# Semi-supervised Segmentation of Optic Cup in Retinal Fundus Images Using Variational Autoencoder

Suman Sedai(✉), Dwarikanath Mahapatra, Sajini Hewavitharanage, Stefan Maetschke, and Rahil Garnavi

IBM Research, Melbourne, VIC, Australia
ssedai@au1.ibm.com

**Abstract.** Accurate segmentation of optic cup and disc in retinal fundus images is essential to compute the cup to disc ratio parameter, which is important for glaucoma assessment. The ill-defined boundaries of optic cup makes the segmentation a lot more challenging compared to optic disc. Existing approaches have mainly used fully supervised learning that requires many labeled samples to build a robust segmentation framework. In this paper, we propose a novel semi-supervised method to segment the optic cup, which can accurately localize the anatomy using limited number of labeled samples. The proposed method leverages the inherent feature similarity from a large number of unlabeled images to train the segmentation model from a smaller number of labeled images. It first learns the parameters of a generative model from unlabeled images using variational autoencoder. The trained generative model provides the feature embedding of the images which allows the clustering of the related observation in the latent feature space. We combine the feature embedding with the segmentation autoencoder which is trained on the labeled images for pixel-wise segmentation of the cup region. The main novelty of the proposed approach is in the utilization of generative models for semi-supervised segmentation. Experimental results show that the proposed method successfully segments optic cup with small number of labeled images, and unsupervised feature embedding learned from unlabeled data improves the segmentation accuracy. Given the challenge of access to annotated medical images in every clinical application, the proposed framework is a key contribution and applicable for segmentation of different anatomies across various medical imaging modalities.

**Keywords:** Semisupervised learning · Variational inference · Optic cup segmentation

## 1 Introduction

Glaucoma will affect approximately 80 million persons worldwide by 2020 [12]. Being asymptomatic, Glaucoma patients are usually ignorant about it until a noticeable visual loss occurs at a later stage. Early detection and treatment are


M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 75–82, 2017.
DOI: 10.1007/978-3-319-66185-8_9

essential to reduce the progression of vision loss. Glaucoma diagnosis is based on medical history, intra-ocular pressure and visual field loss tests together with an assessment of the Optic Disc (OD) through ophthalmoscopy. In 2D color retinal fundus images, the OD can be divided into two distinct regions; the central bright optic cup (OC), and the peripheral neuroretinal rim. The loss in optic nerve fibers leads to the enlargement of cup region called cupping. One of the important indicators of glaucoma is the enlargement of the cup with respect to OD which can be measured as the vertical cup to disc ratio CDR. Quantification of CDR requires accurate delineation of the boundaries of the optic disc and cup.

There are many automated methods for segmentation of optic disc [10,14] in the literature. However, only a few have tackled optic cup segmentation, since ill-defined and in-homogeneous boundaries make its segmentation very challenging. Existing approaches of optic cup segmentation are based on level sets [6], super-pixels classification [15] and sparse dictionary learning [1]. In another method [2], fusion of cup segmentation from multi-view fundus images was used to improve the performance.

Fully supervised approaches require large numbers of annotated images to achieve reasonable robustness and accuracy, which is often difficult to obtain as it can be time-consuming and costly. Semi-supervised approaches tackle this problem by leveraging large number of unlabeled data along with the labeled data to improve the performance. For example, semi-supervised approach has been applied in different medical imaging tasks like, brain MRI segmentation [11], lung nodule detection [7] and retina vessel segmentation [16].

In this paper, we propose a novel approach which leverages unlabeled images to segment the optic cup in retinal fundus images. The proposed method is based on learning a generative model from the unlabeled data and utilizing the feature embedding provided by the trained generative model. We propose to use the variational-autoencoder (VAE) [5] as a generative model which learns the feature embedding as a latent variable without assumption of specific distance measure. Although, VAE have been extended to semi-supervised classification [3, 4,9], it has not been applied on the segmentation task. Our approach is based on first learning the feature embedding using VAE from large number of unlabeled images. We then train the segmentation autoencoder that maps the image to the segmentation mask by transferring the properties of the learned feature embedding through end-to-end training.

## 2 Proposed Semi-supervised Segmentation Method

In the proposed semi-supervised segmentation method, a generative model is trained from a large number of unlabeled data. The feature embedding from the generative model is then incorporated in the segmentation model so that the segmentation model can be trained from limited number of labeled images. We use variational autoencoder (VAE) [5] as the generative model which models each observation in terms of a low dimensional latent variable. VAE has two

parts; encoder network which maps the input image to the continuous latent variable $\mathbf{z}$, and the decoder network which uses the latent variable $\mathbf{z}$ to reconstruct the image.

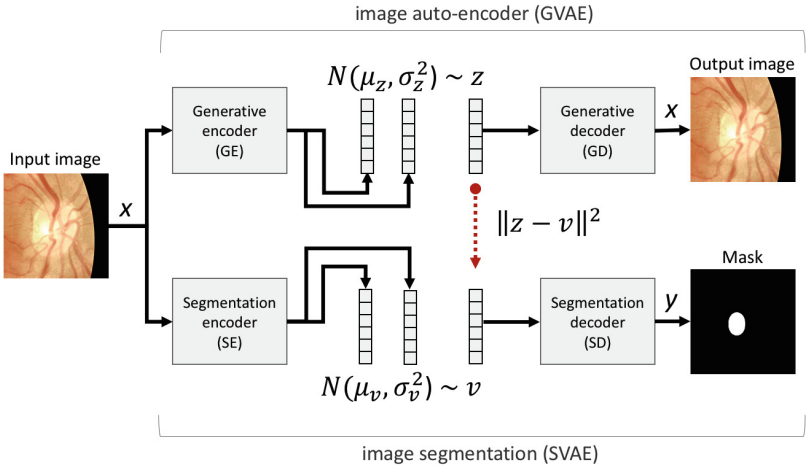


**Fig. 1.** Proposed segmentation framework: the segmentation model (SVAE) contains segmentation encoder (SE) and segmentation decoder (SD). The generative model (GVAE) is used in the training of SVAE and is not required in testing phase.

As shown in Fig. 1, our proposed semi-supervised learning approach for optic cup segmentation consists of two main components; generative VAE (GVAE) and segmentation VAE (SVAE), details of each is provided in the following.

### 2.1 GVAE: Generative Variational Autoencoder

GVAE models the probability distribution of the image using neural networks, and is composed of two parts; generative encoder (GE), and generative decoder (GD). GE takes an image $x$ as input and outputs the mean $\mu_g$ and standard deviation $\sigma_g$. The latent representation $\mathbf{z}$ of the generative model is constructed by sampling from the distribution $q_\phi(\mathbf{z}|\mathbf{x}) = N\left(\mu_z, \sigma_z^2\mathbf{I}\right)$ where $\phi$ is the parameter of GE network. GE is modeled using convolution neural network with five convolution layers where each convolution layer is followed by a max-pooling layer which effectively reduces the size of feature response by half. Two dense layers are then attached to the features response from the last layer to output $\sigma_z$ and $\mu_z$. The GD network consist of five in-network deconvolution layers [8] which takes the latent representation $\mathbf{z}$ and reconstructs the image $\mathbf{x}$. GVAE is trained using the following loss function given by the variational lower bound [5]:

$$\mathcal{L}\left(\theta; \phi; \mathbf{x}\right) = -D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) + E_{q_\phi(z|x)}[\log p_\theta(\mathbf{x}|\mathbf{z})] \tag{1}$$

where the first term is the negative KL-divergence between the posterior approximation $q_\phi(\mathbf{z}|\mathbf{x})$ to the prior $p(\mathbf{z})$, and the second term is expected reconstruction error obtained from the GD network and $\theta$ is the parameter of GD network. The reconstruction error can be computed using binary crossentropy. The prior $p(\mathbf{z})$ is a spherical Gaussian distribution $p(\mathbf{z}) = N(0, I)$ and the posterior distribution $q_\phi(\mathbf{z}|\mathbf{x})$ which is the output of the encoder network, is also Gaussian. The KL part can then be written in analytical form as:

$$-D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) = \frac{1}{2}\Sigma_{j=1}^{J}(1 + \log((\sigma_z^j)^2) - (\sigma_z^j)^2 - (\mu_z^j)^2) \quad (2)$$

The input to the decoder is the random sample generated from the posterior $q_\phi(\mathbf{z}|\mathbf{x})$, yet back-propagation is not possible through random sampling. To overcome this obstacle, re-parametrization trick have been used [5].

The GVAE model can be used to estimate the probability density of data from which unseen samples can be generated. The GE part of the GVAE model provides the latent representation of the input image which we use as the feature embedding to improve the segmentation performance.

### 2.2 SVAE: Segmentation Variational Autoencoder

The goal of SVAE is to predict the segmentation mask of the optic cup from the given image by leveraging the feature embedding learned by GVAE. Similar to GVAE, SVAE consists of two parts; segmentation encoder (SE) and segmentation decoder (SD). SE is modeled using five blocks of convolution and max-pool layers followed by two dense layers which outputs $\sigma_v$ and $\mu_v$. The latent representation $\mathbf{v}$ of the segmentation model is obtained by sampling from the distribution $q_\alpha(\mathbf{v}|\mathbf{x}) = N\left(\mu_v, \sigma_v^2\mathbf{I}\right)$ where $\alpha$ is the parameter of SE network. SD consist of five deconvolution layer which takes the latent representation $\mathbf{v}$ as input and outputs the segmentation mask $\mathbf{y}$.

In order to leverage the information within unlabeled data for segmentation, SVAE model is trained to reconstruct not only the segmentation mask but also the latent representation learned by GVAE. Given an image $\mathbf{x}$, the corresponding latent representation $\mathbf{z}$ is generated from conditional distribution $p_\phi(\mathbf{z}|\mathbf{x})$ given by GE. The SVAE network is then trained using following loss function:

$$\mathcal{L}(\alpha; \beta; \mathbf{x}) = -D_{KL}(q_\alpha(\mathbf{v}|\mathbf{x}) \parallel p(\mathbf{v})) + E_{q_\alpha(v|x)}[\log p_\beta(\mathbf{x}|\mathbf{v})] + \parallel \mathbf{z} - \mathbf{v} \parallel^2 \quad (3)$$

where the first term is the negative KL-divergence between the posterior approximation $q_\alpha(\mathbf{v}|\mathbf{x})$ to the prior $p(\mathbf{v}) = N(0, I)$ which can be computed in analytical form as

$$-D_{KL}(q_\alpha(\mathbf{v}|\mathbf{x}) \parallel p(\mathbf{v})) = \frac{1}{2}\Sigma_{j=1}^{J}(1 + \log((\sigma_v^j)^2)) - (\sigma_v^j)^2 - (\mu_v^j)^2). \quad (4)$$

The second terms of Eq. 3 denotes the expected reconstruction error of the segmentation mask which can be computed using binary crossentropy, $\beta$ denotes the parameter of SD network and the third term is the Euclidean distance loss

**Algorithm 1.** Training of the proposed semi-supervised method of segmentation of optic cup in retinal fundus images.

1. Given an unlabeled data $X_u = \{x_1, \cdots x_{Nu}\}$, train a GVAE model to reconstruct the image using the loss function given by Eq. 1.
2. Given the labeled samples given by images $X_l = (x_1, \cdots x_{Nl})$ and corresponding segmentation mask $Y_l = (y_1, \cdots, y_{Nl})$, train the SVAE model as follows:
3. For each minibatch, until the validation loss converges:
   (a) Compute the latent representation $\mathbf{z}$ by sampling from the posterior $q_\phi(\mathbf{z}|\mathbf{x}) = \mathbf{N}(\mu_\mathbf{z}, \sigma_\mathbf{z}^\mathbf{2})$
   (b) Compute the latent representation of segmentation $\mathbf{v}$ by sampling from the posterior $p_\alpha(\mathbf{v}|\mathbf{x}) = N(\mu_v, \sigma_v^2)$.
   (c) Compute the loss using Eq. 3.
   (d) Update the parameters of $\alpha$ of SE and $\beta$ of SD using backpropagation.

between the latent codes produced by GE and SE. Therefore, the training process transfers the latent representation learned from the generative model into the segmentation model. Our entire algorithm is summarized in Algorithm 1.

## 3 Experiment Results

The dataset used in this research is provided by EyePACS[1] and contains 12000 high resolution fundus images. We select 600 images to create a labeled set, where the ground truth has been obtained by manual delineation of the cup regions of all images by a clinician. The remaining 11400 images are used as the unlabeled set. Since our goal is to segment the optic cup, we cropped the optic disc region of all images by first segmenting the optic disc using the approach of [10] and rescaled to $128 \times 128$ dimension. Therefore, images in both labeled and unlabeled sets are disc cropped.

We split the labeled set into 400 training and 200 test sets. We train the network parameters using the training set, and then evaluate the final model on the test set. In order to evaluate the effect of number of training samples, we further divide the training set into four subsets containing 50, 100, 200 and 400 samples. Data augmentation is an important step in training deep networks. We augment the training images and corresponding label masks in each subset through a mirror-image reflection and rotation at 6 different angles, leading to 12x the original set.

We first train the generative model (GVAE) network using the unlabeled set for 20000 iterations. We have used mini-batch gradient descent using the RMSprop algorithm with momentum and a batch size of 50. The learning rate is set to 0.001 which is decreased by one tenth after 10000 iterations of the training process. Figure 2 shows examples of images reconstructed using GVAE. It can be seen that the reconstruction preserves the optic disc and optic cup structure

[1] http://www.eyepacs.com/.

in images, even though the surrounding vasculature is not clear. This shows that the latent representation obtained from GVAE is able to capture the structural information of the cup region.

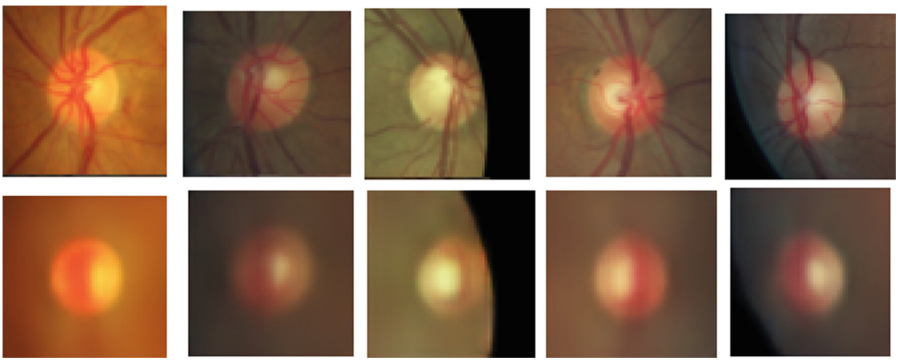

**Fig. 2.** Random images from the test set (top row) and the corresponding reconstructed images from using GVAE (bottom row)

We then train the segmentation model (SVAE) using the labeled set by leveraging the trained GVAE as described in Algorithm 1. The SVAE network is trained for 10000 iteration using mini-batch gradient descent and the Adam optimizer with learning rate of 0.0001. In the testing phase, the output of the SE (i.e. the mean value of the latent code $\mu_v$) is directly fed to SD to obtain the segmentation mask, as the sampling is only required in the training phase. For both GVAE and SVAE we set the dimension of latent variable $\mathbf{z}$ and $\mathbf{v}$ to $J = 100$. We compare our method with Unet network [13] which is the state-of-the-art method in biomedical image segmentation. The architecture of the Unet is kept similar to SE and SD except that in the Unet there are skip connections between SE and SD. The Unet is trained using binary crossentropy loss using the same parameters as of SVAE. We also compared the result of SVAE with plain SVAE (SVAE-Plain) which does not take into account the feature embedding from GVAE, i.e., it does not include the third term of the loss in Eq. 3.

Table 1 compares the average Dice coefficient (DC) between the ground truth and predicted segmentation for the proposed SVAE, the Unet methods and the SAVE-Plain. The proposed method SVAE resulted in average DC of 0.80 when trained on the full training set, slightly improving over the Unet and SVAE-Plain. However, when we use less number of training samples, the proposed method improves significantly over both Unet and SVAE-Plain. This demonstrates that the proposed approach improves the segmentation performance when the number of labeled images are limited. This also demonstrates that our approach can leverage the information from unlabeled samples by first learning the encoding by GVAE and training SVAE with encoding samples generated from GVAE. Figure 3 shows the examples of segmentation produced by the proposed semi-supervised method.

**Table 1.** Optic cup segmentation performance of the proposed method compared with the Unet and SVAE-Plain for different training sizes.

| | Dice coefficient | | |
|---|---|---|---|
| #Images | Proposed SVAE | Unet | SVAE-Plain |
| 50 | $0.73 \pm 0.06$ | $0.69 \pm 0.07$ | $0.70 \pm 0.07$ |
| 100 | $0.77 \pm 0.04$ | $0.74 \pm 0.06$ | $0.75 \pm 0.05$ |
| 200 | $0.78 \pm 0.03$ | $0.77 \pm 0.04$ | $0.78 \pm 0.04$ |
| 400 | $0.80 \pm 0.03$ | $0.79 \pm 0.03$ | $0.79 \pm 0.03$ |

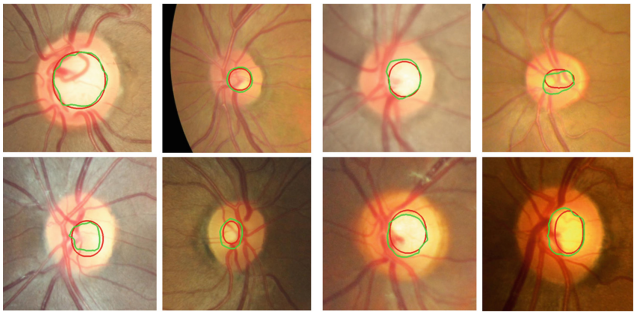

**Fig. 3.** Examples of the optic cup segmentation produced by the proposed method. The red color indicates the ground truth cup region (contour). The green color indicates the predicted optic cup region (contour)

## 4 Conclusion

In this paper, we have presented a novel semi-supervised segmentation algorithm based on variational autoencoder (VAE) to segment optic cup in retinal fundus images. The generative VAE was trained using large number of unlabeled images. The segmentation VAE, which maps the image to the segmentation mask, was then trained using limited number of labeled images by leveraging the feature embedding provided by the generative VAE. We have demonstrated the effectiveness of our proposed method using limited number of labeled samples to the challenging task of segmentation of optic cup in retinal fundus images. We have demonstrated that the proposed method improves the segmentation performance when the number of labeled images is limited. Therefore, our approach is useful in clinical applications where the availability of annotated images in limited. Although, we have applied our approach in segmentation of cup in fundus image, we believe that our method is equally applicable to other modalities.

## References


1. Chakravarty, A., Sivaswamy, J.: Coupled sparse dictionary for depth-based cup segmentation from single color fundus image. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 747–754. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_93
2. Joshi, G.D., Sivaswamy, J., Krishnadas, S.R.: Depth discontinuity-based cup segmentation from multiview color retinal images. IEEE Trans. Biomed. Eng. **59**(6), 1523–1531 (2012)
3. Sønderby, C.K., Raiko, T., Maaløe, L., Sønderby, S.K., Winther, O.: Ladder variational autoencoders. arXiv e-prints, February 2016
4. Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M.: Semi-supervised learning with deep generative models. arXiv e-prints, June 2014
5. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes (2013). CoRR, abs/1312.6114
6. Liu, J., Wong, D.W.K., Lim, J.H., Li, H., Tan, N.M., Zhang, Z., Wong, T.Y., Lavanya, R.: ARGALI: an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing. In: Lim, C.T., Goh, J.C.H. (eds.) 13th International Conference on Biomedical Engineering. IFMBE Proceedings, vol. 23, pp. 559–562. Springer, Heidelberg (2009). doi:10.1007/978-3-540-92841-6_137
7. Liu, Y., Xing, Z., Deng, C., Li, P., Guo, M.: Automatically detecting lung nodules based on shape descriptor and semi-supervised learning. In: 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), vol. 1, pp. V1-647–V1-650, October 2010
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on CVPR, pp. 3431–3440 (2015)
9. Maaløe, L., Sønderby, C.K., Sønderby, S.K., Winther, O.: Auxiliary deep generative models. arXiv e-prints, February 2016
10. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 140–148. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_17
11. Portela, N.M., Cavalcanti, G.D.C., Ren, T.I.: Semi-supervised clustering for MR brain image segmentation. Expert Syst. Appl. **41**(4), 1492–1497 (2014)
12. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. Br. J. Ophthalmol. **90**(3), 262–267 (2006)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation (2015). CoRR, abs/1505.04597
14. Sedai, S., Roy, P., Mahapatra, D., Garnavi, R.: Segmentation of optic disc and optic cup in retinal fundus images using coupled shape regression. In: Proceedings of the OMIA Workshop, pp. 1–8 (2016)
15. Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D.W.K., Wong, T.Y., Liu, J.: Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 788–795. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_98
16. You, X., Peng, Q., Yuan, Y., Cheung, Y., Lei, J.: Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. Pattern Recogn. **44**(10–11), 2314–2324 (2011)