3D Conditional Generative Networks for Detection of Pulmonary Nodules

Nhan Le, Timothy Chu, Amos Newswanger, Max Torop, Haofu Liao, Jiebo Luo, and Chenliang Xu
Department of Computer Science
University of Rochester
Rochester, NY 14627
nle2@u.rochester.edu
tchu5@u.rochester.edu

Abstract

Recent breakthroughs in deep learning have promoted research interest in automated nodule detection in pulmonary computed tomography (CT) scans. Standard neural network architectures like fully convolutional networks (FCNs) and U-Nets have long served as the foundation behind semantic segmentation research. With recent applications of conditional generative adversarial networks (cGANs) producing state-of-the-art results for 2D medical imaging problems, we sought to test the generalizability of adversarial training to the 3D realm. Using the LUNA16 dataset, we trained encoder-decoder FCNs and U-Nets alone and as generator networks within cGANs to segment pulmonary nodules. We hypothesized that adversarial training would provide superior results to non-adversarial training. Our results show that basic networks improved performance when subjected to adversarial training, and thus, we propose that future deep learning architectures for 3D medical image segmentation problems be considered as generator networks in cGANs.

1. Introduction

Lung cancer is one the most prevalent health issues in the world, and as with all cancers, early diagnosis is ideal for good prognosis. Lung cancer screenings often use computed tomography (CT) scans. CT works by measuring X-ray absorption from different angles to produce 2D cross-sections that can be combined in 3D visualization.

In CT scans, nodules appear alongside healthy tissues but have distinctive round shapes. The ability for doctors to detect nodules vary between individuals based on experience. In theory, computer-aided detection bypasses human error by using objective probabilistic inferencing. In practice, the 3D semantic segmentation task of nodule detection is extremely difficult. Approaches for 2D segmentation may be unfeasible for 3D segmentation due to memory constraints, and the added complexity of the third dimension may lead to either overfitting or underfitting. Since research into automated diagnosis is still new, nodule detection, as the first step in the process, must be refined before we tackle the problem of determining nodule malignancy.

In this paper, we explore whether neural network performance for nodule detection would improve with adversarial training. Because generative adversarial networks (GANs) teach neural networks to adapt to task-specific parameters, we hypothesize that the abilities of our 3D fully convolutional network (FCN) and our 3D U-Net would both improve when trained as a generator in a conditional GAN (cGAN) [4, 6].

Our main contribution is to demonstrate the effectiveness of adversarial training for 3D semantic segmentation. As pulmonary nodule detection is a computer vision problem made difficult by modern hardware limitation, if the computationally-inexpensive incorporation of semantic segmentation architectures as cGAN generators can improve results, then we propose adversarial training as a low-cost method of improving any semantic segmentation architecture.

2. Related Works

2.1. Deconvolutional Networks and U-Nets

FCNs are convolutional neural networks that are trained from end to end without any fully connected layers. Being able to take in arbitrarily-sized inputs and produce correspondingly-sized outputs with great efficiency has made FCN-based encoder-decoder networks common for 2D semantic segmentation tasks [2, 5, 11]. The deconvolutional network (deconvNet) is one FCN architecture that has been successfully employed for 2D semantic segmentation [2]. A deconvNet consists of a contracting path with convolution and pooling layers and an expanding path with convolution and upsampling layers [2].

U-Nets are deconvNets with skip connections between contracting layers and expanding layers that serve to concatenate upsampled images in the expanding path with corresponding feature maps from the contracting path (Figure 1) [8]. The benefits of deconvNets are preserved in U-Nets but there is an added benefit of low-level information flowing to high-level information [3].

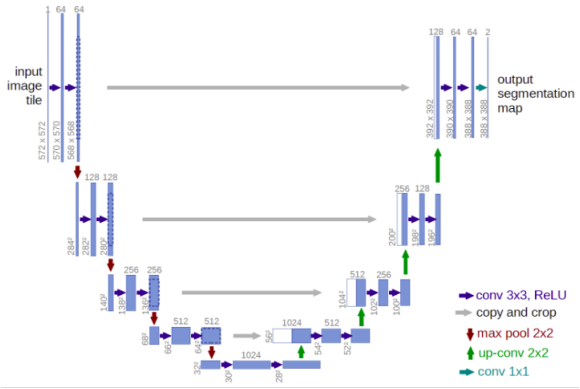

Figure 1: Vanilla U-Net [8].

2.2. Conditional Generative Adversarial Networks

GANs are robust architectures for estimating cost functions via adversarial training, in which a generating network and a discriminating network incrementally improve by being trained against one another [4]. The concept is that the generator creates fake outputs while the discriminator attempts to differentiate between real and fake outputs. The feedback loop is used to teach the generator to tweak the parameterization for the current task until fake outputs approximate real outputs.

cGANs are extensions of GANs in which training is done on conditional data (Figure 6) [6]. The main advantage of cGANs over vanilla GANs is that conditional training allows inputs and outputs to be paired one-to-one.

2.3. Adversarial Training for Semantic Segmentation

We were inspired by current research using cGANs for semantic segmentation in 2D images [9, 10]. Many recent breakthroughs in computer vision stem from adversarial approaches, and we were interested in determining whether the superiority of such methods would generalize to 3D images. We have come across one study [7] in which a cGAN was used for brain nodule segmentation, but because there was no control method of non-adversarial training, we are unable quantitatively ascertain know how adversarial training approaches improve deep learning architectures. To our knowledge, a cGAN has not been used for a lung nodule detection task.

3. Methods

3.1. Dataset

The repository by LUng Nodule Analysis 2016 (LUNA16) provided the dataset for our experiments [1]. The dataset consists 888 three-dimensional, MetaImage-format low-dose CT scans with 1186 nodules that were identified as such by at least three out of four radiologists. The LUNA16 data is a filtered subset of the Lung Image Database Consortium (LIDC-IDR) in which scans with a slice thickness of greater than 2.5 mm and nodules with a diameter of less than 3 mm were removed.

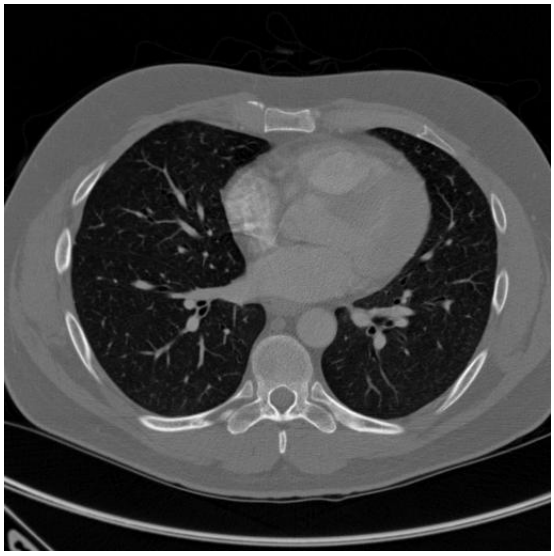

Figure 2: A 2D slice of a 3D CT scan.

3.2. Preprocessing

3.2.1 Hounsfield Unit Conversion

Raw medical images were converted into Hounsfield units (HUs), which is an important step in medical imaging. The Hounsfield scale is the standard scale for describing radiodensity.

3.2.2 Lung Mask Extraction

Because pulmonary CT scans contain information of irrelevant body substances and outside environments, isolating lung information is necessary to produce accurate results. We created lung masks by iterating through 2D slices and filtering out irrelevant body substances by binarizing with a -400 HU threshold below which lung tissues register and outside environments by removing connected components

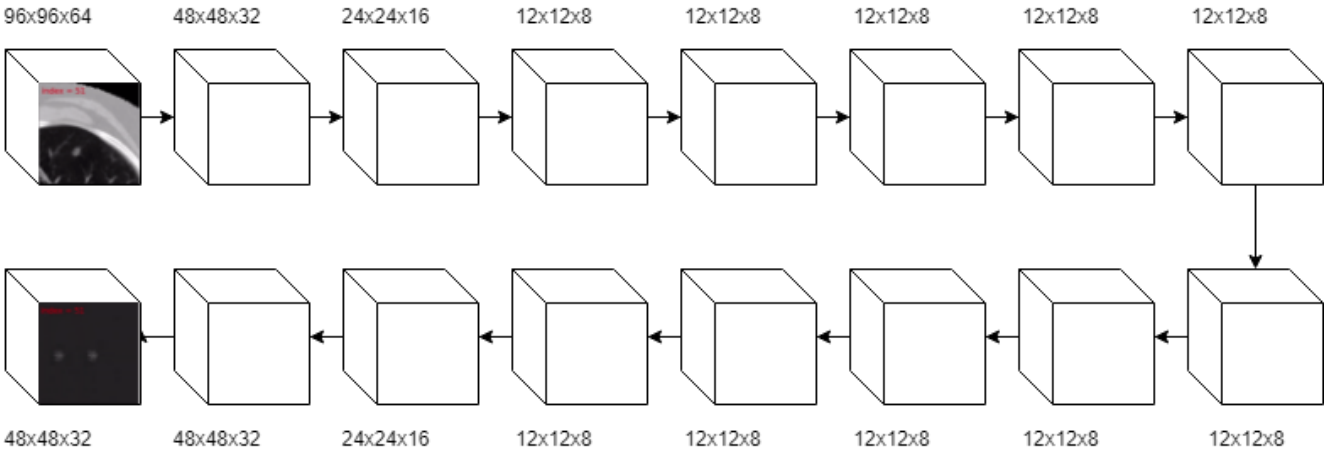

Figure 3: Our deconvNet model.

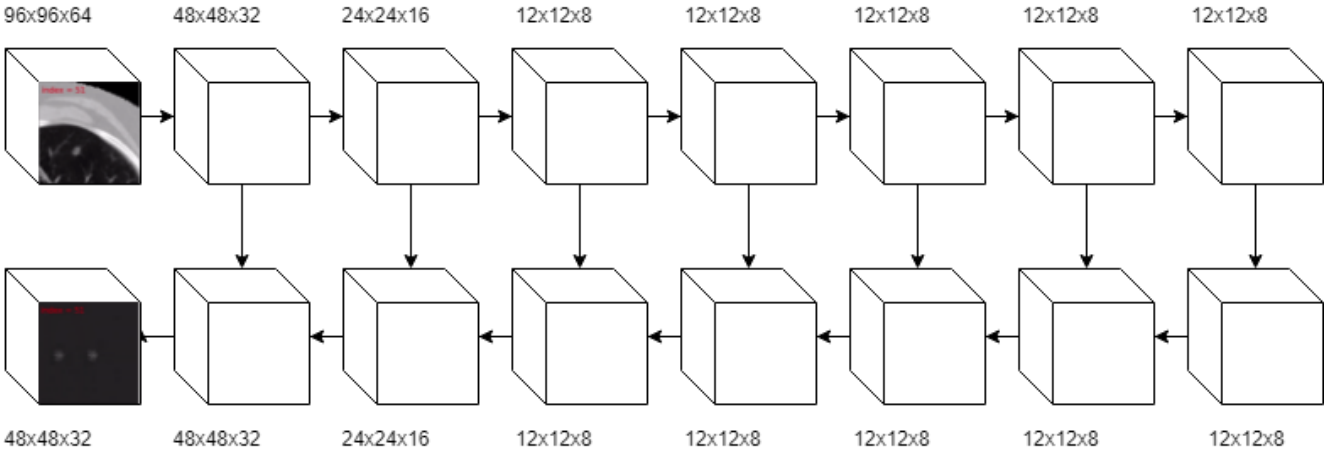

Figure 4: Our U-Net model.

in which the minimum distance from the image center exceeds 60 mm and the area exceeds 6000 mm. The masks were multiplied with the scans.

3.2.3 Spacing Normalization

Because the axis spacing varies from image to image, we normalized the spacing to $0.56 \times 0.56 \times 1$ ($x \times y \times z$).

3.2.4 Nodule Heatmap Generation

We created real binary heatmaps from random samples containing true nodules by creating a mask in which true nodules are represented by values of 1 and everything else is 0. The goal was to train our generator to produce fake binary heatmaps that are comparable to our real binary heatmaps.

3.2.5 Random Sample Extraction

Because of memory constraints, we were unable to use entire images as inputs. Our solution was to random sample $96 \times 96 \times 64$ cubes to be inputted into the network. Cube portions that exceed scan portions were padded with values of 170. We filtered our random samples to overcome class imbalance such that our input 70% contained at least one whole nodule and 30% reflected true random sampling.

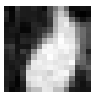

Figure 5: A 2D slice of a 3D random sample.

3.2.6 Random Sample Augmentation

Because overfitting is common with deep learning, extensive data augmentation was performed. We produced additional samples for each augmentation method: flipping the image with respect to an axis chosen at random, rotating the image by an amount from 90, 180, and 270 degrees chosen at random, and if applicable, shifting the nodule between ± 1 and ± 3 pixels along an axis chosen at random.

3.3. DeconvNet Structure

Our deconvNet (Figure 3) follows the typical encoder-decoder structure. The encoding path consists of the input layer and 7 hidden layers, the first 3 consisting of $2 \times 2 \times 2$ max pooling, ReLu activation, and convolution and the last 4 consisting of just convolution. The decoding path consists of the output layer and 7 hidden layers, the first 4 consisting of just convolution and the last 3 consisting of $2 \times 2 \times 2$ upsampling and convolution. The encoding path and decoding path are connected by convolution.

3.4. U-Net Structure

The U-Net structure (Figure 4) is the same as the FCN structure but with skip connections between each layer i and $n - i$ layer, for which n is the total layer count.

3.5. Adversarial Training

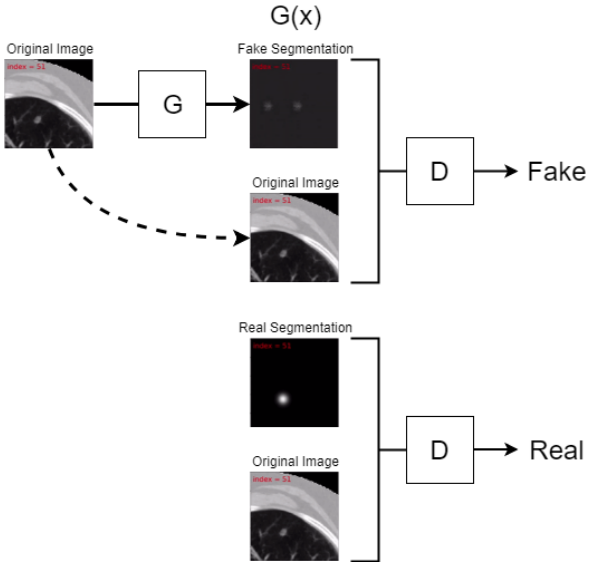

Figure 6: Conditional adversarial network. We created a cGAN for which G was a deconvolutional neural network and a cGAN for which G was a U-Net.

Because we wanted to test the effectiveness of conditional adversarial training on the performance of our deconvNet and U-Net, we used an deconvNet generator and a U-Net generator as part of our deconvolutional-network-based cGAN and U-Net-based cGAN, respectively.

3.6. Experiments

We trained four different network architectures: FCNs, FCN-based cGANs, U-Nets, and U-Net-based cGANs.

3.7. Training Phase

We trained our models for 375 epochs using instance normalization, Adam optimization, stochastic gradient descentand a starting learning rate of 0.0001 that decreased tenfold with every 75 epochs.

3.8. Testing Phase

The trained model is used to predict locations of nodules in our images. For each image, the testing phase outputs a list of n nodule predictions in the form of $\{x_i, y_i, z_i, r_i, p_i\}$, where x_i, y_i, and z_i are the normalized coordinates of the nodule center, r_i is the nodule radius, and p_i is the prediction confidence for nodule i.

4. Results

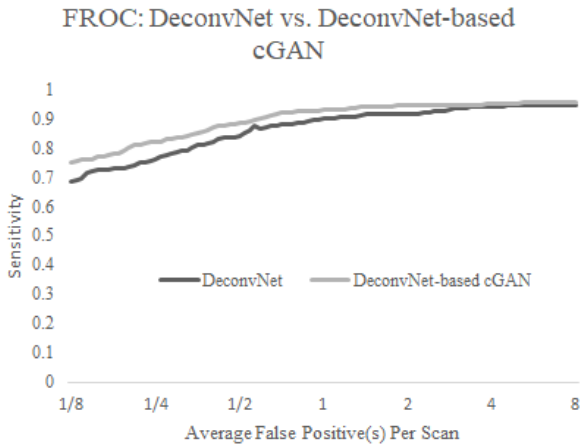

Figure 7: The nodule detection FROC curves for FCN and FCN GAN.

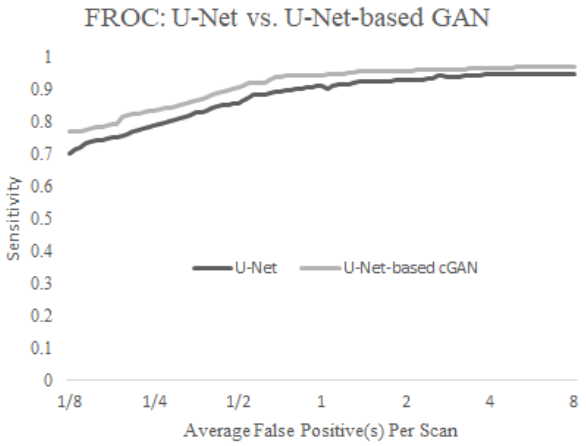

Figure 8: The nodule dtection FROC curves for U-Net and U-Net GAN.

Figures 7 and 8 display the Free Response Operating Characteristic (FROC) curves for the four tested network architectures average over five trials. The average sensitivity values across 1/8, 1/4, 1/2, 1, 2, 4, and 8 false positive(s) per scan are 0.855, 0.890, 0.868, and 0.907 for deconvNet, deconvNet-based cGAN, U-Net, and U-Net fGAN, respectively. Overall, adversarial training increases semantic segmentation sensitivities by approximately 4%.

5. Conclusion

Our results suggest that conditional adversarial training approaches can be improve 3D semantic segmentation results of various neural networks.

References

[1] Lung nodule analysis 2016.

[2] B. H. Hyeonwoo Noh, Seunghoon Hong. Learning deconvolution network for semantic segmentation.

[3] B. H. Hyeonwoo Noh, Seunghoon Hong. Learning deconvolution network for semantic segmentation.

[4] M. M. B. X. D. W.-F. S. O. A. C. Y. B. Ian J. Goodfellow, Jean Pouget-Abadie. Generative adversarial nets.

[5] E. S. Jonathan Long and T. Darrell. Fully convolutional networks for semantic segmentation.

[6] S. O. Mehdi Mirza. Conditional generative adversarial nets.

[7] W. G. T. K. M. F. H. Y.-C. M. Mina Rezaei, Konstantin Harmuth. Conditional adversarial network for semantic segmentation of brain tumor.

[8] T. B. Olaf Ronneberger, Philipp Fischer. U-net: Convolutional networks for biomedical image segmentation.

[9] C. C. Pauline Luc and S. Chintala. Semantic segmentation using adversarial networks.

[10] H.-P. S. K. Y. M. H. B. H. J.-P. R. Simon Kohl, David Bonekamp and K. Maier-Hein. Adversarial networks for the detection of aggressive prostate cancer.

[11] A. K. Vijay Badrinarayanan and I. Roberto Cipolla, Senior Member. Segnet: A deep convolutional encoder-decoder architecture for image segmentation.