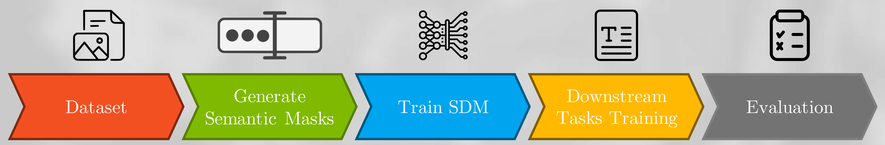
High-Fidelity Image Synthesis from Pulmonary Nodule Lesion Maps using Semantic Diffusion Model

Xuan Zhao[1], Benjamin Hou[1,2]

[1] Biomedical Image Analysis Group/Department of Computing, Imperial College London, UK | [2] National Institutes of Health, Bethesda, Maryland, USA

Project Pipeline



- LUNA16 [1]
- LIDC-IDRI dataset [2]

- Spherical crop
- OTSU threshold
- Largest connected region

- Generate synthetic samples using masks

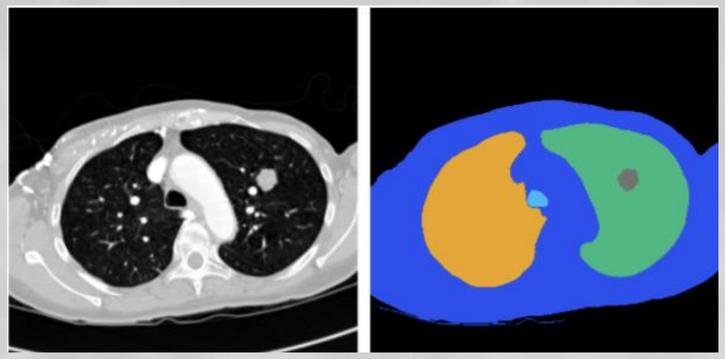
- Nodule detection and classification

- Accuracy, Precision, Recall, F1-Score; Avg. Precision/Recall, Wilxocon Rank-sum Test

- LUNA16 is used to generate synthetic samples, while LIDC mainly serves for looking up additional information other than annotations.
- The spherical crop is based on the annotation from LUNA16, which includes the center location and diameter of each nodule.
- Nodule detection and classification tasks are based on the mmPreTrain framework, all configs are default from mmPreTrain setup.

Semantic Mask Generation

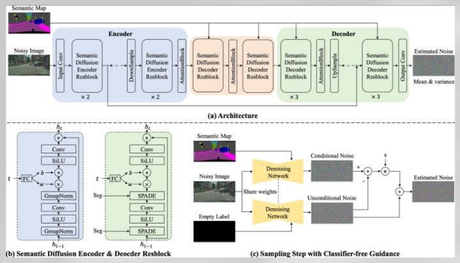
For all experiments, the CT window is set between [-1000,400], and operations are performed in 2D. Slices are considered only if they contain lung structure, as nodules do not exist outside these regions. To create nodule masks, the nodules are first cropped spherically based on the centroid and diameter information from the provided annotations. A manual OTSU threshold is then applied to each cropped region-of-interest to get the final masks. The intensities of the cropped pixels are clustered using the K-Means algorithm with two centers, and the threshold is selected as the average of these centers. Additionally, a 'body mask' is also generated, which comprises the entire patient's body. Each slice is intensity thresholded at 127, followed by a morphological hole fill process. The largest connected region is then selected. The final mask is then composed of the structures in this order: background, left lung, right lung, trachea, body mask, and nodule if one is present.



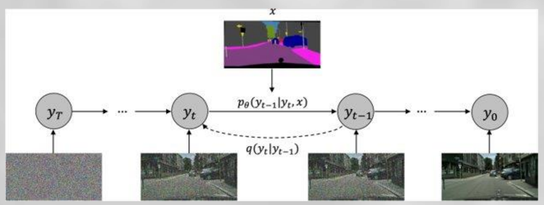
In the CT segmentation mask: black - background, orange - left lung, green – right lung, cyan - trachea, blue - body mask.

Semantic Diffusion Model

The semantic diffusion model [3] processes input image and semantic masks separately, which leverages the input semantic masks well. It feeds noisy image to the encoder of the U-Net structure while the semantic layout to the decoder by multi-layer spatially-adaptive normalization operators.



(a) Architecture of conditional denoising network. The denoising network takes the noisy image as input and estimates the involved noise under the guidance of the semanticlabel map. (b) The detailed structure of semantic diffusion encoder and decoder resblock. (c) The sampling procedure with classifier-free guidance.



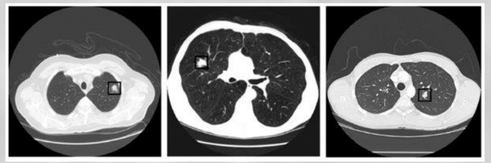
The diffusion process of SDM with semantic information used in leveraging the diffusion process

Examples of quality and diversity of images generated by SDM over 4 different datasets. Left-top is the input mask, and the rest of the three images are examples of generation via SDM

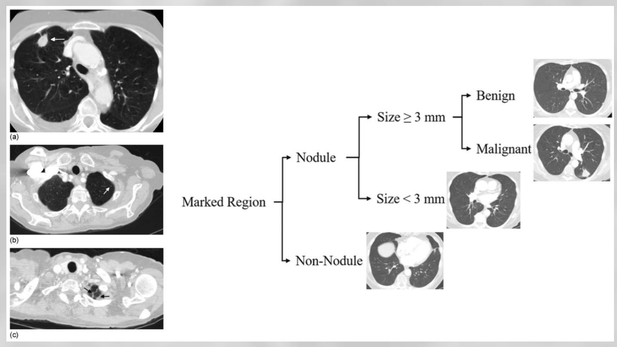
Downstream Models

Task	Architecture	Training Config
Nodule Detection	SE-ResNet [4]: seresnet101_8xb32_in1k	50 epochs with 128 iterations per epoch, learning rate set to 0.1 and uses staircase decay, batch size sets to 8
Nodule Localization	Faster R-CNN [5]: faster_rcnn_x101_32x4d_fpn_1x_coco	12 epochs with 510 iterations per epoch, learning rate is set to 0.1 and linear decay at rate of 0.1, batch size is set to 2

LUNA16 and LIDC Dataset

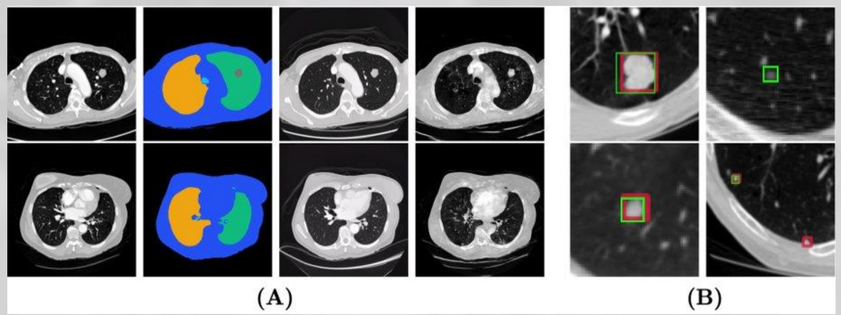


Sample CT scans from LIDC-IDRI dataset containing pulmonary nodules



Left: Example of lesions considered to satisfy the LIDC/IDRI definition of a) nodule ≥ 3mm, b) nodule < 3mm, and c) non-nodule ≥ 3mm. Right: The categorization process of pulmonary nodules in LIDC/IDRI. Nodules larger then 3mm will then be further categorized into benign or malignant tumors.

Results and Evaluation



(A) Example images generated by SDM and SPADE. L-to-R: CT image, CT mask, SDM image, SPADE image. Top: Nodule slice. Bottom: Nodule-free slice. (B) Localization downstream task. Top: SDM, Bottom: SPADE. Left: Correctly identified nodules, Right: False Negative/Positive detections. Green box is ground truth and red box is prediction.

	Accuracy (%)	Precision (%)	Rec./Sen. (%)	Specificity (%)	F1	p-value
I,A	85.76 ± 1.69	$\mathbf{85.63 \pm 3.66}$	86.42 ± 5.05	85.03 ± 6.10	85.83 ± 1.61	-
I,B	88.99 ± 1.32	83.3 ± 2.74	$\mathbf{90.64 \pm 2.86}$	87.86 ± 2.98	86.76 ± 1.31	82.0×10^{-6}
I,C	$\mathbf{89.72 \pm 1.26}$	85.09 ± 2.21	90.37 ± 2.14	$\mathbf{89.29 \pm 1.93}$	$\mathbf{87.61 \pm 1.23}$	$\mathbf{9.54 \times 10^{-6}}$
	AP_{50} (%)	AP_{60} (%)	AR_{50} (%)	AR_{60} (%)	AR_{70} (%)	p-value
II,A	80.26 ± 5.62	73.73 ± 6.03	89.23 ± 3.85	83.62 ± 4.12	64.96 ± 4.39	-
II,B	80.04 ± 4.60	72.37 ± 5.14	90.18 ± 3.52	83.52 ± 4.10	66.24 ± 3.50	0.985
II,C	$\mathbf{88.75 \pm 3.21}$	$\mathbf{84.80 \pm 3.72}$	$\mathbf{95.02 \pm 2.15}$	$\mathbf{91.55 \pm 2.49}$	$\mathbf{78.08 \pm 2.96}$	$\mathbf{4.825 \times 10^{-4}}$

Relevant metrics on 4 experiments: I: Nodule detection task. II: Nodule localization task. A: Without synthetic images in train set. B: With SPADE images in train set. C: With diffusion images in train set. p-value is generated between A (control experiment) and other experiments (B or C). AP and AR are Average Precision and Recall, and the subscript denotes the IoU% used.

About Us

Read the paper | Download the code | LinkedIn | BioMedIA

References

[1] Setio, Arnaud Arindra Adiyoso, et al. "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge." Medical image analysis 42 (2017): 1-13.

[2] 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. doi: https://doi.org/10.1118/1.3528204.
URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3528204.

[3] Weilun Wang, Jianmin Bao, Wengang Zhou, Dongdong Chen, Dong Chen, Lu Yuan, and Houqiang Li. Semantic image synthesis via diffusion models. CoRR, abs/2207.00050, 2022.

[4] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

[5] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

Acknowledgements:
This work is only based on open public dataset, and no private clinical data has been used.
Training of models have taken place on GPUs provided by college as well as private GPUs thanks to Peter Zhang from Cambridge.
This project exists as part of the final year individual project at the Department of Computing, Imperial College London.