Multiple Organ detection in CT Volumes Using Random Forests

Daniel Donenfeld, Sarfaraz Hussein, Ulas Bagci, Mubarak Shah

Abstract

We propose a method for the localization of organs in 3D computed tomography (CT) scans. Our algorithm worked on a supervoxel segmentation to localize the heart, liver, and kidneys. Voxel labeling, a method used to segment organs and other anatomical structures in medical images, was extended to supervoxel labeling, where each supervoxel was given a class label. This reduced the search space significantly, and increased the perceptual meaning, as the supervoxels adhered to edges, and were more uniform. Hand crafted features were extracted from the supervoxel, and classified using a random forest.

Introduction

Localizing anatomical structures in medical images is the process of finding and determining the extent of the structure. The approach described in this paper was designed for 3D CT images, so it returns a rectangular bounding volume which surrounds the organ. This problem is important as a step in other pipelines, such as organ segmentation, in pathology detection, and in fat quantification. The current state of the art approach uses regression forests to use each voxel to contribute to a prediction of the location of organs. It is a fast algorithm, taking around 6 seconds, and has better accuracy than a registration based approach. The approach we propose is a classification based method, which works on a supervoxel segmentation. Voxels do not have perceptual meaning, and looking at a single voxel on its own would not give any indication what that voxel was a part of. Supervoxels are an over-segmentation technique which creates groupings of similar voxels to create regions which have meaning and are uniform in intensity. Supervoxels, as they are uniform regions, adhere to boundaries between objects in images. The larger size and adherence to edges makes them contain more relevant information in each supervoxel compared to a voxel. There is a reduction in search space, as there are far fewer supervoxels, and as each supervoxel adheres to intensity edges, each voxel contained is the same classification. Local gradient and contextual features were used to classify the supervoxels.

Method

The data used was 30 CT images, stored as 3D arrays of intensity values, which were annotated with tight bounding boxes around the liver, heart, left kidney, and right kidney. The data was split into 20 for training, and 10 for testing.

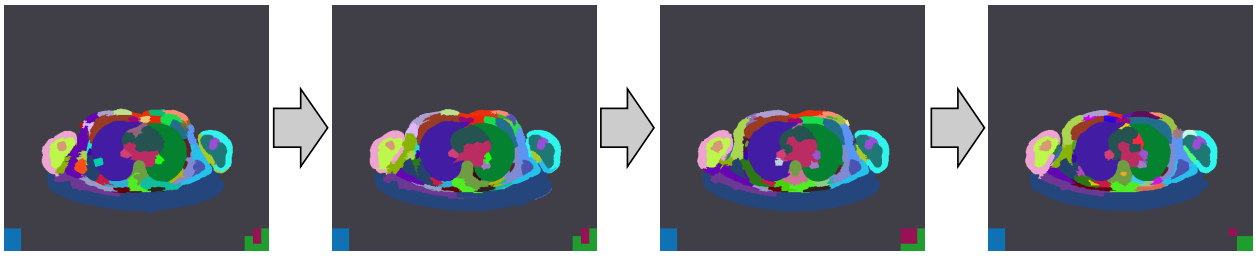


Figure 1: Progression of supervoxel segmentation across slices containing the heart

Each image was first over-segmented into 3000 supervoxels, using the method described in *Spatio-Temporal Object Detection Proposals.*[1] An example of the supervoxels generated can be seen in figure 1. From each supervoxel, features were extracted. The local gradient features examined were Histogram, 3D Haar-like[2], 3D SIFT[3], Gray Level Co-Occurrence Matrix Features[4]. All of these features were used in addition to the center of the supervoxel appended to the feature vector. For the histogram features, mean, variance, minimum and maximum were used. The non-local features were the mean of patches surrounding the supervoxel.[5] To extract contextual features, a standard set of 3D rays was generated at random, all of which were in a sphere of specified radius. To extract the contextual feature vector for a supervoxel, the vectors were positioned at the center of the supervoxel, and a 5x5x5 voxel patch was placed at the end of each one. The mean of each patch made up the contextual features. From the ground truth data, each supervoxel was given a class label if it was inside one of the four organ bounding boxes, otherwise it was given a background label. A random forest classifier was trained on the features with the ground truth labels, and was tested on the testing set.

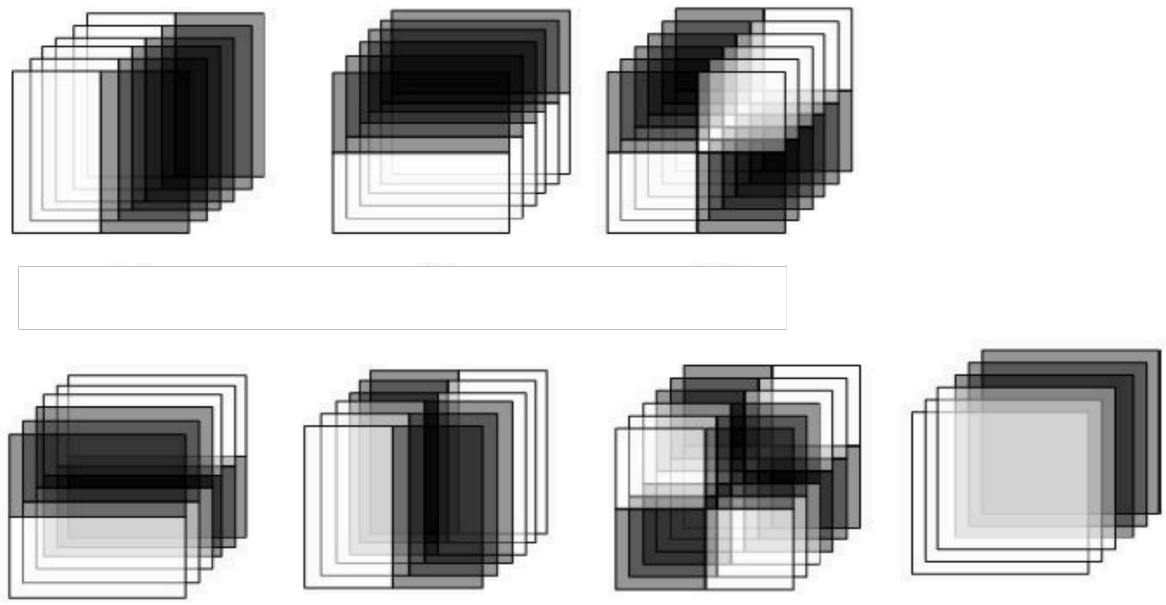


Figure 2: The masks used for the 3D Haar-like features

Results

The classifier was first trained on each feature separately, and with 3D Haar-like, 3D SIFT, and Gray Level Co-Occurrence Matrix Features, all the supervoxels were classified as background. The histogram correctly classified less than 20 percent of each organ correctly. Weights were added to the classifier, and it increased the classification in the organs slightly with the histogram features and 3D SIFT features. For the 3D Haar-like features and Gray Level Co-Occurrence Matrix Features, the percentage of liver went up significantly, while the percentage of background classified correctly went down significantly. To improve the results further, contextual features were added. In all cases, the contextual features significantly improved the results, for the liver and heart around 50%, and for the kidney around 15% percent, seen in figure 3. The 3D Haar-like features, displayed in figure 2, had the best results of the features tested. A visual of the classification results is shown in figure 3, showing the correct classification of the kidneys, with the boundary extending past the organ.

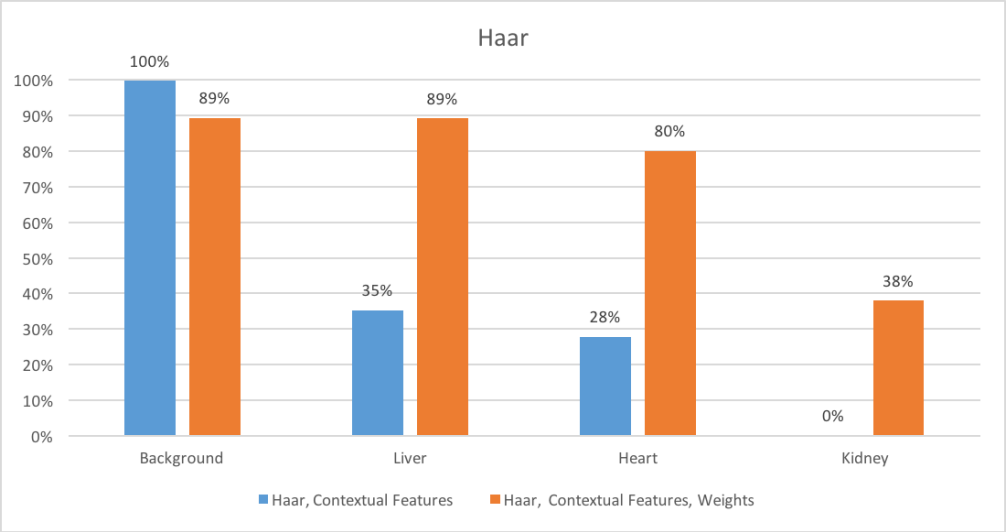


Figure 3: Classification results using 3D Haar-like features with 20 training patients and 10 testing patients

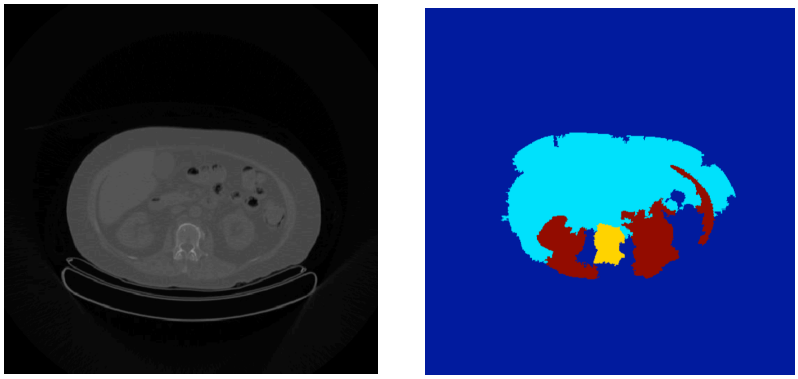


Figure 4: An example of classification. Light blue is liver, red is kidney, and yellow is heart

Discussion

Using only a local feature caused most of the organs to be misclassified. This is due to the small ratio of organ to background data, which was around 3 percent. Removing the background and training and testing yielded reasonable classification results, which ruled out the features being non-discriminative. Using weights only marginally improved the problem, and adding contextual features significantly improved the results. The contextual features included relative positions of the organs, which the classifier was able to train on.

There are few misclassifications far away from the organs, however near the organs many supervoxels are misclassified. This caused the bounding boxes not to tightly surround the organs as intended. This is most likely due to the bounding volume used for ground truth supervoxel labeling, included many supervoxels not contained in the

organ. For example the bounding box for the heart contains some supervoxels from the lung, so the results indicate some supervoxels in the lung are misclassified as heart, instead of background. One potential way to fix this would be to, use organ segmentations for the training data instead of bounding boxes. Another method to fix it could be a preprocessing of the training data, potentially by performing clustering.

Conclusion

The localization of organs is achieved using a classification based approach, in which the search space is reduced. The algorithm is effective at determining the location of organs, but overestimates the extent of the organ, as described above. Future work could be to improve the supervoxel segmentation, as it was a key part of the algorithm. Another future work could be to reduce the search space using hierarchical anatomical structure.

References

[1] Dan Oneata, Jerome Revaud, Jakob Verbeek, Cordelia Schmid. Spatio-Temporal Object Detection Proposals. David Fleet; Tomas Pajdla; Bernt Schiele; Tinne Tuytelaars. ECCV 2014 - European Conference on Computer Vision, Sep 2014, Zurich, Switzerland. Springer, 8691, pp.737-752, <10.1007/978-3-319-10578-9 48>. <hal-01021902v2>

[2] Cui, X., Liu, Y., Shan, S., Chen, X., & Gao, W. 3D Haar-Like Features for Pedestrian Detection. Multimedia and Expo, 2007 IEEE International Conference on Multimedia and Expo

[3] Paul Scovanner, Saad Ali, and Mubarak Shah, A 3-Dimensional SIFT Descriptor and its Application to Action Recognition, ACM MM 2007.

[4] 3D statistical texture algorithm [Computer software]. (2008). Retrieved July 11, 2015, from http://www.mathworks.com/matlabcentral/fileexchange/19058-cooc3d/content/cooc3d.m

[5] Dang, K., Yuan, J., & Tiong, H. (2013). Voxel labelling in CT images with data-driven contextual features. 2013 IEEE International Conference on Image Processing.