Detection of Lymph Nodes in T2 MRI Using Neural Network Ensembles

Tejas Sudharshan Mathai[1(✉)], Sungwon Lee[1], Daniel C. Elton[1], Thomas C. Shen[1], Yifan Peng[2], Zhiyong Lu[2], and Ronald M. Summers[1]

[1] Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD, USA
tejas.mathai@nih.gov

[2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Abstract. Reliable localization of abnormal lymph nodes in T2 Magnetic Resonance Imaging (MRI) scans is needed for staging and treatment of lymphoproliferative diseases. Radiologists need to accurately characterize the size and shape of the lymph nodes and may require an additional contrast sequence such as diffusion weighted imaging (DWI) for staging confirmation. The varied appearance of lymph nodes in T2 MRI makes staging for metastasis challenging. Moreover, radiologists often times miss smaller lymph nodes that could be malignant over the course of a busy clinical day. To address these imaging and workflow issues, in this pilot work we aim to localize potentially suspicious lymph nodes for staging. We use state-of-the-art detection neural networks to localize lymph nodes in T2 MRI scans acquired through a variety of scanners and exam protocols, and employ bounding box fusion techniques to reduce false positives (FP) and boost detection accuracy. We construct an ensemble of the best detection models to identify potential lymph node candidates for staging, obtaining a 71.75% precision and 91.96% sensitivity at 4 FP per image. To the best of our knowledge, our results improve upon the current state-of-the-art techniques for lymph node detection in T2 MRI scans.

Keywords: MRI · T2 · Lymph node · Detection · Deep learning

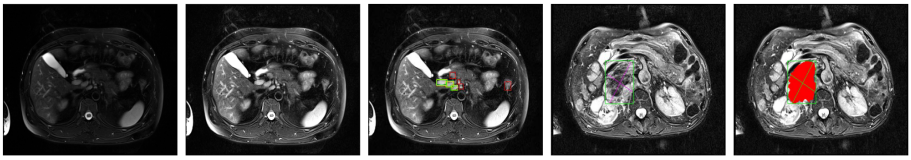
1 Introduction

Lymph nodes (LN) are a part of the lymphatic system, and they contain immune cells to help the body fight infection by filtering foreign substances in the lymphatic fluid that flows through them. Localization of lymph nodes in the abdomen is crucial as it allows enlarged and metastatic lymph nodes to be distinguished from non-metastatic lymph nodes [1–3]. It is especially paramount to identify enlarged LN if they are found at sites that do not correspond to the first site of lymphatic spread as this signals distant metastasis. Nodal location

C. Lian et al. (Eds.): MLMI 2021, LNCS 12966, pp. 682–691, 2021.
https://doi.org/10.1007/978-3-030-87589-3_70

and size are important aspects of the AJCC tumor, lymph node, metastasis (TNM) system [2], which provides guidelines for the diagnosis, management and treatment strategy of cancer and lymphoproliferative disorders [1,2].

Multi-parametric MRI images are becoming widely used for cancer and lymph node staging. Among the various MRI sequences, T2 and Diffusion Weighted Imaging (DWI) are commonly preferred to image the LN [1,3–5]. Unfortunately, accurate staging is difficult due to their irregular shapes, heterogeneous anatomical location, diverse yet similar appearance to other tissue structures (e.g. fat), and size. As shown in Fig. 1(a), lymph nodes in T2 MRI appear iso-intense (or slightly hyper-intense) relative to the surrounding fat [1] making it difficult to distinguish them, and contrast agents are often administered to reveal intensity differences between metastatic and normal LN tissue [1,3]. Further confounding the analysis of LN is the multitude of MRI scanners and exam protocols being used at different institutions to visualize abnormalities.



(a) (b) (c) (d) (e)

Fig. 1. (a) A T2 MRI image and (b) its normalized and histogram equalized counterpart are shown. There are 3 annotated LN regions as shown in (c), which is the result of our proposed ensemble detector; green boxes are the ground truth, yellow are the true positives, and red are the false positives. (d) A large lymph node annotated with both short and long axis diameters. (e) Failed segmentation result shown in red after generating pseudo-mask using GrabCut from the annotations. Note the incomplete segmentation of LN boundary (bottom left) that provides wrong training signals to a detection + segmentation network. (Color figure online)

Nodal size is the widely used criteria to determine malignancy [1,6]. Radiologists in clinical practice first localize the lymph nodes by scrolling back and forth through the T2 MR slices, and measure an identified node by using two orthogonal lines: the long and short axis diameters (LAD and SAD) [6]. Lymph nodes with a mean SAD of 3–10 mm are considered normal when imaged using conventional MRI, while those with a SAD of $\geq$10 mm are deemed suspicious for metastasis [1]. At our institution however, the accepted guideline is for radiologists to complete nodal sizing with either both diameters or the LAD only. This guideline can vary across institutions, thereby adding another obfuscating element to the analysis of LN. Furthermore, as radiologists size lymph nodes often during the course of a busy clinical day, despite their diligence, small LN might be missed ($\geq$3 mm) [3]. Thus, given the morbidity surrounding lymphoproliferative

diseases and the imaging and workflow challenges, there is a need for automated LN detection in T2 MRI images for subsequent sizing and staging.

Prior research has focused on the identification of lymph nodes in CT scans [9–15] with limited work done on MRI scans to localize lymph nodes [3,7,8]. In [3], a Mask-RCNN network was used to identify and segment lymph node regions in a multi-parametric pelvic MRI image (2 DWI + 1 T2 image). They achieved a sensitivity of 62.6% on their external testing dataset obtained from different hospitals and different scanners with 8.2 FP being identified per volume. In [7], initial lymph node candidate regions were detected using a GentleBoost classifier trained on image features followed by a convolutional neural network (CNN) to reduce false positives. Multiple views were incorporated into the CNN training to provide 3D structure, and a sensitivity of 85% was achieved at the rate of 5–10 FP per image. Finally, a Faster-RCNN was used in [8] to identify metastatic pelvic LN with an AUC of 91.2%.

As T2 MRI is arduous enough to identify and stage lymph nodes with a combination of T2 and DWI scans being preferred clinically, radiologist complete their annotations on one of the two (or more) sequences. This requires the registration of one sequence to another (e.g. DWI to T2) for reliable annotation usage in algorithmic development. However, as documented in [3], insufficient registration can result in errors that can lower the accuracy of LN detection. Therefore, in this pilot work, we first opt to use challenging T2 MRI sequences acquired using different scanners and exam protocols, and detect lymph nodes with state-of-the-art detection networks [16–23]. The intention is that the reasonable T2-based detection performance can be supplemented by the addition of accurately registered DWI scans. We improve upon the detection performance by using bounding box fusion techniques [24] that significantly reduce the number of FP detected by these networks. We detail results that either improve upon or are on par with the performance of previously published lymph node detection approaches, despite a lower quantity of training data. Finally, to replicate clinical use, we propose an ensemble of the best detection models and show its applicability towards lymph node detection.

Contribution. 1) For the LN detection task, we use state-of-the-art models to localize lymph nodes in T2 MRI and employ bounding box fusion techniques to reduce false positives. 2) To mimic clinical usage, we propose a ensemble of the best detection models to identify potential lymph node candidates for staging.

2 Methods

State-of-the-Art Object Detectors. We first quantified the performance of state-of-the-art object detectors on the LN detection task in T2 MRI: 1) Faster R-CNN [16], 2) DetectoRS [17], 3) YOLOv3 [18], 4) SSD [19], 5) RetinaNet [20], 6) FCOS [21], 7) FoveaBox [22], and 8) VFNet [23]. We further subdivided them into two categories: two-stage and one-stage detectors. Two-stage detectors (e.g. Faster R-CNN, DetectoRS) are region-based; the first stage generates region proposals corresponding to objects of interest, and the second stage classifies these region proposals and regresses the object bounding box coordinates.

These types of detectors are slow and computationally expensive. DetectorRS uses a Recursive Feature Pyramid (RFP) to add extra gradient feedback at the Feature Pyramid Network (FPN) [27] bottom-up layers, and Switchable Atrous Convolutions (SAC) to achieve peak object detection accuracy.

On the other hand, one-stage detectors [18–23] are faster as they skip the region proposal stage, densely sample all possible locations for objects, and directly predict the bounding box coordinates and class probabilities for different categories in a single pass. One-stage detectors can further be subdivided into anchor-based and anchor-free methods; anchor-based methods include YOLOv3 [18], SSD [19], and RetinaNet [20], while anchor-free methods include FCOS [21], FoveaBox [22], and VFNet [23]. YOLOv3 and SSD were attempts at fast and efficient object detection in images, but they were not suitable for detecting small objects. RetinaNet [20], on the other hand, overcame the common class imbalance problem plaguing detection tasks by using a focal loss function, along with the FPN to represent objects at different scales. However, these methods (including Faster R-CNN) require optimization of the sizes, aspect ratios and number of anchor boxes to achieve optimal object detection performance, involving additional computation and hyperparameter optimization [21–23].

FCOS [21] navigates away from anchor-based configurations by incorporating a FPN-based multi-level prediction inside the Fully Convolutional Network (FCN) [28], and a centerness score computed from the classification score to reduce the FP that are far away from the target object center. VFNet combines FCOS (without centerness branch) with an Adaptive Training Sample Selection (ATSS), sets the IoU between the ground truth and the prediction as the classification score, integrates it into a novel IoU-aware Varifocal loss, develops a star-shaped bounding box representation, and refines the box predictions. FoveaBox consists of a backbone to compute features from the input and a fovea head network that estimates the object occurrence possibility through per pixel classification on the backbone's output, and predicts the box at each position in the image that may be potentially covered by an object. We chose the anchor-free detectors in our experiment as they achieve superior results over anchor-based and even two-stage detectors.

Weighted Boxes Fusion. Object detectors typically predict an object's location with bounding box coordinates and provide confidence scores for them. Often times, ensembling over multiple selected epochs of a single detector or over different object detection models adds generalization [24] to a detection task, and yields more accurate results in contrast to a single model. Many strategies [24,31,32] have been proposed to combine the predicted boxes to yield accurate results, and we utilize weighted boxes fusion [24] as a post-processing step.

Ensemble of Best Detection Models. In order to add generalizability to the predictions of the detection networks, a threshold of 60% on the precision was set. Models that met the cut-off were ensembled together, leading to a higher detection performance in comparison to a single model.

Data Description. The lymph node dataset consisted of abdominal MRI studies that were acquired between January 2015 and September 2019, and they

were downloaded from the National Institutes of Health (NIH) Picture Archiving and Communication System (PACS). Initially, 584 T2-weighted MRI scans and associated radiology reports from different patients ($n = 584$) were downloaded. As mentioned in Sect. 1, radiologists completed nodal sizing with both the SAD and LAD, or with just the LAD. The lymph node size measured and annotated by the radiologist on a slice in a scan is linked to the radiology report using "bookmark" hyperlinks. NLP was used to find only lymph node related bookmarks [33]. The lymph node extent and size measurements were extracted from these bookmarks and considered the gold standard annotation.

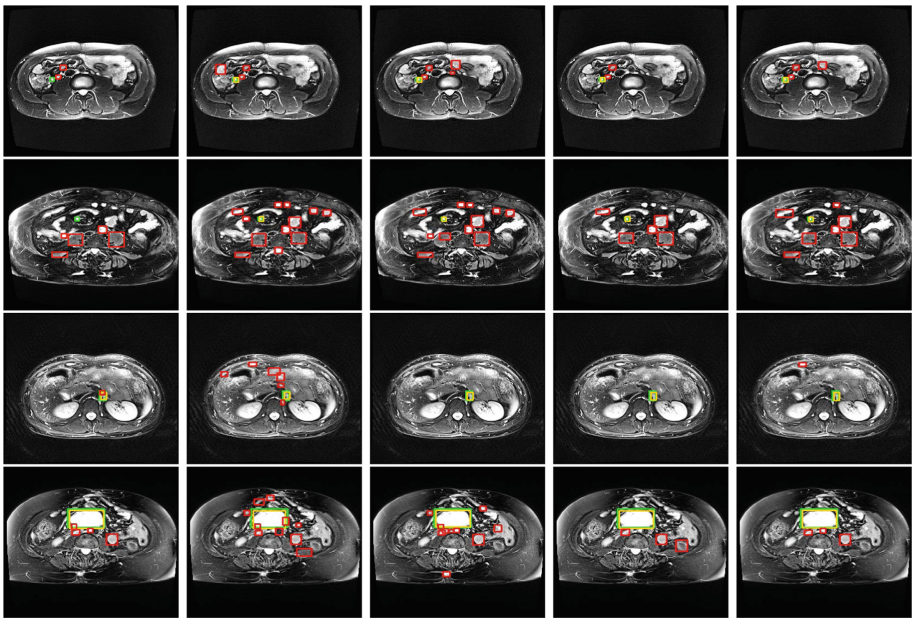
Next, an experienced radiologist checked the collected data and filtered those patients with scans that had incorrect annotations (e.g. kidney masses). Patients with scans that also contained only LAD measurements were removed; the intent was to standardize the data and analysis with the AJCC TNM guidelines [2]. This process resulted in a total of 376 T2 MRI scans ($n = 376$ patients) with 520 distinct lymph nodes that had been annotated with both the LAD and SAD measurements. The voxels in the scans were normalized to be within the 1%–99% of their intensity range [25]. As seen in Fig. 1, it increases the contrast between the bright and dark structures in T2 MRI scans, and mitigates the effect of outlier intensities from the image sensors present in various MRI scanners.

Following normalization, contrast enhancement of important structures, such as lymph nodes, was achieved through histogram equalization [26] without the excessive enhancement of image noise. The final dataset was then randomly divided into training (60%, 225 scans), validation (20%, 76 scans), and test (20%, 75 scans) splits at the patient-level. The resulting scans had dimensions in the range from (256–640) $\times$ (192–640) $\times$ (18–60) voxels. In contrast to prior detection and segmentation approaches [3,29], we do not rely on pseudo masks (e.g. generated through GrabCut); as seen in Fig. 1, their synthesis from radiologist annotations can often be incorrect and affect the training of the detection and segmentation networks, such as Mask RCNN [30].

Implementation Details. Radiologists focus their LN sizing efforts over a few slices of specific MRI sequences (e.g. T2), and generally corroborate their finding with another sequence (e.g. DWI). Prior work suggests that either 1- or 3- slice(s) are sufficient for the sizing task, with in-plane slice providing the most information [7]. With this in mind, we generated 3-slice T2 MRI images with the center slice containing the radiologist annotated LN, and used these 3-slices as the input for the detection networks. We used the framework proposed in [34] to implement the various one-stage and two-stage detectors used in this work. ResNet50 was used as the backbone for Faster R-CNN, DetectoRS, RetinaNet, FCOS, FoveaBox, and VFNet, while YOLOv3 used DarkNet53 and SSD used the VGG16 backbone. All the models were trained starting with the pre-trained MS COCO weights. Data augmentation included random flipping, random crops, random shifts and rotations in the range of [0, 32] pixels, and [0, 10] degrees respectively. A grid search was run on the batch size and learning rate for each model resulting in a batch size of 2 samples. The learning rate for YOLOv3, DetectoRS, SSD, FCOS, FoveaBox, and VFNet was 1e-3, while it was 25e-4 for

Faster R-CNN and RetinaNet. Each model was trained for 24 epochs, and the 5 epochs with the lowest validation loss were ensembled together and used for LN detection. In contrast, the models in the final detection ensemble only had 1 epoch with the lowest validation loss. These results are shown in Table 1. All experiments were run on a workstation running Ubuntu 16.04LTS and containing a NVIDIA Tesla V100 GPU.

Baseline Comparisons. We compare our lymph node detection results against those obtained by [3] on multi-parametric pelvic MRI, [7] on T1-weighted pelvic MRI, [9,10] on mediastinal LN in chest CT, and [11] on abdominal and mediastinal LN in CT.



(a) DetectoRS (b) FCOS (c) FoveaBox (d) VFNet (e) Ensemble

Fig. 2. Columns (a)–(d) show the lymph node detection results of the best detectors (DetectoRS, FCOS, FoveaBox, VFNet) incorporated into our ensemble on four different MRI images. Column (e) displays the result of our ensemble detector following Weighted Boxes Fusion. Green boxes are the ground truth, yellow are the true positives, and red are the false positives. Note that some detectors (e.g. DetectoRS) miss detecting lymph nodes (rows 1 and 2, col 1), but the ensemble detector benefits from the prediction of the remaining networks to yield the final detections.

3 Results and Discussion

Results. A clinically acceptable result [3] for lymph node detection is a precision of ≥60%, sensitivity of 85%, and 4–6 FP per image. Missing potentially

Table 1. Detection performance of various state-of-the-art detectors and our proposed ensemble method. "S" stands for Sensitivity @[0.5, 1, 2, 4, 6, 8, 16] FP

Method	mAP	S@0.5	S@1	S@2	S@4	S@6	S@8	S@16
Faster R-CNN [16]	58.72	61.90	70.23	80.95	83.33	83.33	83.33	83.33
DetectoRS [17]	61.83	66.67	73.81	77.38	80.95	80.95	80.95	80.95
YOLOv3 [18]	56.42	65.47	66.67	77.38	77.38	77.38	77.38	77.38
SSD [19]	40.21	35.71	57.14	70.23	79.76	79.76	79.76	82.14
RetinaNet [20]	57.36	53.57	65.47	77.38	83.33	86.90	88.09	89.28
FCOS [21]	60.09	61.90	77.38	83.33	88.09	89.28	89.28	89.28
FoveaBox [22]	61.67	61.90	76.19	79.76	84.52	88.09	89.28	89.28
VFNet [23]	63.91	67.85	75	80.95	83.33	83.33	83.33	83.33
Ensemble (VFNet + FoveaBox + FCOS + DetectoRS)	**71.75**	**73.81**	**79.76**	**85.71**	**91.66**	**91.66**	**91.66**	**91.66**
Ensemble (SAD $<$ 10 mm)	61.77	64.51	77.41	87.09	87.09	87.09	87.09	87.09
Ensemble (SAD $\geq$ 10 mm)	74.30	77.35	81.13	84.91	94.34	94.34	94.34	94.34

metastatic lymph nodes would be problematic, so we strive for a reasonable trade-off between precision and recall. Detection results are presented in Fig. 2 and Table 1 in terms of mean average precision (mAP) and sensitivity on the test dataset. YOLOv3 and SSD perform the worst in terms of LN detection with a precision of $\leq$60% and sensitivity of $\leq$80% indicating that they had significant issues with detecting LN $\leq$10 mm. RetinaNet and Faster R-CNN had higher sensitivities ($\geq$80%) at 4 FP per image, but their mAP values were still low and below our precision cut-off for ensembling. The detectors surpassing the $\geq$60% precision threshold included FCOS, FoveaBox, VFNet, and DetectoRS, with the maximum mAP of 63.91% achieved by VFNet. The one-stage anchor-free detectors significantly outperformed the anchor-based and two-stage detectors. Although promising, it was difficult to establish a clear winner amongst the anchor-free detectors. To this end, we ensembled the models passing the cut-off to obtain the highest LN detection performance with a mAP of 71.75% and 91.66% sensitivity at 4 FP per image. Against [3], our mAP is 71.75% vs 64.5%, and recall is 91.66% vs 62.6% at 8 FP. Against [7], our sensitivity is 91.66% vs 80% at 8 FP per image. Compared with [9–11], we obtain sensitivities of 91.66% at 4 FP vs 66% at 4 FP, 88% at 6 FP, and 90% at 6 FP respectively.

We further analyzed the behavior of the ensemble detector on lymph nodes when they were stratified by size. The ensemble model posts a moderate performance of 61.77% mAP and sensitivity of 87.09% at 4 FP per image when tested on LN with a SAD of $\leq$10 mm. One reason for the lower mAP is the ensemble detects many FP (as seen in Fig. 2, 2^{nd} row) due to the small size of the LN, thereby lowering the mAP. Our results are similar to those of [3] at $\sim$65% sensitivity, but we still achieve a significantly higher recall *without the use of DWI sequences*. On the other hand, we attain a mAP of 74.30% and sensitivity of 94.34% at 4 FP per image on lymph nodes with SAD $\geq$10 mm. These results are again consistent with past literature [3,12,13], yet we considerably outperform their results with a $\geq$10% increase in sensitivity. Our ensemble detector executes in 285 ms/22 s per image/volume vs. 862 ms/67 s from DetectoRS, 218 ms/17 s from FCOS, 134 ms/10 s from FoveaBox, and 276 ms/22 s from VFNet respectively.

Discussion. As described above, the two-stage Faster R-CNN and one-stage anchor-based detectors (YOLOv3, SSD, RetinaNet) resulted in significantly inferior detection performance and did not match our clinical implementation standards. Potential reasons for their performance include anchor ratio, size, and box optimization, and detection difficulties when encountering LN candidates with small size ($\leq$5 mm). In contrast, two-stage DetectoRS and the anchor-free detectors (FCOS, FoveaBox, VFNet) met our clinical use goals, and yielded significantly improved detection results as shown in Table 1. Weighted Boxes Fusion was the only post-processing step that was undertaken as it was necessary to fuse multiple detections from different models, but it is a small price to pay for the substantial increase in mAP and sensitivity. Of note, our results were achieved by taking off-the-shelf object detectors, and retraining them on a difficult T2 MRI dataset. The results were unencumbered by optimization of anchor parameters and free of region proposal generation. These make the networks simpler with fewer network weights to be learned, thereby reducing the chances of overfitting to the small training dataset. In fact, the only major parameter tuning that was done was with respect to the batch size and optimizer learning rate. We believe that this shows the power of the anchor-free detection networks without the need for a complicated training pipeline design [35].

We do not use DWI scans and this is a limitation of our work as it is routine practice for radiologists to confirm a finding on T2 MR with DWI. The ensemble performs satisfactorily on lymph nodes of size ($\leq$10 mm), but smaller lymph nodes ($\leq$3 mm) can be potentially metastatic [36], and a multi-parametric MRI input to the network might boost the detection performance of smaller LN. Future work is oriented towards the utilization of different MRI sequences (e.g. DWI and T1) to localize LN better. A structured report on the lymph node characteristics can also be created detailing the location, exam protocol, malignancy status, etc., thereby reducing this tedious analysis workload of the radiologists.

4 Conclusion

In this paper, we first quantify the performance of state-of-the-art object detectors on localizing lymph nodes in challenging T2 MRI scans. Weighted Boxes Fusion was used to then fuse the bounding box predictions from multiple epochs of a model to boost the detection performance. Consequently, a max mAP of 63.91% for VFNet and max sensitivities of 89.28% were achieved by FCOS and FoveaBox respectively. Next, we ensembled the best performing detectors together, which were VFNet, FCOS, FoveaBox and DetectoRS, to yield the highest mAP of 71.65% and sensitivity of 91.66% at 4 FP per image. We also stratified the results based on lymph node size, and found the ensemble performance to be significantly better than the current state-of-the-art in lymph node detection.

Acknowledgements. This work was supported by the Intramural Research Programs of the NIH National Library of Medicine and NIH Clinical Center. We also thank Jaclyn Burge for the helpful comments and suggestions.

References

1. Taupitz, M.: Imaging of lymph nodes — MRI and CT. In: Hamm, B., Forstner, R. (eds.) MRI and CT of the Female Pelvis. MR, pp. 321–329. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68212-7_15
2. Amin, M.B., et al.: 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 (2017)
3. Zhao, X., et al.: Deep learning based fully automated detection and segmentation of lymph nodes on multiparametric MRI for rectal cancer: a multicentre study. eBioMedicine **56** (2020)
4. Caglic, I., et al.: Diffusion-weighted imaging (DWI) in lymph node staging for prostate cancer. Transl. Androl. Urol. **7**(5), 814–823 (2018)
5. Heijnen, L.A., et al.: Diffusion-weighted MR imaging in primary rectal cancer staging demonstrates but does not Characterise lymph nodes. Eur. Radiol. **23**, 3354–3360 (2013)
6. Ganeshalingam, S., et al.: Nodal staging. Cancer Imaging **9**(1), 104–111 (2009)
7. Debats, O.A., et al.: Lymph node detection in MR lymphography: false positive reduction using multi-view convolutional neural networks. PeerJ **7**, e8052 (2019)
8. Lu, Y., et al.: Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. Cancer Res. **78**(17), 5135–5143 (2018)
9. Liu, J., et al.: Mediastinal lymph node detection and station mapping on chest CT using spatial priors and random forest. Med. Phys. **43**, 4362–4374 (2016)
10. Seff, A., Lu, L., Barbu, A., Roth, H., Shin, H.-C., Summers, R.M.: Leveraging mid-level semantic boundary cues for automated lymph node detection. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 53–61. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_7
11. Roth, H.R., et al.: A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 520–527. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_65
12. Barbu, A., et al.: Automatic detection and segmentation of lymph nodes from CT data. IEEE Trans. Med. Imaging **31**, 240–250 (2012)
13. Feulner, J., et al.: Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. **17** (2012)
14. Feuerstein, M., et al.: Automatic mediastinal lymph node detection in chest CT. Med. Imaging (2009)
15. Kitasaka, T., et al.: Automated extraction of lymph nodes from 3-D abdominal CT images using 3-D minimum directional difference filter. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4792, pp. 336–343. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75759-7_41
16. Ren, S., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE PAMI **39**(6), 1137–1149 (2017)
17. Qiao, S., et al.: DetectoRS: detecting objects with recursive feature pyramid and switchable Atrous convolution, arXiv (2020)
18. Redmon, J., et al.: YOLOv3: an incremental improvement, arXiv (2018)
19. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2

20. Lin, T.Y., et al.: Focal loss for dense object detection. In: ICCV, pp. 2999–3007 (2017)
21. Tian, Z., et al.: FCOS: fully convolutional one-stage object detection. In: ICCV, pp. 9627–9636 (2019)
22. Kong, T., et al.: FoveaBox: beyond anchor-based object detector, arXiv (2019)
23. Zhang, H., et al.: VarifocalNet: an IoU-aware dense object detector. In: CVPR, pp. 8514–8523 (2021)
24. Solovyev, R., et al.: Weighted boxes fusion: ensembling boxes from different object detection models. Img. Vis. Comp. **107** (2021)
25. Kociolek, M., et al.: Does image normalization and intensity resolution impact texture classification? Comput. Med. Imaging Graph. **81** (2020)
26. Chen, C.-M., et al.: Automatic contrast enhancement of brain MR images using hierarchical correlation histogram analysis. J. Med. Biol. Eng. **35**(6), 724–734 (2015). https://doi.org/10.1007/s40846-015-0096-6
27. Lin, T.Y., et al.: Feature pyramid networks for object detection. In: CVPR, pp. 2117–2125 (2017)
28. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
29. Zlocha, M., Dou, Q., Glocker, B.: Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 402–410. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_45
30. Yan, K., et al.: MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 194–202. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_22
31. Gidaris, S., et al.: Object detection via a multi-region and semantic segmentation-aware CNN model. In: ICCV, pp. 1134–1142 (2015)
32. Jaeger, P., et al.: Retina U-net: embarrassingly simple exploitation of segmentation supervision for medical object detection. In: ML4H at NeurIPS, pp. 1–12 (2019)
33. Peng, Y., et al.: Automatic recognition of abdominal lymph nodes from clinical text. Clin. Nat. Lang. Proc. (2020)
34. Chen, K., et al.: MMDetection: open MMLab detection toolbox and benchmark, arXiv (2019)
35. Yan, K., et al.: Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT. IEEE Trans. Med. Imag. (2020)
36. Langman, G., et al.: Size and distribution of lymph nodes in rectal cancer resection specimens. Dis. Col. Rec. **58**(4), 406–414 (2015)