SIIM 2018 ANNUAL MEETING
INSPIRING COLLABORATION
May 31 – June 2 | Gaylord National Resort | National Harbor, MD

Improving the Robustness of Deep Learning Models: Application in Lung Nodule Detection

William Hsu, PhD, University of California, Los Angeles; Shiwen Shen, MS; Alex Bui, PhD

Introduction

Automated detection of lung nodules in low-dose computed tomography (LDCT) is a time-consuming and effort-intensive task. Computer-aided detection algorithms have been proposed to automatically identify suspicious nodules. The performance of these algorithms, including recent developments using convolutional neural networks (CNN) and other deep models, may vary significantly when deployed at different institutions. We explored an approach for generating robust lung nodule classification models using a deep CNN, using augmentation methods to improve model transferability.

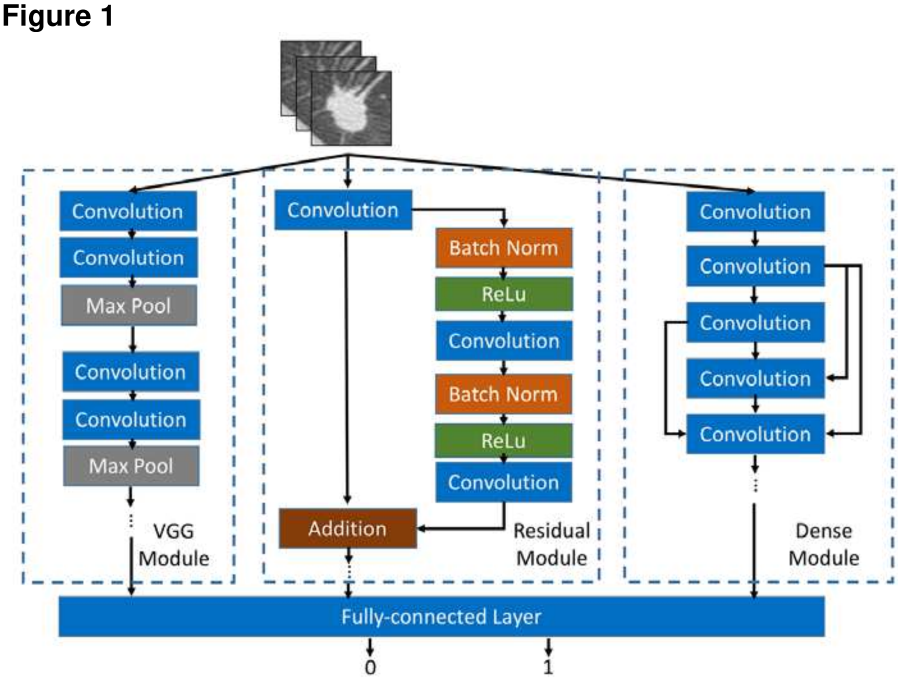
Hypothesis

Introducing simulated variations in data quality during model training improve model robustness.

Methods

The deep CNN was trained on the Lung Image Database Consortium (893 diagnostic chest CT scans) with 6,776 annotated nodules and evaluated on local UCLA cases (158 diagnostic chest CTs) with 158 annotated nodules. Images were normalized to the range of (0,1) from (-1000, 500 HU). Nodule candidates were extracted using multi-level thresholding and morphological operations. A 3D cube of 40 × 40 × 40 mm is generated for each nodule candidate. We designed a hybrid ensemble CNN structure consisting of a VGG module, a residual module, and a dense module. One convolution (conv) unit in VGG module comprises two 3× 3 conv layers and one 2 × 2 max pooling layer. Three conv modules were used in total. The residual module enforces the structure to learn residual functions with reference to the layer inputs. A 49-layer residual structure is used. The dense module consists of three connected convolution blocks, each made up of 12 batch normalizations + conv layer combinations. The output of each module was linked to a fully-connected layer with 1024 rectified linear units and connected to a binary output: nodule/non-nodule. During training, we evaluated different combinations of online data augmentation to introduce noise, rescaling, reslicing, and rotation using data from each 3D cube. This process enhances the inherent variability in the dataset and yields different views of the nodule, mimicking differences in acquisition (e.g., dosages, slice thicknesses).

Figure 1



Results
The models were trained and validated using 686 CT scans from the Lung Image Database Consortium (LIDC). LIDC subset of 207 scans (1,262 nodules and 8,281 non-nodules) was used as a test set. An external evaluation was performed using the UCLA dataset, consisting of 158 CT scans (158 nodules and 3,938 non-nodules). The model achieved an area under the curve (AUC) of 0.994 and accuracy (ACC) of 97.4% on the LIDC data. In the external validation, the model achieved an AUC of 0.955 and ACC of 94.2%. Without the use of data augmentation, the model achieved an AUC of 0.78 and ACC of 74%.

Conclusion
We describe a deep learning-based lung nodule detection approach, investigating the impact of data augmentation to improve model robustness. Training models on simulated examples that introduce noise, rotations, and varying resolutions aids with model transferability.

Statement of Impact
Introducing training cases of variable quality (noise, slice thickness, the field of view) yields models that are better at handling acquisition differences across institutions.

Keywords
deep learning, model performance, data augmentation, pulmonary nodule detection