Automatic Detection of Multiple Organs Using Convolutional Neural Networks

Elizabeth Cole
University of Massachusetts Amherst
Amherst, MA
ekcole@umass.edu

Sarfaraz Hussein
University of Central Florida
Orlando, FL
shussein@knights.ucf.edu

***Abstract*—We aim to automatically localize multiple organs in a variety of three-dimensional full body CT volumes. We propose performing feature extraction on the CT volumes from the last linear layer of the deep convolutional neural network GoogLeNet, pre-trained on the dataset from the ILSVRC 2014 classification challenge, with subsequent SVM classification. We manually annotated tight bounding boxes around the organs for each patient to use as our ground truth. This method does well when each slice from the CT volumes is divided into large patches and labelled according to their level of intersection with the ground truth. This project has real world applications in fat quantification, radiology, and organ segmentation.**

Keywords—convolutional neural networks; medical imaging; CT; GoogLeNet; SVM; deep learning; organ detection

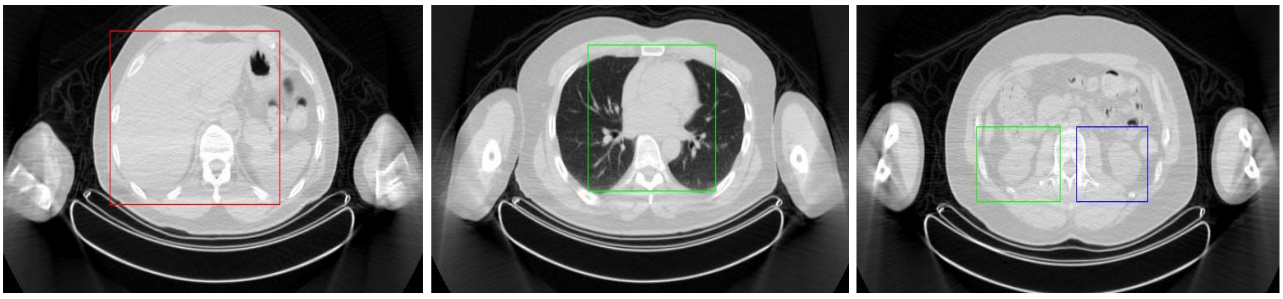
I. Introduction

This paper aims to solve the problem of detecting the location of the liver, heart, and left and right kidneys within a three-dimensional Computed Tomography (CT) volume. Specifically, we seek to return a three-dimensional tight bounding box around each organ for each patient in our testing dataset. We want to distinguish the structure of these four organs from each other and from the rest of the patient, denoted as "background". To do this, we divide each slice of our patients into patches, label these patches, extract features from these patches using the pre-trained convolutional neural network GoogLeNet, and train and test a Support Vector Machine (SVM) using these labels and image features. Our method performs far better when done on larger patches.

II. Dataset

Our dataset is comprised of 44 full-body three-dimensional CT scans, obtained from a hospital. We used 30 patients for training our SVM and 14 patients for testing our SVM. In each volume, the liver, heart, and right and left kidneys were manually annotated using the 3D medical imaging software platform Amira. Tight, three-dimensional bounding boxes were drawn around each organ to denote the ground truth. Each organ's box spanned multiple slices of each patient, so that each slice that contained an organ displayed a rectangular annotation around that organ. In Figure 1, three example slices from one person of their liver, heart, and kidneys are displayed with the bounding box around them, drawn in MATLAB.

Figure 1



Liver | Heart | Right and Left Kidneys

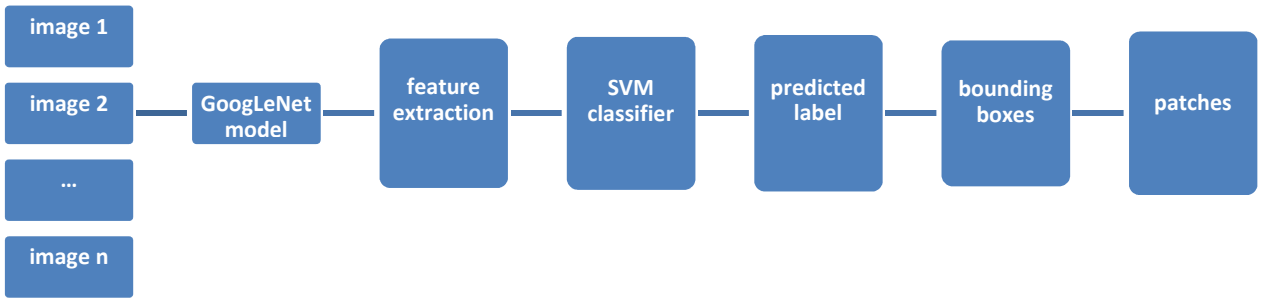
Multiple challenges arise due to the limited size and uniqueness of our dataset. There is not currently a standard dataset for medical imaging, and data is hard to obtain. To only have 44 subjects is an extremely small dataset in comparison to the million or so images other projects and papers utilize. Additionally, these images are very unique as the vast majority of pre-trained convolutional neural networks are trained on more everyday images such as people, cars, and animals.

III. Methodology

A. *Overview*

The pipeline we established to reach our goal in this project involved splitting each slice from the image into patches. We then labelled each patch as liver, heart, right kidney, left kidney, or background. These patches were passed into the pre-trained deep convolutional neural network GoogLeNet. Using GoogLeNet, image features were extracted from the linear layer, which is the last layer before the classification layer. This created a 1 x 1000 feature vector for each patch. A SVM classifier was then trained and tested on these feature vectors and labels. This pipeline is shown in Figure 2.

Figure 2



B. *Software Platforms*

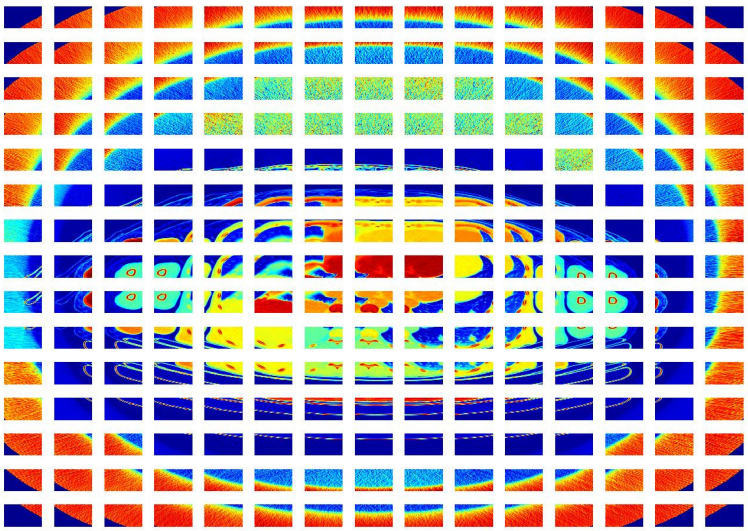
Throughout the course of this project, MATLAB and the deep learning framework Caffe were primarily used. Other experiments were made with Python and the MATLAB toolbox MatConvNet.

C. *Patch Division*

Because this project incorporated the detection and localization of multiple organs, we divided each slice of our CT scans into patches in order to better localize the placement of each organ. We experimented with different sized patches to achieve our goal.

Initially, each slice from every CT scan was uniformly divided into 64 x 64 patches with 50% overlap in the X and Y directions. These patches, if classified correctly, would allow us to draw a tight bounding box around each organ. The patch division of one slice from a single patient that displays the heart is shown in Figure 3.

Figure 3



Each patch was labelled as one of the four organs we are attempting to detect based on if it overlapped 60% or more with the ground truth bounding box. If a patch overlapped less than 60% with any ground truth bounding box, that patch was labelled as background.

After this method of patch division was tested, we settled on using a

different patch size depending on what organ we were searching for. Figure 4 shows the different size of patches based on what organ they correspond to. These patches also have a 50% overlap. These larger patches were now labelled as an organ if it intersected more than 70% of that organ's bounding box.

Figure 4

Organ	Patch Size
Liver	160 x 210
Heart	140 x 140
Right Kidney	110 x 110
Left Kidney	110 x 110

Figure 5

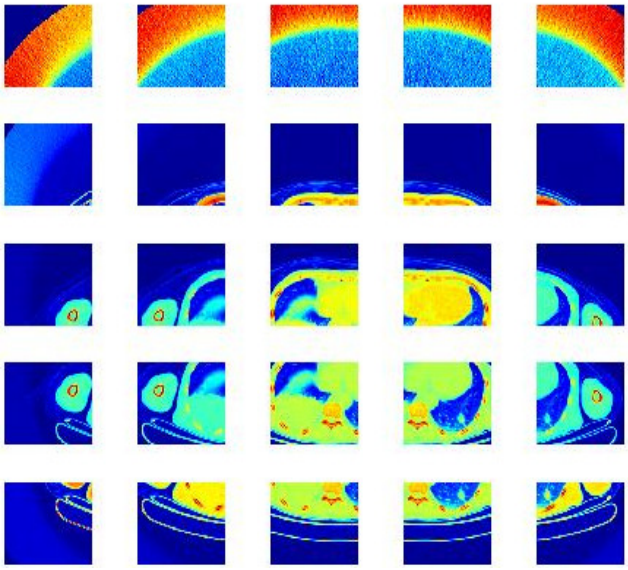
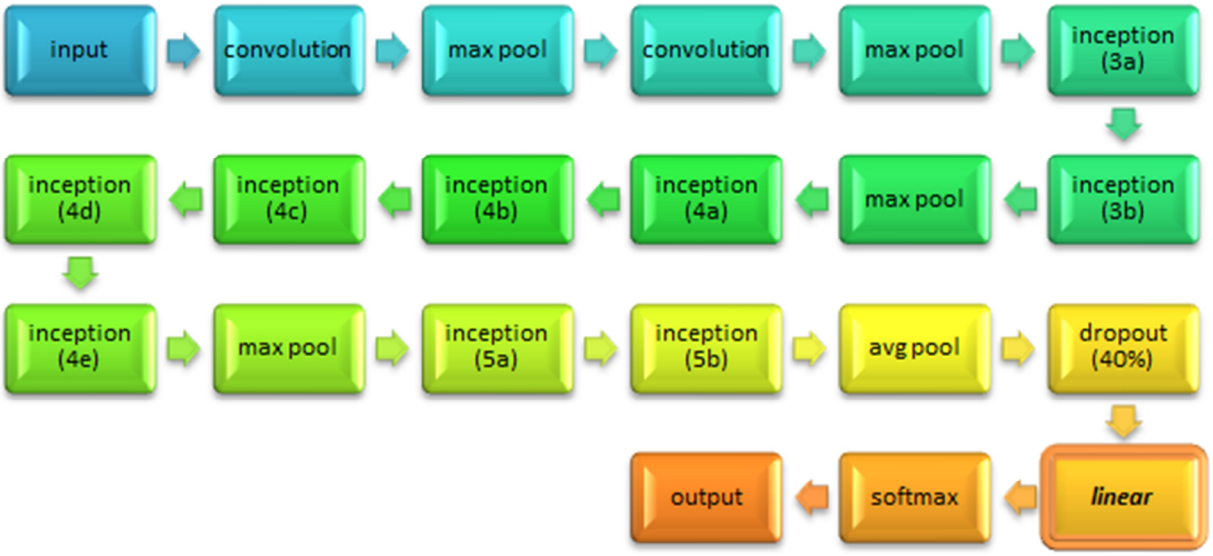


Figure 5 shows an example of these patches for the heart, with the heart displayed in the yellow and green center of the figure.

D. GoogLeNet Structure

The pre-trained convolutional neural network we used for feature extraction is GoogLeNet, which is produced by Google and is 22 layers deep. This model has the current best performance on the ILSVRC 2014 image classification challenge, which contributed to our decision to use this model. We extracted image features from our patches using the second to last layer of this network, which is linear and produces a 1 x 1000 vector output. Figure 6 shows the structure of this deep neural network.

Figure 6



E. Feature Visualization

Using the deep learning framework Caffe, we were able to visualize how different organ patches displayed different filter activations. Figures 7 and 8 show some of the different features for one patient when all patches from

one organ class are passed into GoogLeNet. Figure 7 shows activations from the second convolutional layer, and Figure 8 shows activations from the inception 3a layer.

Figure 7

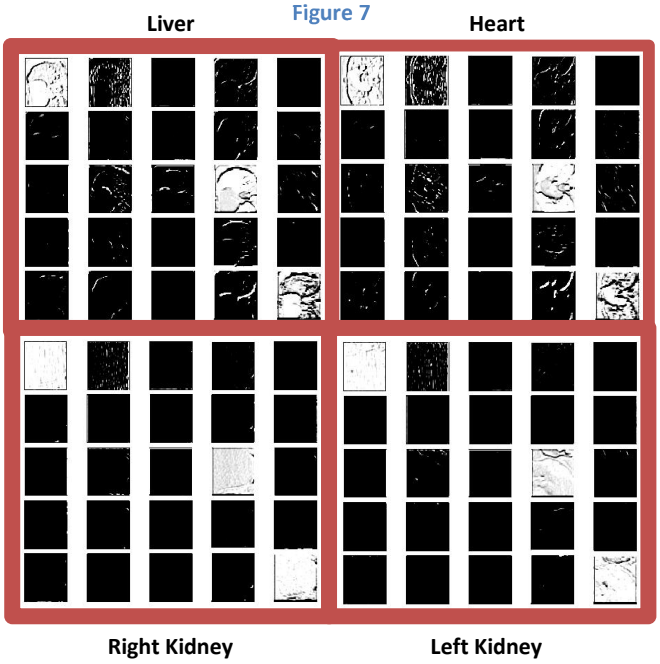
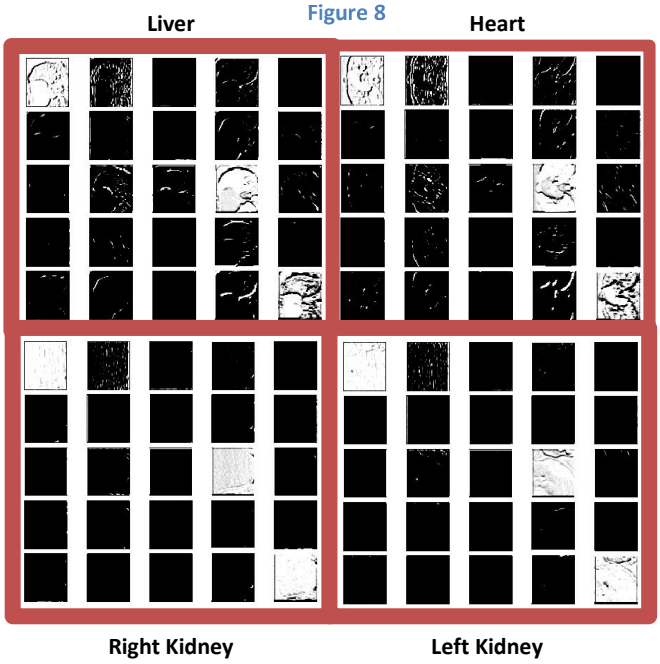


Figure 8



F. *SVM Training and Testing*

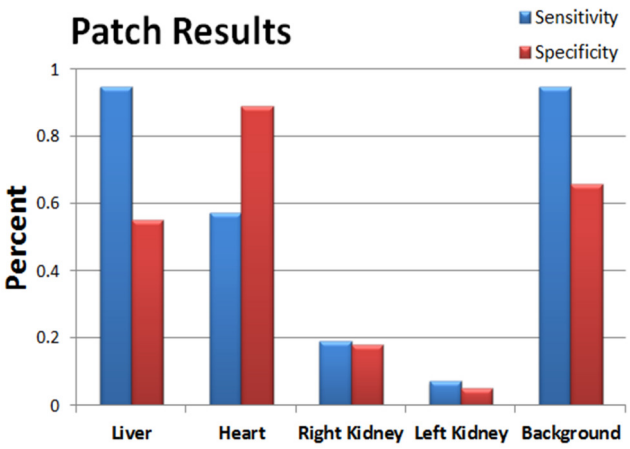
The final step in our pipeline was to train and test an SVM using the feature vector extracted from the last linear layer of GoogLeNet and the label originally given to the patch denoting which organ it displayed. The LibSVM package, along with 30 training patients and 14 testing patients, were used to complete this task.

IV. Results/Discussion

A. *Initial Patch Results*

Figure 9 shows our initial results for the first type of patch division. The blue bars represent sensitivity, or true positive rate, and the red bars represent specificity, or true negative rate. This method does fairly well, over 50% true positive and true negative rate, for the liver, heart, and background patches. However, this method does not perform well at all for the right and left kidneys, with true positive and true negative rates for both kidneys falling under 20%.

Figure 9



B. *Larger Patch Results*

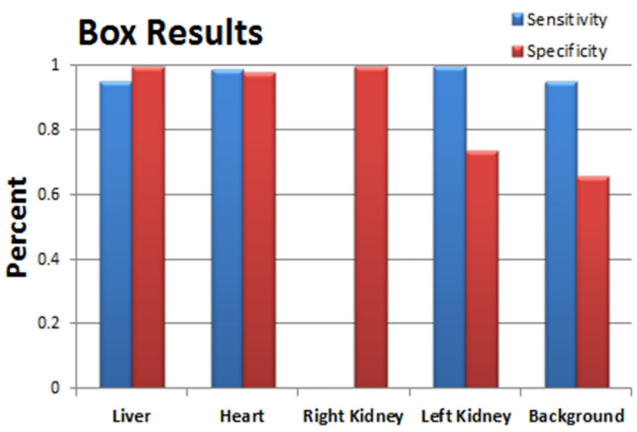
Figure 10 shows our secondary results for the larger type of patch division. Our true positive and true negative rates for every organ are much improved, while our true positive and true negative rates for background stay about the same. Unfortunately, all right kidneys were classified as left kidneys. However, total kidney accuracy greatly improved. This could be due to the kidneys looking very similar to each other, or due to the smaller amount of kidney data compared to other organ data, due to their smaller size than other organs. This made sense in the context of our dataset as most likely the features

extracted from the whole organ would be more discriminative than the features extracted from a small patch of an organ.

C. *Improved Patch Results*

After realizing that a larger patch size provides far more accurate results, we attempted to use the SVM trained on 64x64 patches and test it on the patches now classified as organs, when divided into 64x64 patches. This gave the SVM less data to search through. We did this because, in many cases, too much of the patch classified as an organ could have too little of an intersection with the ground truth bounding box. However, this only improved the patch results slightly, as shown in Figure 11.

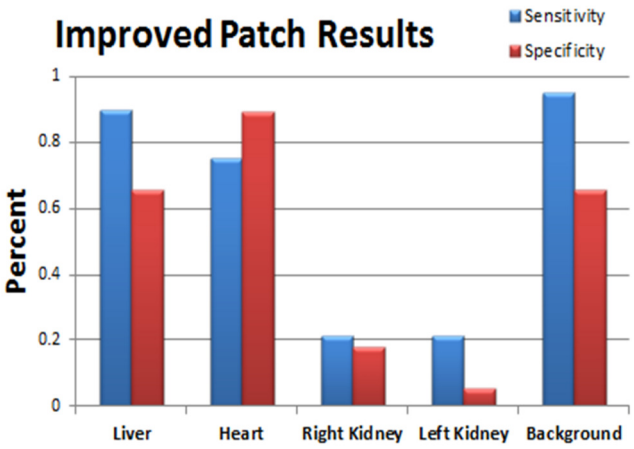
Figure 10



D. *Conclusion*

Over the course of this project, a working model of automatic multiple organ detection was created. Extracting features from the last linear layer of GoogLeNet from patches that encompassed the entire organs we were searching for gave us the best results, over other models and other layers of GoogLeNet. These results could have been better, had we possessed more annotated data.

Figure 11



V. Future Work

This project could go in many directions in terms of improvements. Using this past work, the confidences of the two-dimensional patch results could be fused using Conditional Random Fields. Contextual information, such as distance priors, could be used to improve accuracy. An additional idea we had to improve the results was to use the GoogLeNet deep learning features with superpixels, or to extend this into the third-dimension and use supervoxel segmentation.

We did not have enough data to train a convolutional neural network ourselves and have it produce favorable results. However, with more data, a two-dimensional and three-dimensional convolutional neural network could be trained and tested to see if this produces better results than the ones we discovered. Challenges arise with training a three-dimensional convolutional network in terms of Caffe supporting three-dimensional convolutions.

VI. References

[1] "Alzheimer's Disease Neuroimaging Study Launched Nationwide by the National Institutes of Health." PsycEXTRA Dataset (2006).

[2] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J., "Region-based Convolutional Networks for Accurate Object Detection and Segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions.*

[3] Ji, Shuiwang, Wei Xu, Ming Yang, and Kai Yu. "3D Convolutional Neural Networks for Human Action Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal. Mach. Intell. 35.1 (2013): 221-31.

[4] Roth, Holger. "A New 2.5D Representation for Lymph Node Detection Using Random Sets O." F Deep Convolutional Neural Network Observations.

[5] Schroff Florian, James Philbin, and Dmitry Kalenichenko. "FaceNet: A Unified Embedding for Face Recognition and Clustering."

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. "Going Deeper with Convolutions." CVPR (2015): Computer Vision Foundation.