Asymmetric Matching in Abdominal Lymph Nodes of Follow-up CT Scans

Yiji Mao[1], Yi Zhang[1], Xinyu Zou[1], Yuling Zheng[1], Hao Huang[2], and Haixian Zhang[1](✉)

[1] Machine Intelligence Lab, School of Computer Science and Technology, Sichuan University, Chengdu, Sichuan, China zhanghaixian@scu.edu.cn
[2] Colorectal Cancer Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China

Abstract. Accurate tracking of abdominal lymph nodes (LN) across follow-up computed tomography (CT) scans is crucial for colorectal cancer staging and treatment response evaluation. However, establishing reliable LN correspondences remains underexplored due to challenges including scale variations, low resolution, difficulty distinguishing nodes from adjacent structures, inability to handle tissue deformation, and dynamic visibility. To address these challenges, we propose an asymmetric matching framework that strikes a balance between enhancing LN specificity and contextual correlations. For specificity, we achieve cross-dimensional feature consistency and generate discriminative LN features via self-supervised learning on orthogonal 2D projections of 3D node volumes. For correlation, we develop a graph model capturing lymphatic topology within scans, reinforced by temporal contrastive learning that encourages consistency between matched node pairs across CT. To balance specificity and correlation, we propose a multi-module architecture that integrates volumetric LN features with projection embeddings through attention-based fusion, enabling confidence-calibrated similarity assessment across temporal scans. Experimental results demonstrate that our solution provides reliable lymph node correspondence for clinical follow-up and disease monitoring. Code is available at `https://github.com/maoyij/Asymmetric-Matching`.

Keywords: Lymph node matching · Abdominal CT scans · Spatiotemporal feature · Graph Neural Network (GNN) · Contrastive learning

1 Introduction

Lymph node (LN) metastasis critically determines colorectal cancer prognosis and therapeutic strategies [1, 2], and longitudinal abdominal LN monitoring in serial computed tomography (CT) scans is essential for detecting metastatic progression during postoperative surveillance [3, 4]. However, anatomical complexity and temporal morphological variations of LNs hinder precise cross-scan correlation, requiring over 15 minutes per case. Retrospective analyses indicate that 22.2-56.8% of recurrence cases experience postponed treatment adjustments [5] due to undetected LN progression. Such diagnostic delays pose a risk to clinical outcomes by exceeding the optimal intervention windows, thereby underscoring the necessity for standardized automated monitoring systems to enhance matching accuracy and longitudinal consistency.

The clinical challenges in longitudinal LN monitoring originate from distinctive complexities in abdominal lymph node data. Metastatic nodes frequently demonstrate non-rigid deformations, with diameters that can reach 10-fold across sequential scans. Simultaneously, temporal scans further reveal dynamic shifts, where nodes new/disappear unpredictably, creating spurious matches. Furthermore, anatomical ambiguity in the abdomen further compounds these challenges. All these characteristics indicate that maintaining both the specificity between different LNs and the correlation within the same LN is very difficult. This causes problems both clinically and technically, leading to issues with asymmetric matching.

Existing research on abdominal LN primarily focuses on detection [6–8], segmentation [9], and metastasis prediction [10]. However, automatic abdominal LN matching remains underdeveloped, and current methods for monitoring tissues face challenges in addressing the asymmetric matching problem. The existing approaches for monitoring tissues typically adopt two main paradigms: image registration and universal lesion monitoring. Conventional registration techniques, including affine transformations [11] and deformable models [12], demonstrate efficacy in aligning major anatomical landmarks, but are inadequate for tracking small and morphologically labile targets. While universal lesion tracking systems have achieved clinical utility for relatively stable structures such as lung nodules [13], liver lesions [14], and brain anatomy [15], their performance degrades significantly when applied to LNs exhibiting non-rigid deformations within complex anatomical environments. Recent deep learning trackers that employ similarity metrics [16], multi-scale feature fusion [17], and cross-attention have shown potential for stationary lesions but fail to address the difficulty of tracking small targets and handling LNs' new/disappear. Even advanced self-supervised paradigms [18–20] that refine voxel-wise feature learning remain constrained by the inherent heterogeneity and non-rigid deformations of LNs, particularly when distinguishing LNs from adjacent vascular structures.

To address the asymmetric matching problem, we propose a framework that integrates spatial and temporal information. The main contributions of our work are as follows: (1) Multilevel Spatial-Temporal Matching from global to local to address small target matching in complex anatomical environments; (2) Self-Supervised Feature Extraction harnessing 2D projections to amplify LN-specific features obscured by anatomical ambiguity; (3) Balanced Feature Integration employing adaptive fusion to reconcile specificity and contextual correlations.

2 Method

Given a series of longitudinal 3D CT scans $\{I_1, I_2, \ldots, I_T\}$, each scan I_T contains n_T lymph node regions of interest (ROIs), denoted as $\mathcal{L}_T = \{l_1^T, l_2^T, \ldots, l_{n_T}^T\}$,

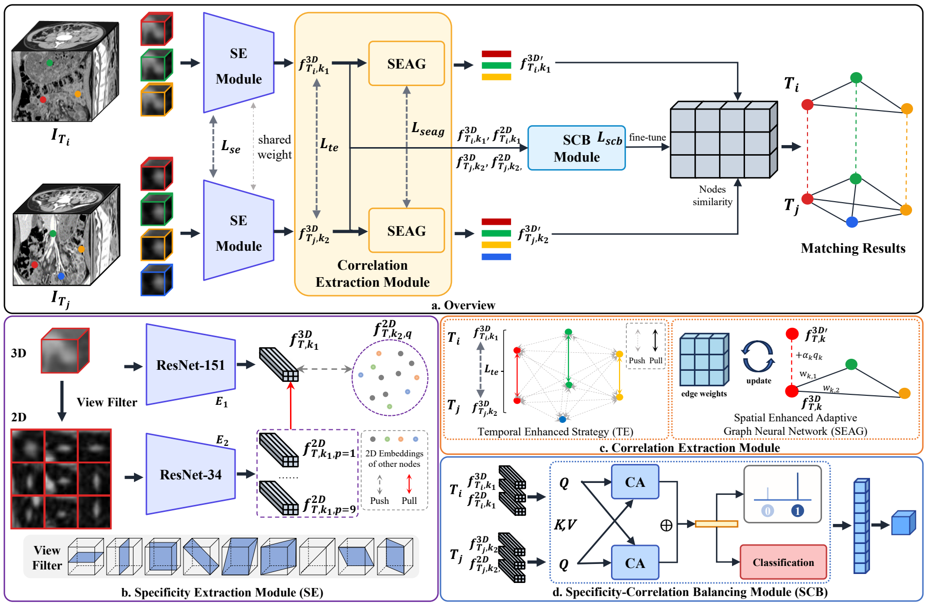

Fig. 1. (a) Overview of the asymmetric matching framework. (b) SE module. (c) Correlation Extraction module. (d) SCB module.

where l_k^T represents the k-th lymph node in scan I_T. The asymmetric matching problem aims to establish a mapping $\mathcal{M}_{I_{T_i} \to I_{T_j}} : \mathcal{L}_{T_i} \to \mathcal{L}_{T_j}$ between two scans I_{T_i} and I_{T_j} $(1 \leq T_i < T_j \leq T)$. The mapping $\mathcal{M}_{I_{T_i} \to I_{T_j}}$ is constructed to: (1) identify persistent nodes through one-to-one correspondences $\mathcal{M}(l_{k_1}^{T_i}) = l_{k_2}^{T_j}$; (2) classify unmatched ROIs as new/disappear.

As shown in Fig. 1, the asymmetric matching framework consists of three main components: specificity extraction module, correlation extraction module, and specificity-correlation balancing module. The specificity extraction module extracts distinctive features from individual LNs to differentiate between similar nodes. The correlation extraction module captures relationships among LNs and temporal correlations of the same LN undergoing non-rigid deformations. The specificity-correlation balancing module reconciles these features and fine-tunes the final results.

2.1 Specificity Extraction Module (SE)

The SE module focuses on extracting discriminative features from each individual lymph node through multi-view representation learning. For every node, nine 2D views (three orthogonal, six diagonal/oblique)[21] are extracted from its 3D cube by projecting the 3D volume along different directions. Each 2D view is encoded into a 128-dimensional vector using a shared ResNet-34 encoder, while the 3D volume is processed by a ResNet-151 backbone to extract 3D features.

The 2D-3D pairs from the same node are treated as positives, while 2D views from other nodes serve as negatives for contrastive loss.

For the k_1-th lymph node at time points T, we represent its features as $f^{2D}_{T,k_1,p} \in \mathbb{R}^{128}$ for the p-th 2D projection and $f^{3D}_{T,k_1} \in \mathbb{R}^{128}$ for its 3D features. The 2D-3D contrastive loss function is:

$$L_{se} = -\log\left(\frac{\sum_{p=1}^{9} e^{s^{+}_{k_1,p}}}{\sum_{p=1}^{9} e^{s^{+}_{k_1,p}} + \sum_{k_2 \neq k_1} \sum_{q=1}^{9} e^{s^{-}_{k_1,k_2,q}}}\right), \tag{1}$$

where $s^{+}_{k_1,p} = \text{sim}(f^{3D}_{T,k_1}, f^{2D}_{T,k_1,p})/\tau$, $s^{-}_{k_1,k_2,q} = \text{sim}(f^{3D}_{T,k_1}, f^{2D}_{T,k_2,q})/\tau$, $\text{sim}(\cdot)$ computes cosine similarity, and τ is the temperature parameter.

2.2 Correlation Extraction Module

The correlation extraction module captures temporal and spatial correlations among lymph nodes through two complementary components: the TE extracts longitudinal correlations between the same lymph node across inter-CT scans, while the SEAG captures horizontal correlations by leveraging contextual relationships among lymph nodes within intra-CT scans.

Temporal Enhanced Strategy (TE) TE extracts longitudinal correlations by learning consistent representations of the same lymph node across different temporal CT scans. It introduces a contrastive loss based on cross-CT node pairs to encourage temporal consistency and feature discrimination. TE directly leverages information from both CT scans to learn core features that remain consistent despite non-rigid deformations, but does not alter the feature extraction or aggregation process. The temporal contrastive loss is formulated as:

$$L_{te} = -\frac{1}{N}\sum_{k_1=1}^{N} \log\left(\frac{e^{s^{+}_{k_1,k_1}}}{e^{s^{+}_{k_1,k_1}} + \sum_{k_2 \neq k_1} e^{s^{-}_{k_1,k_2}}}\right), \tag{2}$$

where $s^{+}_{k_1,k_1} = \text{sim}(f^{3D}_{T_1,k_1}, f^{3D}_{T_2,k_1})/\tau$, $s^{-}_{k_1,k_2} = \text{sim}(f^{3D}_{T_1,k_1}, f^{3D}_{T_2,k_2})/\tau$, $f^{3D}_{T_1,k_1}$ and $f^{3D}_{T_2,k_1}$ represent the 3D features of the k_1-th lymph node at time points T_1 and T_2 respectively, and N is the number of matched lymph node pairs.

Spatial Enhanced Adaptive Graph Neural Network (SEAG) SEAG captures horizontal correlations by modeling the relational context among lymph nodes within a single CT scan using a two-layer graph neural network architecture. Each CT is represented as a fully connected graph, where nodes are lymph nodes represented by 128-dimensional features f^{3D}_{T,k_1} extracted from the SE module and the initial edge weights are determined by coordinate differences between nodes, $w_{k_1,k_2} = \exp(-\|c_{k_1} - c_{k_2}\|)$. SEAG leverages contextual relationships among lymph nodes within the same CT scan to update node features, capturing anatomical proximity and structural patterns to address non-rigid deformations.

The first layer updates node features through a two-step process: (1) Neighbor Aggregation: For each lymph node, it first collects information from all neighboring nodes by computing a weighted sum $q_{k_1} = \sum_{k_2} w_{k_1,k_2} h_{k_2}^{(0)}$. (2) Attention-based Update: It then computes self-attention weights $\alpha_{k_1} = \text{softmax}\left(\frac{q_{k_1} Q^\top}{\sqrt{d_k}}\right)$ to determine the importance of the aggregated information, and finally updates the node representation with $h_{k_1}^{(1)} = h_{k_1}^{(0)} + \alpha_{k_1} q_{k_1}$, where Q is a learnable parameter matrix.

Before the second layer, the edge weights are dynamically updated based on the new node representations $h^{(1)}$ as $w_{k_1,k_2}^{(1)} = \exp(-\frac{\|h_{k_1}^{(1)} - h_{k_2}^{(1)}\|_2^2}{\sigma^2})$, where σ is a learnable parameter. The second layer takes the features $h^{(1)}$ and $w_{k_1,k_2}^{(1)}$ as input to generate the final node representations $h^{(2)}$, which involves computing $q_{k_1}^{(1)} = \sum_{k_2} w_{k_1,k_2}^{(1)} h_{k_2}^{(1)}$ and then $h_{k_1}^{(2)} = \text{ReLU}(h_{k_1}^{(1)} + \sum_{k_2} \alpha_{k_1,k_2}^{(1)} q_{k_2}^{(1)})$.

After processing both CT scans, we compute the similarity matrix $S \in \mathbb{R}^{N_1 \times N_2}$ between all node pairs using cosine similarity. The SEAG module is optimized using binary cross-entropy loss.

2.3 Specificity-Correlation Balancing Module (SCB)

The SCB module enables cross-CT matching through cross-attention mechanism, performing multi-view fusion and classification to balance specificity and correlation between lymph node pairs. The strategy operates through three integrated components to learn spatiotemporal similarity: discriminative view selection, cross-modal feature fusion, and threshold similarity evaluation.

For discriminative view selection, a learnable gate network computes importance scores for each of the nine 2D views. The top-2 views with highest scores are selected to reduce computational complexity while maintaining discriminative power. For each lymph node pair (k_1, k_2) from different time points (T_i, T_j), we consider their 2D features f_{T_i,k_1,p_1}^{2D}, f_{T_j,k_2,p_2}^{2D} and 3D features f_{T_i,k_1}^{3D}, f_{T_j,k_2}^{3D} extracted from the SE module. Since each lymph node has one 3D feature and two selected 2D features, we generate nine combined features through all possible feature combinations between the two nodes: $f_{k_1,k_2}^{(p_1,p_2)} = \text{Fuse}(f_{T_i,k_1,p_1}^{2D}, f_{T_j,k_2,p_2}^{2D}, f_{T_i,k_1}^{3D}, f_{T_j,k_2}^{3D})$, where $p_1, p_2 \in \{0, 1, 2\}$ represent the feature types (0 for 3D, 1,2 for selected 2D projections) for nodes k_1 and k_2 respectively. Each fused feature undergoes nonlinear similarity assessment through a threshold scoring mechanism:

$$s_{k_1,k_2}^{(p_1,p_2)} = \sigma(W f_{k_1,k_2}^{(p_1,p_2)} + b) \cdot \mathbb{I}_{[\sigma(W f_{k_1,k_2}^{(p_1,p_2)} + b) \geq 0.5]}, \tag{3}$$

where the indicator function $\mathbb{I}$ suppresses low-confidence matches by nullifying scores below 0.5. The final similarity metric aggregates evidence across all nine feature combinations: $s_{k_1,k_2}^{\text{final}} = \frac{1}{9} \sum_{p_1=0}^{2} \sum_{p_2=0}^{2} s_{k_1,k_2}^{(p_1,p_2)}$. Nodes with $s_{k_1,k_2}^{\text{final}} <$

0.5 are classified as new/disappear instances. The model is optimized using binary cross-entropy loss.

2.4 Training Strategy and Inference Formulation

The model is trained end-to-end in a joint manner. At each training step, the input data is first processed by the SE module to extract features, which are then fed into the SEAG and SCB modules. The total loss combines multiple components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{se}} + \mathcal{L}_{\text{te}} + \mathcal{L}_{\text{seag}} + \mathcal{L}_{\text{scb}}. \tag{4}$$

During inference, all modules contribute to the final matching results. The SE module performs feature extraction for each lymph node. The SEAG module computes pairwise similarities between lymph nodes across CT scans and applies the Hungarian algorithm for initial matching. The SCB module independently computes its own similarity scores and uses these scores to correct the initial matching assignments from SEAG, particularly for identifying new/disappear instances and resolving mismatches.

3 Experiments

3.1 Experimental Settings

Dataset Our study utilized a combined dataset from two cohorts, totaling 167 patients. The first cohort included 120 colorectal cancer patients (83 with 3 scans, 37 with 2 scans), providing 1,339 matched node pairs and 101 unmatched nodes. The second cohort consisted of 47 patients with recurrence (2 scans each, acquired within 6 months post-surgery), contributing an additional 356 matched pairs and 55 unmatched nodes. Negative nodes were defined by size consistency of less than 10 mm across scans, while positive nodes exhibited more typical non-rigid deformations, with a maximum observed size of 49.62 mm. To evaluate performance, we employed a five-fold cross-validation scheme on the entire 167-patient dataset.

Implementation Details All scans were standardized to $0.7 \times 0.7 \times 2$ mm voxel spacing, with nodes cropped as $24 \times 24 \times 24$ volumes (50% expanded radii around radiologist-annotated centers). The model was implemented in PyTorch and trained on 2 NVIDIA 4090 GPUs using Adam optimizer (lr=0.0001, batch=4) for 100 epochs. Training took approximately 10 hours, with early stopping applied when validation loss plateaued for 20 epochs.

Evaluation Metrics In addition to standard classification metrics Precision (Pre) and Recall (Rec), the metrics involved in this paper include:

- **New/Disappear Lymph Node Rate (NLR/DLR)**: Specifically designed to evaluate asymmetric matching challenges, this metric measures the accuracy of identifying nodes that new/disappear between scans: Rate $= \frac{|S_{\text{correct}}|}{|S_{\text{gt}}|}$, where $|S_{\text{correct}}|$ and $|S_{\text{gt}}|$ are the correctly identified and ground-truth counts for new/disappear nodes respectively.
- **Center Point Matching (CPM) Accuracy**: The proportion of correctly matched nodes where the distance between the predicted center ($pred$) and ground-truth center (gt) is within a tolerance, defined as: $\|pred - gt\|_2 < \min(10\,\text{mm}, \text{radius})$.
- **Mean Euclidean Distance (MED)**: The average Euclidean distance between the centers of predicted and ground-truth matched nodes, calculated as $\frac{1}{N}\sum_{i=1}^{N}\|pred_i - gt_i\|_2$.

3.2 Ablation Study Results and Discussion

The ablation study in Tab. 1 demonstrates progressive performance improvements as each module is added. The baseline model achieves only 75.68% precision, highlighting the fundamental challenges of asymmetric matching. The non-monotonic trends in NLR and DLR reflect inherent balance challenges between detecting new/disappear nodes and maintaining matching accuracy.

Table 1. Ablation Study Results of the Asymmetric Matching Framework.

SEAG	**TE**	**SE**	**SCB**	**Pre ↑**	**Rec ↑**	**NLR ↑**	**DLR ↑**
				75.68 ± 5.62	74.92 ± 4.71	33.33 ± 11.20	42.86 ± 14.98
✓				96.63 ± 1.46	96.51 ± 1.49	51.35 ± 16.46	78.38 ± 7.92
✓	✓			97.32 ± 1.10	97.19 ± 1.02	67.10 ± 7.92	**81.57 ± 9.49**
✓	✓	✓		97.82 ± 1.01	97.69 ± 1.13	**71.88 ± 5.02**	75.81 ± 12.31
✓	✓	✓	✓	**97.99 ± 0.89**	**97.82 ± 1.08**	70.88 ± 7.74	79.24 ± 11.44

In heatmap Fig. 2, the baseline model exhibits unstable attention due to anatomical ambiguity. The SE module enhances central node stability but increases sensitivity to non-rigid deformations. The correlation extraction module improves spatial guidance yet over-prioritizes environmental context. The SCB module addresses these limitations, achieving 97.99% precision while stabilizing attention patterns evident in cross-temporal heatmaps. However, challenges persist in context-scarce scenarios where nodes lack sufficient spatial context.

3.3 Comparative Experiment Results and Discussion

Conventional MOT tasks track rigid objects in videos, while our study targets dynamic 3D LNs in CT with non-rigid deformations. We adapt a graph neural network method in MOT [22] for our dataset, but as shown in Tab. 2, it performs less effectively on medical images due to domain differences.

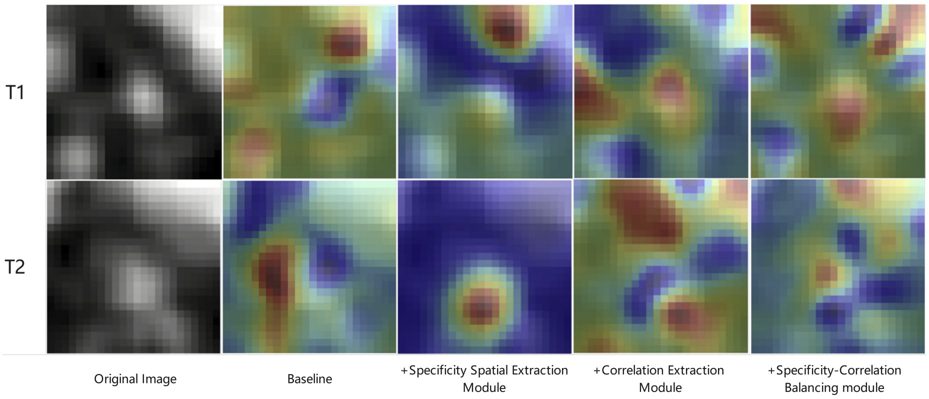

Fig. 2. Heatmaps track non-rigid deformations in a patient's LN across scans.

Table 2. Comparison with Multiple Object Tracking (MOT) Method.

Method	**Pre** ↑	**Rec** ↑	**NLR** ↑	**DLR** ↑
LGM[22]	82.26 ± 1.50	82.19 ± 1.44	52.81 ± 3.07	56.08± 14.74
Ours	**97.99 ± 0.89**	**97.82 ± 1.08**	**70.88 ± 7.74**	**79.24 ± 11.44**

Table 3. Comparison of Generalized Lesion Tracking Methods.

Method	**CPM@Radius** ↑	**MED (mm)** ↓
TLT[17]	49.54	15.98 ± 21.63
SAM[18]	9.67	39.79 ± 38.77
UAE[19]	13.89	25.65 ± 17.3

Asymmetric lymph node matching represents a fundamentally distinct task from conventional center-point tracking. The results in Tab. 3 demonstrate that generalized lesion tracking methods, while effective for their intended applications, are not well-suited for our asymmetric matching problem in abdominal lymph nodes. We identified several limitations in model performance. The small size of LNs resulted in minimal error tolerance, leading to lower prediction accuracy for positive LNs compared to negative cases, particularly due to non-rigid deformations. Additionally, the model faces challenges in tracking the new and dsappear LNs. Furthermore, two common failure scenarios were observed: (1) reduced detection sensitivity when LNs shrink in size, and (2) increased false positive predictions in regions with densely clustered LNs.

4 Conclusion

This study formulates the asymmetric lymph node matching task and proposes an framework to address the challenges of abdominal LN. Abdominal LNs are naturally small and located in a complex anatomical environment. Furthermore, they may experience changes in size, shape, and number across follow-up CT scans. We define this scenario in medical target tracking as an asymmetric matching problem. The solution we propose not only addresses this issue but also offers an approach for matching other similar small targets. While the framework demonstrates matching capabilities, it is still constrained by its dependence on pairwise CT scan annotations. These annotations require significant clinical effort and may suffer from inconsistencies in coordinating longitudinal data. As a result, our methodology currently operates at a semi-automated level and cannot yet fully eliminate the need for physician involvement. Current efforts are focused on developing integrated detection-matching solutions to improve clinical applicability. The established framework lays a crucial foundation for advancing annotation-efficient tracking methods in medical image analysis.

Acknowledgments. We thank for the support of National Natural Science Foundation of China under Grant 62476183 and Natural Science Foundation from the Science & Technology Department of Sichuan Province under Grant 2024NSFTD0051.

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

References

1. N. Hoshino, K. Murakami, K. Hida, T. Sakamoto, and Y. Sakai, “Diagnostic accuracy of magnetic resonance imaging and computed tomography for lateral lymph node metastasis in rectal cancer: a systematic review and meta-analysis,” *International Journal of Clinical Oncology*, vol. 24, pp. 46–52, 2019.
2. K. Ichimasa, S.-e. Kudo, H. Miyachi, Y. Kouyama, K. Mochizuki, Y. Takashina, Y. Maeda, Y. Mori, T. Kudo, Y. Miyata *et al.*, “Current problems and perspectives of pathological risk factors for lymph node metastasis in t1 colorectal cancer: systematic review,” *Digestive Endoscopy*, vol. 34, no. 5, pp. 901–912, 2022.
3. Rebecca L. Siegel, Nikita Sandeep Wagle, Andrea Cercek, Robert A. Smith, and Ahmedin Jemal, "Colorectal cancer statistics, 2023," *CA: a cancer journal for clinicians*, vol. 73, no. 3, pp. 233–254, 2023.
4. Jian Hu, "Postoperative Follow-up Should Not Be Neglected," *Zhongguo fei ai za zhi= Chinese Journal of Lung Cancer*, vol. 21, no. 3, pp. 205–205, 2018.
5. Y. Wang, "Imaging Evaluation of Lateral Lymph Node Metastasis in Rectal Cancer," in *Chinese Journal of Practical Surgery*, vol. 38, no. 10, pp. 1132–1136, 2018.
6. Han Wang, Hao Huang, Jingling Wang, Mingtian Wei, Zhang Yi, Ziqiang Wang, and Haixian Zhang, "An intelligent system of pelvic lymph node detection," *International Journal of Intelligent Systems*, vol. 36, no. 8, pp. 4088–4116, 2021.
7. Zhen Pan, Shuo Huang, Han Wang, Mingtian Wei, Junjie Cui, Ziqiang Wang, and Haixian Zhang, "Multi-context 3D ResNet for small-size false positive reduction in pelvic lymph node detection," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1773–1780, 2021.
8. Yi Zhang, Jiayue Li, Xinyang Li, Min Xie, Md Tauhidul Islam, and Haixian Zhang, "FAOT-Net: A 1.5-Stage Framework for 3D Pelvic Lymph Node Detection With Online Candidate Tuning," *IEEE Transactions on Medical Imaging*, 2023.

9. Han Wang, Fasheng Yi, Jingling Wang, Zhang Yi, and Haixian Zhang, "RECIST-Sup: Weakly-supervised lesion volume segmentation using RECIST measurement," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1849–1861, 2022.
10. Min Xie, Yi Zhang, Xinyang Li, Jiayue Li, Xingyu Zou, Yiji Mao, and Haixian Zhang, "An Intelligent System of Predicting Lymph Node Metastasis in Colorectal Cancer Using 3D CT Scans," *International Journal of Intelligent Systems*, vol. 2024, no. 1, pp. 7629441, 2024.
11. Kasper Marstal, Floris Berendsen, Marius Staring, and Stefan Klein, "SimpleElastix: A user-friendly, multi-lingual library for medical image registration," *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 134–142, 2016.
12. Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.
13. Xavier Rafael-Palou, Anton Aubanell, Ilaria Bonavita, Mario Ceresa, Gemma Piella, Vicent Ribas, and Miguel A González Ballester, "Re-identification and growth detection of pulmonary nodules without image registration using 3D Siamese neural networks," *Medical Image Analysis*, vol. 67, pp. 101823, 2021.
14. A. Gomariz, W. Li, E. Ozkan, et al., "Siamese networks with location prior for landmark tracking in liver ultrasound sequences," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 1757-1760.
15. Benjamin Billot, Douglas N. Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V. Dalca, Juan Eugenio Iglesias, et al., "SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining," *Medical Image Analysis*, vol. 86, pp. 102789, 2023, Elsevier.
16. Jinzheng Cai, Youbao Tang, Ke Yan, Adam P. Harrison, Jing Xiao, Gigin Lin, and Le Lu, "Deep lesion tracker: monitoring lesions in 4D longitudinal imaging studies," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15159–15169, 2021.
17. Wen Tang, Han Kang, Haoyue Zhang, Pengxin Yu, Corey W. Arnold, and Rongguo Zhang, "Transformer lesion tracker," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 196–206, 2022, Springer.
18. Ke Yan, Jinzheng Cai, Dakai Jin, Shun Miao, Dazhou Guo, Adam P. Harrison, Youbao Tang, Jing Xiao, Jingjing Lu, and Le Lu, "SAM: Self-supervised learning of pixel-wise anatomical embeddings in radiological images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2658–2669, 2022.
19. X. Bai, F. Bai, X. Huo, J. Ge, J. Lu, X. Ye, K. Yan, and Y. Xia, "UAE: Universal Anatomical Embedding on Multi-Modality Medical Images," 2023.
20. A. Vizitiu, A. T. Mohaiu, I. M. Popdan, et al., "Multi-scale Self-Supervised Learning for Longitudinal Lesion Tracking with Optional Supervision," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer Nature Switzerland, 2023, pp. 573-582.
21. P. Zhai, H. Cong, E. Zhu, G. Zhao, Y. Yu, and J. Li, "MVCNet: Multiview Contrastive Network for Unsupervised Representation Learning for 3-D CT Lesions," in *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
22. J. He, Z. Huang, N. Wang, and Z. Zhang, "Learnable Graph Matching: Incorporating Graph Partitioning with Deep Feature Learning for Multiple Object Tracking," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 526–535, June 2021. doi: https://doi.org/10.1109/cvpr46437.2021.0052610.1109/cvpr46437.2021.00526.