International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-023-02954-7

ORIGINAL ARTICLE

Universal detection and segmentation of lymph nodes in multi-parametric MRI

Tejas Sudharshan Mathai[1] · Sungwon Lee[1] · Thomas C. Shen[1] · Daniel Elton[1] · Zhiyong Lu[2] · Ronald M. Summers[1]

Received: 16 January 2023 / Accepted: 5 May 2023

Abstract

Purpose Reliable measurement of lymph nodes (LNs) in multi-parametric MRI (mpMRI) studies of the body plays a major role in the assessment of lymphadenopathy and staging of metastatic disease. Previous approaches do not adequately exploit the complementary sequences in mpMRI to universally detect and segment lymph nodes, and they have shown fairly limited performance.

Methods We propose a computer-aided detection and segmentation pipeline to leverage the T2 fat-suppressed (T2FS) and diffusion-weighted imaging (DWI) series from a mpMRI study. The T2FS and DWI series in 38 studies (38 patients) were co-registered and blended together using a selective data augmentation technique, such that traits of both series were visible in the same volume. A mask RCNN model was subsequently trained for universal detection and segmentation of 3D LNs.

Results Experiments on 18 test mpMRI studies revealed that the proposed pipeline achieved a precision of $\sim 58\%$, sensitivity of $\sim 78\%$ at 4 false positives (FP) per volume, and dice score of $\sim 81\%$. This represented an improvement of $\geq 12\%$ in precision, $\geq 15\%$ in sensitivity at 4 FP/volume, and $\geq 14\%$ in dice score, respectively, over current approaches evaluated on the same dataset.

Conclusion Our pipeline universally detected and segmented both metastatic and non-metastatic nodes in mpMRI studies. At test time, the input data used by the trained model could either be the T2FS series alone or a blend of co-registered T2FS and DWI series. Contrary to prior work, this eliminated the reliance on both the T2FS and DWI series in a mpMRI study.

Keywords MRI · T2 · Lymph node · Detection · Segmentation deep learning

Introduction

Lymph nodes (LNs) are small structures scattered throughout the body and are a part of the lymphatic system. Lymphocytes (immune cells) in LNs remove foreign material in the body. Patients with an abnormal rise in lymphocytes have swollen and enlarged nodes (lymphadenopathy), and this is typically due to infection, autoimmune disease, or malignancy. In current clinical practice, it is important to distinguish enlarged and metastatic nodes from the non-metastatic LNs [1, 2]. Radiologists identify suspicious nodes through nodal size measurement with the help of established guidelines, such as the tumor, node, and metastasis (TNM) criteria [1]; this helps them manage the therapeutic pathway for patients. Among the various radiological imaging modalities to visualize the LNs, multi-parametric MRI (mpMRI) is usually preferred due to the superior soft tissue resolution and improved contrast between fat and water [2]. Multiple sequences are generally acquired during a mpMRI study including T2-weighted fat suppression (T2FS) and diffusion-weighted imaging (DWI) series among others. A node is considered enlarged if its smallest diameter (along the short axis) is greater than 10 mm on an axial MRI slice.

Radiologists routinely locate and manually measure the nodal size on T2FS series and often refer to different sequences (DWI) for confirmation. This process can be cumbersome and time-consuming during a busy clinical workday. Moreover, LNs can straddle major anatomical structures (e.g., liver, bowel, blood vessels) while having diverse shapes and non-homogeneous appearances, and a variety of MRI

✉ Tejas Sudharshan Mathai
tejas.mathai@nih.gov

[1] National Institutes of Health (NIH) Clinical Center, Bethesda, MD, USA

[2] National Library of Medicine, NIH, Bethesda, MD, USA

Published online: 16 June 2023

imaging scanners from different manufacturers are also used by institutions across the world; both these factors exacerbate the workload for radiologists. To relieve the tediousness, many automated LN measurement approaches have been proposed [3–9]. Some focus on detecting LNs in specific regions of body, such as the pelvis [4, 5] and rectum [3]. Other works [6–9] detect nodes only in T2FS MRI volumes.

Very few approaches exploit the complementary sequences in mpMRI studies [3, 4]. Both T2FS and DWI series were used by Zhao et al. [3]; a mask RCNN model [10] was trained for LN detection and segmentation. The authors investigated various combinations of T2FS and DWI slices (e.g., 2 T2FS slices + 1 DWI slice) as input. In [4], a faster RCNN model was trained on T2FS and DWI slices that were *not* co-registered. It is important to note that diffusion sequences may not always be acquired in practice, and thus, it may not be ideal to rely heavily on their presence while developing an approach to detect LNs in mpMRI. Therefore, an automated computer-aided detection and segmentation pipeline to universally localize nodes in mpMRI studies of the body while accounting for the workflow-based issues is highly desirable.

In this paper, we tackled these imaging and workflow challenges and present a pipeline to universally detect and segment LNs in mpMRI studies of the body. Figure 1 shows an overview of the pipeline. The mpMRI studies were acquired at our institution using various MR scanners (Siemens, GE, Philips) and a variety of exam protocols. The T2FS and DWI series in a study were co-registered and then linearly interpolated together to create a blended volume using a selective data augmentation technique. The blended volume contained traits of both series, such as fat suppression and diffusion restriction. The blending interpolation factor was drawn from a beta distribution. A mask RCNN model was trained on the blended volumes to universally detect and segment LNs. During the testing phase, data presented to the trained model could come from either the T2FS series alone or from blending any available T2FS and DWI series that were co-registered. In this manner, the model did not rely on the presence of both T2FS and DWI series in the mpMRI study. In contrast to prior work [3, 8], we used full-size inputs for evaluation, and achieved a mean average precision (mAP) of $\sim 58\%$, sensitivity of $\geq 78\%$ at 4 FP/volume, and a dice score of $\sim 81\%$. The use of mpMRI increased the sensitivity at 4 FP/volume by $\sim 18\%$ when compared to only using the T2FS series as in prior works [3, 8]. Compared to previous work [3, 8, 9], we simultaneously detect and segment LNs in mpMRI studies.

Methods

Data The Picture Archiving and Communication System (PACS) at our institution was queried for patients who had undergone MRI imaging between January 2015 and September 2019. Initially, a total of 383 patients (224 males and 159 females with ages between 6 and 85 years) and 500 mpMRI studies were identified. The radiology report associated with a study was obtained, and a natural language processing algorithm [11] extracted the presence of metastatic and/or non-metastatic LNs, extent, and size measurements. Each study contained various series such as T2-weighted (T2WI) series, T2 fat-suppressed (T2FS) series, diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps. However, the studies did not always contain DWI and ADC series, and they were acquired using a variety of MR scanners (GE, Philips, Siemens) and exam protocols. At our institution, radiologists sized the LNs by scrolling back and forth across the slices in the T2FS series, matched the appearance of suspicious nodes in the DWI series, and measured the largest LN extent present on a single 2D slice in the T2FS series according to the routine clinical protocol for measuring LNs. LNs were measured with either the long-axis diameter (LAD) or short-axis diameter (SAD), or both simultaneously. As it is cumbersome for radiologists to measure the full 3D extent of suspicious LNs during a busy clinical day, the primary measurement of LAD and SAD was prospectively made only on a single 2D slice. If only a single measurement (LAD or SAD) was done, a radiologist conducted a quality check to ensure that both LAD and SAD measurements were available.

Next, 62 mpMRI studies from 55 different patients (34 males, 21 females, aged between 9 and 80 years) were used in this work. They included 39 chest, abdomen and pelvis studies, and 23 abdomen and pelvis studies. One radiologist manually segmented the full 3D extent of LNs in the T2FS series. Since LNs with a SAD ≥ 8 mm should be mainly considered as suspicious for metastasis [2], we used this reported range for detecting and segmenting LNs. In these studies, there were often multiple DWI sequences (minimum 1, maximum 3) acquired with low (0–200 s/mm^2), intermediate (400–800 s/mm^2), and high (800–1400 s/mm^2) *b*-values. For our work, we exploited all the available DWI sequences with different b-values. As the LNs in only the T2FS series were fully segmented, the DWI series were co-registered to T2FS to transfer the LN annotations using an Insight Toolkit (ITK)-based rigid registration algorithm [12]. The studies were randomly divided on a patient-level into $\sim$ 70% train (38 patients, 38 studies), $\sim$ 10% validation (6 patients, 6 studies), and $\sim$ 20% test (11 patients, 18 studies) splits. Fivefold cross-validation was conducted with the train and validation sets, and the test set was held out for evaluation. N4 bias normalization [13] was subsequently performed on the registered sequences, followed by normalization to [1%, 99%] of the voxel intensity range [14], and histogram equalization [15] to boost the contrast between bright and dark structures in the volumes. The resulting series had various dimensions

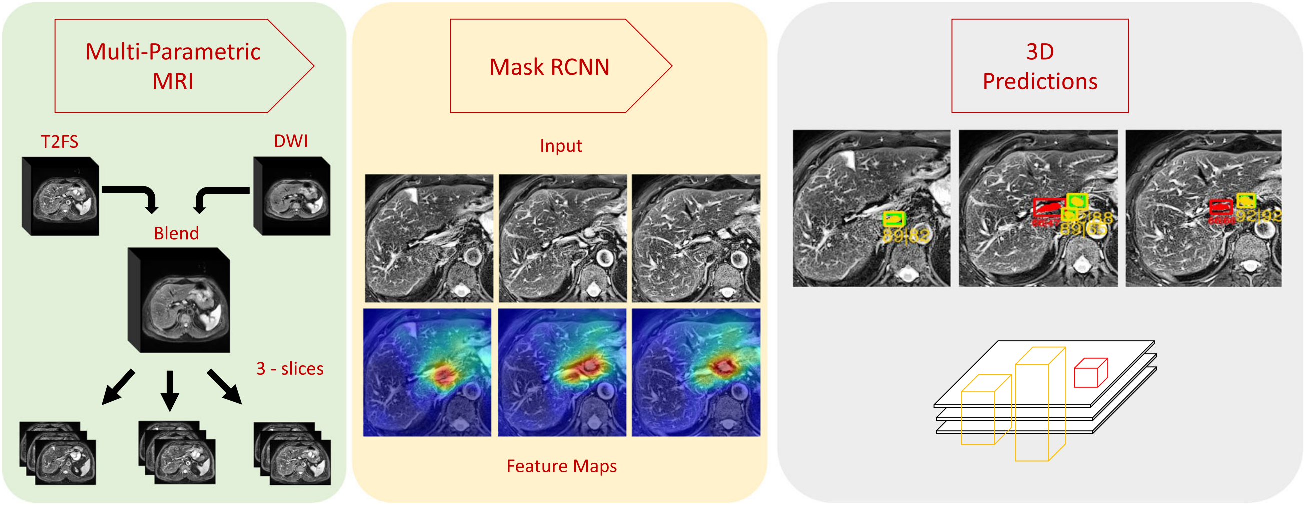


Fig. 1 Flowchart of the proposed computer-aided detection (CAD) pipeline. First, T2FS and DWI series in an mpMRI study were co-registered and then blended together. Next, three consecutive slices from the resulting volume were collated to form a three-channel image. The images were then fed to mask RCNN to detect and segment potential LNs in each slice. Green boxes: ground truth, yellow: true positives, red: false positives. The 2D LN candidates were then merged into 3D based on their confidence scores as well as their IoU overlaps with boxes in adjacent slices. The text below each detected box, e.g., "89|82", describes the highest confidence score across all elements of the 3D prediction followed by the confidence score of the candidate box detected in the current slice. The figure is best viewed in color in the PDF

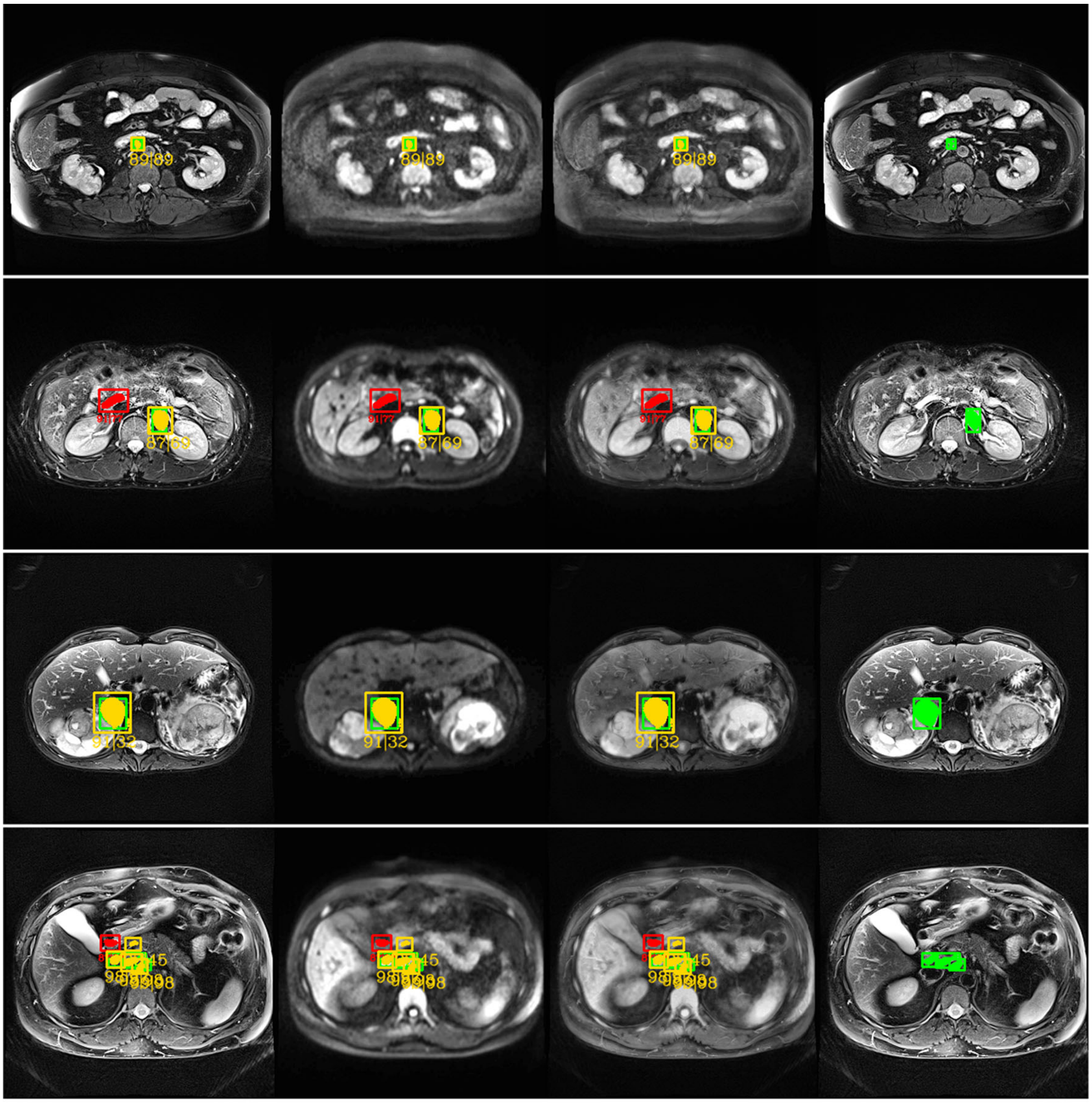
in the range of $(256 \sim 640) \times (192 \sim 640) \times (18 \sim 60)$ voxels.

Selective augmentation In prior works [3, 4], the presence of both the T2FS and DWI sequences was required for their approaches to work. However, the reliance on both sequences cannot always be guaranteed as a complete MRI workup is not always deemed necessary by the referring physician or the radiologist, and thus certain sequences may not be acquired. To exploit the available sequences in an mpMRI study, we used a recently proposed method by Yao et al. [16] to learn invariant representations via selective augmentation (LISA). LISA interpolated the training data samples which have the same label, but were sampled from different domains (T2FS and DWI). Co-registration of the T2FS and DWI series in our work enabled the labels (bounding boxes and masks) from the T2FS series to be transferred to the DWI series. Crucially, we used Intra-Label LISA (ILL) to blended the T2FS and DWI sequences together, such that traits of both sequences (e.g., diffusion restriction and fat suppression) are visible in the same volume. The blending is rooted in the MixUp [17] and CutMix [18] techniques, which linearly interpolate training data samples and remove any correlations [19] between the domain and labels. Through our simple trick, the mask RCNN model learned invariant predictors for LNs. Formally, we assume that two data samples (x_i, y_i, d_i) and (x_j, y_j, d_j) are drawn from two distinct domains d_i and d_j. The two samples can be linearly interpolated according to:

$$x_m = \lambda x_i + (1-\lambda)x_j \quad \text{and} \quad y_m = \lambda y_i + (1-\lambda)y_j \qquad (1)$$

$$\hat{\theta} := \underset{\theta \in \Theta}{\operatorname{argmin}}\ \mathbb{E}_{\{(x_i,y_i,d_i),(x_j,y_j,d_j)\sim \hat{P}\}}\left[l(f_\theta(x_m), y_m)\right] \qquad (2)$$

where $\lambda \in [0, 1]$ is the interpolation ratio sampled from a beta distribution $Beta(\alpha, \beta)$ and dictates the strength of volumetric blending. Since the label (bounding box and mask) is the same $y_i = y_j$ for the co-registered T2FS and DWI sequences, interpolation of the data samples results in volumes where characteristics of both domains are partially present and any spurious correlations that exist between the domains and labels are removed. Examples are shown in Figs. 1 and 2. This gives rise to an empirical risk minimization setting as in Eq. 2 where given a training distribution P_{tr}, a loss function l is used to train a model f_θ to optimize its parameters $\theta \in \Theta$. As the parameters of the beta distribution govern the blending ratio λ, it permitted the use of either the T2FS sequence alone or a combination of T2FS and DWI series. If a study did not contain the DWI series, then ILL was not applied. The model encountered diverse examples during training to enhance robustness against noise at the test time. Experiments conducted in "Experiments and results" section attest to the advantage of using our elegant selective augmentation approach for LN detection. For training the model, 2.5D images were extracted from the blended volume and each image contained three consecutive slices from the volume with the annotated LNs present in the slice in the middle.



(a) T2FS (b) DWI (c) Blend (d) GT

Fig. 2 LN detection and segmentation results for different slices obtained from distinct mpMRI studies used in this work. The detections and segmentations are shown for a small LN (8 mm), a medium sized LN (1.2 cm), and a large LN (2.3 cm). Green—ground truth, red—false positives, and yellow—true positives. The text below each detected box, e.g., "91|32", describes the highest confidence score across all elements of the 3D prediction followed by the confidence score of the candidate box detected in the current slice

Mask RCNN The standard design of the mask RCNN [10] framework consists of a backbone network, Feature Pyramid Network (FPN) [20], Region Proposal Network (RPN) [21], and a network head. Traditionally, the backbone has been based on ResNet models [22] (e.g., ResNet-50 or ResNet-101), but in our work we replaced it with the general-purpose Swin Transformer backbone [23]. The rationale behind this was due to the hierarchical feature maps computed by the Swin transformer; they have the same dimensions as those obtained from standard backbones, e.g., ResNet. These hierarchical feature maps were estimated by first splitting the input image into small-sized image patches in shallower layers and merging neighboring patches in deeper layers. A fixed number of image patches were then taken to constitute a window, within which self-attention was computed locally. Shifted window partitioning also introduced connections across neighboring windows that increased the representation modeling power while also maintaining linear computation complexity. For more details on the implementation, we refer the reader to [23]. Through experiments described in "Experiments and results" section, we show that the mask RCNN model was able to achieve higher precision and sensitivities at different FP with the Swin transformer backbone. The FPN of mask RCNN was used to detect LNs appearing at multiple scales, and the RPN shared the extracted features of the full slice with the head network, which obtained effective target LN candidates for detection and segmentation. Moreover, as mask RCNN can generate numerous false positives (FP), Hard Negative Example Mining [24] was employed to reduce the FP number and boost performance. After the model had been trained, Weighted Boxes Fusion [25] was used to combine the various predictions from the checkpoint with the lowest validation loss during each cross-validation run of the mask RCNN model.

3D proposal generation The 2D proposals generated by mask RCNN in every slice of the blended volume were later post-processed into 3D predictions. We followed the Kalman filter-based bounding box tracking approach [26, 27] to obtain the 3D proposals. The 2D predictions were first filtered to keep boxes with scores $\geq$ 10% and eliminate those with lower scores. Next, we stacked the 2D predictions together from pairs of adjacent slices when their IoU score was $\geq$ 25% to create the 3D proposals; we chose this threshold as it accounted for large variations in voxel sizes (especially along the z-axis) for volumes acquired by different scanners. Finally, we filtered the clusters based on the maximum confidence score available in that cluster and removed those that did not cross a confidence threshold of 30%. We chose this value in order to keep the number of 3D predictions manageable.

Experiments and results

Baseline comparisons We coined our main experiment for LN detection and segmentation E_{SA}; we used selective augmentation to create a blended volume with co-registered T2FS and DWI series based on $Beta(60, 10)$, and a Swin transformer backbone. In our first experiment, we compared our results against those obtained by Wang et al. [8], wherein a mask RCNN model for universal LN detection and segmentation was trained with only T2FS volumes. For the next series of experiments, we contrasted our results against the various data combinations proposed by Zhao et al. [3]. These included: (1) 3-slices of only T2FS (E_T), (2) 3-slices of only DWI (E_D), (3) 1-slice of T2FS and 2-slices of DWI (E_{12}), and (4) 2-slices of T2FS and 1-slice of DWI (E_{21}). A mask RCNN model was also used by Zhao et al. [3]. We also compared the *detection-only* performance for the ensemble-based approach proposed by Mathai et al. [9] for T2FS-only data; segmentation dice scores were unavailable as this approach purely detected LNs in T2FS only. Furthermore, we assessed one of the main contributions in our work: selective data augmentation through blending. As the Beta distribution governed the blending process, we evaluated the effect of the choice of its parameters. These included $Beta(\alpha, \beta)$ with: (1) $Beta(2, 2)$, (2) $Beta(1, 1)$, (3) $Beta(4, 4)$, and (4) $Beta(60, 10)$, which heavily favored the T2FS series. Next, we also evaluated the performance with the use of a ResNet-50 backbone instead of the Swin transformer backbone. Finally, we compared the performance of our main experiment E_{SA} against a baseline approach (Swin transformer backbone) trained and tested only on T2FS series.

Metrics Prior approaches [3, 8] operated on 2D measurements of LAD and SAD, and they reported their results based on these 2D measurements. This did not reflect the true performance of a LN detection and segmentation model. Any correct predictions made by the network on any unmeasured node in the same slice or the adjacent slice(s) would be counted as false positives when they should actually be counted as true positive predictions instead. Furthermore, we believe that automated methods should process the entire volume and report results on a volumetric level as opposed to a slice-based level. These results will reflect the true nature of the nodal detection performance, and to that end, we adopt the pseudo-3D (P3D) IoU metric proposed in [27]. Specifically, we denote the slice containing the 2D annotation by the radiologist as z and the corresponding bounding box that resulted from that measurement as (x_1, x_2, y_1, y_2, z, z). We represent a 3D prediction by (x_1^*, x_2^*, y_1^*, y_2^*, z_1^*, z_2^*). The P3D IoU metric assigns a 3D prediction as a true positive if and only if $z_1^* \leq z \leq z_2^*$ and the $\mathrm{IoU}\big[(x_1, x_2, y_1, y_2), (x_1^*, x_2^*, y_1^*, y_2^*)\big] \geq 50\%$. Otherwise, the 3D prediction was a false positive. For more information regarding the P3D IoU metric, we refer the reader to [27]. We also used mean average precision (mAP) and the dice similarity coefficient (DSC) score to quantify the detection and segmentation performance, respectively.

Implementation details For consistent comparison across all works, we did not crop our slices and used the full-sized images as training inputs. Training of the mask RCNN model was accomplished with the mmDetection framework [28]. Outside of selective augmentation, standard data augmentation strategies were also done (e.g., random horizontal flips, crops, rotations etc.). Pre-trained weights were used for all models to speed up convergence during training. A grid search run across training hyper-parameters yielded optimal values for the learning rate (1e−5), optimizer (AdamW), batch size (2), and number of iterations (36). Fivefold cross-validation was performed for each comparative method. For each cross-validation fold, the checkpoint with lowest validation loss was chosen for testing. An ensemble of the 5 checkpoints from the 5 runs was used during prediction/test phase to locate and segment the LNs. All experiments were run on a workstation running Ubuntu 18.04LTS with 4 NVIDIA Tesla V100 GPUs. Evaluation was always performed at an IoU threshold of 25% to be consistent with prior work [8].

Results Based on prior work by [3, 8, 9], a clinically acceptable result for LN detection meant a sensitivity of 65% at 4–6 FP per volume. Table 1 summarizes the detection and segmentation performance for the mask RCNN model across all experiments. Our main experiment E_{SA} (row 6: selective augmentation with $Beta(60, 10)$ and a Swin Transformer backbone) achieved the best LN detection and segmentation performance of $\sim$ 58% mAP, $\sim$ 78% sensitivity at 4 FP/volume, and dice score of ~81% over other approaches. The experiment E_T that used T2FS-only provided a sensitivities of $\sim$ 52% and $\sim$ 56% at 4 FP/vol for [8] and [3], respectively. The experiment E_{D} that used DWI-only alone yielded low LN detection sensitivities of $\sim$ 48% at 4 FP/vol. These results are not surprising as the tissue structures in DWI series appear diffused with poor spatial resolution in contrast to the T2FS series.

Zhao et al. [3] found that their experiment E_{12} with a data combination of 1 T2FS slice and 2 DWI slices worked best. Contrary to their findings, we observed that the experiment E_{21} with the data combination of 2 T2FS slices and 1 DWI slice generally performed better than E_{12}, E_{T} and E_{D}. This may be due to the images being cropped around the rectum in their work [3], thereby limiting the contextual information available to the network from the provision of full-sized input images. Comparing the *detection-only* performance against the ensemble method proposed by [9], we see that our precision and sensitivity at 4 FP/vol improved by $\sim$ 4% and $\sim$ 3%, respectively. Overall, our main results with the selective augmentation experiment E_{SA} showed a marked improvement in LN detection and segmentation sensitivities.

Moreover, Table 2 shows the comparative results where we used our trained mask RCNN model to predict on solely T2FS input data (row 2). We noticed that the results were similar in comparison to using blending with both T2FS and DWI series. These results support our idea that the model could be trained on studies containing both T2FS and DWI series, but it did not require the DWI series to be present at test time. Comparing our results against the baseline mask RCNN (Swin transformer backbone) trained and test on T2FS-only data (row 3), we found that the precision, sensitivity at 4 FP/vol, and dice score improved by $\sim$ 5%, $\sim$ 7% and $\sim$ 10%, respectively. Additionally, there was a decrease in mAP (−13%), sensitivity at 4 FP/vol (−18%), and dice score (−16%), respectively, with the use of the ResNet-50 backbone (row 4) instead of the Swin Transformer. Finally, experimenting with different beta distribution parameters (rows 5–7) yielded lower results for all tested parameters except for $Beta(60, 10)$; we believe that this is due to the probabilities being drawn from a distribution that heavily favored the T2FS series.

Discussion and conclusion

Our approach exploits complementary sequences in mpMRI to universally detect and segment lymph nodes, whereas previous approaches either did not use diffusion sequences [8, 9] and the detection and segmentation performance was not adequate for clinical needs [3]. In current practice, localization and measurement of LNs in mpMRI studies is a repetitive task that is routinely performed by radiologists. Universal detection and segmentation of LNs with an automated pipeline, such as the one proposed in our work, can speed up the nodal localization with SAD $\geq$ 8 mm as the ensuing measurements help to differentiate metastatic from non-metastatic nodes.

Patient studies used in this work were acquired with different imaging scanners and exam protocols, but the full mpMRI workup (including diffusion sequences) were not always acquired. To handle such scenarios, our pipeline was trained on mpMRI studies containing T2FS and DWI series, but it did not require the diffusion sequence to be available at test time. The T2FS and DWI series were co-registered, and selective data augmentation blended the two series together using an interpolation ratio λ that was drawn from a Beta distribution $Beta(60, 10)$. Blending the two series together promoted the use of complementary information available in both series, such as fat suppression and diffusion restriction. It also closely mimicked the current practice where radiologists referred to co-registered DWI sequences for confirmation of LN presence in the T2FS series.

Our work stands in comparison to prior works [3, 4], where various data combinations were necessary to achieve reasonable performance. Utilizing the different data combinations (e.g., E_{21}, E_{12}) proposed by Zhao et al. [3] was inelegant in contrast to the blending-based approach proposed in our work. We have also observed that the ensemble-based detection method proposed by Mathai et al. [9] holds promise; the ensemble of networks in their work detected LNs with reasonable improvements over [3, 8]. A similar ensemble-based approach for simultaneous detection and segmentation of LNs could potentially improve performance. Finally, training and testing on the mask RCNN (with Swin transformer backbone) on T2FS-only data yielded lower detection and segmentation performance in contrast to training with both T2FS + DWI series. Despite structures in the DWI series appearing diffused with poor spatial resolution, our results show that the inclusion of the DWI sequence does provide some additional supervision during training for the mask RCNN model.

However, our results do indicate some false positives shown by the red boxes in Figs. 1 and 2. Insufficient registration of the volumes is a potential reason for false positives as we only rigidly register the T2FS and DWI series to roughly align them and to have consistent spacing, origin,

Table 1 Performance comparison of approaches on the test set

#	Method	Exp	Mode	mAP	S@0.5	S@1	S@2	S@4	Dice
1	Wang 2022 (2D) [8]	E_T	T2FS only	43.6	17.8	23.3	34.2	51.7	57.4
2	Zhao 2020 (3D) [3]	E_T	T2FS only	39.5	23.9	31.5	46.5	56.2	54.3
3	Zhao 2020 (3D) [3]	E_D	DWI only	37.7	24.1	30.1	39.7	47.9	47.6
4	Zhao 2020 (3D) [3]	E_{12}	NSA	39.8	21.9	34.3	43.8	60.2	59.7
5	Zhao 2020 (3D) [3]	E_{21}	NSA	43.8	26.1	36.6	47.9	61.4	62.2
6	Mathai 2022 (3D) [9]	E_T	NSA	53.4	36.3	51.9	62.6	75.2	–
7	Mask RCNN (Ours)	E_{SA}	ILL	**57.7**	**39.7**	**53.4**	**64.4**	**78.1**	**81.2**

“Exp” stands for the abbreviation of the experiment name. “Mode” describes the {T2FS, DWI} data combination mode. “SA” and “NSA” indicate selective and no selective augmentation, respectively. “ILL” stands for Intra-Label LISA. “S” describes the sensitivity @[0.5, 1, 2, 4] FP/volume. Bold indicates best results

Table 2 Comparative experiments of the mask RCNN model

#	Method	Exp	mAP	S@0.5	S@1	S@2	S@4	Dice
1	Ours (Swin + SA)	E_{SA}	**57.7**	**39.7**	**53.4**	**64.4**	**78.1**	**81.2**
2	Swin + SA (test on T2FS only)	E_{SA}	57.1	38.2	50.7	63.1	76.3	80.5
3	Swin (train/test on T2FS only)	E_T	52.3	33.6	47.1	58.3	71.8	70.9
4	ResNet-50	E_{SA}	44.9	27.4	34.3	50.7	60.3	64.6
5	$Beta(1, 1)$	E_{SA}	52.5	35.6	41.1	47.9	69.9	72.3
6	$Beta(2, 2)$	E_{SA}	52.7	27.4	45.2	58.9	71.2	74.4
7	$Beta(4, 4)$	E_{SA}	51.9	34.5	45.2	60.3	73.9	71.6

“Exp” stands for the abbreviation of the experiment name with “SA” indicating selective augmentation. “S” describes the sensitivity @[0.5, 1, 2, 4] FP/volume. Bold indicates best results

and dimensions. Other factors include the similar intensity (iso-intensity) of the LN on high b-value DWI to surrounding structures, such as the bowel, and the overlap of LN with vessels that contributed to the partial volumetric averaging of such regions into the LN areas. However, when our results were contrasted against those results generated through different data combinations (e.g., E_{21} in “Experiments and results” section), we found our results to outperform current state-of-the-art methods. Additionally, the training data used in this work was limited; while selective data augmentation was conducted in this paper to increase data diversity, additional data should be acquired to account for robustness. For future work, we plan employ self-supervision in the model, whereby additional LNs are mined in the unannotated studies obtained from PACS and integrated into the training dataset to further improve the detection and segmentation performance.

Funding This work was supported by the Intramural Research Programs of the NIH National Library of Medicine and NIH Clinical Center (Project Number 1Z01 CL040004). We also thank Jaclyn Burge for the helpful comments and suggestions.

Declarations

Conflict of interest RMS receives royalties from iCAD, Philips, PingAn, ScanMed, and Translation Holdings. His lab received research support from PingAn. The authors have no additional conflicts of interest to declare.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this study, informed consent was not required.

References

1. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP (2017) The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J Clin 67(2):93–99
2. Taupitz M (2007) Imaging of lymph nodes—MRI and CT. Springer, Berlin, pp 321–329
3. Zhao X, Xie P, Wang M, Pickhardt PJ, Xia W, Xiong F, Zhang R, Xie Y, Jian J (2020) Deep learning based fully automated detection and segmentation of lymph nodes on multiparametric MRI for rectal cancer: a multicentre study. EBioMedicine 56:102780
4. Lu Y, Yu Q, Gao Y, Zhou Y, Liu G, Dong Q, Ma J, Ding L, Yao H, Zhang Z, Xiao G, An Q, Wang G, Xi J, Yuan W-T, Lian Y, Zhang D, Zhao C-G, Yao Q, Liu W, Zhou X, Liu S, Wu Q, Xu W, Zhang J, Wang D, Sun Z, Gao Y, Zhang X, Hu J, Zhang M, Wang G, Zheng X, Wang L, Zhao J, Yang S (2018) Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. Can Res 78(17):5135–5143

5. Debats OA, Litjens GJS, Huisman HJ (2019) Lymph node detection in MR lymphography: false positive reduction using multi-view convolutional neural networks. Peer J 7:e8052
6. Mathai TS, Lee S, Elton DC, Shen TC, Peng Y, Lu Z, Summers RM (2021) Detection of lymph nodes in T2 MRI using neural network ensembles. In: Lian C, Cao X, Rekik I, Xu X, Yan P (eds) Machine learning in medical imaging. Springer, Cham, pp 682–691
7. Mathai TS, Lee S, Elton DC, Shen TC, Peng Y, Lu Z, Summers RM (2021) Lymph node detection in T2 MRI with transformers. arXiv:2111.04885
8. Wang S, Zhu Y, Lee S, Elton DC, Shen TC, Tang Y, Peng Y, Lu Z, Summers RM (2022) Global-local attention network with multi-task uncertainty loss for abnormal lymph node detection in MR images. Med Image Anal 77:102345
9. Mathai TS, Lee S, Shen TC, Lu Z, Summers RM (2022) Universal lymph node detection in T2 MRI using neural networks. Int J CARS 18(2):313–318
10. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-CNN. In 2017 IEEE international conference on computer vision (ICCV), pp 2980–2988
11. Peng Y, Lee S, Elton DC, Shen T, Tang Y, Chen Q, Wang S, Zhu Y, Summers R, Lu Z (2020) Automatic recognition of abdominal lymph nodes from clinical text. In: Proceedings of the 3rd clinical natural language processing workshop, Online, November. Association for Computational Linguistics, pp 101–110
12. McCormick M, Liu X, Jomier J, Marion C, Ibanez L (2014) ITK: enabling reproducible research and open science. Front Neuroinform 8:13
13. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29(6):1310–1320
14. Kociołek M, Strzelecki M, Obuchowicz R (2020) Does image normalization and intensity resolution impact texture classification? Comput Med Imaging Graph 81:101716
15. Chen C-M, Chen C-C, Ming-Chi W, Horng G, Hsien-Chu W, Hsueh S-H, Ho H-Y (2015) Automatic contrast enhancement of brain MR images using hierarchical correlation histogram analysis. J Med Biol Eng 35:724–734
16. Yao H, Wang Y, Li S, Zhang L, Liang W, Zou J, Finn C (2022) Improving out-of-distribution robustness via selective augmentation. In: International conference on learning representations
17. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2018) mixup: beyond empirical risk minimization. In: International conference on learning representations
18. Yun S, Han D, Chun S, Oh S, Yoo Y, Choe J (2019) Cutmix: regularization strategy to train strong classifiers with localizable features. In: 2019 IEEE/CVF international conference on computer vision (ICCV), Los Alamitos, CA, USA. IEEE Computer Society, pp 6022–6031
19. Cramér H (2016) Mathematical Methods of Statistics (PMS-9), vol 9. Princeton University Press, Princeton
20. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 936–944
21. Ren S, He K, Girshick R, Sun J (2015) Faster r-CNN: towards real-time object detection with region proposal networks. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R (eds) Advances in neural information processing systems, vol 28. Curran Associates, Inc., Red Hook
22. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
23. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV), p 10012–10022
24. Tang Y, Yan K, Tang Y-X, Liu J, Xiao J, Summers RM (2019) ULDor: a universal lesion detector for CT scans with pseudo masks and hard negative example mining. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), pp 833–836
25. Solovyev R, Wang W, Gabruseva T (2021) Weighted boxes fusion: ensembling boxes from different object detection models. Image Vis Comput 107:104117
26. Wang Z, Li Z, Zhang S, Zhang J, Huang K (2019) Semi-supervised lesion detection with reliable label propagation and missing label mining. In: Lin Z, Wang L, Yang J, Shi G, Tan T, Zheng N, Chen X, Zhang Y (eds) Pattern recognition and computer vision. Springer, Cham, pp 291–302
27. Cai J, Harrison AP, Zheng Y, Yan K, Huo Y, Xiao J, Yang L, Le L (2021) Lesion-harvester: iteratively mining unlabeled lesions and hard-negative examples at scale. IEEE Trans Med Imaging 40(1):59–70
28. Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X, Sun S, Feng W, Liu Z, Xu J, Zhang Z, Cheng D, Zhu C, Cheng T, Zhao Q, Li B, Lu X, Zhu R, Wu Y, Dai J, Wang J, Shi J, Ouyang W, Loy C, Lin D (2019) MMdetection: open mmlab detection toolbox and benchmark