Yang, X., Qiao, Y., MacCormick, I.J., Huang, X., Zheng, Y.: Graph-based region and boundary aggregation for biomedical image segmentation. IEEE Transactions on Medical Imaging **41**(3), 690–701 (2021)
17. Ren, J., Zhang, X., Zhang, L.: Hifiseg: High-frequency information enhanced polyp segmentation with global-local vision transformer. IEEE Access (2025)
18. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. arXiv preprint arXiv:2402.02491 (2024)
19. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 618–626 (2017)
20. Tragakis, A., Liu, Q., Kaul, C., Roy, S.K., Dai, H., Deligianni, F., Murray-Smith, R., Faccio, D.: Glfnet: Global-local (frequency) filter networks for efficient medical image segmentation. In: IEEE International Symposium on Biomedical Imaging. p. 1–5. IEEE (May 2024)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
22. Wang, F., Cheng, C.T., Peng, C.W., Yan, K., Wu, M., Lu, L., Liao, C.H., Zhang, L.: A cascaded approach for ultraly high performance lesion detection and false positive removal in liver ct scans. arXiv preprint arXiv:2306.16036 (2023)
23. Wang, R., Zheng, G.: Cycmis: Cycle-consistent cross-domain medical image segmentation via diverse image augmentation. Medical Image Analysis **76**, 102328 (2022)
24. Wang, R., Zhou, Q., Zheng, G.: Few-shot medical image segmentation regularized with self-reference and contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 514–523. Springer (2022)
25. Wei, Y., Yang, M., Zhang, M., Gao, F., Zhang, N., Hu, F., Zhang, X., Zhang, S., Huang, Z., Xu, L., et al.: Focal liver lesion diagnosis with deep learning and multistage ct imaging. Nature communications **15**(1), 7040 (2024)

26. 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)
27. 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)
28. Yin, H.: Tensor sparse representation for 3-d medical image fusion using weighted average rule. IEEE Transactions on Biomedical Engineering **65**(11), 2622–2633 (2018)
29. 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)
30. Zhang, J., Chao, H., Dasegowda, G., Wang, G., Kalra, M.K., Yan, P.: Revisiting the trustworthiness of saliency methods in radiology ai. Radiology: Artificial Intelligence **6**(1), e220221 (2023)
31. Zhang, J., Chao, H., Dhurandhar, A., Chen, P.Y., Tajer, A., Xu, Y., Yan, P.: Spectral adversarial mixup for few-shot unsupervised domain adaptation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 728–738. Springer (2023)