

DOPNET: DENSELY ORIENTED POOLING NETWORK FOR MEDICAL IMAGE SEGMENTATION

Mourad Gridach¹ Irina Voiculescu²

¹ Department of Computer Science, University of Ibn Zohr, Morocco

² Department of Computer Science, University of Oxford, UK.

ABSTRACT

Since manual annotation of medical images is time consuming for clinical experts, reliable automatic segmentation would be the ideal way to handle large medical datasets. Deep learning-based models have been the dominant approach, achieving remarkable performance on various medical segmentation tasks. There can be a significant variation in the size of the feature being segmented out of a medical image relative to the other features in the image, which can be challenging. In this paper, we propose a Densely Oriented Pooling Network (DOPNet) to capture variation in feature size in medical images and preserve spatial interconnection. DOPNet is based on two interdependent ideas: the dense connectivity and the pooling oriented layer. When tested on three publicly available medical image segmentation datasets, the proposed model achieves leading performance.

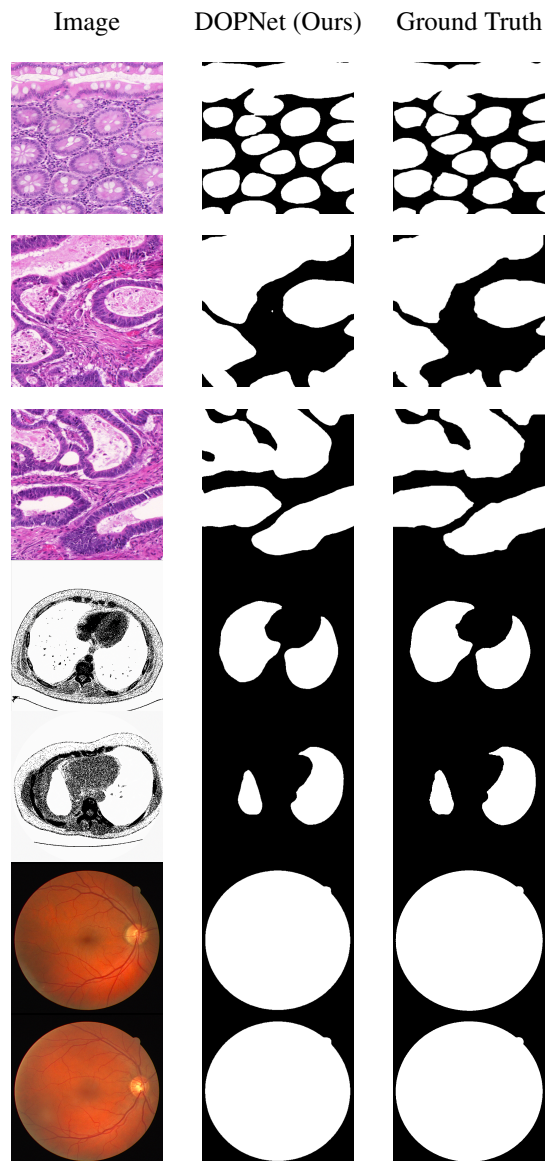
Index Terms— Medical Image Segmentation, Deep Learning, Densely Oriented Pooling Network

1. INTRODUCTION

Deep neural networks (DNNs) have been used successfully in medical image segmentation. Most approaches [1, 2] are based on the well-known U-Net architecture [3], also known as an encoder-decoder structure. It consists of a contracting path to capture context and a symmetric expanding path to enable precise localization. Successful U-Net based approaches still suffer from the same fundamental problem, namely that some of the spatial interconnection of the labelled features are lost. This is caused by the consecutive use of pooling layers and strided convolutions. While this invariance is useful for tasks such as classification and object detection, it harms dense prediction tasks like image segmentation which usually needs more spatial interconnection detail.

Existing approaches suffer from two main limitations: (1) with complex and large variations of feature size relative to the other features in the image, a single receptive field size fails to deal with such variations; (2) some methods lose sight of the global context of the image, thereby being unable to associate any semantic meaning to the segmented features.

Fig. 1: Examples of raw data samples and masks predicted by DOPNet, each with corresponding *GT* mask. Histology gland data (3 rows), lung CTs (2 rows), retinal scans (2 rows)



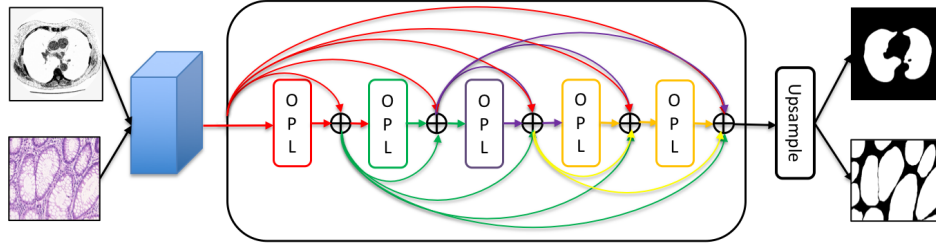


Fig. 2: DOPNet architecture consists of dense connectivity and OPL modules.

To concretize the first drawback, the first three rows of Fig. 1 present two very different examples of gland segmentation: the benign case shown in the first row can be covered by a small (128×128) receptive field, whereas the malignant cases in the second and third rows require a larger receptive field because, in adenocarcinomas, glands are physically stretched, resulting in larger segmented features relative to the image size. Therefore, it is important to design a network architecture which is able to cover feature size variations in medical images and which captures other contextual features to help improve the segmentation performance.

Rows four and five of Fig. 1 illustrate a case of the need for global context: a conventional network would have no understanding that there are a left and a right lung.

2. METHOD

In order to address both these issues, we propose a novel Densely Oriented Pooling Network (DOPNet). It contains two major ideas: a novel Oriented Pooling Layer (OPL) and an overall dense connectivity. On the one hand, the OPL module enables the model to capture both the appropriate contextual information and the variations in object sizes in medical images, which directly influence the final prediction and make it more accurate. On the other hand, inspired by the recent success of densely-connected networks, we add dense connections to capture the advantages of DenseNets namely: (a) parameter efficiency (our model has fewer parameters compared to its predecessors), and (b) feature reuse (since every layer has access to all the preceding layers, the information and gradient flow are improved across the entire network, thereby lightening the load of the training process).

2.1. Oriented Pooling Layer

Returning to the stretched gland example in Fig. 1, it is essential for a convolutional layer to detect such variations in feature or lesion size. It is our model’s OPL, summarised in Fig. 3, which solves this challenge.

Within OPL, the input feature map is divided into multiple pooled regions of different sizes from finer to coarser levels, and combines local features within each. Depending on

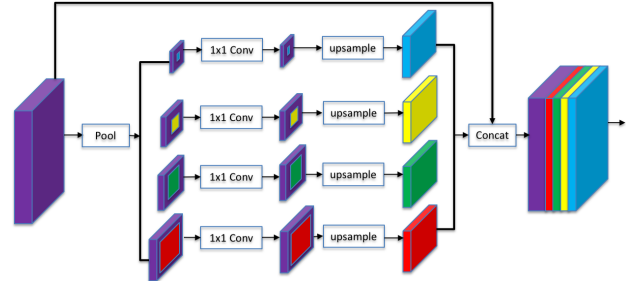


Fig. 3: OPL architecture takes feature map as input

the size (resolution) of the feature map (tiny, small, medium, large), corresponding pooled bins are implemented and used in parallel. The pooled representation of each region is then responsible for handling lesions of various sizes. Notably, the oriented pooling structure is able to extract different scale of feature information and to increase the receptive field effectively at pixel level.

If S denotes the number of channels in a given feature map, we reduce the dimension of context representation to $\frac{1}{S}$ of the original size by applying a 1×1 convolution layer after each pooled bin. We then use bilinear interpolation to upsample the resulting feature maps in order to recover feature maps the same size as the original one. Finally, we concatenate different output feature maps to get the final global feature. By using the parallel oriented pooling layer, our network captures different sized features. This is achieved using pooled bins of various sizes. These pooled bins are able to cover the whole, half of, or a small section of the image, and then fuse these to construct coherent information for the final segmentation.

2.2. DOPNet: Densely Oriented Pooling Network

The features extracted by a single OPL module are not sufficiently dense to cover the range of potential variations in feature size, shape or location. As sketched in Fig. 2, each layer is connected to all subsequent layers in a feed-forward manner, which explicitly differentiates between information that is added to the network and information that is preserved from earlier stages.

Dense connections provide several advantages. First, since each layer has direct access to all subsequent layers, it improves the flow of information and gradients across the entire network, which immediately accelerates the training. Second, the short paths to all feature maps lead to an implicit deep supervision. Lastly, dense connectivity has a regularizing effect which, on tasks with smaller training set sizes, reduces the risk of overfitting. Due to the lack of annotation time, medical image datasets are often quite small.

The main building block of DOPNet is the OPL module. We use a pretrained ResNet model to extract the feature map. Then, we apply a sequence of feature concatenations where the output of each layer (or block) is concatenated with the input feature map and all the output feature maps from previous layers and the result of concatenation is fed to the next layer. The dense connections allow the model to capture information from multiple scales, effective information propagation from one layer to another and adjust any possible misclassification errors introduced by the previous layers. Consequently, DOPNet acquires ‘knowledge’ of the global context such as the relative positions of organs or other features, thereby preserving spatial interconnection between pixels belonging to the same contour.

3. EXPERIMENTS

3.1. Datasets and Baselines

Our experiments have been conducted on three segmentation datasets: Warwick [4], Lung¹ and retinal scans [5]. The former is a small-sized standard benchmark with two test sets (A and B); the latter is a challenging large-scale set. We chose them for three reasons: (1) they are of different sizes; (2) they were acquired through different modalities, histology and CT; (3) they are used in the literature. We thereby demonstrate the effectiveness of our model on both large and small datasets of different kinds. We also compare our DOPNet algorithm against existing methods such as U-Net [3], CE-Net [1] and Suggestive Annotation [6].

3.2. Training Settings and Metrics

We implemented DOPNet using PyTorch; the code will be made available after publication.² We use an Adam optimizer with the default initial learning rate of $3 \cdot 10^{-3}$ and weight decay of 10^{-4} . We use the poly learning rate policy by multiplying the initial rate with $(1 - epoch / maxEpochs)^{0.9}$. The growth rate $k = 32$ was determined empirically. We use the Dice coefficient loss with regularization (weight decay). The model trained on an Intel Xeon E5-2698 with a 32GB GPU.

¹<https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/>

²<https://github.com/MouradGridach?tab=repositories>

Table 1: Lung data comparison, averaged over three runs

Methods	Sen	Spec	F1	Acc
U-Net [3]	0.938	0.9672	0.9458	0.975
Backbone [1]	0.967	-	-	0.967
CE-Net [1]	0.980	0.9783	0.9601	0.990
DOPNet	0.981	0.9945	0.9796	0.9907

Table 2: Gland segmentation comparison, using common measures, on both A & B test sets, averaged over three runs

Datasets	Glands A	Glands B
Methods	F1	F1
Deep Multichannel [8]	0.8930	0.8430
CUMedVision1 [2]	0.8666	0.8001
CUMedVision2 [2]	0.8974	0.7810
Suggestive annotation [6]	0.9210	0.8550
DOPNet	0.9357	0.9229

A selection of frequently-used evaluation measures are reported for the purpose of direct comparisons. But measures of the overlap between the ground truth (*GT*) and the machine segmentation (*MS*) will only tell part of the story [7]. For this reason, we think it appropriate to also report the extent of match along the boundaries, expecting an overall match of 50% to represent adequate performance.

3.3. Quantitative Results and Discussion

Table 1 summarizes a direct comparison, with DOPNet results being averaged over three training episodes. For the gland segmentation, Table 2 summarizes the comparison results of our DOPNet model against three algorithms namely Deep Multichannel, DCAN (CUMedVision1 and CUMedVision2) and Suggestive Annotation method. We note that the gland segmentation dataset contains two test sets (A and B), and that F1 is the only evaluation measure common amongst all tests. The experiments show DOPNet outperforming previous models by a significant ($> 7\%$) margin. These results illustrate our model’s strong learning ability even from small sets of training data. This is also evident from other evaluation measures on our model, shown in Table 3.

Table 3: DOPNet performance on gland and retina dataset

Datasets	Sen	Spec	Acc	Dice
glands A	0.9162	0.9609	0.937	0.9357
glands B	0.9253	0.8756	0.9097	0.9229
retina	0.9992	0.9857	0.9949	0.9963

Table 4: DOPNet performance on boundary measures

Datasets	DBD _G	DBD _M	SBD
lung	0.8194	0.7476	0.7815
retina	0.8692	0.6490	0.7590
glands A	0.5148	0.7108	0.6024
glands B	0.5656	0.6670	0.6119

Table 5: Computation cost comparison

Methods	Backbone	Param.
Dense-UNet [3]	DenseNet	45.082 M
Inf-Net [9]	Res2Net	33.122 M
CE-Net [1]	ResNet	29.039 M
DOPNet	ResNet	24.133 M

3.4. Efficiency

In terms of efficiency, each training session of the glands dataset has finished within 100 mins. The retina dataset, as the smallest, trains within 20 mins. The lung dataset, the most complex of the three, trains in 150 mins. We use the number of training parameters as a measure of computational cost. Table 5 compares the numbers of training parameters in methods that make their code available for this purpose.

3.5. Qualitative Results

To measure the effectiveness of a segmentation result, it is always important to look at the qualitative merits of the images. We have selected various images from the test sets of both datasets, both test sets (part A and part B) in the gland segmentation dataset and both cases including benign and malignant for gland segmentation. Fig. 1 has already given a preview of the raw data, followed by the predictions of DOPNet followed by the corresponding *GT* mask. For gland segmentation, our model can precisely locate and segment the gland objects in both benign and malignant cases in many challenging situations, such as varied size, touching and separating gland objects, etc.

For lung segmentation, we selected two different images from the test set with different lung sizes. The next two rows of Fig. 1 show segmentation results of our model. DOPNet is able to segment the lungs with high precision, regardless of the feature size in each image. For the retina dataset, the last two rows in Fig 1 show our model’s prediction. It is clear that DOPNet is able to capture the orientation of the eye with high accuracy, which confirm the quantitative results discussed in the previous section.

Finally, we note that our model is capable of segmenting medical images from different types of acquisition methods (histology and CT images), which is another important advantage of our DOPNet model.

4. CONCLUSION AND FUTURE WORK

As well as being able to process different kind of acquisition protocols and different classes of features in images, what DOPNet is best at is capturing both variations in feature size and multi-scale information. The model gets a sense of the global picture by preserving any spatial interconnection of the pixels that it discovers at any stage of the process. This increases its robustness over different classes of image with complex features made up of several connected components.

5. ACKNOWLEDGMENTS

We acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. <http://dx.doi.org/10.5281/zenodo.22558>. There are no conflicts of interest to declare.

6. REFERENCES

- [1] Zaiwang Gu et al., “Ce-net: Context encoder network for 2D medical image segmentation,” *IEEE TMI*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [2] Hao Chen et al., “Dcan: deep contour-aware networks for accurate gland segmentation,” in *Proc IEEE CVPR*, 2016, pp. 2487–2496.
- [3] Olaf Ronneberger et al., “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*. Springer, 2015, pp. 234–241.
- [4] Korsuk Sirinukunwattana et al., “Gland segmentation in colon histology images: The glas challenge contest,” *Medical image analysis*, vol. 35, pp. 489–502, 2017.
- [5] J Staal et al., “Ridge-based vessel segmentation in color images of the retina,” *IEEE TMI*, vol. 23, no. 4, pp. 501–509, 2004.
- [6] Lin Yang et al., “Suggestive annotation: A deep active learning framework for biomedical image segmentation,” in *MICCAI*. Springer, 2017, pp. 399–407.
- [7] Varduhi Yeghiazaryan et al., “Family of boundary overlap metrics for the evaluation of medical image segmentation,” *SPIE JMI*, vol. 5, no. 1, pp. 015006, 2018.
- [8] Yan Xu et al., “Gland instance segmentation using deep multichannel neural networks,” *IEEE Trans Biomed Eng*, vol. 64, no. 12, pp. 2901–2912, 2017.
- [9] Deng-Ping Fan et al., “Inf-net: Automatic COVID-19 lung infection segmentation from CT images,” *IEEE TMI*, 2020.