Duke + DataScience

Classifying Lesions in CT Imagery with Deep Learning

Ricardo Henao, Ben Alexander, Patrick Botros, Amey Chaware, Jerry Liu

Abstract

Since its advent in 1972, computed tomography (CT) imaging has become a standard practice in the treatment of cancer, the assessment of blunt trauma, and imaging of soft tissue generally. Typically, this radiological data must be analyzed manually by trained professionals. One of the principal goals of these radiologists is to identify lesions in different regions of the body. Our research examines whether this process can be automated, in order to provide physicians with a "second opinion" and lead to better diagnoses.

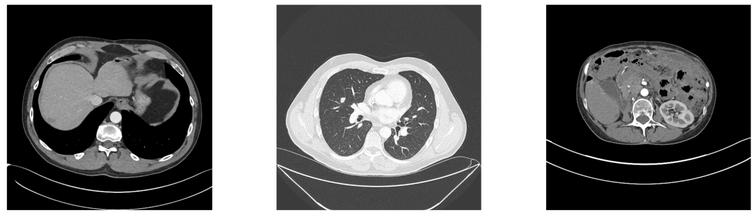
In this Duke +DataScience project, deep learning methods were applied to the DeepLesion dataset [1], a collection of 32,000 CT images with annotated lesions, nearly 10,000 of which are classified by region. Methods were applied to the predictive network to account for the imbalance of lesion types. This predictive model may be combined with a lesion-detection model to aid health professionals in analyzing radiological data.

Objectives

- To train a model that can accurately classify lesions in CT imagery into eight categories (corresponding to the location of the lesion).
- To contribute to the body of research on applying deep learning techniques to medical imaging, which may one day allow automated approaches to aid clinicians in identifying lesions in CT imagery.

Dataset

The DeepLesion dataset, provided by the National Institutes of Health (NIH), contains approximately 10,000 CT scans that are labeled as possessing one of eight possible lesion types: abdomen, bone, liver, lung, pelvis, kidney, mediastinum, and soft tissue lesions. Some examples of these CT scans are displayed below.



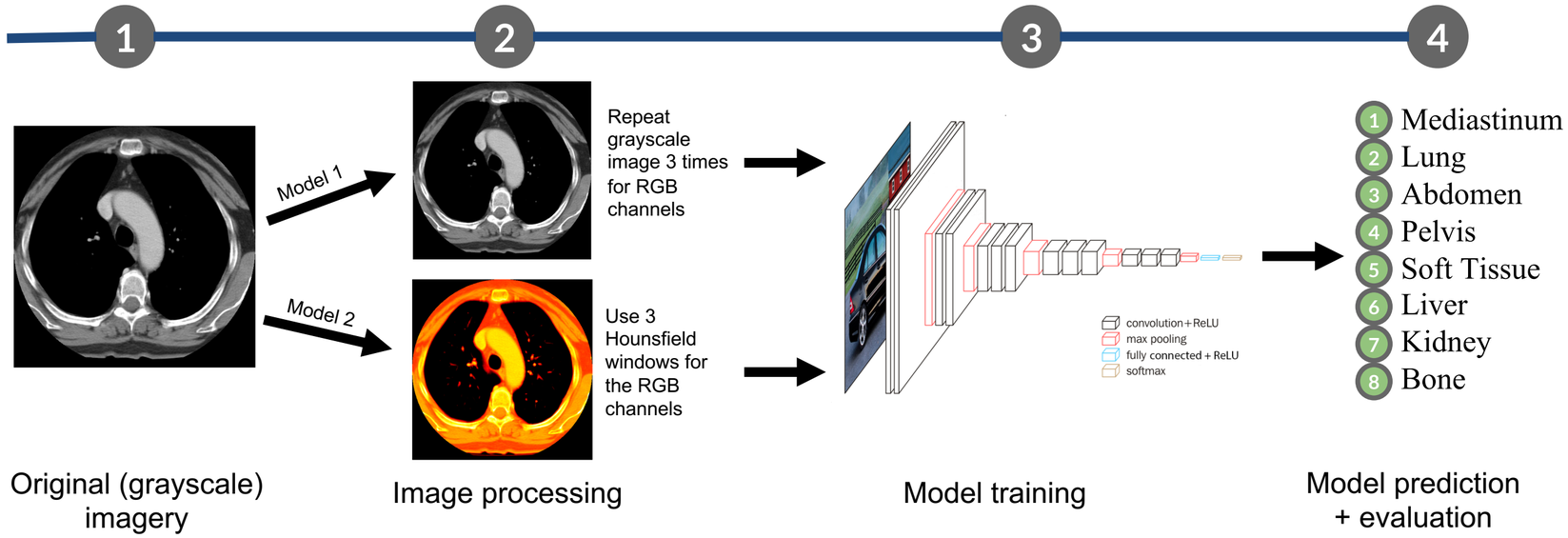
Methods

Our deep learning-based model uses transfer learning to leverage the model weights from a pre-trained convolutional neural network (CNN), known as VGG-16 [2], which was pre-trained on the ImageNet dataset.

Downstream of the VGG-16 layers is a fully-connected (dense) layer, a dropout layer (to prevent overfitting), and finally, a second fully-connected layer to generate the prediction. In essence, our model borrows low-level feature extractors from the VGG-16 network (which identify shapes such as edges, corners, etc.) and then trains a few layers at the end, which use the extracted features to classify the image into the eight categories.

Labeled, unique DeepLesion images were randomly assigned to training (6515), validation (1471), and testing sets (1443), using a 70%-15%-15% train-validation-test split. We trained two main models. The first model repeats the original grayscale CT images three times, once for each RGB channel. The second model uses three different Hounsfield windows, extracted from the original image, as the three RGB channels. Both models use a weighted loss function to adjust for class imbalance.

Project Overview



Results

Because the base VGG-16 model was pre-trained on RGB images, we needed to create three-channel images from the DeepLesion raw data, which are originally grayscale (one-channel) images.

Model 1:

For the first model, we repeated the original grayscale image three times, once for each of the three RGB channels.

Model 2:

For the second model, we used three different Hounsfield windows as the three RGB channels.

Results:

Lesion Type	True Class Proportion	Model 1 Accuracy	Model 2 Accuracy
1 Mediastinum	23.4%	65.9%	79.3%
2 Lung	22.6%	72.5%	81.6%
3 Abdomen	17.4%	61.5%	43.0%
4 Pelvis	14.3%	84.3%	79.4%
5 Soft tissue	8.4%	61.5%	30.7%
6 Liver	6.9%	64.6%	60.0%
7 Kidney	4.8%	43.4%	50.0%
8 Bone	2.3%	50.0%	23.5%

In the table, we have also included the true class proportions, to give a sense of which types of lesions were most prevalent in the ground-truth dataset.

As we can see from the results above, both Model 1 and Model 2 were able to do a fairly good job of identifying lesions of each type. Although we expected Model 2 to perform better than Model 1 due to the extra information provided by the three Hounsfield windows, the results for Model 2 were not clearly better than the results from Model 1.

Conclusions

- **Manual identification** and characterization of lesions in CT scans is **difficult** and **time-consuming** for radiologists. A deep learning model can help physicians make **more frequent** and **more accurate** diagnoses.
- **Two methods** for extracting three-channel images from the raw data are applied, one using the Hounsfield window suggested in the DeepLesion paper and the other using three different Hounsfield windows.
- We obtain overall accuracies of **67.4%** and **65.5%**, respectively, which likely could be improved with further fine-tuning of hyperparameters.
- Since the dataset used is **unevenly balanced**, and because some lesions are easier to identify than others, the accuracies for each type of lesion differ significantly.
- Surprisingly, including information from **multiple Hounsfield windows did not directly improve the performance** of the model as expected, but instead shifted the performances in each lesion type.
- **Future work** could focus on ways to improve overall accuracy **without sacrificing performance** on specific categories. Additionally, future work could **further explore the effect of different Hounsfield windows**, such as trying different thresholds to see whether some thresholds improve performance on certain lesion types.

References

1. Ke Yan, Xiaosong Wang, Le Lu, Ronald M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," J. Med. Imag. 5(3) 036501 (20 July 2018)
2. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

Acknowledgements: This work was supported by the Duke Plus DS (+DS) Program. We thank Duke Research Computing for the use of the Duke Compute Cluster for high-throughput computation. The Duke Data Commons storage is supported by the National Institutes of Health (1S10OD018164-01). Health Data Science at Duke is supported by the Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002553.

Student Showcase: +DS Projects in Medicine and Health Data Science Internship Program

April 25, 2019, Duke University, Gross Hall