Shape-based Segmentation of Retinal Layers and Fluids in OCT Image Data

Timo Kepp[a,b], Julia Andresen[b], Claus von der Burchard[c], Johann Roider[c], Gereon Hüttmann[d], and Heinz Handels[a,b]

[a]German Research Center for Artificial Intelligence, Lübeck, Germany
[b]Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
[c]Department of Ophthalmology, University of Kiel, Kiel, Germany
[d]Institute of Biomedical Optics, University of Lübeck, Lübeck, Germany

ABSTRACT

The main application of optical coherence tomography (OCT) is in the field of ophthalmology, where it is used for diagnosis of various eye diseases. The automatic segmentation of the individual retinal layers as well as pathological structures in OCT scans is helpful for clinical examination and treatment planning. Current methods often do not consider the strict arrangement of retinal layers. Although graph-based methods are suitable for correcting topology errors, their applicability is costly and complex, especially in the presence of pathologies. In this work, a segmentation method is proposed that utilizes additional shape information of the retinal layers to provide improved topology preservation while maintaining simple applicability. For this purpose, a U-Net-based network architecture is extended to a multi-task approach to allow regression of the shape information of the retinal layers in addition to pixel-wise classification. This introduces spatial regularization and allows the generation of plausible segmentations. A consistency term ensures agreement between the classification and regression task, which also allows for semi-supervised training. In a comprehensive evaluation, the performance of the proposed multi-task approach is investigated, using OCT image data from patients with diabetic macular edema. The results demonstrate that the integration of shape information improves the preservation of the topology of the retinal layers. Moreover, the use of a semi-supervised training scheme via a consistency term improves the robustness and refines the fluid delineation of the proposed method.

Keywords: DME, OCT, Retina, Fluid, Segmentation, Deep Learning

1. PURPOSE

Diabetic macular edema (DME) is the most common cause of blindness in patients with diabetes.[1] As a complication of diabetic retinopathy, DME leads to the growth of abnormal, leaky blood vessels in the macula as a result of high blood glucose levels. This results in edema within the macula, causing swelling of the affected area that significantly impairs the sharp central vision. Hemorrhage leads to increased intraretinal fluid near the inner nuclear as well as outer plexiform layer. The swelling caused by the fluid also results in morphological deformation or severing of the retinal layers. In the early stages of the disease, retinal damage often remains unnoticed by the patient, thus regular examinations by the ophthalmologist are crucial. Aside from funduscopy, OCT is utilized as an imaging modality both for diagnosis and therapy monitoring.[2] OCT enables high-resolution cross-sectional imaging of the retina, allowing detailed visualization of pathological changes as well as their quantitative evaluation. Here, both the condition of the retinal layers and the fluid accumulations represent important OCT-based biomarkers that serve as functional predictive factors in DME therapy.[3]

The retina is composed of several layers that follow a strict hierarchy. To ensure a correct topology, some early works use graph-based segmentation approaches.[4–6] However, the conception of graph models with varying constraints for individual retinal layers and pathologies is complex and thus reduces the flexibility of segmentation methods. Current deep learning methods such as fully convolutional networks (FCNs) offer a simpler applicability together with high segmentation accuracy.[7]

Send correspondences to Timo Kepp: timo.kepp@dfki.de

Despite their ability to learn highly complex contextual information, topological contexts are not considered by most current FCN-based methods. To cope with this problem, several FCN-based methods have been proposed to flexibly incorporate shape and topology information into the segmentation process, reducing costly post-processing.[8–10] To account for the topological locations of individual retinal layers in relation to each other, Wei et al. propose mutex Dice loss.[8] He et al. use a cascade of two U-Nets to segment several retinal surfaces as well as macular edema.[9] While the first network performs pixel-wise classification, the second is used for post-processing to ensure a correct topology of the retinal surfaces. Fully connected layers at the network's output directly regress the distances between the retinal surfaces. In a subsequent work, He et al. combine these two steps into a unified framework by directly modeling the distribution of surface positions.[10]

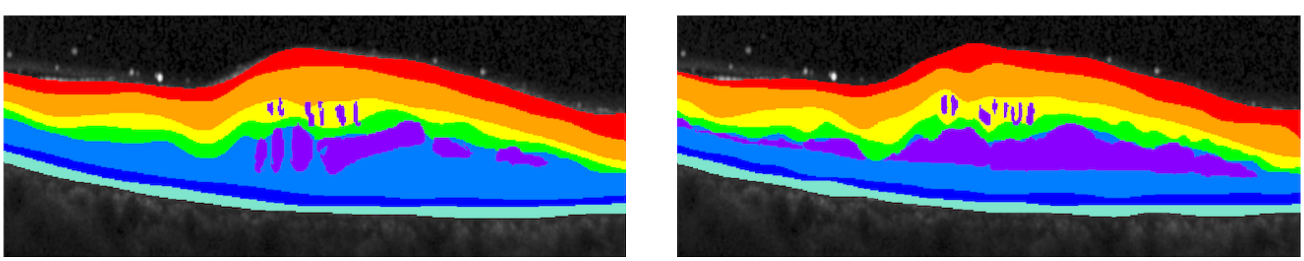
Several works have shown that the inherent shape information of distance maps can be used to regularize the segmentation process of FCNs.[11] Unlike (binary) segmentation masks, whose pixels are divided into foreground and background, intensities of a distance map correspond to the respective distances to the nearest object boundary. Often, signed distance maps (SDMs) are used, which depict negative distances within the object contour. By using SDMs for the segmentation of the synaptic cleft in electron microscope images, Heinrich et al. achieved the first place on the leaderboard in the CREMI challenge (Circuit Reconstruction from Electron Microscopy Images, `https://cremi.org`).[12] By regressing SDMs for multi-class segmentation of several retinal layers and edema in OCT image data, Kepp et al. show that topological errors can be reduced.[13] Inspired by this approach, further work uses SDMs to segment retinal layers. Cao et al. propose a multi-task framework in their work that regresses retinal layers in the form of SDMs in addition to a pixel-wise classification.[14] In a subsequent work, Liu et al.[15] suggest an adversarial learning strategy to focus the FCN on difficult regions in the OCT image.

Although SDM-based regression improved topology preservation in our previous work,[13] it results in fuzzy fluid boundaries. To address this problem, this paper proposes a multi-task framework using an additional classification path in parallel with the SDM regression of retinal layers to provide a sharper delineation of fluid regions. We adapt the dual-task consistency[16] (DTC) to increase the consistency between regression and pixel-wise classification task. The DTC allows a semi-supervised learning strategy, which significantly increases the amount of available training image data. To avoid stability problems during training, we use a special loss function.[17]

2. METHODS

2.1 Dataset

The proposed approach was developed and evaluated using the publicly available Duke OCT dataset introduced in.[18] The dataset consists of a total of ten OCT scans (Heidelberg Spectralis SD-OCT scanner) of patients with DME. Every scan consists of 61 B-scans, each with a size of 512×740 pixels. Two medical experts manually annotated seven retinal layers and fluids in eleven B-scans centered at the fovea (distance to foveal B-scan: 0, ± 2, ± 5, ± 10, ± 15, ± 20) for each OCT scan. This results in a total of 110 annotated B-scans per expert. In addition to the fluids (■), the following retinal layers were segmented: retinal nerve fiber layer (RNL ■), ganglion cell layer and inner plexiform layer (GCL-IPL ■), inner nuclear layer (INL ■), outer plexiform layer (OPL ■), outer nuclear layer and (ONL-ISM ■), inner segment ellipsoid (ISE ■), outer segments of the photoreceptor cells



(a) Expert 1 (b) Expert 2

Figure 1: Segmentation example of a B-scan by both experts.

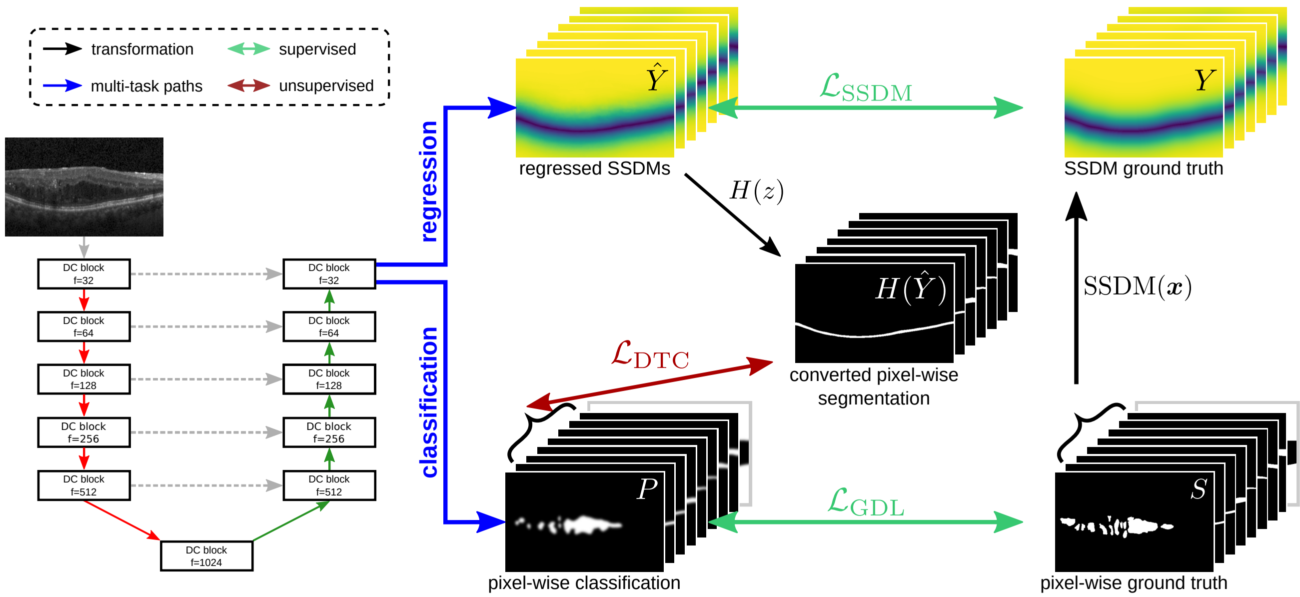


Figure 2: Overview of our proposed MTL segmentation method. In addition to pixel-wise classification, regression of retinal layers is performed in the form of scaled SDMs (SSDMs). The DTC provides consistency between tasks and enables semi-supervised training.

and retinal pigment epithelium (OS-RPE ▢). Combined with the background, the ground truth thus consists of $C = 9$ classes. Figure 1 shows an example B-scan with the manual segmentations of both experts. Further details regarding the dataset are available in the work of Chiu et al.[18]

2.2 Multi-task Learning for Shape-based Segmentation

By solving multiple tasks with a single FCN, more meaningful feature representations are learned in early layers, while task-specific representations can be learned in the separate parts of the network. Thus, multi-task learning (MTL) can achieve higher generalizability than single-task learning (STL). In order to both maximize the number of shared feature representations and also keep the sum of learnable network parameters small, a multi-head FCN architecture is used for the proposed MTL segmentation method. A schematic overview of the method is shown in Fig. 2. As a basis, we use the densely connected U-Net (DCU-Net) proposed in a previous work.[19] Compared to the standard U-Net, the DCU-Net incorporates additional skip connections in its DC blocks for improved feature reuse. Two $1{\times}1$ convolutional layers at the end of the network realize task-specific paths of the FCN.

During the training phase, the segmentation framework receives single B-scans as input and outputs a pixel-wise classification for each class as well as a regressed scaled SDMs (SSDM) for each retinal layer. Pixel-wise classification assigns each image pixel $\boldsymbol{x}$ of the input B-scan I to one of $C = 9$ classes consisting of seven retinal layers, fluids, and background. Generalized Dice loss ($\mathcal{L}_{\mathrm{GDL}}$) is used to train the classification path, which is computed between the softmax probability P and the ground truth segmentation S. SSDMs of the retinal layers are estimated via the regression path ($\mathcal{L}_{\mathrm{SSDM}}$), which has a regularizing effect on the segmentation and should lead to better preservation of the topological relations. For this purpose, the individual binary masks of the retinal layers are selected in the ground truth S (see curly brackets in Fig. 2) and transformed into SSDMs Y. Fluid and background classes are not regressed using SSDMs, since they do not follow a specific shape. Both task-specific paths are connected via the DTC ($\mathcal{L}_{\mathrm{DTC}}$), which provides consistency between both classified and regressed retinal layers. For this purpose, the regressed retinal SDMs are converted into pixel-wise segmentations using an approximated Heaviside function $H(z)$. This allows the use of unlabeled B-scans via a semi-supervised training strategy. At test time, only the classification path is used to generate the final segmentation S, which is determined by applying the $\arg\max$ operator on the pixel-wise class probabilities.

The proposed method is described in more detail in the following sections, including the regression of (S)SDMs, the dual-task consistency and the semi-supervised training process.

2.2.1 Regression of the Signed Distance Maps

The regression of the SDMs is performed using the ground truth masks S_l. For each class label l, the corresponding SDM can be computed using one-hot encoding as follows:

$$\mathrm{SDM}(\boldsymbol{x}) = \begin{cases} -\inf_{\boldsymbol{y}\in\mathcal{S}} \|\boldsymbol{x}-\boldsymbol{y}\|_2, & \boldsymbol{x}\in\Omega_{\text{in}} \\ 0, & \boldsymbol{x}\in\mathcal{S} \\ \inf_{\boldsymbol{y}\in\mathcal{S}} \|\boldsymbol{x}-\boldsymbol{y}\|_2, & \boldsymbol{x}\in\Omega_{\text{out}} \end{cases} \tag{1}$$

where the areas inside and outside the object are represented by Ω_{in} and Ω_{out}. $\mathcal{S}$ represents the object's contours and $\|\boldsymbol{x}-\boldsymbol{y}\|_2$ denotes the Euclidean distance between the pixels $\boldsymbol{x}$ and $\boldsymbol{y}$. The algorithm of Maurer et al.[20] is used for an efficient computation of the SDM. SDMs encode segmentation masks into a higher-dimensional space that contains more complex information about the object's shape. For example, negative values denote image areas within the object, while positive values represent background areas. Values near or equal to zero denote positions near or on the object contour. While small changes in the segmentation mask affect only local points, they have a larger impact on the SDM and thus lead to larger errors. As a result, SDMs are much more sensitive compared to (binary) segmentation masks and penalize outliers more severely.[21] Since large distances to the respective retinal layers are not meaningful, scaled SDMs (SSDMs) are used as suggested by Heinrich et al.:[12]

$$\mathrm{SSDM}(\boldsymbol{x}) = \tanh\left(\mathrm{SDM}(\boldsymbol{x})\cdot s\right). \tag{2}$$

While s is used to scale the SDM, the tanh nonlinearity effectively saturates all distance values in the range $[-1, 1]$. For our experiments, we set $s = 0.02$.

SSDMs of retinal layers are regressed using the corresponding path in the FCN model (see Fig. 2). For each retinal layer, a corresponding SSDM is learned. The tanh function is used as a nonlinear activation to constrain the output to the desired value range. For regression of the SSDMs, the L1 loss function is applied. Regressing multiple structures with the L1 loss ($\mathcal{L}_{\text{L1}}$) may result in unstable training. To solve this problem, Xue et al. propose a complementary product-based loss function:[17]

$$\mathcal{L}_{\text{Product}} = -\sum_{l=1}^{C-2} \frac{\sum_{\boldsymbol{x}} \hat{Y}_l(\boldsymbol{x})Y_l(\boldsymbol{x}) + \varepsilon}{\sum_{\boldsymbol{x}} \hat{Y}_l(\boldsymbol{x})Y_l(\boldsymbol{x}) + \sum_{\boldsymbol{x}} \hat{Y}_l^2(\boldsymbol{x}) + \sum_{\boldsymbol{x}} Y_l^2(\boldsymbol{x}) + \varepsilon}. \tag{3}$$

Here, Y describes the SSDM ground truth and $\hat{Y}$ the prediction for the regression path. The parameter ε ensures numeric stability and is set to 10^{-5}. This product penalizes a wrong sign in the regressed SSDM $\hat{Y}$.

A combination of both loss functions is used for the SSDM regression: $\mathcal{L}_{\text{SSDM}} = \mathcal{L}_{\text{Product}} + \mathcal{L}_{\text{L1}}$, which results in larger gradients, thus improving the regression results and making the training more robust.[17] Note that $\mathcal{L}_{\text{SSDM}}$ is calculated only for the class labels of the retinal layers $l \in \{1, \ldots, C-2\}$ and not for the background ($l = 0$) or the fluids ($l = C-1$).

2.2.2 Semi-supervised Training

As a result of both tasks, the retinal layers are output as class probabilities on the one hand, and as higher-dimensional SSDMs on the other. Therefore, it is important that both predictions are consistent with each other. For this purpose, we adopt the dual-task consistency (DTC) proposed by Luo et al.[16] Using the Heaviside function, the values of the regressed SSDMs can be converted into probabilities. Since the exact Heaviside function is not differentiable, a smooth approximation is applied:

$$H(z) = \frac{1}{1+e^{-z\cdot k}}, \tag{4}$$

where z represents the SSDM value at pixel position $\boldsymbol{x}$ and k specifies the slope of the function. According to Xue et al.,[17] we set $k = 1500$ for our experiments. The DTC is realized via an L1 loss function ($\mathcal{L}_{\text{DTC}}$) computed between $H(\hat{Y}_l(\boldsymbol{x}))$ and $P_l(\boldsymbol{x})$.

A major advantage of DTC is that no ground truth is needed, since $\mathcal{L}_{\text{DTC}}$ is determined between the (transformed) predictions of the retinal layers of the regression and classification path (see Fig. 2). This allows for a semi-supervised training procedure where the 50 unlabeled B-scans of each OCT scan (see Sec. 2.1) can be used for training in addition to the eleven labeled ones. Pixel-wise classification ($\mathcal{L}_{\mathrm{GDL}}$) as well as regression of the SSDMs ($\mathcal{L}_{\mathrm{SSDM}}$) follows supervised training, whereas DTC ($\mathcal{L}_{\mathrm{DTC}}$) runs completely unsupervised. The overall loss function is composed as:

$$\mathcal{L} = \underbrace{\mathcal{L}_{\mathrm{GDL}} + \beta\mathcal{L}_{\mathrm{SSDM}}}_{\text{supervised}} + \underbrace{\gamma_t\mathcal{L}_{\mathrm{DTC}}}_{\text{unsupervised}} \ , \tag{5}$$

with weighting factors β and γ_t. We set a value of five for β in the experiments. In contrast, γ_t represents a time-dependent weighting function that controls the balance between the supervised and unsupervised loss functions. It is defined as follows:

$$\gamma_t(t) = e^{\left(-5\left(1-\frac{t}{t_{\max}}\right)^2\right)}. \tag{6}$$

Here, t is the current training epoch and $t_{\max}$ is the total number of training epochs. As a result of $\gamma_t(t)$, the network is trained in a supervised manner in the first epochs. Otherwise, the DTC would prevent the learning of the subtasks at the beginning of the training. In later stages of training, $\mathcal{L}_{\mathrm{DTC}}$ gradually receives more weight.

3. RESULTS

For evaluation of the proposed framework, we compare it with several methods. To investigate the advantages of MTL, the individual subtasks of the presented framework are realized as STL versions. For this purpose, separate STL frameworks are built based on both the classification path ($\mathrm{STL}_{\mathrm{Class.}}$) and the regression path ($\mathrm{STL}_{\mathrm{Regr.}}$). Since the SSDMs can be transformed into pixel-wise classifications using Eq. 4, it seems obvious to perform regression and classification sequentially without separate task paths. This is consistent with the methodology of Xue et al.[17] and will be referred to as $\mathrm{MTL}_{\mathrm{Xue}}$. Our proposed method and the associated semi-supervised learning strategy using DTC are examined via an ablation study. Accordingly, the proposed multi-head FCN is evaluated without ($\mathrm{MTL}_{\mathrm{MH}}$) and with the DTC component ($\mathrm{MTL}_{\mathrm{MH}_{\mathrm{DTC}}}$).

The five methods described above are evaluated using leave-one-patient-out cross-validation using the image dataset described in Sec. 2.1. Since two separate expert annotations are available, we cross-validate twice, each time using a different ground truth. All B-scans are cropped to exclude irrelevant regions of the image and to fit the manual annotations that occupy only a smaller portion of the image. For each cross-validation run, the OCT scans are divided into non-overlapping training, validation, and test datasets in a ratio of 7–2–1. To cope with the limited amount of available training samples, we perform data augmentation. It consists of the following components whose parameterization is random: horizontal flip, rotation, translation, scaling and non-linear intensity shift. To reduce the influence of atypical morphologies, such as in the regions with edema, we use elastic transformations[22] in addition. For quantitative analysis of the segmentation performance, we determine Dice similarity coefficient (DSC), average symmetric surface distance (ASSD) and Hausdorff distance (HD). Mean values of these metrics are listed for each structure in Tab. 1 and their distributions are illustrated as boxplots in Fig. 3. The inter-rater reliability (IRR) between the two ground truths for each metric is also reported. Moreover, three qualitative segmentation examples are shown in Fig. 4.

All methods achieve high accuracy in segmenting the retinal layers and achieve expert level (IRR) regarding DSC and ASSD. However, segmentation in the foveal region is difficult because the retinal layers are rather thin in this region (see Fig. 4). Segmenting the INL, OPL, and ONL-ISM proves to be even more challenging due to the accumulation of edema fluid in their region. Although the segmentations of the ONL-ISM layer show high DSC values, their ASSD and HD are considerably larger compared to the other retinal layers, which is due to its relatively large size. With respect to their layer thickness, both ISE and OS-RPE can be segmented most robustly, which is also indicated by the small interquartile distances of the corresponding boxplots. In contrast to the retinal layers, fluid areas are segmented with considerably less accuracy by all methods. The $\mathrm{STL}_{\mathrm{Class.}}$ method shows the highest average DSC value of 0.553 for this class. The DSC values of the other methods are on average 0.522, which corresponds to the IRR level. Despite the separation into two different task paths, topological errors are mostly avoided. Finally, the DTC achieves smaller interquartile ranges for the DSC values (see Fig. 3).

As demonstrated in our previous work, sole regression ($\mathrm{STL}_{\mathrm{Regr.}}$) achieves a better preservation of the topological order of the retinal layers (see Fig. 4). The obtained results are on average inferior in comparison to the

$STL_{Class.}$ method and small retinal layer variations are not captured, leading to higher ASSD values. Edema fluids are difficult to segment with sole regression ($STL_{Regr.}$) leading to a poor separation and diffuse boundaries. $STL_{Class.}$ method shows the highest DSC values for the fluid class. However, severe topology errors often occur in larger fluid regions (see Fig. 4).

By using a subsequent classification (MTL_{Xue}), the segmentation accuracy regarding NFL, GCL-IPL, and INL can be significantly improved compared to the $STL_{Regr.}$ method. Significant improvements can also be observed affecting HD values for the GCL-IPL, INL, and OPL compared to $STL_{Class.}$. Qualitatively, it can also be shown that fluids accumulations are better separated and their boundaries are less diffusely segmented while considering the topological structure.

Both proposed multi-head approaches (MTL_{MH} and $MTL_{MH_{DTC}}$) provide the best segmentation results for the retinal layers. With only a few exceptions, their results differ significantly from those obtained using the STL methods. The use of DTC reduced the HD values, which are significantly different except for NFL and INL. The $MTL_{MH_{DTC}}$ method shows significantly smaller HD values than all STL methods, which only applies to the first two to four retinal layers (starting at NFL) for the MTL_{MH} variant. In contrast, the ASSD values of both MTL_{MH} variants did not differ. Regarding the DSC values, significant differences are observed only for the retinal layers OPL, ONL-ISM and ISE. Fluids are segmented with the MTL_{MH} methods at a similar level as with the sole regression ($STL_{Regr.}$). In contrast, the qualitative results (see Fig. 4) show that the use of the DTC results in a more distinct delineation of fluids.

Table 1: Averaged results of the two cross-validations based on annotations of the respective expert. The best results are shown in bold.

	structures → methods ↓	NFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
DSC	$STL_{Class.}$	0.861	0.896	0.782	0.740	0.872	0.859	0.842	**0.553**
	$STL_{Regr.}$	0.844	0.888	0.779	0.739	0.876	0.857	0.848	0.519
	MTL_{Xue}	0.856	0.896	0.784	0.745	0.877	0.859	0.845	0.520
	MTL	0.869	0.900	0.791	**0.758**	**0.882**	0.863	0.852	0.530
	MTL_{DTC}	**0.871**	**0.902**	**0.796**	0.746	0.877	**0.866**	**0.853**	0.522
	IRR	0.846	0.888	0.770	0.724	0.868	0.847	0.842	0.522
ASSD	$STL_{Class.}$	1.46	1.64	1.96	2.07	2.36	1.13	1.18	–
	$STL_{Regr.}$	2.70	1.80	2.27	2.16	2.63	1.16	1.16	–
	MTL_{Xue}	2.49	1.64	2.06	2.09	2.43	1.17	1.20	–
	MTL	1.58	1.56	2.12	1.99	2.52	1.10	**1.13**	–
	MTL_{DTC}	**1.40**	**1.53**	**1.93**	**1.91**	**2.32**	**1.06**	1.14	–
	IRR	1.70	1.76	1.92	1.87	2.17	1.17	1.19	–
HD	$STL_{Class.}$	13.80	12.13	15.34	15.83	19.04	5.97	5.86	–
	$STL_{Regr.}$	21.00	11.00	14.39	15.91	20.20	5.34	4.67	–
	MTL_{Xue}	17.98	9.71	13.74	15.25	19.23	6.29	5.40	–
	MTL	13.16	9.20	13.29	14.84	20.10	5.05	5.27	–
	MTL_{DTC}	**11.83**	**8.55**	**12.22**	**12.98**	**17.80**	**4.62**	**4.30**	–
	IRR	13.60	9.33	11.78	11.20	17.83	4.48	4.38	–

4. CONCLUSIONS

The evaluations show that the topology of the retinal layers is better preserved by SSDM regularization than by the sole classification ($STL_{Class.}$). However, using only regression ($STL_{Regr.}$) leads to a decrease in segmentation accuracy. The results indicate that this is often due to segmentation defects in the compressed upper retinal layers caused by large edema swellings. Also, finely defined fluid inclusions tend to be segmented as a diffusely

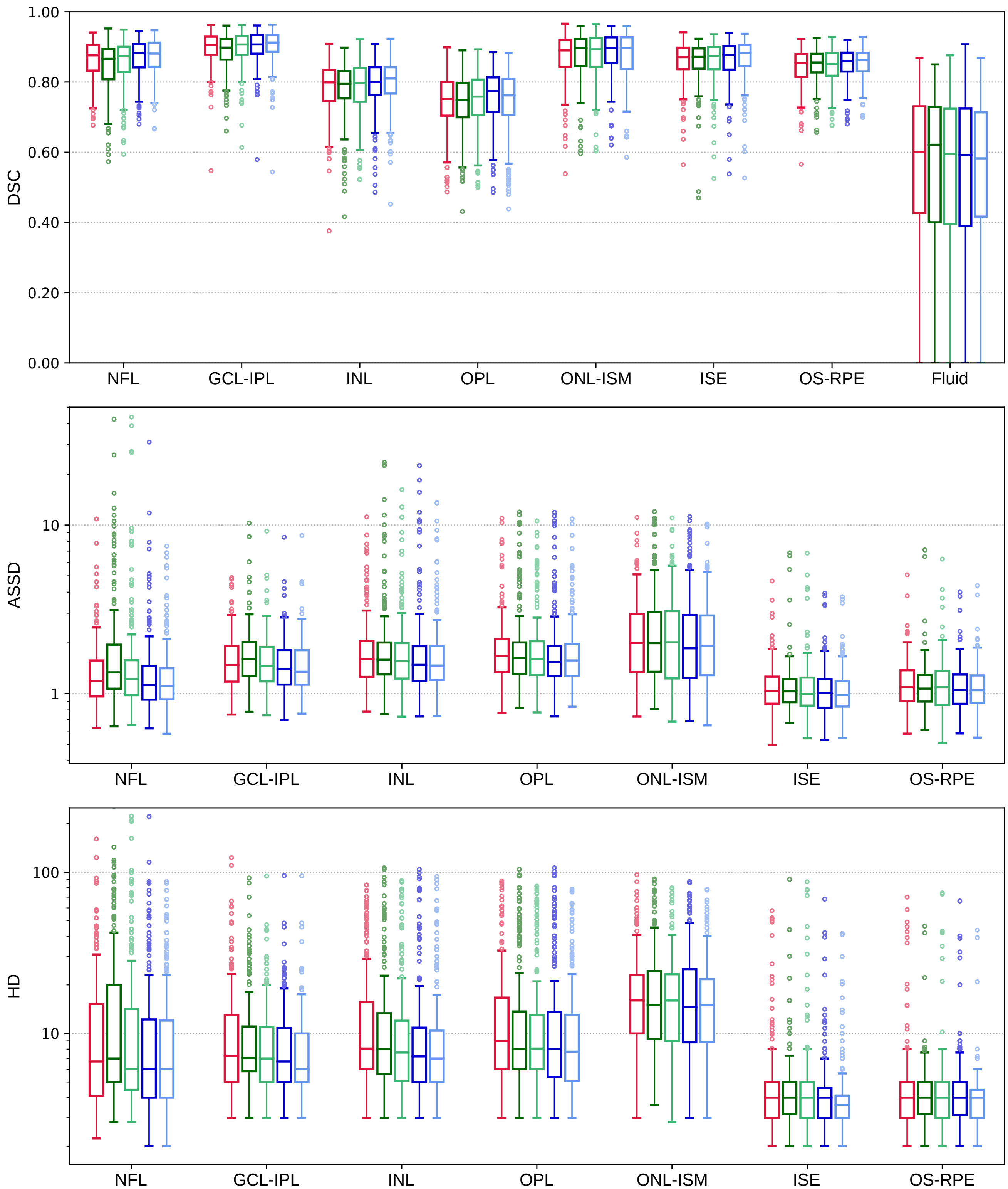


Figure 3: Boxplots of DSC, ASSD and HD values of both cross-validations. The color coding is as follows: $\mathrm{STL}_{\mathrm{Class.}}$ (red), $\mathrm{STL}_{\mathrm{Regr.}}$ (dark green), $\mathrm{MTL}_{\mathrm{Xue}}$ (green), $\mathrm{MTL}_{\mathrm{MH}}$ (blue), $\mathrm{MTL}_{\mathrm{MH}_{\mathrm{DTC}}}$ (light blue).

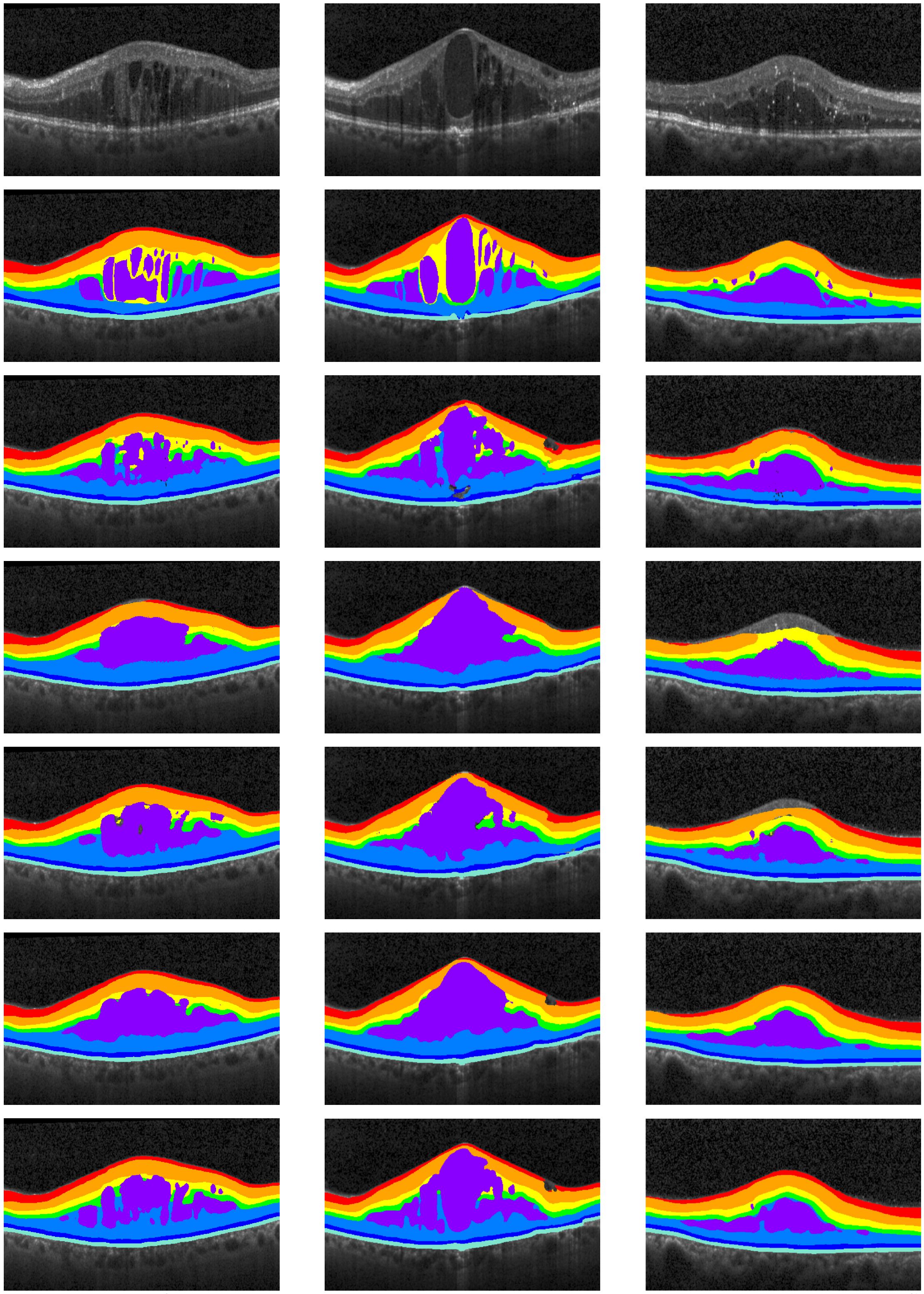


Figure 4: Visualization of three exemplary segmentations. From top to bottom: B-scan, ground truth, $\mathrm{STL}_{\mathrm{Class.}}$, $\mathrm{STL}_{\mathrm{Regr.}}$, $\mathrm{MTL}_{\mathrm{Xue}}$, $\mathrm{MTL}_{\mathrm{MH}}$, $\mathrm{MTL}_{\mathrm{MH}_{\mathrm{DTC}}}$.

connected area. A reasonable explanation for this is that fluids, unlike retinal layers, do not follow a specific shape and are expressed differently depending on their location.

Improved fluid delineation could be demonstrated by adding another classification task ($\mathrm{MTL}_{\mathrm{Xue}}$). The segmentation quality of the retinal layers could also be optimized by this method. It appears that the direct coupling to the SSDMs utilizing the approximated Heaviside function leads to similar problems, as the upper retinal layers are often incompletely segmented.

On average, the MTL methods achieve the best results in terms of retinal layer segmentation. A higher generalizability is achieved by learning through separate task paths, which prevents overfitting. At the same time, feature representations are learned that are preferred by both tasks. This focuses on relevant features, especially when there is only a small amount of available training image data that also shows large variations due to the DME disease. With the help of the DTC, consistency between both tasks can be established, and reciprocal support can be ensured. This coupling is not as strict as in the $\mathrm{MTL}_{\mathrm{Xue}}$ method, allowing flexibility in the individual tasks. The application of DTC demonstrates that the use of additional unlabeled image data can improve robustness and refine fluid delineation. Segmentation results of the $\mathrm{MTL}_{\mathrm{MH}_{\mathrm{DTC}}}$ method have the least amount of topology errors, which otherwise result in costly post-processing.

In general, fluid segmentation in OCT image data remains a challenging task. While problems such as strong class imbalance can be addressed by special error functions[23] or sophisticated sampling strategies, residual uncertainty remains in the delineation of individual fluid regions. Due to a weak image contrast in these regions, even manual identification is often ambiguous. This is also confirmed by the very different segmentations of the fluids of the two experts (see Fig. 1) resulting in a low IRR. Similar results can be found in the evaluations of the RETOUCH fluid segmentation task.[24] Therefore, the determination of segmentation uncertainties is a useful extension. For clinical application, uncertain segmentation regions can be identified and serve as complementary feedback for interpretation.[25] In conclusion, this work provides an important contribution to the challenging problem of retinal layer segmentation in the presence of fluid inclusions. Our proposed method allows for improved segmentation accuracy while effectively accounting for the topological properties of the retinal layers.

REFERENCES

[1] Bhagat, N., Grigorian, R. A., Tutela, A., and Zarbin, M. A., "Diabetic macular edema: pathogenesis and treatment," *Survey of Ophthalmology* **54**(1), 1–32 (2009).

[2] Spital, G., "Therapie des diabetischen makulaödems," *Der Diabetologe* **14**(8), 577–589 (2018).

[3] Zur, D., Iglicki, M., Busch, C., Invernizzi, A., Mariussi, M., Loewenstein, A., Cebeci, Z., Chhablani, J. K., Chaikitmongkol, V., Couturier, A., et al., "OCT biomarkers as functional outcome predictors in diabetic macular edema treated with dexamethasone implant," *Ophthalmology* **125**(2), 267–275 (2018).

[4] Li, K., Wu, X., Chen, D. Z., and Sonka, M., "Optimal surface segmentation in volumetric images-a graph-theoretic approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(1), 119–134 (2006).

[5] Garvin, M. K., Abràmoff, M. D., Wu, X., Russell, S. R., Burns, T. L., and Sonka, M., "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Transactions on Medical Imaging* **28**(9), 1436–1447 (2009).

[6] Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A., and Farsiu, S., "Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation," *Optics Express* **18**(18), 19413–19428 (2010).

[7] Roy, A. G., Conjeti, S., Karri, S. P. K., Sheet, D., Katouzian, A., Wachinger, C., and Navab, N., "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomedical Optics Express* **8**(8), 3627–3642 (2017).

[8] Wei, H. and Peng, P., "The segmentation of retinal layer and fluid in sd-oct images using mutex dice loss based fully convolutional networks," *IEEE Access* **8**, 60929–60939 (2020).

[9] He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Deep learning based topology guaranteed surface and mme segmentation of multiple sclerosis subjects from retinal oct," *Biomedical Optics Express* **10**(10), 5042–5058 (2019).

[10] He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Structured layer surface segmentation for retina oct using fully convolutional regression networks," *Medical Image Analysis* **68**, 101856 (2021).

[11] Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., Gaoxiang, C., Liu, J., Peng, C., Wang, L., et al., "How distance transform maps boost segmentation cnns: an empirical study," in [*Medical Imaging with Deep Learning (MIDL)*], 479–492, PMLR (2020).

[12] Heinrich, L., Funke, J., Pape, C., Nunez-Iglesias, J., and Saalfeld, S., "Synaptic cleft segmentation in non-isotropic volume electron microscopy of the complete drosophila brain," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*], 317–325, Springer (2018).

[13] Kepp, T., Ehrhardt, J., Heinrich, M. P., Hüttmann, G., and Handels, H., "Topology-preserving shape-based regression of retinal layers in oct image data using convolutional neural networks," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)* , 1437–1440 (2019).

[14] Cao, J., Liu, X., Zhang, Y., and Wang, M., "A multi-task framework for topology-guaranteed retinal layer segmentation in oct images," in [*2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*], 3142–3147, IEEE (2020).

[15] Liu, X., Cao, J., Wang, S., Zhang, Y., and Wang, M., "Confidence-guided topology-preserving layer segmentation for optical coherence tomography images with focus-column module," *IEEE Transactions on Instrumentation and Measurement* (2020).

[16] Luo, X., Chen, J., Song, T., Chen, Y., Wang, G., and Zhang, S., "Semi-supervised medical image segmentation through dual-task consistency," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **35**(10), 8801–8809 (2021).

[17] Xue, Y., Tang, H., Qiao, Z., Gong, G., Yin, Y., Qian, Z., Huang, C., Fan, W., and Huang, X., "Shape-aware organ segmentation by predicting signed distance maps," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **34**(07), 12565–12572 (2020).

[18] Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A., and Farsiu, S., "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomedical Optics Express* **6**(4), 1172–1194 (2015).

[19] Kepp, T., Droigk, C., Casper, M., Evers, M., Hüttmann, G., Salma, N., Manstein, D., Heinrich, M. P., and Handels, H., "Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks," *Biomedical Optics Express* **10**(7), 3484–3496 (2019).

[20] Maurer, C. R., Qi, R., and Raghavan, V., "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(2), 265–270 (2003).

[21] Ma, J., He, J., and Yang, X., "Learning geodesic active contours for embedding object global information in segmentation cnns," *IEEE Transactions on Medical Imaging* (2020).

[22] Simard, P. Y., Steinkraus, D., Platt, J. C., et al., "Best practices for convolutional neural networks applied to visual document analysis," in [*ICDAR*], **3**(2003), Citeseer (2003).

[23] Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., and Martel, A. L., "Loss odyssey in medical image segmentation," *Medical Image Analysis* , 102035 (2021).

[24] Bogunović, H., Venhuizen, F., Klimscha, S., Apostolopoulos, S., Bab-Hadiashar, A., Bagci, U., Beg, M. F., Bekalo, L., Chen, Q., Ciller, C., et al., "Retouch: The retinal oct fluid detection and segmentation benchmark and challenge," *IEEE Transactions on Medical Imaging* **38**(8), 1858–1874 (2019).

[25] Kepp, T., Andresen, J., Sudkamp, H., von der Burchard, C., Roider, J., Hüttmann, G., Ehrhardt, J., and Handels, H., "Epistemic and Aleatoric Uncertainty Estimation for PED Segmentation in Home OCT Images," in [*Bildverarbeitung für die Medizin 2022*], Springer Vieweg (2022).