Spatially-aware Graph Geometric Representation for Lung Nodules Detection

Luis Guayacán[1], Santiago Leal[1], Diego Patiño[2], and Fabio Martinez[1]✉

[1] BIVL[2]ab, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia
https://bivl2ab.uis.edu.co
{luis.guayacan, famarcar}@saber.uis.edu.co
santiago2201095@correo.uis.edu.co

[2] Department of Computer Science and Engineering, University of Texas at Arlington, Texas, USA
diego.patino@uta.edu

Abstract. In clinical practice, expert radiologists localize peripheral or central features in lung tissue, including vessels, bronchi, and other structures, to identify and categorize pulmonary nodules (PN). However, current computational PN detection methods rely on standard convolutional and transformer-based models without exploiting the irregular anatomical and contextual coherence within the parenchyma, which may be key to detecting nodular lesions at early stages. Emulating expert analysis, this work proposes a hierarchical, geometric graph representation that considers non-local and non-rigid relationships among lung regions. Using this representation, we learn a differentiable soft cluster assignment via a graph pooling operation, converging into a binary background/nodule representation. We encode each node's features with a second-order embedding to capture local inter-feature covariances, enabling spatially and contextually coherent detection. Moreover, we introduce an orthogonality regularization to enforce dissimilarity between the two embeddings in the grouping-matrix space. We evaluated our method on the LIDC-IDRI dataset, achieving a CPM of 0.51, which outperforms baselines such as YOLOv5, Faster R-CNN, DETR, and Grounding-DINO. In a controlled subset designed to validate the contribution of lung structures to the localization task, our method achieved a notable CPM of 0.74 (IoU=0.1), supporting the initial hypothesis.

Keywords: Pulmonary Nodules · Lung Nodules Detection · Graph Neural Networks · Superpixels · Diffpooling.

1 Introduction

Lung cancer remains one of the deadliest cancers worldwide, with an estimated 2.2 million new cases and 1.8 million deaths reported in 2022 [6]. Detecting pulmonary nodules (PNs) in CT scans is critical for early lung cancer diagnosis and effective intervention. However, PN identification is still challenging due to

their small size (3–30 mm), variable attenuation, and morphological heterogeneity [10]. Additionally, PNs often resemble normal structures such as blood vessels and are easily overlooked in routine readings [18]. Moreover, radiological assessments depend on expert visual inspection, introducing intra- and inter-observer variability.

Most computer-aided detection systems for PNs are based on convolutional anchor-based object detectors [1] that use fixed grids and receptive fields, which often fail to capture small nodules or irregular anatomical structures. Particularly in lung CTs, a single feature vector may mix heterogeneous tissues such as air, vessels, and nodules [5]. More recently, transformer-based detectors have been explored, leveraging global attention mechanisms but still relying on fixed patch tokenization, which may poorly align with anatomical boundaries. Both paradigms typically incorporate post-processing steps to reduce an initially large number of false positives, thereby increasing pipeline complexity [21,7]. In contrast, radiologists routinely interpret nodules, considering factors such as their central or peripheral location and proximity to critical anatomical structures, e.g., blood vessels, bronchi, or the pleura. For instance, distinguishing juxtavascular nodules or masses in contact with the pleural surface often requires detailed structural and contextual reasoning [14]. Graph-based learning offers a flexible alternative by modeling non-Euclidean structures that consider anatomical geometry and context, adapting to the underlying anatomy by representing regions as nodes and facilitating information flow along meaningful topologies.

In this work, we propose a graph-based framework for pulmonary nodule detection that models lung parenchyma as a graph of anatomically meaningful superpixels, each encoded with second-order descriptors capturing local interchannel dependencies. This graph is hierarchically clustered using differentiable pooling, producing nodule and background groups guided by a soft assignment matrix. To enhance cluster separation, we introduce an orthogonality regularization term that improves discrimination in the grouping space. To isolate the graph contribution from 3D continuity, we evaluate the model in a 2.5D setting. This design tests whether whole-slice non-local relationships help distinguish nodules from other lung structures. This lightweight design enables broader clinical applicability and facilitates integration into existing workflows. Hence, the proposed architecture is modality-agnostic and can be extended to other detection tasks.

2 Related Work

PN detection has been extensively explored through 2D deep learning approaches, particularly convolutional object detectors. Classic single-stage architectures such as YOLO and SSD have been adapted to medical images, achieving strong performance when paired with preprocessing strategies such as parenchyma masking and multi-slice stacking. Two-stage 2D methods, including adaptations of Faster R-CNN, have also been used to improve detection accuracy by separating proposal generation and classification [4]. Other works have improved 2D detection

with deformable convolutions [7] or synthetic nodule augmentation [8] to address sensitivity and false-positive issues. While 2D methods are less computationally demanding and easier to train than volumetric models, they typically rely on carefully designed context strategies to compensate for the lack of full 3D information. On the other hand, many recent benchmarks have chosen 3D models due to their ability to leverage volumetric information and inter-slice continuity, including 3D CNNs [19,24] and hybrid architectures incorporating attention [20].

While most detection approaches rely on Euclidean representations, graph-based approaches have been introduced in different tasks to capture structural and contextual relationships. Some methods used nonlocal pixel-level or attribute-based graphs [15,25], while others integrated radiomic features and clinical descriptors into graph representations [12,13,16,11,17]. Hierarchical graph pooling methods like DiffPool [22] and its variants [3,23] have been employed in non-medical domains to enable scalable, interpretable graph compression, motivating their adaptation to medical tasks. However, to the best of our knowledge, no existing approaches localize medical structures by learning hierarchical graphs that pool contextual information until converging on regions of interest.

3 Proposed Approach

We propose a graph-based framework for PN detection that computes a superpixel graph and encodes each superpixel using second-order features via regularized covariance matrices. Later, a hierarchical GNN with DiffPool layers [22] compresses the initial graph into a reduced version containing only background and nodule nodes. From this reduced graph, we derive a *noduleness map* that guides the anchor bounding boxes to high-likelihood regions. We refine these bounding boxes with a detection head that uses information from the reduced graph representation. Our unified design combines superpixel-based modeling, second-order feature encoding, hierarchical pooling, and anchor-based detection to achieve precise and semantically coherent nodule localization. Moreover, we introduce an orthogonality regularization term that penalizes correlations among superpixels merged into different final clusters, thereby encouraging clear separation between background and nodule representations.

Lung Graph Representation and Second-Order Modeling. Unlike grid-based representations, graphs are not constrained by a fixed input size and naturally adapt to anatomical variability in lung shape, size, and internal organization. We herein leverage this property of GNNs to better capture structural heterogeneity and contextual relationships among lung regions, which is critical for discriminating nodules from surrounding tissue. In our method, we start by passing the input image $\mathcal{I}$ obtained from stacking three CT slices through a CNN backbone, producing a dense feature map $\boldsymbol{h} \in \mathbb{R}^{H \times W \times C}$, with C the number of feature channels, and $H \times W$ the spatial dimensions. Then, we apply the SLIC algorithm to over-segment the lung region into a set of superpixels, such that $\mathcal{P}_s = \{\boldsymbol{p}_i\}_{i=1}^{|\mathcal{P}_s|}$ denotes the set of pixels belonging to superpixel $s = 1, 2, ..., N$. We denote by $\boldsymbol{f}_p \in \mathbb{R}^C$ the feature vector at pixel $\boldsymbol{p} \in \mathcal{P}_s$. We define an ini-

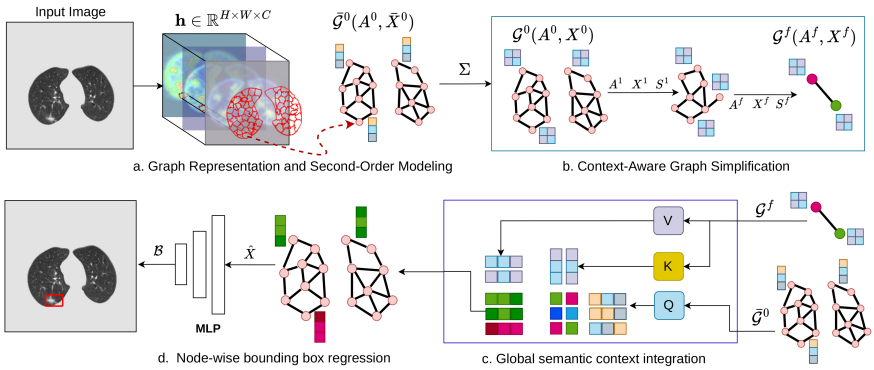

Fig. 1. Method overview. We compute a superpixel-based graph on the input image, extract features, and aggregate them over superpixels to get a node-embedded graph. A GNN with DiffPool layers reduces the graph to background and nodule clusters. Finally, we integrate semantic context into the original graph and compute the final bounding boxes.

tial superpixel graph $\bar{\mathcal{G}}^{(0)} = (\boldsymbol{A}^{(0)}, \bar{\boldsymbol{X}}^{(0)})$ with adjacency matrix $\boldsymbol{A}^{(0)} \in \mathbb{R}^{N\times N}$ based on spatial adjacency between neighboring superpixels. This process yields an irregular graph structure that conforms to the underlying lung anatomy (Fig. 1.a). Next, we compute second-order statistics on each $\mathcal{P}_s$ by assuming that its pixels follow a normal distribution and computing their mean and covariance matrix as

$$\boldsymbol{\mu}_s = \frac{1}{|\mathcal{P}_s|}\sum_{\boldsymbol{p}\in\mathcal{P}_s} \boldsymbol{f}_p, \qquad \boldsymbol{\Sigma}_s = \frac{1}{|\mathcal{P}_s|-1}\sum_{\boldsymbol{p}\in\mathcal{P}_s} (\boldsymbol{f}_p - \boldsymbol{\mu}_s)(\boldsymbol{f}_p - \boldsymbol{\mu}_s)^\top. \quad (1)$$

To ensure numerical stability and enforce symmetric positive definiteness (particularly for small superpixel regions), we apply a regularization term, setting $\boldsymbol{\Sigma}_s^{\text{spd}} = \boldsymbol{\Sigma}_s + \varepsilon \boldsymbol{I}$, where $\varepsilon > 0$ is a small constant and $\boldsymbol{I}$ is the identity matrix.

Because $\boldsymbol{\Sigma}_s^{\text{spd}}$ is symmetric, we reduce redundancy by retaining only its lower triangular part (including the diagonal) and flattening it into a vector using the vech$(\cdot)$ operator: $\boldsymbol{x}_s^{(0)} = \text{vech}\big(\boldsymbol{\Sigma}_s^{\text{spd}}\big) \in \mathbb{R}^{\frac{C(C+1)}{2}}$. We set this vector as the node embedding of $\mathcal{P}_s$, encoding inter-channel dependencies and local texture statistics beyond first-order averaging. By stacking all superpixels' embeddings, we define the node feature matrix as $\boldsymbol{X}^{(0)} = \left[\boldsymbol{x}_1^{(0)} \,\|\, \boldsymbol{x}_2^{(0)} \,\|\, \ldots \,\|\, \boldsymbol{x}_N^{(0)}\right]^\top \in \mathbb{R}^{N\times \frac{C(C+1)}{2}}$, which defines the enriched graph $\mathcal{G}^{(0)} = (\boldsymbol{A}^{(0)}, \boldsymbol{X}^{(0)})$. The resulting features are projected to the graph embedding dimension and processed by the GNN and the hierarchical pooling modules.

Context-Aware Graph Simplification. Lung nodules typically occupy very small regions of lung tissue, resulting in a severe class imbalance where nodule nodes, i.e., node superpixels, are vastly outnumbered by non-nodule nodes. This imbalance can bias the model, hindering correct nodule identification.

To address this issue, we apply graph pooling to $\mathcal{G}^{(0)}$ to reduce its size, grouping strongly related regions while preserving relevant structural information. This hierarchical reduction also helps mitigate class imbalance by progressively aggregating contextual information around potential nodules (Fig. 1.b). The graph pooling method in our framework is DiffPool [22], a hierarchical, differentiable pooling strategy that enables multi-level graph coarsening. Given a graph at layer l with node features $\boldsymbol{X}^{(l)} \in \mathbb{R}^{N_l \times D_l}$ and adjacency matrix $\boldsymbol{A}^{(l)} \in \mathbb{R}^{N_l \times N_l}$, DiffPool computes a pooled graph $\mathcal{G}^{(l+1)} = (\boldsymbol{A}^{(l+1)}, \boldsymbol{X}^{(l+1)})$ such that $\boldsymbol{X}^{(l+1)} = {\boldsymbol{S}^{(l)}}^\top \boldsymbol{Z}^{(l)} \in \mathbb{R}^{N_{l+1} \times D_l}$ and $\boldsymbol{A}^{(l+1)} = {\boldsymbol{S}^{(l)}}^\top \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \in \mathbb{R}^{N_{l+1} \times N_{l+1}}$. Here, $\boldsymbol{S}^{(l)} \in \mathbb{R}^{N_l \times N_{l+1}}$ is the assignment matrix encoding soft cluster memberships of nodes at layer l, and $\boldsymbol{Z}^{(l)}$ is a learned transformation of $\boldsymbol{X}^{(l)}$ computed via an independent GNN. In our model, $\boldsymbol{S}^{(l)}$ and $\boldsymbol{Z}^{(l)}$ are generated by separate GNN branches, allowing independent optimization of grouping structure and feature abstraction.

To obtain $\boldsymbol{S}^{(l)}$, we use a GNN that takes $\boldsymbol{X}^{(l)}$ and the adjacency matrix $\boldsymbol{A}^{(l)}$ as inputs. The GNN outputs a set of row-wise softmax logits that represent soft cluster assignments. We use standard GNN layers computed as $\boldsymbol{X}^{(l+1)} = \boldsymbol{\sigma}\Big(\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}}\, \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{X}^{(l)} \boldsymbol{W}^{(l)}\Big)$, with $\tilde{\boldsymbol{A}} = \boldsymbol{A}^{(l)} + \boldsymbol{I}$, $\tilde{\boldsymbol{D}}$ the corresponding degree matrix, $\boldsymbol{W}^{(l)}$ a learnable weight matrix, and $\boldsymbol{\sigma}(\cdot)$ a non-linear activation function.

Nodule Detection Head. To generate bounding box predictions, we use the enriched node representations from both the initial graph $\bar{\mathcal{G}}^{(0)}$ with N superpixels and the final two-node graph output. Each superpixel node is enriched via cross-attention with the two global node embeddings, allowing it to integrate global semantic context. Specifically, the query matrix $\boldsymbol{Q}$ is derived from the initial node features, while the key and value matrices $\boldsymbol{K}$, $\boldsymbol{V}$ are obtained from the final binary representation. The matrix $\hat{\boldsymbol{S}}$ modulates the attention weights, biasing the interaction toward the most relevant global node according to the learned grouping structure, following $\boldsymbol{\alpha} = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^\top}{\sqrt{d}} + \log(\hat{\boldsymbol{S}} + \varepsilon)\right)$, where $\boldsymbol{\alpha}$ represents the attention weights, influenced by the logarithmic prior from $\hat{\boldsymbol{S}}$ (Fig. 1.c). This process encourages alignment between the final two-node graph and the spatially aware initial graph representation. After merging local representations with global-aware embeddings, we pass the resulting representations $\hat{\boldsymbol{X}}$ through an MLP to predict offsets and confidence scores. Since each superpixel corresponds to one anchor region, this process yields N bounding box predictions (Fig. 1.d), which we later filter using non-maximum suppression.

Regularization for Structured Cluster Separation. While DiffPool enables hierarchical graph coarsening through soft assignments, it does not prevent degenerate solutions in which multiple clusters capture highly correlated information. This is a common issue in PN detection due to severe class imbalance, where the background class might dominate the learned assignments, hindering the formation of discriminative nodule representations.

Because our architecture employs K DiffPool layers that progressively reduce the graph to two final nodes, we define the global assignment matrix $\hat{\boldsymbol{S}}$ as the

Table 1. Comparative results with state-of-the-art methods.

Metric	YOLOv5s	Ground. DINO	RT-DETR	Faster R-CNN	Ours
mAP (↑)	0.419	0.434	0.240	0.407	**0.532**
CPM (↑)	0.445 ± 0.011	0.251 ± 0.115	0.106 ± 0.080	0.206 ± 0.125	**0.505 ± 0.210**

product of the intermediate assignments such that $\hat{\boldsymbol{S}} = \boldsymbol{S}^{(0)}\boldsymbol{S}^{(1)}\cdots\boldsymbol{S}^{(K-1)} \in \mathbb{R}^{N\times 2}$. Each row $\hat{\boldsymbol{s}}_s \in \mathbb{R}^2$ encodes the soft membership of superpixel s to the final background and nodule clusters. To enforce a structurally clean separation at the pooling level, we regularize $\hat{\boldsymbol{S}}$ by penalizing similarity between assignment vectors of superpixels belonging to different clusters. Let $\mathcal{C}_0$ and $\mathcal{C}_1$ denote the sets of superpixels assigned to the background and nodule classes, respectively. We define the orthogonality regularization as

$$\mathcal{L}_{\text{ortho}} = \frac{1}{|\mathcal{C}_0|\,|\mathcal{C}_1|} \sum_{p\in\mathcal{C}_0} \sum_{q\in\mathcal{C}_1} \left(\frac{\hat{\boldsymbol{s}}_p^\top \hat{\boldsymbol{s}}_q}{\|\hat{\boldsymbol{s}}_p\|_2 \, \|\hat{\boldsymbol{s}}_q\|_2} \right)^2 . \tag{2}$$

This regularization encourages assignment vectors of background and nodule superpixels to become mutually orthogonal, reducing membership overlap and preventing trivial pooling solutions in which both final nodes encode highly correlated information.

4 Experiments

Experimental Setup. We evaluate on the LIDC-IDRI dataset [2] (1,018 CT scans and 2,742 nodules), annotated by four radiologists. We split the data into 70% for training, 10% for validation, and 20% for testing. We preprocessed axial slices with nodules (2,510 total) using lung segmentation [9] and 3-slice stacking for 3D context. We report mAP and the CPM score, computed as the average sensitivity at 7 FP/scan levels, and counted predictions with IoU ≥ 0.25 after NMS as true positives. We used DarknetFPNx3 as our backbone network with 64 output feature channels. We regularize covariance descriptors with $\varepsilon = 10^{-3}$, and set the embedding dimension to $d = 64$ across all graph layers. The graph processor consists of two GNN layers with 64 hidden units each, followed by DiffPool operations to reduce the graph from N nodes to 16 and 2 nodes, respectively. We have released the source code publicly to ensure reproducibility[3].
Results and Discussion. We compare our method against state-of-the-art 2D object detectors. We include YOLOv5s, Faster R-CNN, RT-DETR, and Grounding-DINO as baselines, trained and evaluated under the same conditions on axial CT slices from the LIDC-IDRI dataset. We report mean Average Precision (mAP) and the Competition Performance Metric (CPM), with detections evaluated at an IoU threshold of 0.25. We summarize our results in Tab. 1.

[3] Code: `https://gitlab.com/luis.guayacan1/miccai_2026-nodulesdetection`

Table 2. Ablation study. We evaluated the contribution of the second-order descriptor (Covariance) and the orthogonality regularization term. Additionally, we evaluate our method under small, controlled architectural variations.

	Configuration	**mAP** (↑)	**CPM** (↑)
Feature representation	No covariance	0.488	0.473 ± 0.185
	Covariance	0.495	0.466 ± 0.170
	Covariance + orthogonality reg.	**0.532**	**0.505 ± 0.210**
Layer architecture	[32, 32, 32]	0.477	0.456 ± 0.168
	[64, 64, 64]	**0.532**	**0.505 ± 0.210**
	[32, 64, 128]	0.474	0.453 ± 0.168

Our method outperforms all baselines, achieving an mAP of 0.532 and a CPM of 0.505. YOLOv5s, the closest baseline, reaches 0.419 mAP and 0.445 CPM, while Faster R-CNN achieves similar mAP (0.407) but much lower CPM (0.206), indicating low sensitivity. Grounding-DINO and RT-DETR perform worse. We hypothesize that DETR's performance results from its high parameter count relative to the dataset size or from the lack of prompt engineering tailored to this medical detection task. These results highlight the importance of incorporating second-order local texture descriptors and hierarchical graph pooling for global context, as their combination enables more accurate and robust nodule detection.

Fig. 2 presents qualitative results from our method across nodules of varying sizes, categorized as small ($d < 6$ mm), medium ($6 \leq d \leq 10$ mm), and large ($d > 10$ mm). Our model accurately localizes most nodules across all size categories, including subtle and low-contrast examples, with minimal false positives. Note that we detect small and medium nodules with high spatial precision. These nodules are typically the most difficult due to limited visual cues. Failures mainly occur near central thoracic structures, where imperfect masking may remove nodular tissue. Some errors reflect the limited axial distinctiveness and z-axis context of the 3-slice input. Extending the model to 3D could reduce these ambiguities and misclassifications.

One of the main interests of this work is to investigate the contribution of contextual lung regions to the localization task by defining a spatially aware graph that evolves into two semantic classes. To validate this hypothesis, we performed inference, considering only slices with nodules, on a controlled subset of 502 images. In this setting, the proposed approach achieves a CPM of 0.745 and mAP of 0.674 at an IoU threshold of 0.1, and a CPM of 0.693 with mAP of 0.626 at an IoU of 0.25. These results highlight the importance of semantic lung structures in guiding nodule detection, even when such structures are not in close spatial proximity to the lesion.

We conducted an ablation study to assess the impact of key architectural components. In the upper part of Tab. 2, we evaluated the contribution of the second-order descriptor (Covariance) and the orthogonality regularization term. The full model, combining both components, achieves the highest performance (mAP = 0.532, CPM = 0.505). These results suggest complementary roles in enhancing local texture encoding and promoting embedding separation. In the

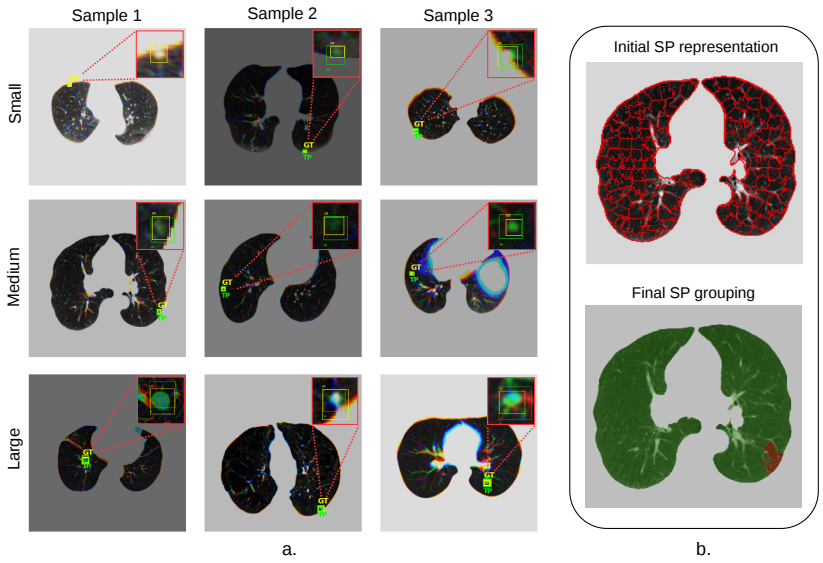

Fig. 2. Visualization of model predictions across varying nodule sizes (a) and corresponding superpixel grouping (b).

lower section of Tab. 2, we explore different GNN layer configurations. We obtained the best results with a uniform embedding size of 64 across layers. Both smaller and pyramidal setups yielded lower CPM scores, indicating reduced sensitivity, supporting the use of a consistent intermediate representation size to balance feature richness and stability during pooling.

5 Conclusions

We presented a graph-based framework for PN detection that supports radiological reasoning through spatially coherent, anatomically grounded lung representations. Using superpixels, second-order descriptors, and hierarchical graph pooling, the model captures local and global context, improving mAP and CPM on real patient data. Additionally, we introduced an orthogonality regularization loss to enforce structurally independent clusters and prevent collapse during hierarchical aggregation. These results demonstrate that explicitly modeling anatomical structure and context enhances detection performance and supports more interpretable diagnostic systems. Computational cost is concentrated in SLIC graph construction. Nevertheless, inference is competitive at 0.034 s/image versus 0.008 for YOLOv5s and 0.038 for RT-DETR. Future work will explore volumetric extensions that leverage more complex 3D geometries.

Acknowledgments. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational re-

sources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu

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

References

1. Ansari, M., Lodi, K.: A survey of recent trends in two-stage object detection methods. In: Renewable Power for Sustainable Growth. pp. 669–677. Springer (2021). https://doi.org/10.1007/978-981-33-4080-0_64
2. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics **38**(2), 915–931 (2011). https://doi.org/10.1118/1.3528204
3. Bianchi, F.M., Grattarola, D., Alippi, C.: Spectral clustering with graph neural networks for graph pooling. In: International conference on machine learning. pp. 874–883. PMLR (2020), https://proceedings.mlr.press/v119/bianchi20a.html
4. Cao, H., Liu, H., Song, E., Ma, G., Xu, X., Jin, R., Liu, T., Hung, C.C.: A two-stage convolutional neural networks for lung nodule detection. IEEE Journal of Biomedical and Health Informatics **24**(7), 2006–2015 (2020). https://doi.org/10.1109/JBHI.2019.2963720
5. Chen, F., Wu, F., Xu, J., Gao, G., Ge, Q., Jing, X.Y.: Adaptive deformable convolutional network. Neurocomputing **453**, 853–864 (2021). https://doi.org/10.1016/j.neucom.2020.06.128
6. Guo, L., Zhu, C., Cai, L., Zhang, X., Fang, Y., Chen, H., Yang, H.: Global burden of lung cancer in 2022 and projected burden in 2050. Chinese Medical Journal **137**(21), 2577–2582 (2024). https://doi.org/10.1097/CM9.0000000000003268
7. Haiying, Y., Zhongwei, F., Ding, D., Zengyang, S.: False-positive reduction of pulmonary nodule detection based on deformable convolutional neural networks. In: 2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB). pp. 130–134. IEEE (2021). https://doi.org/10.1109/ICBCB52223.2021.9459209
8. Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., Umemoto, K., Li, Y., Nakayama, H.: Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. In: 2019 International Conference on 3D Vision (3DV). pp. 729–737. IEEE (2019). https://doi.org/10.1109/3DV.2019.00085
9. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. European radiology experimental **4**(1), 50 (2020). https://doi.org/10.1186/s41747-020-00173-2
10. Larici, A.R., Farchione, A., Franchi, P., Ciliberto, M., Cicchetti, G., Calandriello, L., Del Ciello, A., Bonomo, L.: Lung nodules: size still matters. European Respiratory Review **26**(146) (2017). https://doi.org/10.1183/16000617.0025-2017
11. Li, M.M., Huang, K., Zitnik, M.: Graph representation learning in biomedicine and healthcare. Nature biomedical engineering **6**(12), 1353–1369 (2022). https://doi.org/10.1038/s41551-022-00942-x

12. Li, R., Zhou, L., Wang, Y., Shan, F., Chen, X., Liu, L.: A graph neural network model for the diagnosis of lung adenocarcinoma based on multimodal features and an edge-generation network. Quantitative Imaging in Medicine and Surgery **13**(8), 5333–5348 (2023). https://doi.org/10.21037/qims-23-2
13. Ma, L., Wan, C., Hao, K., Cai, A., Liu, L.: A novel fusion algorithm for benign-malignant lung nodule classification on CT images. BMC Pulmonary Medicine **23**, 474 (2023). https://doi.org/10.1186/s12890-023-02708-w
14. Mazzone, P.J., Lam, L.: Evaluating the patient with a pulmonary nodule: a review. JAMA **327**(3), 264–273 (2022). https://doi.org/10.1001/jama.2021.24287
15. Mou, L., Lu, X., Li, X., Zhu, X.X.: Nonlocal graph convolutional networks for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **58**(12), 8246–8257 (2020). https://doi.org/10.1109/TGRS.2020.2973363
16. Patiño, D., Avendaño, J., Branch, J.W.: Automatic skin lesion segmentation on dermoscopic images by the means of superpixel merging. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. pp. 728–736. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_83
17. Patiño, D., Ceballos-Arroyo, A.M., Rodriguez-Rodriguez, J.A., Sanchez-Torres, G., Branch-Bedoya, J.W.: Melanoma detection on dermoscopic images using superpixels segmentation and shape-based features. In: 15th International Symposium on Medical Information Processing and Analysis. vol. 11330, pp. 370–378. SPIE (2020). https://doi.org/10.1117/12.2545300
18. Rubin, G.D.: Lung nodule and cancer detection in computed tomography screening. Journal of Thoracic Imaging **30**(2), 130–138 (2015). https://doi.org/10.1097/RTI.0000000000000140
19. Song, W., Tang, F., Marshall, H., Fong, K.M., Liu, F.: A multiscale 3D network for lung nodule detection using flexible nodule modeling. Medical Physics **51**(10), 7356–7368 (2024). https://doi.org/10.1002/mp.17283
20. Tang, J., Chen, X., Fan, L., Zhu, Z., Huang, C.: LN-DETR: An efficient transformer architecture for lung nodule detection with multi-scale feature fusion. Neurocomputing **633**, 129827 (2025). https://doi.org/10.1016/j.neucom.2025.129827
21. Xiong, Y., Deng, L., Wang, Y.: Pulmonary nodule detection based on model fusion and adaptive false positive reduction. Expert Systems with Applications **238**, 121890 (2024). https://doi.org/10.1016/j.eswa.2023.121890
22. Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. In: Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc. (2018), https://proceedings.neurips.cc/paper_files/paper/2018/hash/e77dbaf6759253c7c6d0efc5690369c7-Abstract.html
23. Yuan, H., Ji, S.: StructPool: Structured graph pooling via conditional random fields. In: International Conference on Learning Representations (2020), https://openreview.net/forum?id=BJxg_hVtwH
24. Zhang, H., Peng, Y., Guo, Y.: Pulmonary nodules detection based on multi-scale attention networks. Scientific Reports **12**, 1466 (2022). https://doi.org/10.1038/s41598-022-05372-y
25. Zhang, X., Wang, K., Zhang, X., Huang, S.: Pulmonary Nodule Classification of CT Images with Attribute Self-guided Graph Convolutional V-Shape Networks. In: PRICAI 2021: Trends in Artificial Intelligence. Lecture Notes in Computer Science, vol. 13031, pp. 280–292. Springer (2021). https://doi.org/10.1007/978-3-030-89188-6_21