Ground Glass Opacity Detection Using Fully Convolutional Neural Networks

Victoria Mazo[1], Itamar Tamir[2], Eyal Toledano[1], and Eldad Elnekave[1]

[1] Zebra Medical Vision, Shfaim, Israel,
victoria.mazo@zebra-med.com,
[2] Rabin Medical Center, Campus Golda, Israel

Abstract. Ground Glass Opacity (GGO) is a radiologic feature which appears as a hazy region with vague boundaries on Chest Computed Tomography (CT) scans. GGO is a non-specific sign seen in various pathologies, most commonly: alveolar inflammation, infection, hemorrhage, edema or cancer. Relative to more discreet findings, GGO is often considerably more subtle and thus overlooked. We present an automated method for detection of GGO in CT scans based on a Fully Convolutional Neural Network. We utilized segmentation of axial CT reconstructed images to reduce the number of CT studies required as training data in order to obtain high accuracy of GGO detection (96.9%) using a Deep Learning technique. We explore two architectures of Fully Convolutional Neural Networks: U-Net and Fully Convolutional DenseNet. DenseNet-like network is first applied to the Medical Imaging domain and achieves superior detection accuracy due to a higher layer connectivity within a network. We report results of GGO binary classification per axial slice and measurement of slice segmentation goodness (Dice score). The algorithm is constructed to be applicable to any Chest CT scan, allowing for variations in data acquisition protocols such as inspiration/expiration imaging and technical acquisition variations which may result in the appearance of the lung tissue.

Keywords: Ground Glass Opacity (GGO), Deep Learning, Medical Imaging, Computer aided detection (CAD), Segmentation, Chest CT

1 Introduction

Ground Glass Opacity (GGO) is a radiologic feature which appears as a hazy, vaguely bounded region in the lungs on Chest Computed Tomography (CT) scans. GGO is a non-specific sign which indicates partial replacement of alveolar air by low-density material such as inflammatory fluid, infection, hemorrhage or cancer cells. GGO is also frequently seen in various progressive interstitial lung diseases [1, 2]. Focal regions of GGO are particularly suspicious for Bronchoalveolar carcinoma (BAC), persistent regions of GGO are associated with malignancy in up to 75% of cases [2].

Relative to more discreet and solid lung nodules, small regions of GGO are often considerably more subtle and thus prone to being overlooked [3]. Changes

in diffuse GGO patterns in the lungs are relevant for pulmonary management but difficult to characterize due to the irregular shape and boundaries of the involved tissue. It is important to emphasize the challenge in identifying GGO, with its varied appearances, in the context of physiologic variations of lung aeration. Degrees of lung inspiration and even patient positioning can create non-pathological regional variations in lung density. Within these, abnormal GGO must be isolated. Here we describe an automatic method to detect GGO densities within lung tissue on CT using a convolutional neural network based approach. The network design challenge is thus to allow for training on small positive patches within large heterogenous lung regions, while avoiding the anticipated false positives.

Traditional computer-aided diagnosis (CAD) medical imaging systems are based on manually engineered features defined predominantly by shape and other basic image qualities as derived from a limited (usually < 200 images) training sample [4]. This relatively narrow scope of feature identification and limited incorporation of anatomic and physiologic context results in overfitting and hindered applicability of first generation CAD in clinical practice.

In recent years, Deep Learning (DL) has emerged as a powerful approach to learning imaging representations directly from large volumes of data thus reducing the need to hand-engineer predictive features [5, 6]. Convolutional neural networks (CNNs) demonstrate the ability to learn hierarchically organized low to high-level features directly from raw images [7, 8]. Prior efforts to achieve GGO detection using CNNs [9, 10] were based upon several hundred CT slices from the Interstitial Lung Disease dataset (ILD) [11]. Anthimopoulos et al. [10] performed patch classification with a small convolutional network and Shin et al. [9] reported accuracy for slice classification using GoogLeNet at 57%.

The aim of this article is to explore the feasibility of Fully Convolutional Networks (FCNs) for GGO detection. FCNs have been successfully applied to the challenges of medical imaging segmentation as diverse as identification of neuronal structures in electron microscopic recordings with U-Net[12] and multi-slice MRI cardiac segmentation with Recurrent U-Net[13]. FCNs were introduced in the literature [14] as a natural extension of CNNs to tackle per pixel prediction problems such as semantic image segmentation. They have received increasing interest lately as they unify object localization and segmentation in a single process by effectively extracting both global and local context [12, 14]. FCNs add upsampling layers to standard CNNs to recover the spatial resolution of the input at the output layer. As a consequence, FCNs can process images of arbitrary size. In order to compensate for the resolution loss induced by pooling layers, FCNs introduce skip connections between their downsampling and upsampling paths [12], which help the upsampling path recover fine-grained information from the downsampling layers. For better filters reuse, additional skip connections might be added within blocks of layers [15], which improves accuracy of semantic segmentation, as shown for Fully Convolutional DenseNet (FC-DenseNet) [16].

The present work represents a first-time application of Fully Convolutional Neural Networks for GGO detection of any size (see Fig. 1). The FCNs were

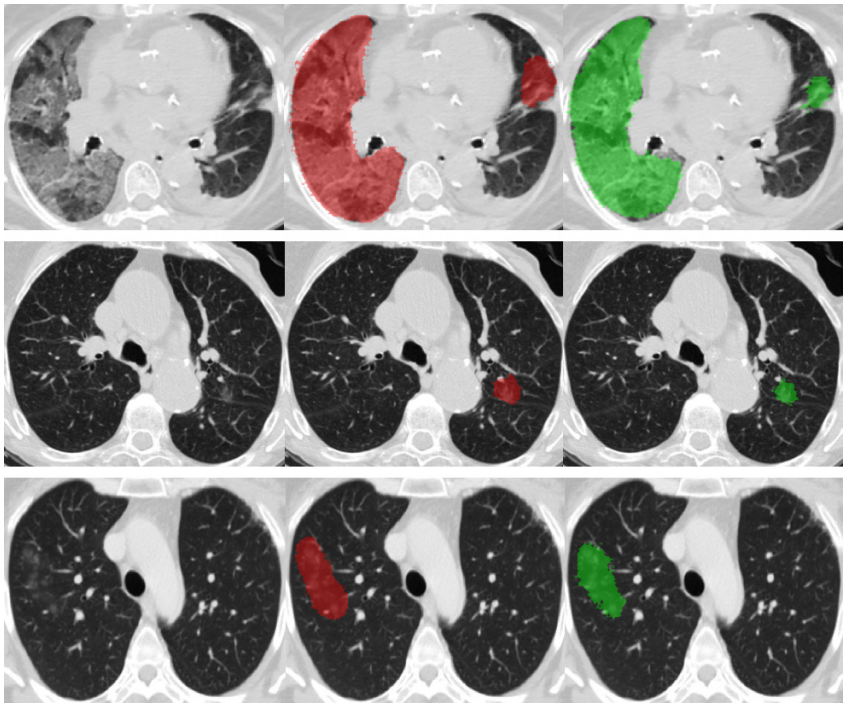

Fig. 1. Examples of GGO segmentation. *Left* - axial CT slice, *middle* - segmentation by a radiologist, *right* - predicted segmentation by FC-DenseNet.

trained and evaluated on the largest dataset described to-date (4107 segmented slices with GGO and 21778 non-GGO slices), and achieve superior accuracy of binary classification at the slice level. In a search of the best model, we compare U-Net- and DenseNet-like architectures of varying depth. Our end-to-end system does not require any special preprocessing (automatic lung segmentation for lung slice selection is optional but not required) and is applicable to any Chest CT scans.

2 Dataset

We assembled a dataset from 153 patients who had undergone CT scans of Chest for any clinical indication. Out of these, 86 CTs contained GGO and 67 were of healthy individuals (see Table 1). In CT scans with GGO every axial slice with ground glass signs was manually segmented by an expert radiologist (in total 4107 slices were segmented). All healthy studies were verified by a radiologist not to contain GGO, although they might have had contained small lung nodules (< 3mm).

Table 1. Number of CTs and slices in training/validation/test subsets

Subset	CT scans with GGO			Healthy CT scans	
	CTs	Segmented slices	All slices	CTs	All slices
train	67	3076	7279	13	2462
validation	7	380	944	18	3059
test	12	651	1496	36	6538

The dataset contains only lung slices (on average, 142 slices per patient), which were detected using an automatic lungs segmentation (non-lung regions have *not* been masked). We chose only one series per patient. The dataset includes GGOs of all types: small nodule-like regions, medium areas and GGOs large enough to involve the majority of a lung volume. It is worth mentioning that we did not exclude studies on the basis of data acquisition protocols, e.g. inspiration/expiration imaging or technical acquisition variations, despite that such variations could alter the CT appearance of the lung tissue. In allowing for data variations we aimed to develop an algorithm geared toward broader clinical application.

3 Fully Convolutional Neural Networks

The main idea underlying FCNs is extension of a contracting path, in which a sequence of pooling operators progressively reduces the size of the network, by adding successive layers where pooling operators are replaced by upsampling operators. Both in U-net [12] and in DenseNet [16] variations of a FCN, which we explore in this paper, the expanding path is characterized by a large number of feature channels, allowing the network to propagate context information to higher resolution layers. In order to localize, high resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information. The expansive path is more or less symmetric to the contracting path, and yields a u-shaped architecture.

U-Net with 18 convolutional layers was first described by Ronneberger et al. [12], and here we consider a generalized U-Net architecture illustrated in Fig. 2(a). The network consists of $2n+1$ blocks of k_i 3x3 convolution-dropout-batch normalization-ReLU sandwich layers with $i = 1, ..., 2n+1$. In the downsampling path a 2x2 max pooling with stride 2 follows every block, which halves the feature map width and height. After each block, the number of feature channels doubles. Between the downsampling and upsamling paths there is a bottleneck block. Every step in the upsampling path consists of a 2x2 transposed convolution with stride $\frac{1}{2}$, which doubles the feature map dimensions, a concatenation with the correspondingly cropped feature map from the contracting path and a convolutional block. At the end of the expanding path, a 1x1 convolution (with

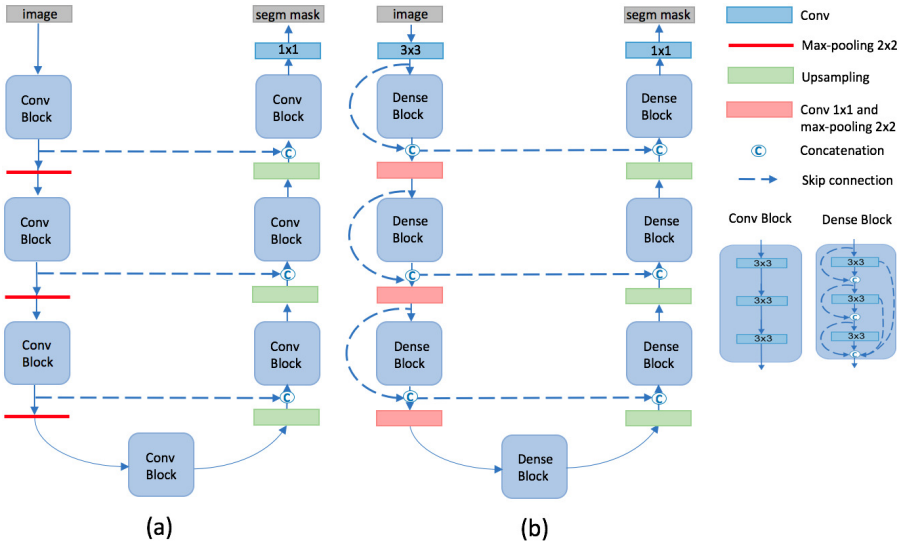

Fig. 2. (a) Architecture of a generalized U-Net, (b) Architecture of a Fully Convolutional DenseNet.

the number of features equals the number of classes) is applied followed by a softmax, which results in a probability map of the same dimensions as the input image. Weighted cross-entropy loss function is applied on the probability map.

Fully Convolutional DenseNet architecture [16] is similar in its block structure to U-Net, but has a higher connectivity within each block, called "dense block", and additional skip connections in the contracting path, which is shown on Fig. 2(b). There is also an additional 3x3 convolutional layer at the beginning of the contracting path and a 1x1 convolution-dropout-batch normalization-ReLU sandwich layer before each max pooling, leading to additional $n + 1$ convolutional layers compared to U-Net with the same block structure. Within a dense block, each layer is connected to all the following layers via concatenation with the skip connections. The number of filters in all 3x3 convolutional layers, except the first one, which has 48 filters, is the same and denoted as g (growth rate). The number of filters in all 1x1 convolutional layers (including transposed convolution), except the last one, equals the number of input channels.

4 Experimental Results

In search of the best model for GGO detection, we have experimented with U-Net and FC-DenseNet architectures of varying length. We report results of GGO detection for the following networks: U-Net18 and FC-DenseNet22 ($g = 12$) with the block structure $\{2\ 2\ 2\ 2\ 2\ 2\ 2\}$, U-Net50 and FC-DenseNet56 ($g = 12$) with

{4 4 4 4 4 4 4 4 4 4 4} and U-Net97 and FC-DenseNet103 ($g = 16$) with {4 5 6 7 8 9 8 7 6 5 4}, where an element in the parenthesis defines number of layers in a block.

Accuracy, sensitivity and specificity are typical statistical measures to evaluate performance of a binary classification test. They are defined as

$$Accuracy = \frac{TP + TN}{N}$$
$$Specificity = \frac{TN}{TN + FP}$$
$$Sensitivity = \frac{TP}{TP + FN},$$

where TP is the number of true positives, TN of true negatives, FP of false positives, FN of false negatives and N is the number of pixels in a slice.

We measure goodness of segmentation with the help of Dice score (DSC) averaged over all positive slices:

$$DSC = 2\frac{TP}{2TP + FP + FN}.$$

The Dice score not only considers how many positives are found, but it also penalizes for the false positives. A pixel is considered positive, if its probability is higher than 0.5 and a slice is considered positive if number of predicted positive pixels is higher than a threshold based upon the best accuracy and sensitivity received on a validation set.

Results are summarized in Table 2. The best accuracy was achieved on the deepest FC-DenseNet (with 103 convolutional layers) - 96.9% (with specificity 98.3% and sensitivity 80.5%), which is expected, since a deeper model can be more efficient at representing some functions than a shallow one [17]. Higher accuracy is obtained from a deep FCN with skip connections, which follows the principle that convolutional networks with skip connections are much easier to optimize than deep CNNs [18], since skip connections diminish the problem of vanishing/exploding gradients in a deep network [19, 20]. The best Dice score, 72.5%, is achieved by U-Net with 97 convolutional layers, which is impressive considering the level of inconsistency in GGO tagging due to high vagueness in the opacity boundaries. Examples of GGO segmentation by FC-DenseNet103 are shown in Fig. 1.

Image intensity was cropped within the window [-1050, 250] in HU. No preprocessing was performed on slices. For augmentation we used random vertical flip and random crop of 55-95% of the original image area, which was resized to 256x256. Cross-entropy loss function was optimized using RMSProp. The initial learning rate was set to 10^{-4} and it is halved every 14 epochs during the training. We trained the networks for 21-69 epochs (with batches of 10) and it took between 6 to 24 hours, depending on number of layers, using a dual NVIDIA TitanX GPU card. Our code is based on Tensorflow.

Table 2. GGO segmentation and binary classification results (per slice)

Model	Accuracy	Specificity	Sensitivity	Dice score
U-Net18	94.7	95.9	81.6	71.0
U-Net50	95.5	96.9	79.1	69.8
U-Net97	95.1	96.7	77.1	**72.5**
FC-DenseNet22	96.4	99.3	63.3	44.6
FC-DenseNet56	95.4	96.7	81.5	55.2
FC-DenseNet103	**96.9**	98.3	80.5	66.8

5 Conclusions

In this manuscript we presented an end-to-end approach for Ground Glass Opacity detection using Fully Convolutional Neural Networks. We explored several architectures of varying depth, U-Net- and DenseNet-like FCNs, and, as expected, the deepest networks showed the best performance. The best accuracy we achieved is 96.9%, which is promising, given the complexity of GGO detection and represents a significant improvement on similar prior endeavors.

6 Acknowledgments

This research was supported by Zebra Medical Vision.

References

1. Jeong, Y. J., Kim, K. I., Seo, I. J., Lee, C. H., Lee, K. N., Kim, K. N., Kim, J. S., Kwon, W. J.: Eosinophilic lung diseases: a clinical, radiologic, and pathologic overview. Radiographics 27(3), 617–637 (2007)
2. Mueller-Mang, C., Grosse, C., Schmid, K., Stiebellehner, L., Bankier, A.: What every radiologist should know about idiopathic interstitial pneumonias. Radiographics 27(3), 595–615 (2007)
3. Mario, S., Nicola, S., Carmelinda, M., Giulio, N., Alfonso, M., Maurizio, Z., Cristina, R., Ugo, P.: Long-term surveillance of ground-glass nodules: evidence from the MILD trial. J. Thorac. Oncol. 7(10), 1541-1546 (2012)
4. Linying, L., et al.: A Review of Ground Glass Opacity Detection Methods in Lung CT Images. Current Medical Imaging Reviews 13, 20–31 (2017)
5. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. arXiv:1206.5538 (2012)
6. Hinton, G. E.: Learning multiple layers of representation. Trends in Cognitive Sciences 11(10), 428-434 (2007)
7. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36(4), 193-02 (1980)
8. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278-2324 (1998)

9. Shin, H. C., et al.: Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans. Med. Imag. 35, 1285–1298 (2016)
10. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. IEEE Trans. Med. Imag. 35, 1207–1216 (2016)
11. Depeursinge, A., et al.: Building a reference multimedia database for interstitial lung diseases. Comput. Med. Imaging and Graph. 36, 227–238 (2012)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 9351, 234–241 (2015)
13. Poudel, R. P. K., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. arXiv:1608.03974 [cs.CV] (2016)
14. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. In: CVPR (2015)
15. Huang, G., Liu, Z., Weinberger, K.Q.: Densely Connected Convolutional Networks. arXiv:1608.06993 [cs.CV] (2016)
16. Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. arXiv:1611.09326 [cs.CV] (2016)
17. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2(1), 1-127 (2009)
18. He, K.,, Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In CVPR (2016)
19. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5(2), 157-166 (1994)
20. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In AISTATS (2010)