Active learning of the ground truth for retinal image segmentation

Nedoshivina Liubov, Lensu Lasse

DOI: 10.1364/JOT.86.000697

UDC 004.93’11

ACTIVE LEARNING OF THE GROUND TRUTH FOR RETINAL IMAGE SEGMENTATION

© 2019 Liubov Nedoshivina, postgraduate master’s student[*], Lasse Lensu, Prof. [**], E-mail: nedshivina@gmail.com, lasse.lensu@lut.fi

* Pavlov Institute of Physiology, Russian Academy of Sciences

** Lappeenranta University of Technology

Different diseases can be diagnosed from eye fundus images by medical experts. Automated diagnosis methods can help medical doctors to increase the diagnosis accuracy and decrease the time needed. In order to have a proper dataset for training and evaluating the methods, a large set of images should be annotated by several experts to form the ground truth. To enable efficient utilization of expert’s time, active learning is studied to accelerate the collection of the ground truth. Since one of the important steps in the retinal image diagnosis is the blood vessel segmentation, the corresponding approaches were studied. Two approaches were implemented and extended by proposed active learning methods for selecting the next image to be annotated. The performance of the methods in the case of standard implementation and active learning application was compared for several retinal images databases.

Keywords: computer aided diagnosis, ·active learning, ·retinal image segmentation.

OCIS Codes: 100.4996, 100.4993, 100.7410, 100.2960, 110.7410, 110.3080, 170.5755

Submitted _____.

Introduction

Various diseases are diagnosed from eye fundus images by medical experts who look for specific lesions in the images. For screening and monitoring a progressive disease, automatic image processing methods are a well-motivated possibility to help a single expert's work or enable a wider screening program [1].

To develop and compare methods for automated image analysis, it is important to have reliable expert knowledge for the image content. In order to have a proper dataset for training and evaluating the methods, a large set of images should be annotated by several experts, and either the annotations should be fused to form the ground truth (GT) for the image content or the level of agreement and performance of the experts should be evaluated to define the gold standard [2].

One of the important steps in the retinal images-based diagnosis is the blood vessel segmentation. The active learning approaches can be applied to form the most informative and compact representation of the ground truth needed for the segmentation.

1. Related work

Active learning uses the idea of the online machine learning [3] where the training data is represented as an ordered sequence during the training. The main element of the active learning algorithms is an active query function by using which one can perform efficient selection or sampling of an object from an unlabeled dataset to a desired training set [4]. The selection of the next image to label and the following model training occur in one iteration of active learning called the active iteration.

A query selection strategy is required. One of the most popular query strategies is uncertainty sampling which was firstly proposed in 1994 by Lewis and Gale [5]. Here an active function queries a pool to find a sample on which classifier produces the most uncertain result.

1.1 Active learning of the ground truth for medical images

In the papers [6] and [7] devoted to the retinopathy diagnosis, C.I. Sánchez et al. pro-posed the uncertainty sampling and compared its performance to the random sampling. The main goal of their classification algorithm was to predict the exact type of the abnormality. The authors preprocessed the input images by using several filters based on Gaussian derivatives and then applied the kNN classifier. Their active learning un-certainty sampling strategy outperformed simple random sampling with area under the Receiver Operating Characteristic curve around 0.88 in the former case against 0.84 in the latter.

1.2 Databases of retinal images with the ground truth

As research in the field of the eye diseases diagnosis have been carried out for a longtime, several large databases of labeled eye fundus images are available. A database of 400 annotated images was created during the STARE project (STructured Analysis of the Retina), which was started in 1975 [8]. It could be applied in the segmentation tasks such as blood vessel segmentation. Also, for the segmentation purposes mainly, DRIVE [9] database was created. 40 segmented images are available in the database. CHASEDB1 [10] consists of the right and left eyes color images and contains 28 images with the ground truth in the form of the segmented images. One of the commonly used datasets which contain relatively large amount of manually segmented retinal images (143) is ARIADB [11]. Key characteristics of the databases can be found in Table 1. Sample images from each considered dataset are presented in Fig. 1.

2. Active learning for retinal image segmentation

Two segmentation methods were selected to examine the effectiveness of different active learning approaches. The both considered methods are supervised and require labeled dataset available.

2.1 Blood vessel segmentation based on Gabor features and supervised Bayesian classification

The method was proposed by Soares et al. [12] (the *Soares method* or the *Soares model*). The first stage of the algorithm is the preprocessing part where the Gabor filtering is applied to an inverted green channel of a retinal image.

The response of the filter is supposed to be a feature description of the input image. The wavelet transform $T\psi$ is calculated pixel-wise for the different scales a and angles θ in the range of [0..170]with the step of 10. The maximum modulus of these values from all the considered rotations is computed to form the feature:

$$M_{\psi}(\mathbf{b}, a) = \max_{\theta} \left| T_{\psi}(\mathbf{b}, \theta, a) \right|. \quad (1)$$

where ψ is the Gabor wavelet, $\mathbf{b}$ is the displacement vector. The obtained pixel features are then normalized with the mean and standard deviation values.

After the filter response is formed, the classification process based on the obtained feature description can be started. In this research, the Gaussian Mixture Model classifier based on the Bayes rule was selected.

As a result of the *Soares method*, a probability map can be obtained, where for each pixel there is a probability value of being a vessel. Based on the output probability map p with size $M \times N$, the overall image logarithm of the likelihood (loglikelihood, LL) can be calculated as follows:

$$LL = \sum_{\substack{0<i<M \\ 0<j<N}} \log p(i, j). \quad (2)$$

The LL value cane be used as a measure of the segmented image uncertainty. Having this criterion, it is possible to apply the active learning to train the model.

2.2 Segmentation based on the U-Net CNN architecture.

A convolutional neural network architecture called U-Net [13] was proposed by O.Ronneberger et al. in 2015. This network is one of the commonly used approaches for biomedical segmentation. The main purpose was to apply it to the

task of image segmentation. The feature of the U-Net is that it is supposed to be trained on a small training set.

The U-Net consists of 23 convolutional layers, where a convolution of an input data with filters of a specific size is performed. This CNN implements the encode-decode architecture [14], which means that it encodes an input image to a feature map and then decodes it to a desired output. The network does not have fully connected layers, as it is in the typical CNN architectures. Each output of the network can be considered as a classifier which predicts if current pixel belongs to a blood vessel or not.

In this work we selected one of the possible ways to apply active learning. The idea was to use the neurons activations to calculate the informativeness of the sample and based on it select the next one. As it was proposed in [15] by A. Kendall et al., having such an encode-decode network architecture the pixel-wise uncertainty can be estimated by using a dropout. The authors proposed to apply the Monte Carlo Dropout during the testing to calculate the uncertainty. This technique allows to approximate the weight distribution by minimizing the Kullback-Leibler divergence between the full posterior distribution and the approximating one. The most uncertain sample is selected based on computing the variance for each pixel for the different predictions.

3. Experiments and result

3.1 Evaluation criteria

For the evaluation of the segmentation results, a standard technique such as the Dice Similarity Coefficient and Accuracy was selected. The Dice Similarity Coefficient (DSC, DC) [16] is defined as:

$$DC(A,B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{3}$$

where $DC \in [0,1]$, A and B are sets which are the image segmented by a model and the image segmented by an expert in case of the segmentation task, $|A|$means the number of elements in the set.

3.2 Parameter selection

Parameters of segmentation based on the Soares method. For the preprocessing step, three wavelet levels should be specified. The optimal wavelet levels were studied in [17] and can be found in Table 2. Firstly, the Soares model was trained on the full available labeled training dataset and then tested. The model obtained after training is called the fully trained model. ARIADB dataset has 143 segmented images. 40 images were selected to compare active learning performance in this dataset with performance in other *smaller* databases. The results of training the Soares model on the full available training set are presented in Table 2.

U-Net segmentation parameters. The U-Net architecture is implemented in Python by using high-level neural network API Keras based on the research [18]. The experiments on the network training were conducted on GPU NVIDIA GeForce TITAN Black, 6 Gb RAM, Intel Xeon CPU E5-2680, 128 Gb RAM. To satisfy these memory resource conditions the optimal image resolution was selected to be 320×320. Also, the sizes of the network layers were reduced taking into account the memory restrictions.

The training parameters can be found in Table 3. The size of the initial labeled set was selected according to the size of the overall training set available.

3.3 Active learning with the Soares model

At first the model was trained on two images from the training set and then retrained with each new frame selected based on the active query. The active learning process in comparison with the performance of the fully trained model sampling is shown in Fig. 2. All the presented results of the Soares method evaluation contain information of a single run. From these results it can be noticed

that the model learns features quite fast and reaches the performance of the fully trained model after the 4-6 active iterations.

Being the uncertainty measure for the Soares method, the LL was calculated for all the segmented images in the unlabeled set in each active iteration. A plot of LL value in each training iteration for the both considered query functions is presented in Fig. 3. In the case of the uncertainty sampling, as it is presented in Fig. 3 the LL gradually increases with each new selected frame whereas in the case of random sampling this does not happen. This fact means that by applying uncertainty sampling model can be trained more efficient.

The visual results in comparison with results obtained with the fully trained model are shown in Fig. 4 (DRIVE database). On the segmented images from the DRIVE in Fig. 4c obtained during the testing of the random sampling the elements of the retina round border can be seen. At the same time, the result of the uncertainty sampling in Fig. 4b is close to the fully trained model results (Fig. 4a) already on the fourth active iteration.

3.4 Active learning with U-Net

The main purpose of this experiment was to explore the possibility to train U-Net with small labeled dataset by means of active learning and achieve the performance level of the model trained in larger set.

In the case of the small amount of training data the preliminary pretraining on the patches from known annotated images can improve learning performance. U-Net is a fully convolutional network, which allows to pretrain it on the set of small patches and then train it on larger images without changes in the architecture. The patch is a small piece of image and corresponding segmentation map. In this research the model pretrained on the set consisted of 100 patches was used. Example of patches and corresponding segmentation maps is presented in Fig. 5. Patch set was prepared based on the 5 segmented maps of DRIVE database since only half of the dataset was included to a training set. DRIVE dataset was selected

from the others because of high visual quality of eye fundus images and segmentation maps. A size of one patch is 96x96 pixels.

To obtain an expected performance level for evaluation active learning of the model, the U-Net was trained on the available training images for each considered dataset.

For the active learning, the pretrained model was used. Pretraining was done with 300 epochs. Then the active learning performance was assessed. Three runs of the model training were conducted. After each active iteration one new frame was annotated and added to the train set.

The only dataset training on which could provide the DC compatible with the Soares model is DRIVE. For this reason, results in this dataset were considered more carefully. The sample testing results from the active model obtained in the fourth iteration are presented in Fig. 6 in comparison with the testing results of the full trained model. The active model segmentation map Fig. 6c has no elements of the round border. The thin vessels are almost absent, but the wide vessels are better segmented and connected. As it can be seen in Fig. 6, segmentation result in Fig. 6b is slightly better in the thin vessel segmentation, but it also contains parts of the round edges in the border.

Conclusions

For the Soares model with the proposed uncertainty measure method, the expected performance level was achieved in 4-6 active iterations depending on the dataset. The best performance and learning rate were achieved while training on DRIVE database. The conducted experiments have shown that by means of the active learning the compact representation of the training set based on the most informative images is possible. The initial results have shown that usage of active learning has an impact on the training process and allows to train the model better, hence it can be effectively applied in the retina blood vessels segmentation task.

As for the future research following measures can be taken. In order to make the query function more accurate, one can use a mask image which can reduce the region of interest when calculating the uncertainty to the retina part of the image only. This mask can be presented in the form of a binary image where 1 corresponds to the retina and 0 to the background. The influence of such mask application on the active learning performance is a question to study further.

Another direction to a more careful research is combining several retinal image databases into one. It could increase the amount of information on which the model learns the features. The conducted experiments on the mixed dataset have shown the necessity to preprocess the images from different databases. Hence, in order to conduct this research, one needs to make the mixed dataset homogeneous, for example, correct the color, illumination and/or resolution.

References

1 *Faust, O., Acharya, R., Ng, E.Y.K., et al.* Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review // Journal of Medical Systems. 2012. V. 36. №1. P. 145–157.

2 *Gulshan, V., Peng, L., Coram, M. et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs // JAMA. 2016. V. 316 №22. P. 2402–2410.

3 *Shalev-Shwartz, S.* Online learning and online convex optimization // Foundations and Trends in Machine Learning. 2012. V. 4. № 2. P. 107–194.

4 *Settles, B.* Active learning // Synthesis Lectures on Artificial Intelligence and Machine Learning. 2012. V. 6. № 1. P. 1–114.

5 *Lewis, D.D., Gale, W.A.* A sequential algorithm for training text classifiers. // Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Springer-Verlag New York, Inc. 1994. P. 3–12.

6 *Sánchez, C.I., Niemeijer, M., Kockelkorn, T. et al.* Active learning approach for detection of hard exudates, cotton wool spots, and drusen in retinal images // Medical Imaging 2009: Computer-Aided Diagnosis. 2009. V. 7260. P. 72601I.

7 *Sánchez, C.I., Niemeijer, M., Abràmoff, M.D. et al.* Active learning for an efficient training strategy of computer-aided diagnosis systems: application to diabetic retinopathy screening // International Conference on Medical Image Computing and Computer-Assisted Intervention. 2010. P. 603–610.

8 *Hoover, A. Goldbaum, M.* Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. // IEEE Transactions on Medical Imaging. 2003. V. 22, №. 8. P. 951-958.

9 *Staal, J., Abràmoff, M.D., Niemeijer, M. et al.* Ridge-based vessel segmentation in color images of the retina // IEEE Transactions on Medical Imaging. 2004. V. 23. №4. P. 501–509.

10 *Owen, C.G., Rudnicka, A.R., Mullen, R., et al*. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. // Investigative Ophthalmology & Visual Science. 2009. V. 50. №5. P. 2004–2010.

11 *Farnell, D.J., Hatfield, F., Knox, P. et al.* Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators // Journal of the Franklin Institute. 2008. V. 345. № 7. P. 748–765.

12 *Soares, J.V., Leandro, J.J., Cesar, R.M. et al.* Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification // IEEE Transactions on Medical Imaging. 2006. V.25. № 9. P. 1214–1222.

13 *Ronneberger, O., Fischer, P., Brox, T.* U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. 2015. pp. 234–241.

14 *Badrinarayanan, V., Kendall, A., Cipolla, R.* Segnet: A deep convolutional encoder-decoder architecture for image segmentation. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. V. 39. № 12. P. 2481–2495.

15 *Kendall, A., Badrinarayanan, V., Cipolla, R.* Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. // 28th British Machine Vision Conference. 2017.

16 *Sorenson, T.* A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. // K Dan Vidensk Selsk Biol Skr. 1948. V. 5, P. 1–34.

17 *Vostatek, P., Claridge, E., Uusitalo, H. et al.* Performance comparison of publicly available retinal blood vessel segmentation methods // Computerized Medical Imaging and Graphics. 2017. V. 55. P. 2–12.

18 *Gorriz M.* Cost-effective active learning for melanoma segmentation // arXiv preprint arXiv:1711.09168. 2017.

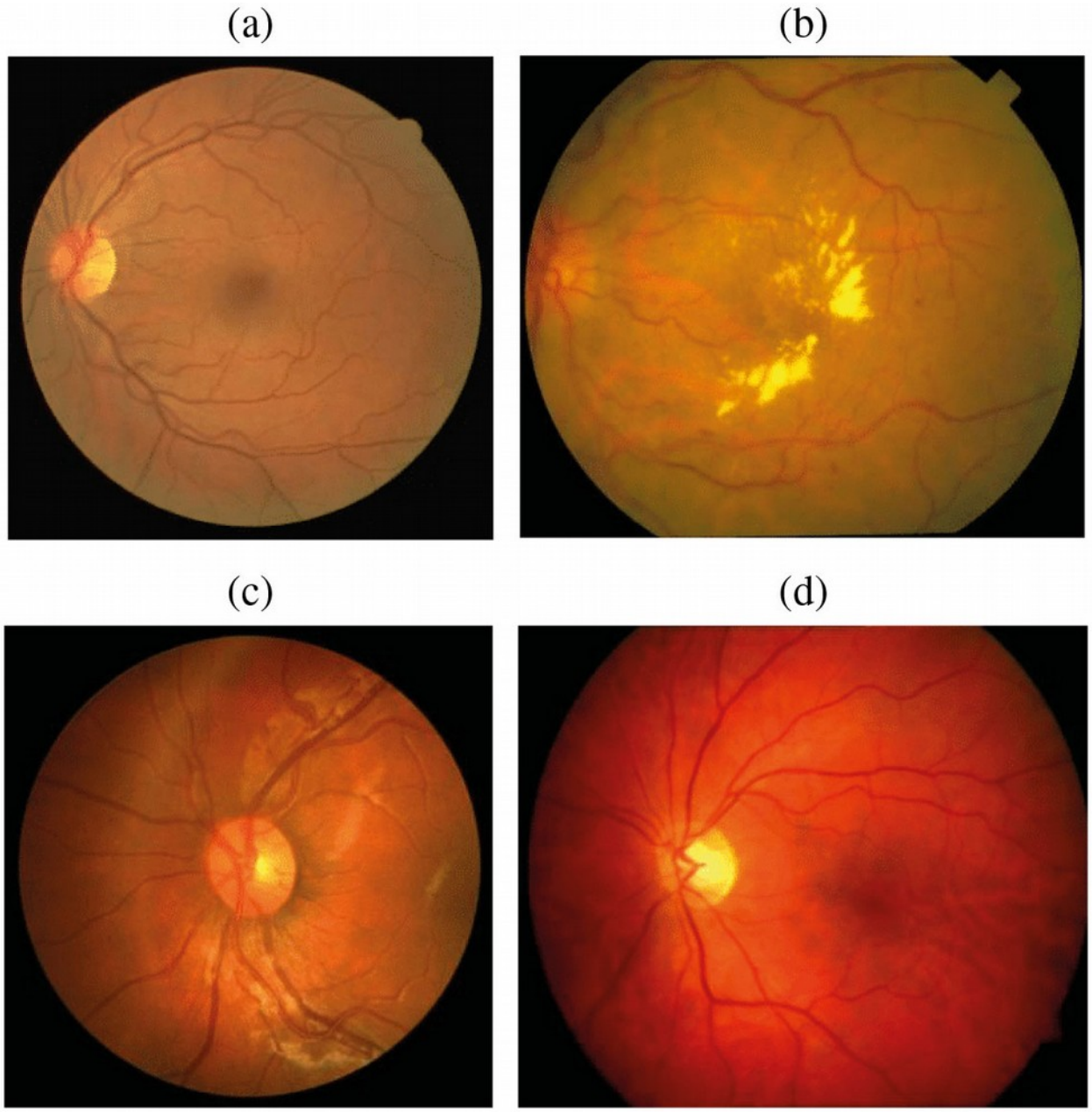

Fig. 1. Sample images from the DRIVE (a), STARE (b), CHASEDB1 (c) and ARIADB (d) datasets.

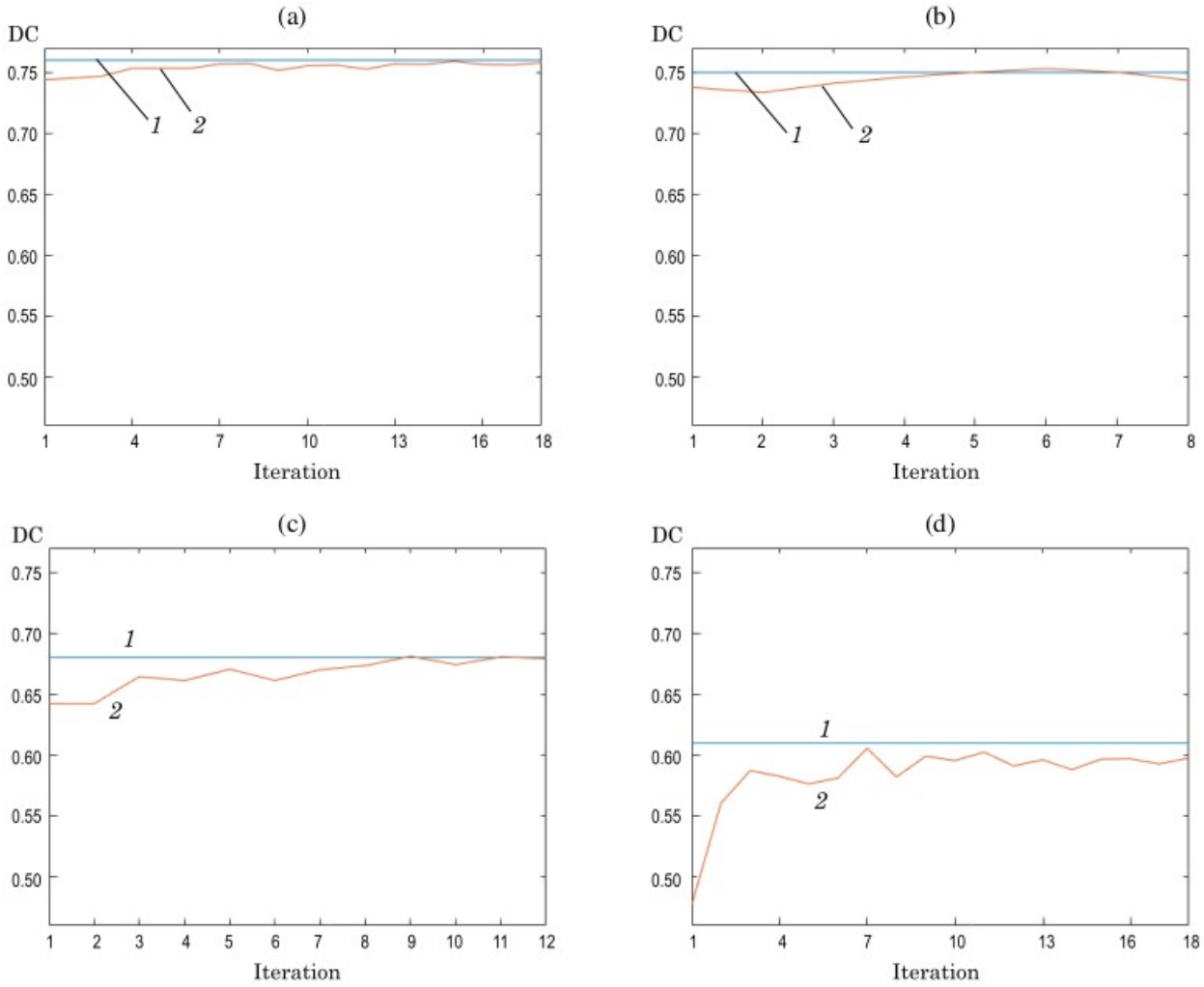

Fig. 2. Comparison of the active learning performance and the performance of the fully trained Soares model: the DRIVE (a), STARE (b); CHASEDB1 (c); ARIADB (d).

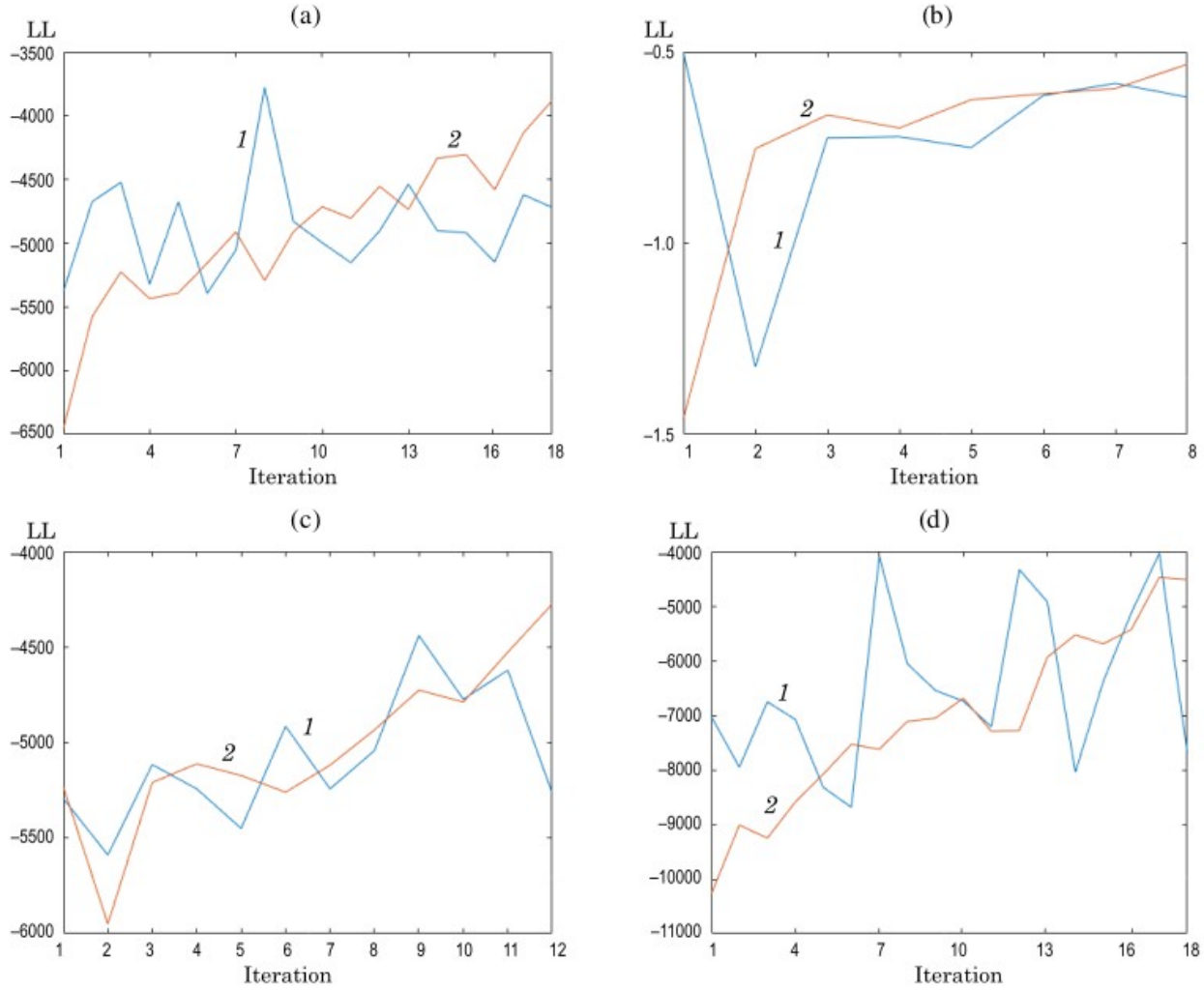

Fig. 3. The LL value among the segmented images from the unlabeled set after each iteration of the Soares model training (DRIVE database).

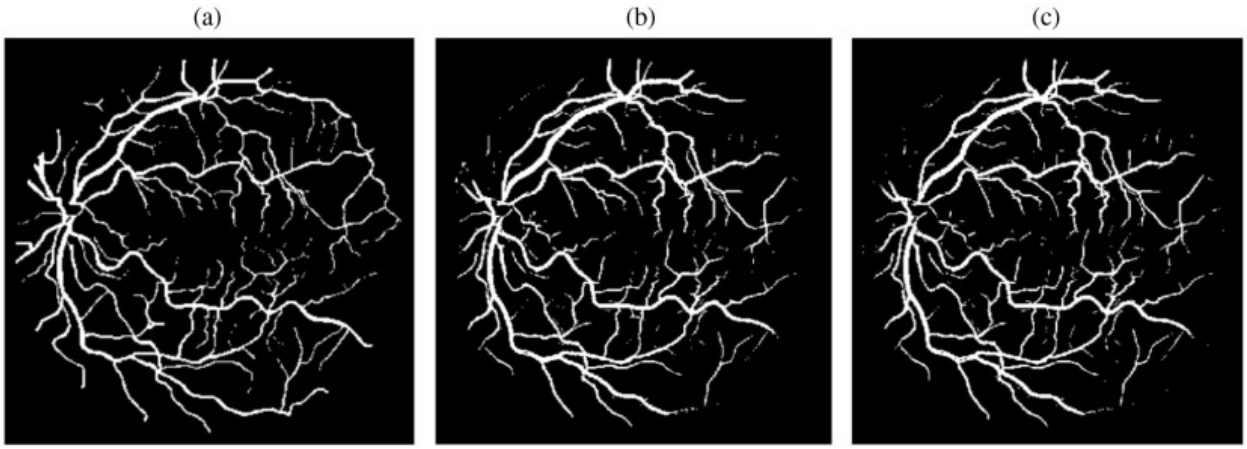

Fig. 4. Example testing results of the Soares model on the DRIVE: (a) the fully trained model, (b) the active trained model with the uncertainty sampling; (c) the active trained model with the random sampling.

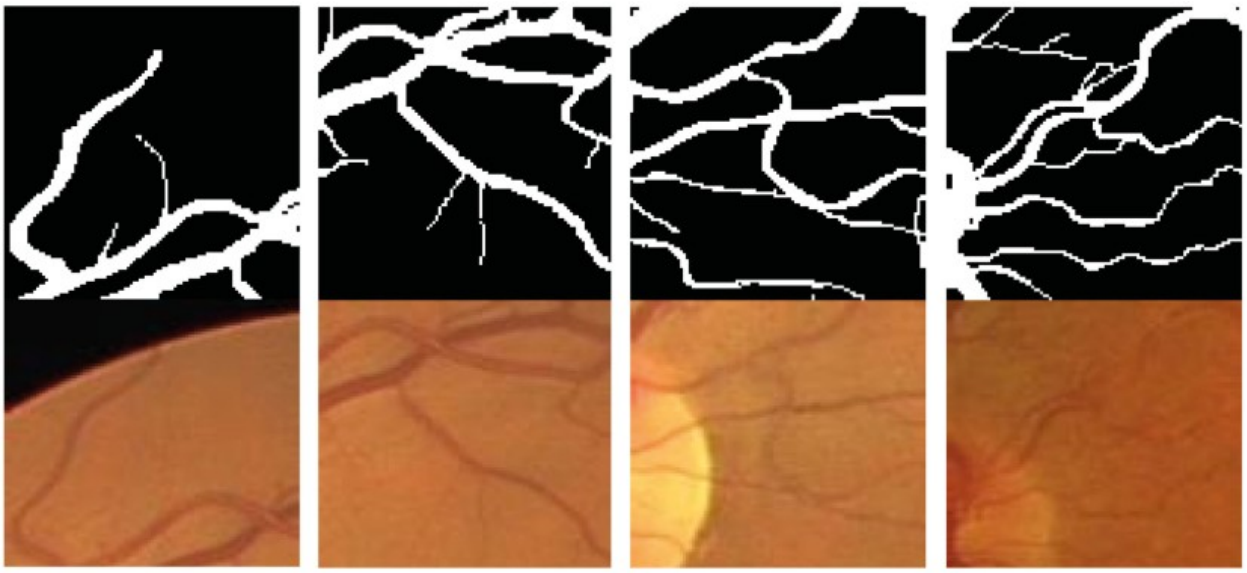

Fig. 5. Example patches for pretraining of the U-Net.

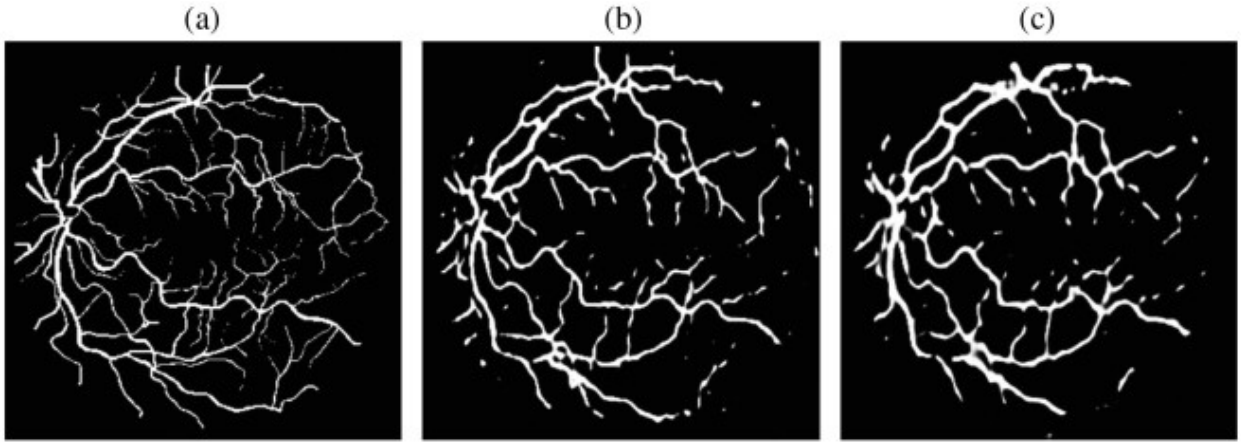

Figure 6. Example testing results of the U-Net mode after four active training iterations and the model trained on the full available DRIVE training set (30 images) with 200 epochs: (a) the manually segmented image; (b) the fully trained model and (c) the active trained model.

Table 1. Selected databases, N_t - the total number of images in the dataset, N_{seg} is the amount of the segmented images (GT), N_{GT} is the amount of GT sets per each image.

Database	N_t	N_{seg}	**Resolution**	N_{GT}
STARE [8]	400	20	700×605	2
ARIADB [11]	143	143	768×576	1
DRIVE [9]	40	40	584×565	2
CHASEDB1 [10]	28	28	500×480 (999×960)	1

Table 2. Results of the Soares method with the model trained on the full training set, where N_{train} and N_{test} are sizes of training and testing sets respectively.

Database	N_{train}	$Ntest$	**Wavelet levels**	**DC**
STARE [8]	10	10	[2, 3, 6]	0.75
ARIADB [11]	20	20	[2, 5, 6]	0.61
DRIVE [9]	20	20	[2, 3, 5]	0.76
CHASEDB1 [10]	14	14	[3, 8, 9]	0.68

Table 3. Training parameters of the U-Net for each considered dataset. N_{train} is the size of the training set, N_{test} is the size of the testing set, E_f is the number of the full training epochs, N_i is the size of the initial labeled set, I_a is the number of active iterations, E_a is the number of training epochs per each active iteration.

Database	N_{train}	N_{test}	E_f	N_i	I_a	E_a
STARE [8]	30	10	200	5	25	6
ARIADB [11]	10	10	200	2	8	6
DRIVE [9]	18	10	200	4	14	6
CHASEDB1 [10]	30	10	200	5	25	6