Ultra-Orthodox Techniques Vs. SegNET for COVID-19 Patients' Lungs CT Scan Semantic Segmentation

Saood, A.
Mechatronics program for the Distinguished Distinction and Creativity Agency
Syria
adnan.saood@outlook.com

***Abstract*—This article discusses two structurally different techniques for semantically segmenting computerized tomography scans of COVID-19 positives: The histogram-dependant global thresholding and OTSU's method, and the DNN structure SegNET. Results display a noticeable overpass of SegNET compared to the other methods in terms of binary and multi-class classification.**

***Index Terms*—COVID-19, Computerized Tomography, SegNET, OTSU, Global thresholding**

I. Introduction

COVID-19, AKA "the pandemic of the century", is a widespread illness causing thousands of death daily worldwide. The early diagnosis of it proved to be one of the most effective methods of infection tree pruning [8]. The large number of patients is rendering health care systems in many countries overwhelmed, thus, a trusted automated technique for identifying and quantifying the pneumonia-effected lung regions would be beneficial.
Radiologists has identified three types of irregularities related to COVID-19: (1) ground glass opacification, (2) consolidations, and (3) pleural effusions [1].

This document demonstrates classic methods of imagery segmentation compared to the novel SegNET deep neural network (DNN). Thresholding is a simple method that has been long used for segmentation. Its an extremely undemanding approach with no training needed what so ever [4]. For an automatic way of threshold selection, both the global thresholding and OTSU's method were implemented [4].

SegNET was originally designed for scene understanding applications [2], which makes the task of training it for the complex task of COVID-19 related pneumonia's fractal-like shape segmentation a novel approach. Reference [3] trained a modified version of SegNET for liver tumor segmentation and reached a close-to-perfect accuracy, yet segmenting liver tumor is easier than pneumonia considering its spatial consistency and convex shape visible in their work.

The dataset and algorithms used are in Section II. The comparison was carried out in two versions: the first version of the systems were designed to segment the CT scans into **(regular, irregular)** regions detailed in sections III-B and IV-A. The second versions are multi-class classifiers detailed in III-C and IV-B.

II. Materials and Method

A. Dataset

The dataset used in this work is produced by the Italian Society of Medical and Interventional Radiology. 100 one-slice CT scans are provided in a resized 512×512 dimensions. not all the CT slices are taken in the exact position regarding the patients' organs, thus some scans show less lung region than others. the labels are already compiled into a NIFTI with proper documentation by the author. See [7].

Classes' pixel count (total number of pixels in a class) and image pixel count (total number of pixels in images that had an instance of a class) show extensive disparity in representation; the dominant class is greater in order of $1E + 3$ than the least represented class. We note here that class C_0 does not represent only the portions of the lungs unaffected bu pneumonia, it also represents the lung-enclosing tissue. See table I.

TABLE I
Dataset Class Metrics

Class Name	Metrics	
	Pixel Count	***Image Pixel Count***
C_0	$2.4394E + 07$	$2.6214E + 07$
C_1	$1.1965E + 06$	$2.5166E + 07$
C_2	$5.8921E + 05$	$2.0447E + 07$
C_3	$3.4265E + 04$	$6.5536E + 06$

The dataset website offered a mask to segment the lungs. Before lungs' extraction a Gaussian Blur of kernel 5×5 was allied following literature suggestion [4]. See Fig. 1 for a sample.

B. Image Thresholding

Programming the thresholding techniques was a straightforward application of their equations.

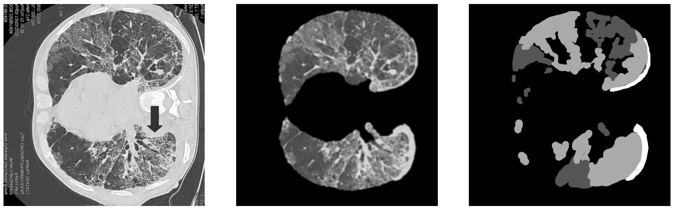

Fig. 1. CT scan (left), masked lungs (middle), labeled classes (right).

1) Global Thresholding: The global thresholding technique depends only on the image's histogram to find a suitable threshold that divides a bimodal histogram into two bells. We first calculate the image's normalized histogram p_i, then an iterative operation takes place to calculate the threshold level that satisfies (1):

$$T = \frac{1}{2}(m_1 + m_2) \tag{1}$$

Where m_1 and m_2 are each class's mean intensity value computed using (2):

$$m_1 = \sum_{i=0}^{T} p_i \quad \text{and} \quad m_2 = \sum_{i=T+1}^{L-1} p_i \tag{2}$$

The term L represents the number of possible intensity values.The iterations continue until reaching a predetermined $\Delta T = T_{i-1} - T_i$

2) OTSU's Thresholding: OTSU's method is an unsupervised method for thresholding images. It is designed to maximize the separability of classes. Although being computationally expensive for multi-class thresholding, it performs well generally. The algorithm solves for the following minimization task:

$$\sigma_b^2(k_1^*, k_2^*, ..., k_{K-1}^*) = \max_{for\ k_i} \sigma_B^2(k_1, k_2, ..., k_{K-1}) \tag{3}$$

Where:

$$\sigma_B^2 = \sum_{k=1}^{K} P_k(m_k - m_G)^2 \tag{4}$$

The term P_k is the probability of class k occurring in the image, m_k is the mean intensity of class k and m_G is the mean intensity of the whole image.
It is obvious that as k (the number of classes to segment to) increases, the minimization problem in (3) gets more computationally expensive.

C. SegNET Deep Neural Network

1) Design: The design of the DNN was carried out using MATLAB's deep network designer. Its a SegNET network with encoder-decoder depth of three. The encoder layers are identical to the convolutional layers in VGG16 network. The creators removed the fully connected layers to reduce complexity. The DNN's architecture is detailed in Fig. 2.

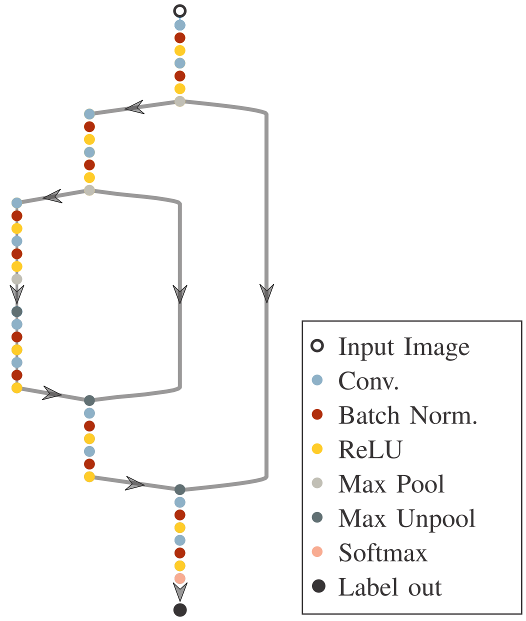

Fig. 2. SegNET of encoder-decoder depth 3.

2) Training: Training the neural network was done using ADAM stochastic optimizer due to its fast convergence rate compared to other optimizers [6]. The input images were resized to 128×128 to reduce training time by a factor of sixteen. The 100 dataset instances were divided to training, validation, and testing with proportions 72%, 10%, and 18% respectively. The Initial learning rate was $1E-3$, with a mini-batch size of eight. Training accuracy and loss are visible in Fig. 3,4.
In spite of class imbalance present in the dataset (see section I), class weights were handed over to the pixel classification layer (last layer) in the network. The weights were calculated using median frequency balancing. Although class weighting helped the DNN converge to an appropriate point,the small number of images in the dataset was always driving the optimizer towards an overfitting weights, an added data augmenter with appropriate preferences was used to help solve this issue.

During the training procedure, the hyper-parameters were tuned heavily to ensure a non-overfitting system, this is interpreted in the results section.

III. Results

The comparison below lays out the performance of algorithms detailed in Section II-A. a binary classification of classes regular, irregular is in subsection B, and the full range of classes were classified in subsection C.

A. Evaluation Criteria

To fully quantify the accuracy of our results, we utilized criteria obtained from literature [3], [7]. They are sensitivity, precision, Sørensen-Dice coefficient, and the visual representation of the normalized confusion matrices. See equations (5) to (7).

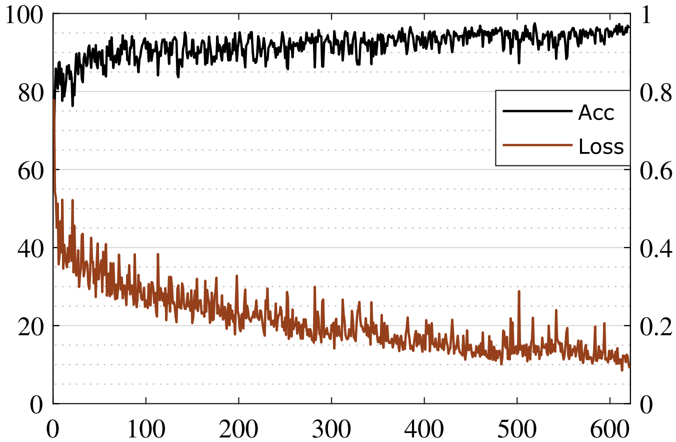

Fig. 3. Training Accuracy and Loss for binary classification

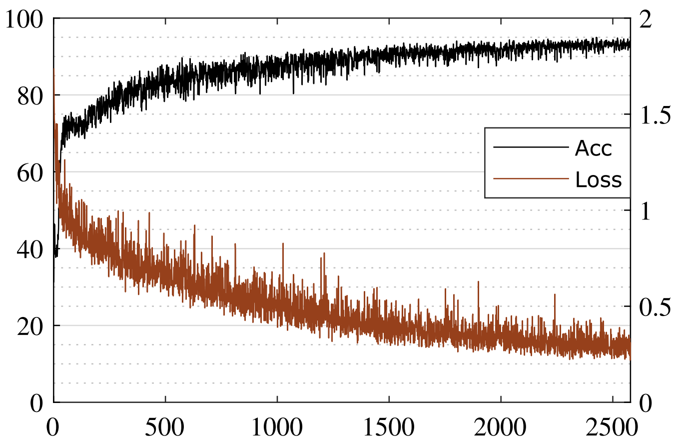

Fig. 4. Training Accuracy and Loss for multi-class classification

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Sørensen-Dice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{7}$$

Those criteria were selected because of the dataset's imbalance nature discussed in II-A.

B. Binary Classification

For the binary classification problem, three classifiers were tested. the results are visible in II. The confusion matrices are computed and plotted in Fig 5 to add a visual sense of performance. We note that the thresholding techniques were tested only against the test data of SegNET classifier.

C. Multi-Class Classification

In order to compare the classifiers (multi-OTSU, SegNET) properly, the multi-OTSU algorithms was benchmarked only on the test data of the SegNET classifier. Results are in table III, and confusion matrices are in Fig. 6.

TABLE II
BINARY CLASSIFIERS EVALUATION METRICS

Classifier Name	Metrics		
	Precision	**Sensitivity**	**Dice**
Global Thresh.	0.779	0.735	0.757
OTSU Thresh.	0.754	0.706	0.730
SegNET	0.941	0.933	0.937

Fig. 5. Confusion Mat. for Global Thresh. (left), OTSU thresh. (middle), SegNET (right)

IV. Discussion

A. Binary Classifiers

It can be seen table II that the SegNET outperforms the classic histogram-dependent methods in terms of precision, sensitivity, and Sørensen-Dice coefficient. Both the OTSU and global thresholders performed exceptionally well in classifying the ”regular” class, but it is expected since the intensity value of that class lies in the lower end of every sample's histogram. A stacked histogram of all the samples was plotted in Fig. 7 supports this.
The results states in a quantifiable manner the reliability of the SegNET in distinguishing between the regular and irregular classes, i.e. ill portions of the lungs. Further tests involving a wider dataset is likely to confirm this. Finally, a multi-parameter thresholding algorithm may be regarded as a viable solution to the binary classification problem.

B. Multi-Class Classification

Fig. 7 strongly express the multi-modality of the samples' histograms. The multi-OTSU classifier was able to identify the true negatives of classes C_1, C_2, C_3 better, that is because class C_0's dominance of order $1E+3$ lies in the lower end of the histogram.
The expected-actual statistics visible in Fig. 6 strongly

TABLE III
BINARY CLASSIFIERS EVALUATION METRICS

Classifier Name	Class Name	Metrics		
		Precision	**Sensitivity**	**Dice**
Multi-OTSU	C_0	0.799	0.789	0.794
	C_1	0.456	0.336	0.386
	C_2	0.382	0.410	0.395
	C_3	0.525	0.630	0.573
SegNET	C_0	0.830	0.932	0.878
	C_1	0.538	0.730	0.620
	C_2	0.340	0.400	0.368
	C_3	0.862	0.293	0.437

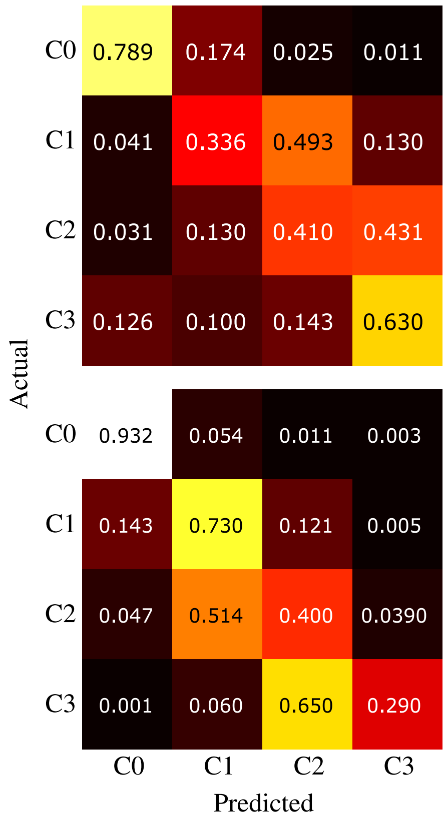

Fig. 6. Confusion Matrices for Multi-OTSU (top), and SegNET (bottom)

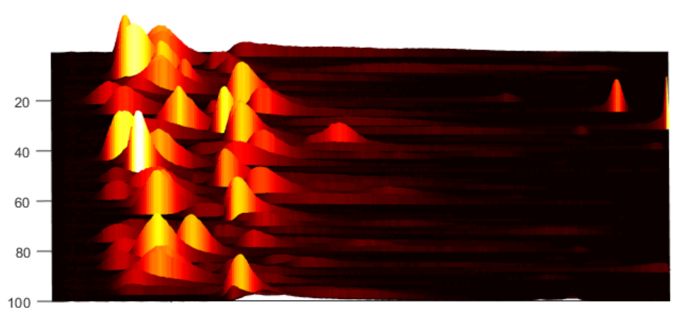

Fig. 7. Stacked histograms of dataset samples

indicates how simple thresholding the picture can extract classes C_0 and C_3. In fact, batch-thresholding the same test data using the arbitrary thresholds of $[69, 110, 171]$ achieved a higher true positive count than the multi-OTSU classifier.

The performance of the SegNET was not satisfactory either, the convergence of a randomly initialized deep neural network by ADAM using 100 data instances was often reaching an under-generalized DNN. Since the hyper-parameters were fine-tuned using validation data, no other conclusion for the low performance was drawn other than the dataset's examples are simply not enough for a multi-class SegNET network, but yielded satisfactory results for training a binary classifier. We strongly advice the use of extra data examples as soon as they get collected from COVID-19 patients which would allow the DNN to converge to a more generalized minima.

Lastly, we present sample of two semantically segmented images in Fig .8. The difference in performance is visually interpretable.

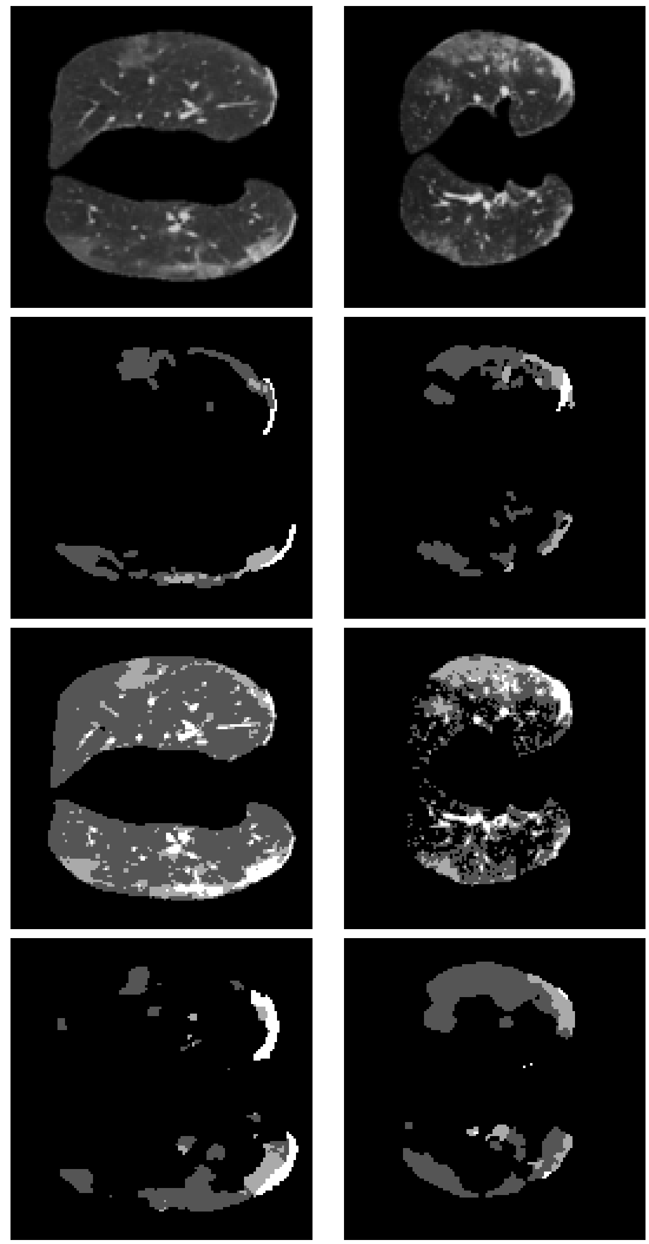

Fig. 8. From top to bottom: CT scans, labels, multi-OTSU's output, SegNET's output.

References

[1] H. Shi et al., "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," The Lancet Infectious Diseases, vol. 20, no. 4, pp. 425–434, Apr. 2020, doi: 10.1016/s1473-3099(20)30086-4.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/tpami.2016.2644615.

[3] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, "Liver Tumor Segmentation in CT Scans Using Modified SegNet," Sensors, vol. 20, no. 5, p. 1516, Mar. 2020, doi: 10.3390/s20051516.
[4] R. Gonzalez and R. Woods, Digital Image Processing (3rd Edition). USA: Prentice-Hall, Inc., 2006.
[5] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[6] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[7] "COVID-19," Medical segmentation. [Online]. Available: http://medicalsegmentation.com/covid19/. [Accessed: 22-Apr-2020].
[8] Meizhu Chen et al., "Key to successful treatment of COVID-19: accurate identification of severe risks and early intervention of disease progression." Cold Spring Harbor Laboratory, 11-Apr-2020, doi: 10.1101/2020.04.06.20054890.