

LN-Transformer: Lung Nodule Transformer for Sparse CT Segmentation

Hooman Ramezani¹, Charlotte Vedrines¹, Dionne Aleman¹, and Daniel Létourneau²

¹ Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Canada
hooman.ramezani@mail.utoronto.ca, dionne.aleman@utoronto.ca,
charlotte.vedrines@gmail.com

² Department of Radiation Oncology, University of Toronto, Toronto, Canada
daniel.letourneau@uhn.ca

Abstract. Accurate segmentation of lung nodules in computed tomography (CT) scans is challenging due to extreme class imbalance, where nodules appear sparsely among healthy tissue. We introduce a novel two-stage approach for lung nodule segmentation, framing it as an anomaly detection problem. The method consists of two stages: Stage 1 employs a custom Detection Transformer architecture with deformable attention and focal loss to generate region proposals, addressing class imbalance and localizing sparse nodules. In Stage 2, the predicted bounding boxes are refined into segmentation masks using a fine-tuned variant of the Segment Anything Model (SAM). To address sparsity and enhance spatial context, a 5mm Maximum Intensity Projection is applied to improve differentiation between nodules, bronchioles, and vascular structures. The model achieves a stage-2 DiceC of 91.4%, with stage-1 yielding an F1 score of 94.2%, 95.2% sensitivity, and 93.3% precision on the LUNA16 dataset despite extreme sparsity, where only 5% of slices contain a nodule, outperforming existing state-of-the-art methods. The model was additionally validated on a privately procured test dataset of 30 patients with significantly different characteristics, achieving a Dice coefficient of 78.3% despite significant distribution drift, demonstrating strong generalization to clinical variability and establishing our approach as the new state-of-the-art for lung nodule segmentation.

1 Introduction

Lung cancer is a leading cause of cancer-related deaths worldwide [3], with early detection being crucial for improving outcomes. Tumor boards, comprising oncologists, radiologists, surgeons, and other specialists, manually segment nodules in CT scans to assess size, location, and spread, a process that is time-consuming and resource-intensive. Implementing a clinical decision support (CDS) system for auto-segmentation can improve efficiency, patient outcomes, and reduce costs [26]. However, existing models struggle with extreme class imbalance, as nodules appear infrequently, vary in size and shape, and often resemble vasculature. We propose LN-Transformer, a two-stage segmentation framework where Stage 1 employs a custom Deformable Detection Transformer for region proposal, leveraging MIP to enhance nodule visibility and focal loss to address class imbalance. Stage 2 refines predictions into pixel segmentation masks using a fine-tuned adaptation Segment Anything Model (SAM). This approach effectively handles nodule sparsity, demonstrating potential for clinical application.

2 Related Work

Thoracic Computed Tomography (CT) involves a series of 2D cross-sectional greyscale images that when combined, form a detailed 3D representation of the patient’s thorax. The LUNA16 dataset, derived from the LIDC-IDRI dataset, consists of 888 thoracic CT scans containing 1,186 annotated lung nodules, annotated by four radiologists with nodules larger than 3mm considered relevant. The challenge lies in the sparse occurrence of nodules, only 0 to 5% of slices contain a nodule [24], and in the variability of voxel sizes and scan resolutions across patients. This dataset serves as a critical benchmark with numerous studies training architectures such as CNNs, 3D-CNNs, and U-Net [10]. However, transformer architectures remain underexplored in this domain.

Class imbalance often biases models toward the majority class, often leading to high accuracy but poor detection capabilities. To mitigate this several strategies exist: oversampling increases the minority class but misrepresents real-world prevalence [20], while class weighting forces models to focus on underrepresented cases but can increase false positives [5, 8]. Focal loss mitigates these issues by dynamically adjusting the loss based on prediction confidence, down-weighting well-classified examples and emphasizing hard-to-classify, often maintaining precision while increasing accuracy [15].

Maximum Intensity Projection (MIP) is a widely used radiology technique [9] that enhances nodule visibility by combining adjacent CT slices into a single 2D image, projecting the highest attenuation voxel from a volume onto a 2D plane to preserve 3D spatial information [9]. MIP helps distinguish nodules, which appear as blobs, from vessels, elongated tube-like structures, and improves detection of 3–10mm nodules [11].

Transformer architectures have emerged as a powerful alternative to CNNs in medical imaging. While CNNs excel at capturing local features, they struggle with long-range dependencies, relationships between distant regions in an image [21]. Transformer self-attention effectively models these dependencies, making it particularly valuable for distinguishing nodules from vessels [21]. DETR, a vision transformer, directly predicts object locations via self-attention, replacing traditional region proposal methods but struggles with slow convergence and small object detection [7, 27]. Deformable-DETR improves this by introducing a deformable attention mechanism which is spatially adaptive and computationally efficient. Unlike standard self-attention, which attends to all pixels in an image, Deformable Attention selectively focuses on a small set of dynamically learned sampling points around a reference location. This allows the model to adaptively refine its receptive field and capture fine-grained details of small objects while significantly reducing computational overhead [27].

Segment Anything Model (SAM), trained on 1 billion masks, enables promptable segmentation and is shown to effectively transfer knowledge to new datasets with fine-tuning [12, 16]. MedSAM fine-tunes SAM on 1.5 million medical image-mask pairs to focus on anatomical complexities in clinical settings [16].

3 Methodology

We present a novel approach to lung nodule segmentation for tumor boards by framing the task as anomaly detection. Our method splits the task into two stages: Stage 1

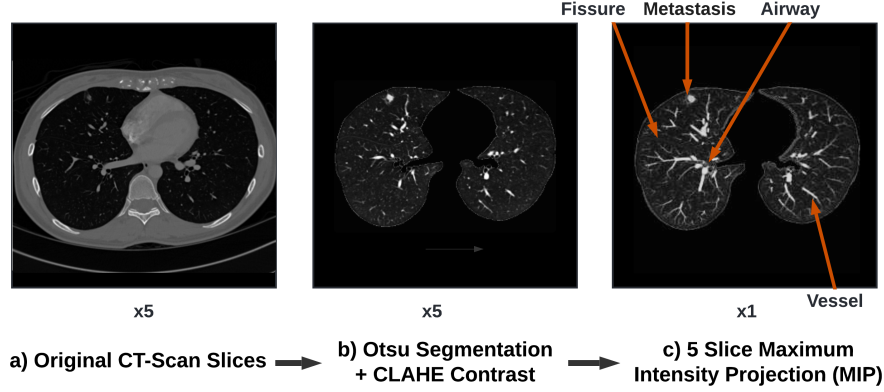


Fig. 1. Data processing pipeline with nodule visible at the top left of lung. a) Original CT slice b) Post Otsu segmentation and CLAHE c) Post 5mm MIP

serves as a region proposal phase to localize sparse nodules, while Stage 2 refines these bounding boxes into pixel-wise segmentation masks. While the building blocks are well-known, our novelty lies in unification of key architectural components such as DETR, SAM along with strategies such as deformable attention, focal loss and MIP into a specialized framework. Our training dataset is based on LUNA16 and consists of 9,676 CT slices, preprocessed to enhance nodule visibility through CLAHE, Otsu’s thresholding, and a customized training regimen to improve convergence. Our model is validated on an independent test set from the University Health Network (UHN), featuring 30 patients with diverse imaging protocols to provide a more robust evaluation.

3.1 Data Preprocessing & Datasets

Our preprocessing pipeline prepares CT scan data for input into our Stage 1 network shown in Figure 1. We first standardize anatomical structures by resampling CT slices to a consistent voxel spacing of $1 \times 1 \times 1$ mm, ensuring uniformity. In order to isolate lung tissue from the surrounding background, we employ Otsu’s method for thresholding[19]. This is followed by morphological operations, including connected component analysis and region erosion to obtain cleanly cropped lung regions. Slices at the superior and inferior cranio-caudal extremes, which provide minimal diagnostic value, are removed based on non-zero area size. This reduces the model’s search space from 15M to 5.25M pixels per patient improving focus on relevant areas. Post segmentation Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to improve the visibility of subtle features like small nodules [23]. Images are cropped to dimensions of 256x256, and a 5mm MIP is finally applied to improve visibility of lung nodules by producing a 2D image that highlights the densest features. This follows the expression $I_{\text{MIP}}(x, y) = \max_z \{I(x, y, z)\}$. $I_{\text{MIP}}(x, y)$ is the 2D image intensity at each (x, y) location, and $I(x, y, z)$ is the intensity of the original 3D image at voxel location (x, y, z) . A slab thickness of 5mm was chosen as a compromise between differentiating the shapes of vessels and nodules and limiting overlap between structures. The final training dataset

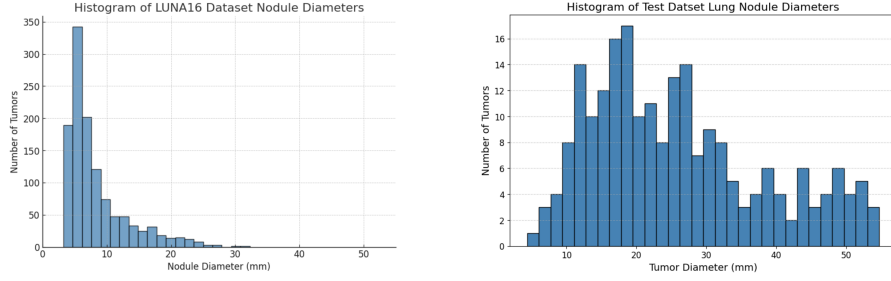


Fig. 2. Comparison of Size Distributions in LUNA16 (left) and UHN Test Dataset (right). LUNA16 contains primarily nodules, while UHN includes nodules and tumors.

consists of 9,676 MIP CT slices, with 1,226 containing nodules, split 70%-20%-10% before augmentation. Non-nodule slices were slightly undersampled during training resulting 12.7% positive class. This adjustment was necessary because lower nodule rates made training less effective due to excessive sparsity. The test set maintained a 5% nodule rate to reflect real-world conditions, with the adjustment leading to higher test accuracy. Additionally, the training set contained augmentations to enhance variability including flips, rotations ($\pm 15^\circ$), brightness shifts ($\pm 15\%$), and Gaussian noise (0.001–0.18% SD).

A supplementary test dataset was obtained in collaboration with University Health Network (UHN), comprising 400 treated patients imaged with a TOSHIBA Aquilion scanner. The CT images have a 3mm slice thickness, 0.781mm pixel spacing, and patient ages ranging from 29 to 90 years (median: 68). All images were segmented by a radiation oncologist. A subset of 30 patients (5,610 CT slices) was randomly selected as a second validation set. Figure 2 highlights nodule diameter variations between UHN and LUNA16. LUNA16 primarily contains small nodules (3–10mm), whereas UHN includes a broader distribution, with many tumors between 25–55mm. UHN’s 3mm slice thickness exceeds LUNA16’s 2.5mm cutoff, potentially reducing nodule visibility. Pixel spacing also differs, with UHN at 0.781mm and LUNA16 varying from 0.46–0.98mm. These differences introduce significant distribution drift, making UHN a strong test for model generalizability.

3.2 Stage One and Two Model Architecture

Figure 3 overviews our two-stage approach, where Stage 1 generates region proposals to localize potential lung nodules in CT scans. Input images pass through a ResNet-50 CNN backbone for multi-scale feature extraction, then augmented with 2D sine-cosine positional encodings and processed by the encoder’s DSA layers, which refine features by attending to a sparse set of learnable sampling points around each nodule. DSA aggregates features as $\mathbf{y}_q = \sum_{m=1}^M \mathbf{W}_m \left(\sum_{k=1}^K A_{mqk} \cdot \mathbf{x}(\mathbf{p}_q + \Delta \mathbf{p}_{mqk}) \right)$, where M is the attention head count, K the sampled points per head, $\mathbf{W}_m$ the projection matrices, and A_{mqk} the attention weight. The learnable offsets allow the model to dynamically adjust its receptive field for small and irregular nodules. Regular self-attention has complexity

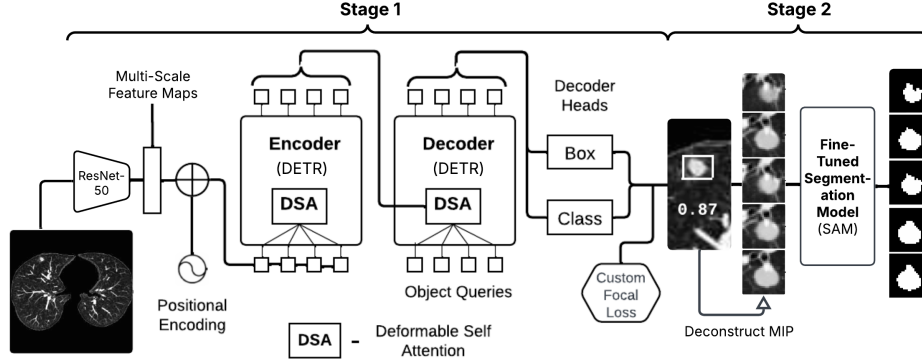


Fig. 3. Overview of the proposed LN-Transformer architecture, illustrating a sample CT MIP with a nodule in the top left corner. MIP deconstructed into its 5 associated slices for Stage 2.

$O(H^2W^2C)$, but DSA reduces this to $O(HWK C)$ where height H , width W , channels C and K representing the number of sampled points per attention head. Final decoder heads refine object queries into bounding boxes and confidence scores, which serve as inputs for Stage 2 segmentation. Stage 1 was trained for 15 epochs using AdamW with a learning rate of 10^{-4} , scheduled to reduce every 10 epochs. Training used an L4 GPU with mixed precision (16-bit), batch size of 4, gradient clipping (0.1), and accumulation over 6 batches. A grid search optimized hyperparameters for stability and performance.

We fine-tuned MedSAM’s pretrained weights on a dataset of 1,400 CT slices, where ground truth bounding boxes were used as prompts for SAM to simulate Stage 1 predictions. These acted as attention cues enabling targeted refinement while maintaining full-image context. The dataset included 1000 slices with ground truth nodule bounding boxes and 400 non-nodule slices to enhance false positive discrimination from Stage 1. Ground truth segmentation masks served as labels, and training was conducted using Dice-CrossEntropy loss with the Adam optimizer. Once trained, the Stage 2 auto-segmentation model processed MIP slices and associated bounding boxes from Stage 1, reconstructing individual CT slices to generate precise pixel-wise segmentation masks as illustrated in Figure 3.

To address class imbalance, we incorporated focal loss into the DETR loss function to enhance nodule detection by down-weighting easy samples and emphasizing hard-to-classify cases [14]. The focal loss is defined as $FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t)$, where p_t is the predicted probability of the correct class, α_t balances positive and negative examples, and γ adjusts focus towards challenging samples. Hyperparameter tuning found $\gamma = 2$ and $\alpha_t = 0.25$ to provide an optimal balance between precision and recall.

4 Results

Table 1 summarizes the performance metrics with nodules size categories: small (up to 7mm), medium (7-15mm), and large (over 15mm). Precision measures the proportion of correctly identified nodules among predictions, while sensitivity captures the percentage

Table 1. Performance Metrics for Region Proposal (Stage 1) and Auto-Segmentation (Stage 2) on LUNA16 and UHN Test Sets. IoU threshold = 0.5.

Metric	LUNA16			UHN Test Set		
	F1/Dice	Precision	Sensitivity	F1/Dice	Precision	Sensitivity
Stage 1: Region Proposal						
F1 Score	94.2%	–	–	79.1%	74.9%	83.5%
Avg IoU (All)	–	93.3%	95.2%	–	–	–
IoU (Small)	–	78.4%	83.3%	–	–	–
IoU (Medium)	–	96.7%	97.0%	–	–	–
IoU (Large)	–	97.8%	99.2%	–	–	–
Slice Accuracy	97.1	–	–	86.8%	–	–
Stage 2: Auto-Segmentation						
Dice Coefficient	91.4%	–	–	78.3%	–	–

Table 2. Comparison of Nodule Segmentation on DiceC, Sensitivity, and Specificity

Author	Architecture	DiceC (%)	Sensitivity (%)	Specificity (%)
Agnes et al. [1]	MRUNet-3D	89.0	94.8	84.2
Bhattacharyya et al. [4]	DB-NET	88.9	90.2	77.9
Song et al. [22]	ConvLSTM	84.0	87.8	81.5
Ma et al. [17]	SW-UNet	84.0	82.0	89.0
Annavarapu et al. [2]	Bi-FPN	82.8	92.2	78.9
Cao et al. [6]	DB-ResNet	82.7	89.4	79.6
Wang et al. [25]	MV-DCNN	77.9	87.0	77.3
Our Method	DETR-SAM	91.4	95.2	93.3

of actual nodules detected. The F1 score balances these metrics, and the Dice Coefficient evaluates the overlap between predicted and actual segmentation masks. Slice accuracy quantifies the proportion of CT slices correctly classified as containing or not containing a nodule, providing a high-level assessment of the ability to distinguish nodule-present and absent slices. For medium and large nodules, the model achieves high precision (96.7% and 97.8%) and recall (97.0% and 99.2%). Stage 2 yields a 91.4% DiceC, with only a 3% drop from the Stage 1 F1-score, indicating strong segmentation accuracy. On the UHN test set, the model attains 86.8% slice-wise accuracy, 79.1% F1-score, and 78.3% Dice, reflecting robustness despite greater tumor heterogeneity (25–55mm) and distribution drift. The model maintains high precision (74.9%) and sensitivity (83.5%), confirming generalization across diverse clinical conditions.

Table 2 presents a comparison of our proposed DETR-SAM approach against comparable models on LUNA16. Our DiceC of 91.4% outperforms next best models such as MRUNet-3D, a multi-resolution U-Net with 3D convolutions (89.0%), and DB-NET, a dual-branch CNN with attention mechanisms (88.9%). For sensitivity and specificity, our scores of 95.2% and 93.3% exceed prior state-of-the-art models such as ConvLSTM (92.2% sensitivity) and SW-UNet, a sliding window U-Net (89.0% specificity).

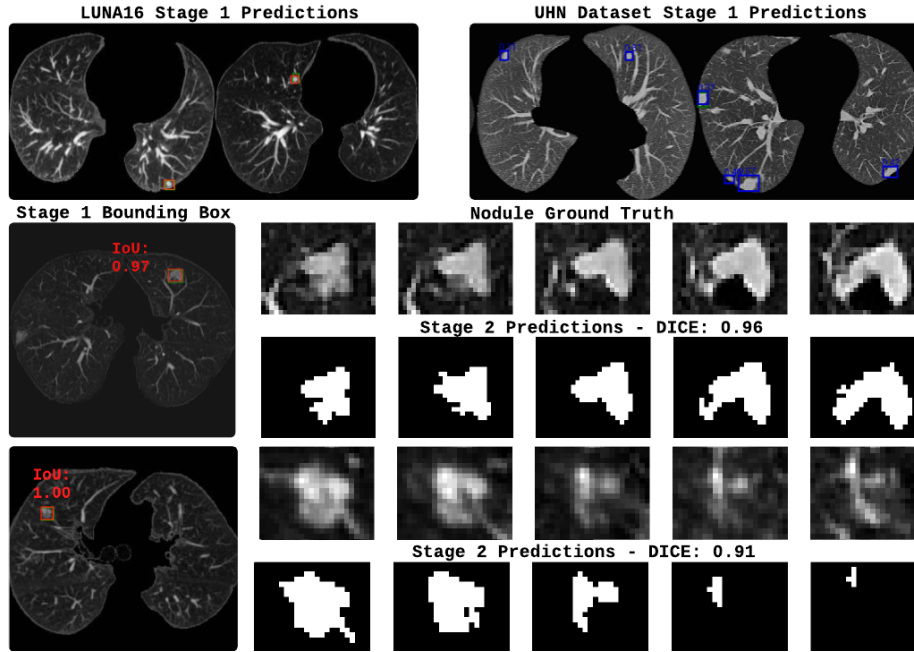


Fig. 4. Qualitative results from Stage 1 (top row) region proposals for LUNA16 (red) and UHN (blue), followed by Stage 2 (middle and bottom rows) masks from deconstructed MIP images.

Figure 4 illustrates full pipeline results, showing Stage 1 bounding box predictions (LUNA16: red, UHN: blue) and Stage 2 segmentation masks from MIP-reconstructed slices. The top slices highlight complex vascular structures and bronchioles that mimic or obscure small nodules, while UHN tumors are larger and more numerous. Bounding boxes exhibit high IoU, leading to accurate Stage 2 segmentation, where most tumors are nearly perfectly delineated. However, the model occasionally misclassifies connected pulmonary vessels as part of the nodule, as seen in the bottom-left mask.

5 Discussion

Our two-stage approach outperforms CNN and U-Net architectures. Models like MRUNet-3D [1], DB-Net [4] enhance feature extraction but are constrained by fixed receptive fields. Hybrid models such as 3D-MSViT [18] improve specificity (97.8%) and sensitivity but focus more on detection than segmentation. Bi-FPN and MV-DCNN emphasize sensitivity but lack a balanced trade-off with specificity. Unlike prior hybrid methods, we train end-to-end transformer models that make predictions without intermediary processing [7, 27]. Other models oversample LUNA16 nodule slices during training, distorting real-world prevalence, and lack external validation [6]. By preserving natural nodule sparsity and validating on independent clinical data, our method surpasses previous models across all metrics and ensures greater generalization to variability.

Table 3. Ablation Study Results: Performance of Different Configurations in the Proposed Pipeline

Configuration	F1 Score (%)	Dice Coefficient (%)
Direct SAM Application	–	3.4
Finetuned SAM Application	–	19.7
Finetuned MedSAM Application	–	26.7
S1: DETR + S2: Finetuned MedSAM	52.1	48.5
S1: Def.-DETR + S2: Finetuned MedSAM	94.2	91.4

While Stage 1 struggles slightly with nodules under 7mm, clinical significance is limited as nodules <6mm rarely warrant follow-up [13]. Stage 1’s ability to sift through highly sparse data and still detect nodules and tumors at state-of-the-art rates is the key contribution of this work. However, errors in Stage 1 propagate to Stage 2, highlighting dependency on precise region proposals. Future work could explore adaptive confidence thresholds to reduce error propagation. Unlike balanced classification, our approach follows an anomaly detection paradigm where accuracy holds greater significance due to the rarity of positive instances. Despite lower performance on UHN, the model generalizes despite large nodule variability, imaging protocols, and distribution shifts.

5.1 Ablation Study

To evaluate each component’s impact, we conducted an ablation study (Table 3). Applying SAM directly to lung CT scans yielded a DiceC of 3.4%, confirming its inability to segment nodules without guidance. MedSAM, despite pre-training on medical images performed similarly achieving 26.7% after fine-tuning, showing that pre-training alone does not ensure generalization. Introducing a detection stage improved results significantly, DETR in Stage 1 raised DiceC to 48.5% by narrowing SAM’s search space but struggled with small nodules (30% detection rate). Replacing DETR with Deformable-DETR more than doubled segmentation accuracy, highlighting the importance of adaptive attention for small object detection.

These results highlight the necessity of a two-stage approach and a robust detection stage to mitigate class imbalance, with DETR outperforming other architectures in detecting sparse, small nodules. Its self-attention mechanism enables flexible, context-aware feature representations, distinguishing nodules from normal structures and potentially learning hierarchical relationships that segmentation models alone cannot.

6 Conclusion

This study introduces LN-Transformer, a two-stage transformer framework tailored for sparse lung nodule segmentation. Key contributions include integrating Deformable-DETR with focal loss, MIP, SAM for mask refinement to address class imbalance and segment nodules. Our method achieves state-of-the-art results on LUNA16 (F1: 94.2%, Dice: 91.4%) and demonstrates robust generalization on an independent clinical dataset (F1: 79.1%, Dice: 78.3%), highlighting its potential for clinical application.

References

1. Agnes, S.A., Anitha, J.: Appraisal of deep-learning techniques on computer-aided lung cancer diagnosis with computed tomography screening. *Journal of Medical Physics* **45**(2), 98–106 (2020). https://doi.org/10.4103/jmp.JMP_101_19
2. Annavarapu, C.S.R., Parisapogu, S.A.B., Keetha, N.V., Donta, P.K., Rajita, G.: A bi-fpn-based encoder-decoder model for lung nodule image segmentation. *Diagnostics (Basel)* **13**(8), 1406 (Apr 2023). <https://doi.org/10.3390/diagnostics13081406>
3. Barta, J.A., Powell, C.A., Wisnivesky, J.P.: Global epidemiology of lung cancer. *Annals of Global Health* **85**(1), 8 (Jan 22 2019). <https://doi.org/10.5334/aogh.2419>, <https://doi.org/10.5334/aogh.2419>
4. Bhattacharyya, D., Thirupathi Rao, N., Joshua, E.S.N., Hu, Y.C.: A bi-directional deep learning architecture for lung nodule semantic segmentation. *The Visual Computer* **39**(11), 5245–5261 (2023). <https://doi.org/10.1007/s00371-022-02657-1>, <https://doi.org/10.1007/s00371-022-02657-1>
5. Buda, M., Maki, A., Mazurowski, M.: A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks: The Official Journal of the International Neural Network Society* **106**, 249–259 (2017). <https://doi.org/10.1016/j.neunet.2018.07.011>
6. Cao, H., Liu, H., Song, E., Hung, C.C., Ma, G., Xu, X., Jin, R., Lu, J.: Dual-branch residual network for lung nodule segmentation. *Applied Soft Computing* **86**, 105934 (2020). <https://doi.org/https://doi.org/10.1016/j.asoc.2019.105934>, <https://www.sciencedirect.com/science/article/pii/S156849461930715X>
7. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. *ArXiv abs/2005.12872* (2020). https://doi.org/10.1007/978-3-030-58452-8_13
8. Chan, R., Rottmann, M., Hüger, F., Schlicht, P., Gottschalk, H.: Metafusion: Controlled false-negative reduction of minority classes in semantic segmentation. *ArXiv abs/1912.07420* (2019)
9. Cody, D.: Aapm/rsna physics tutorial for residents: topics in ct. image processing in ct. *Radiographics : a review publication of the Radiological Society of North America, Inc* **22** 5, 1255–68 (2002). <https://doi.org/10.1148/RADIOGRAPHICS.22.5.G02SE041255>
10. El-bana, S., Al-Kabbany, A., Sharkas, M.: A two-stage framework for automated malignant pulmonary nodule detection in ct scans. *Diagnostics* **10** (2020). <https://doi.org/10.3390/diagnostics10030131>
11. Gruden, J., Ouanounou, S., Tigges, S., Norris, S.D., Klausner, T.: Incremental benefit of maximum-intensity-projection images on observer detection of small pulmonary nodules revealed by multidetector ct. *AJR. American journal of roentgenology* **179**(1), 149–157 (2002). <https://doi.org/10.2214/AJR.179.1.1790149>
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. *arXiv:2304.02643* (2023)
13. 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>, <https://ersjournals.com/content/26/146/170025>
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. pp. 2980–2988 (2017). <https://doi.org/10.1109/ICCV.2017.324>
15. Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. pp. 2999–3007 (2017). <https://doi.org/10.1109/ICCV.2017.324>

16. Ma, J., He, Y., Li, F., et al.: Segment anything in medical images. *Nature Communications* **15**, 654 (2024). <https://doi.org/10.1038/s41467-024-44824-z>
17. Ma, J., Yuan, G., Guo, C., Gang, X., Zheng, M.: Sw-unet: a u-net fusing sliding window transformer block with cnn for segmentation of lung nodules. *Frontiers in Medicine* **10**, 1273441 (2023). <https://doi.org/10.3389/fmed.2023.1273441>, <https://www.frontiersin.org/articles/10.3389/fmed.2023.1273441/full>
18. Mkindu, H., Wu, L., Zhao, Y.: 3d multi-scale vision transformer for lung nodule detection in chest ct images. *Signal, Image and Video Processing* **17**, 2473–2480 (2023). <https://doi.org/10.1007/s11760-022-02464-0>
19. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics SMC-9*(1), 62–66 (1979). <https://doi.org/10.1109/TSMC.1979.4310076>, <https://ieeexplore.ieee.org/document/4310076>
20. Qu, W., Balki, I., Mendez, M., Valen, J., Levman, J., Tyrrell, P.: Assessing and mitigating the effects of class imbalance in machine learning with application to x-ray imaging. *International Journal of Computer Assisted Radiology and Surgery* **15**, 2041 – 2048 (2020). <https://doi.org/10.1007/s11548-020-02260-6>
21. Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H.: Transformers in medical imaging: A survey. *Medical Image Analysis* **88**, 102802 (2023). <https://doi.org/https://doi.org/10.1016/j.media.2023.102802>, <https://www.sciencedirect.com/science/article/pii/S1361841523000634>
22. Song, G., Dai, Q., Nie, Y., Chen, G.: Differential diagnosis of benign and malignant pulmonary nodules in ct images based on multitask learning. *Current Medical Imaging* (2023). <https://doi.org/10.2174/0115734056252399231011042326>
23. Sundaram, M., Ramar, K., Arumugam, N., Prabin, G.: Histogram based contrast enhancement for mammogram images. 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies pp. 842–846 (2011). <https://doi.org/10.1109/ICSCCN.2011.6024667>
24. Walter, J., Heuvelmans, M., de Jong, P.D., Vliegenthart, R., van Ooijen, P.V., Peters, R.B., ten Haaf, K., Yousaf-Khan, U., van der Aalst, C., de Bock, G.D., Mali, W., Groen, H., de Koning, H.D., Oudkerk, M.: Occurrence and lung cancer probability of new solid nodules at incidence screening with low-dose ct: analysis of data from the randomised, controlled nelson trial. *The Lancet. Oncology* **17** 7, 907–916 (2016). [https://doi.org/10.1016/S1470-2045\(16\)30069-9](https://doi.org/10.1016/S1470-2045(16)30069-9)
25. Wang, S., Zhou, M., Gevaert, O., Tang, Z., Dong, D., Liu, Z., Jie, T.: A multi-view deep convolutional neural networks for lung nodule segmentation. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1752–1755 (2017). <https://doi.org/10.1109/EMBC.2017.8037182>
26. Wyatt, D.C., Avery, T., Quan, A., de Waal, P.: Clinical decision support systems and rapid learning in oncology. *BMJ Health & Care Informatics* **26**(1) (2019). <https://doi.org/10.1136/bmjhci-2019-100015>, <https://informatics.bmj.com/content/26/1/e100015>
27. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. *ArXiv abs/2010.04159* (2020)