Segmentation of Optic Disc and Cup: A Comprehensive Survey

Runfeng Zhou
Yupeng Zhang
https://github.com/Ruifeng-Zhou24/SegNet-ADDA

Hongyi Guo
Yuqiao He
https://github.com/gohsyi/SegNet-ADDA

Abstract

Obtaining and analyzing medical images are of great importance in diagnosing and treating many diseases. However, due to restricted medical resources, analysis for medical image is hard to come by, and it is impractical to rely only on human beings to do all the analysis. Application of computer vision techniques in such fields emerge as the situation requires. In this survey, we focus on segmentation of ophthalmoscopy photographs. Specifically, we segment optic disc and cup from each other and from the rest of the image. Such segmentation results can be used to calculate indices such as cup-to-disc ratio, which suggests whether the patience suffers from glaucoma or not. We include various segmentation algorithms and adaptation algorithms that can be used for this task, and list several ophthalmoscopy photograph datasets.

1. Introduction

Glaucoma is a group of eye diseases which result in damage to the optic nerve and vision loss.[9] The most common type is open-angle glaucoma with less common types including closed-angle glaucoma and normal-tension glaucoma.[9] Open-angle glaucoma develops slowly over time and there is no pain.[9] Peripheral vision may begin to decrease followed by central vision resulting in blindness if not treated.[9] Closed-angle glaucoma can present gradually or suddenly.[8] The sudden presentation may involve severe eye pain, blurred vision, mid-dilated pupil, redness of the eye, and nausea.[9][8] Vision loss from glaucoma, once it has occurred, is permanent.[9] Glaucoma has been called the ”silent thief of sight” because the loss of vision usually occurs slowly over a long period of time.[10] Worldwide, glaucoma is the second-leading cause of blindness after cataracts.[8][12] About 6 to 67 million people have glaucoma globally.[8][13]

Open-angle glaucoma is painless and does not have acute attacks, thus the lack of clear symptoms make screening via regular eye check-ups important. The only signs are gradually progressive visual field loss, and optic nerve changes, which means increased cup-to-disc ratio on fundoscopic examination. The cup-to-disc ratio (often notated CDR) is a measurement used in ophthalmology and optometry to assess the progression of glaucoma. The optic disc or optic nerve head is the point of exit for ganglion cell axons leaving the eye. The optic cup is the white, cup-like area in the center of the optic disc. The optic disc is the anatomical location of the eye's ”blind spot”, the area where the optic nerve and blood vessels enter the retina. The optic disc can be flat or it can have a certain amount of normal cupping. But glaucoma, which is in most cases associated with an increase in intraocular pressure, often produces additional pathological cupping of the optic disc. The pink rim of disc contains nerve fibers. The white cup is a pit with no nerve fibers. As glaucoma advances, the cup enlarges until it occupies most of the disc area.[5]

The cup-to-disc ratio compares the diameter of the ”cup” portion of the optic disc with the total diameter of the optic disc. The normal cup-to-disc ratio is 0.3. A large cup-to-disc ratio may imply glaucoma or other pathology.[4] However, cupping by itself is not indicative of glaucoma. Rather, it is an increase in cupping as the patient ages that is an indicator for glaucoma. Deep but stable cupping can occur due to hereditary factors without glaucoma.

According to the research[6], the result shows that the CDR was related to DD(disc diameter) by the equation $CDR = (-1.31 + (1.194 \times DD)) \div DD$. And Using this equation, the sensitivity and specificity of the CDR to identify glaucomatous discs was 62.3% and 98.9% respectively. So its clear that the CDR, relative to disc size, is useful clinically, especially to assist in identifying small glaucomatous discs.

As the development in image processing, its feasible to use segmentation to get CDR. Accurate segmentation of optic disc and cup is in demand, since the result of segmentation can be used in the disgnosis of glaucoma. Apart from SegNet[2], which we use to implement the task, various image semantic segmentation methods are feasible for our tasks.

2. Image Semantic Segmentation

Image Semantic Segmentation: In the project, we use SegNet[2] for Disc and Cup Segmentation. Here we list some other networks and analyze whether they are fitable to do Disc and Cup Segmentation.

2.1. Fully Convolutional Networks[7]

Fully convolutional networks (FCNs) are a special kind of CNN that includes only convolutional layers (no fully connected units). The output of an FCN is a feature map from the last convolutional layer. Given that FCN inputs and outputs are two-dimensional, they are trained end-to-end, pixel-to-pixel, and can be used for dimensionality reduction, regression, or semantic segmentation.

Fully convolutional networks are a rich class of models, of which modern classification convnets are a special case. Recognizing this, extending these classification nets to segmentation, and improving the architecture with multi-resolution layer combinations dramatically improves the state-of-the-art, while simultaneously simplifying and speeding up learning and inference.

FCN shows great effect on semantic segmentation. But various of problems still exists, like accuracy issues, insensitivity to detail, the relationship between pixels and pixels, ignoring spatial consistency, etc. So for the Disc and Cup segmentation, its not a preferred choose to implement on the project. However, since the class amount is small, its also a choice that can be tried.

2.2. Dilated Convolutions

Its put forward in paper: Multi-scale context aggregation by dilated convolutions [16]. It develops a convolutional network module that aggregates multi-scale contextual information without losing resolution or analyzing rescaled images. The module can be plugged into existing architectures at any resolution. Unlike pyramid-shaped architectures carried over from image classification, the presented context module is designed specifically for dense prediction. It is a rectangular prism of convolutional layers, with no pooling or subsampling. The module is based on dilated convolutions, which support exponential expansion of the receptive field without loss of resolution or coverage. The presented module uses dilated convolutions to systematically aggregate multiscale contextual information without losing resolution. The architecture is based on the fact that dilated convolutions support exponential expansion of the receptive field without loss of resolution or coverage.

Dilation is largely the same as run-of-the-mill convolution (frankly so is deconvolution), except that it introduces gaps into its kernels, i.e. whereas a standard kernel would typically slide over contiguous sections of the input, it's dilated counterpart may, for instance, "encircle" a larger section of the image –while still only have as many weights/inputs as the standard form.

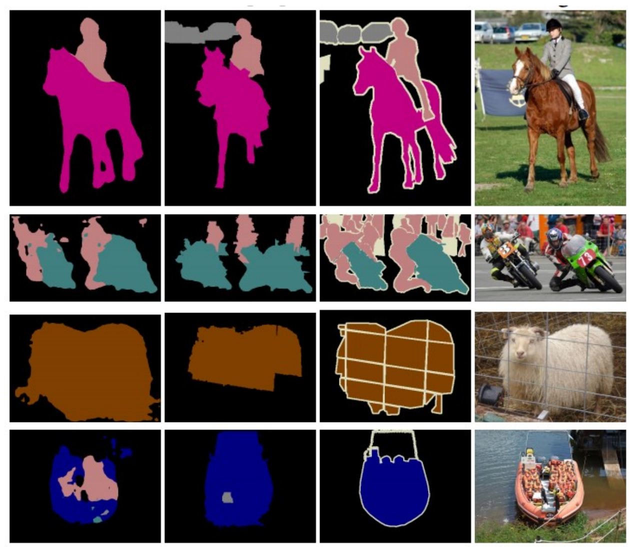

Figure 1. Fully convolutional segmentation nets produce stateof-the-art performance on PASCAL. The left column shows the output of the highest performing net, FCN-8s. The second shows the segmentations produced by the previous state-of-the-art system by Hariharan et al[1]. Notice the fine structures recovered (first row), ability to separate closely interacting objects (second row), and robustness to occluders (third row). The fourth row shows a failure case: the net sees lifejackets in a boat as people.

The convolutional network module can be used for dense predictions. Its more effective in accurate locating and shows high efficiency. So its an optional choice for Disc and Cup segmentation.

2.3. Using Adversarial Networks for Semantic Segmentation

Because that adversarial training has been shown to produce state of the art results for generative image modeling, in paper Semantic Segmentation using Adversarial Networks[11] it proposes an adversarial training approach to train semantic segmentation models. They train a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation maps coming either from the ground truth or from the segmentation network. The motivation for the approach is that it can detect and correct higher-order inconsistencies between ground truth segmentation maps and the ones produced by the segmentation net. The experiments in the paper show that the adversarial training approach leads to improved accuracy on the Stanford Background and PASCAL VOC 2012 datasets. As the result shown, the accuracy is better than FCN. It can be a choice for our segmentation task.

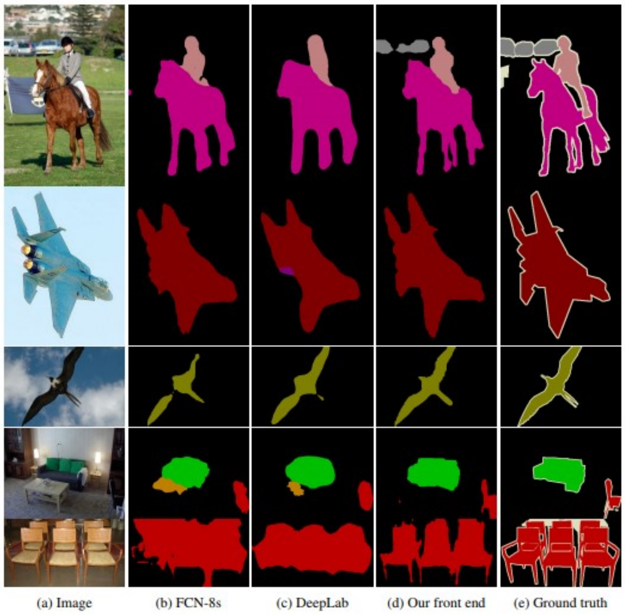

Figure 2. Semantic segmentations produced by different adaptations of the VGG-16 classification network. From left to right: (a) input image, (b) prediction by FCN-8s [7], (c)prediction by DeepLab [3], (d) prediction by the simplified front-end module, (e)ground truth.

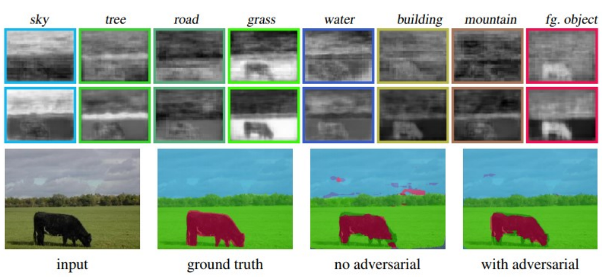

Figure 3. Segmentations on Stanford Background. Class probabilities without (first row) and with (second row) adversarial training. In the last row the class labels are superimposed on the image.

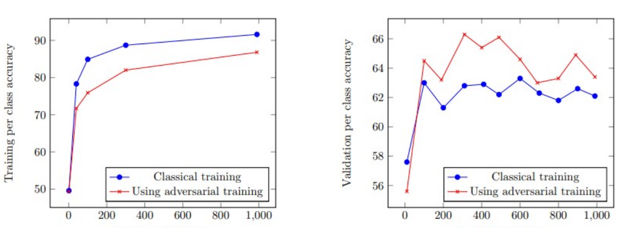

Figure 4. Per-class accuracy across training ephochs on the Stanford Background dataset on train data (left) and validation data (right), with and without adversarial training.

2.4. Conclusion

From previous analysis, we know that Dilated Convolutions and Adversarial Networks can be implemented for our segmentation tasks. Their accuracy performs better than FCN. Considering the small amount of class we aim to detect, FCN can also be selected.

3. Dataset

We list in the appendix A some useful datasets of retinal images we collected on the Internet.

4. Transfer Learning

Transfer Learning is the reuse of a pre-trained model on a new problem. As a particular case of transfer learning, domain adaptation utilizes labeled data in one or more relevant source domains to execute new tasks in a target domain, which can help us to address the lack of massive amounts of labeled data in the field of medical image processing. In this project, we leverage ADDA[14] to handle photos with larger brightness. Here we list some useful domain transfer algorithms in the appendix B and analyze whether they are suitable in our case.

References

[1] R. G. B. Hariharan, P. Arbelaez and J. Malik. Simultaneous detection and segmentation. In European Conference on Computer Vision (ECCV), 2014. 1, 2, 4, 5, 7, 8.

[2] V. Badrinarayanan, A. Handa, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293*, 2015.

[3] P. G. K. I. M. K. Chen, Liang-Chieh and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR, 2015a.

[4] The Eye Digest. The University of Illinois Eye and Ear Infirmary. *Eye Exam and tests for Glaucoma diagnosis*. Archived from the original on 8 July 2012.

[5] The Eye Digest. The University of Illinois Eye and Ear Infirmary. *Glaucoma causes Optic Nerve Cupping (atrophy) and Vision Loss*. Archived from the original on 8 July 2012.

[6] V. A. e. a. Garway-Heath DF, Ruben ST. Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect. Br J Ophthalmol 1998;82:111824.

[7] S. E. . D. T. Long, J. Fully convolutional networks for semantic segmentation. Cvpr 2015 (2014). URL http://arxiv.org/abs/1411.4038. DOI 10.1109/CVPR.2015.7298965.

[8] N. Mantravadi, AV; Vadhar. Glaucoma. Primary Care. 42 (3): 43749. doi:10.1016/j.pop.2015.05.008. ISSN 0095-4543. PMID 26319348 via Elsevier Science Direct.

[9] National Eye Institute. *Facts About Glaucoma*. Archived from the original on 28 March 2016. Retrieved 29 March 2016.

[10] National Eye Institute. *Glaucoma: The 'silent thief' begins to tell its secrets*, January 2014. Archived from the original on 23 July 2015.

[11] S. C. P. Luc, C. Couprie and J. Verbeek. Semantic segmentation using adversarial networks. CoRR, vol. abs/1611.08408, 2016.

[12] D. E. D. K. I. P. R. P. G. P. M. S. P. Resnikoff, Serge; Pascolini. "global data on visual impairment in the year 2002". Bulletin of the World Health Organization. 82 (11): 84451. doi:10.1590/S0042-96862004001100009 (inactive 2018-10-14). PMC 2623053. PMID 15640920. Archived from the original on 12 December 2013.

[13] D. J. . G. . e. Rhee. Philadelphia: Wolters kluwer health/lippincott williams wilkins. p. 180. ISBN 9781609133375. OCLC 744299538.

[14] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. *CoRR*, abs/1702.05464, 2017.

[15] M. Wang and W. Deng. Deep visual domain adaptation: A survey. *CoRR*, abs/1802.03601, 2018.

[16] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

A. Dataset

Name	Time	Size	Samples	Website	Appendix
DRIONS-DB	2008	2.8MB	110	http://www.ia.uned.es/ ejcarmona/DRIONS-DB.html	expert labeled
DeepBlueData/BinRushed	2018	494MB	1670	N/A	
DeepBlueData/Magrabia	2018	2.95GB	668	N/A	
DeepBlueData/MESSIDOR	2018	22.2GB	3221	N/A	
RIGA	1991-2013	N/A	N/A	http://www.fao.org/economic/riga/riga-database/	need appliance

Table 1. Retinal image dataset

B. B.Different Approaches in Domain Adaption[15]

Table 2. DIFFERENT APPROACHES AND USAGE IN DIFFERENT DOMAIN ADAPTATION SETTINGS

One-step DA Approaches	Brief Description	Subsettings	Supervised DA	Unsupervised DA
Discrepancy-based	fine-tuning the deep network with labeled or unlabeled target data to diminish the domain shift	class criterion	✓	
		statistic criterion		✓
		architecture criterion	✓	✓
		geometric criterion	✓	
Adversarial-based	using domain discriminators to encourage domain confusion through an adversarial objective	generative models		✓
		non-generative models		✓
Reconstructionbased	using the data reconstruction as an auxiliary task to ensure feature invariance	encoder-decoder reconstruction		✓
		adversarial reconstruction		✓

Table 3. DIFFERENT DEEP APPROACHES TO MULTI-STEP DA

Multi-step Approaches	Brief Description
Hand-crafted	users determine the intermediate domains based on experience
Instance-based	selecting certain parts of data from the auxiliary datasets to compose the intermediate
Representation-based	freeze weights of one network and use their intermediate representations as input

Table 4. COMPARISON BETWEEN TRANSFER LEARNING AND NON-ADAPTATION LEARNING METHODS

Data Set (reference)	Source vs. Target	Baselines	Deep Domain Adaptation Methods				
Office-31 Dataset ACC (unit:%)		AlexNet	DDC	DAN	RTN	JAN	DANN
	A vs. W	61.6±0.5	61.8±0.4	68.5	73.3±0.3	75.2±0.4	73.0±0.5
	D vs. W	95.4±0.3	95.0±0.5	96.0±0.3	96.8±0.2	96.6±0.2	96.4±0.3
	W vs. D	99.0±0.2	98.5±0.4	99.0±0.3	99.6±0.1	99.6±0.1	99.2±0.3
	A vs. D	63.8±0.5	64.4±0.3	67.0±0.4	71.0±0.2	72.8±0.3	72.3±0.3
	D vs. A	51.1±0.6	52.1±0.6	54.0±0.5	50.5±0.3	57.5±0.2	53.4±0.4
	W vs. A	49.8±0.4	52.2±0.4	53.1±0.5	51.0±0.1	56.3±0.2	51.2±0.5
	Avg	70.1	70.6	72.9	73.7	76.3	74.3
Office-31 Dataset ACC (unit:%)		AlexNet	Deep CORAL	CMD	DLID	AdaBN	DANN
	A vs. W	61.6	66.4	77.0±0.6	51.9	74.2	73
	D vs. W	95.4	95.7	96.3±0.4	78.2	95.7	96.4
	W vs. D	99	99.2	99.2±0.2	89.9	99.8	99.2
	A vs. D	63.8	66.8	79.6±0.6	-	73.1	-
	D vs. A	51.1	52.8	63.8±0.7	-	59.8	-
	W vs. A	49.8	51.5	63.3±0.6	-	57.4	-
	Avg	70.1	72.1	79.9	-	76.7	-
Office-31 Dataset ACC (unit:%)		AlexNet	DLID	DANN	Soft Labels	Domain Confusion	Confusion +Soft
	A vs. W	56.5±0.3	51.9	53.6±0.2	82.7±0.7	82.8±0.9	82.7±0.8
	D vs. W	92.4±0.3	78.2	71.2±0.0	95.9±0.6	95.6±0.4	95.7±0.5
	W vs. D	93.6±0.2	89.9	83.5±0.0	98.3±0.3	97.5±0.2	97.6±0.2
	A vs. D	64.6±0.4	-	-	84.9±1.2	85.9±1.1	86.1±1.2
	D vs. A	47.6±0.1	-	-	66.0±0.5	66.2±0.4	66.2±0.3
	W vs. A	42.7±0.1	-	-	65.2±0.6	64.9±0.5	65.0±0.5
	Avg	66.2	-	-	82.17	82.13	82.22
MNIST, USPS, and SVHN digits datasets ACC (unit:%)		VGG-16	DANN	CoGAN	ADDA		
	M vs. U	75.2±1.6	77.1±1.8	91.2±0.8	89.4±0.2		
	U vs. M	57.1±1.7	73.0±2.0	89.1±0.8	90.1±0.8		
	S vs. M	60.1±1.1	73.9	-	76.0±1.8		