

Towards evaluating FLIM attention regions

Matheus A. Cerqueira*, Bárbara C. Benato†, Alexandru C. Telea‡, and Alexandre X. Falcão*

*Institute of Computing, University of Campinas, Campinas, São Paulo, Brazil

†Institute of Mathematics and Computer Science, University of São Paulo (ICMC-USP), São Carlos, Brazil

‡Department of Information and Computing Sciences, University of Utrecht, Utrecht, The Netherlands

Abstract—FLIM (Feature Learning from Image Markers) has been used to interactively train convolutional networks from user-defined attention regions. Although FLIM has shown competitive results with large deep models, resulting in shallow networks trained with very few weakly annotated images, the performance of FLIM models is related to those attention regions. Much effort has been made to improve the FLIM framework itself, but no investigation has been conducted into these marked locations and their relationship to model performance. In this work, we propose to open ways to evaluate the image markers' location impacts on FLIM. For that, we exploit multiple marker positions, determine their relevance, and create a color code for each location, resulting in a heatmap of marker locations that reflects the FLIM performance from those image regions. From our results, we showed how the FLIM performance differs depending on the background marker location, which provides some trends for better/worse marker scenarios.

I. INTRODUCTION

Even with many advances in computer vision from Convolutional Neural Networks (CNNs) [1] to Visual Transformers (ViT) [2], such techniques still have many challenges. Both models rely on dozens of layers in their architectures, thousands hyper-parameters for tuning [3], and large image annotated datasets [4], [5]. Due to these, the cost for training such models is high, demanding more powerful hardware, and asking for much human effort in data annotation.

Different strategies have been explored to mitigate such problems, as transfer learning, data annotation, and other alternatives to training. Between them, *Feature Learning from Image Markers* (FLIM) have been proposed to interactively extract convolutional filters from *attention regions* of only a few images (1 or 2 per class), and few layers (from 1 to 3) without backpropagation. User-drawn markers define the attention regions. This framework has been explored for image classification [6] and segmentation [7] tasks, resulting in smaller and competitive networks [7], [8]. FLIM considers the user as a key element in the designing of neural networks, *i.e.*, by empowering the user to control from where the model should learn its weights [9], [10].

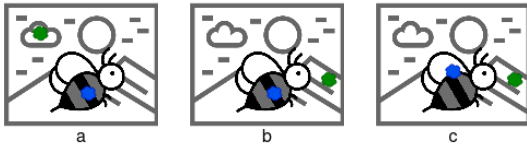


Fig. 1. Schema of different markers position: (a) a marker in the bee stripes and one marker in the cloud, (b) other background marker in the mountains, and (c) other object marker at the edge of stripes and wings.

However, FLIM still has some unaddressed points. Its success directly relies on where the network attention is focused, and because of that, on where the markers are drawn by the user. Figure 1 shows an example of how markers can be placed in different positions in an image. These regions are used to extract information for convolutional filters. Due to this, colors and textures of these regions are used as a step of FLIM. Markers are a core element when designing FLIM networks.

Recent studies have investigated different aspects of such networks. First, some of them have studied how to choose images to add markers can be selected [6], [11]–[13]. Other have proposed different approaches for choosing the convolutional filters [14], [15]. Marker imbalance [15] and marker normalization [12] are also addressed. Different users are evaluated in [6], but their markers are not compared. However, none of these studies have investigated the impact of different markers *positions* in FLIM networks. We are mainly concerned to the segmentation task. The study of classification purposes is left as future work.

In this work, we propose to open ways to evaluate image marking positions when using FLIM networks. Different questions related to image marking are raised when designing such networks as follows.

- *RQ1*: Where the user should draw markers?
- *RQ2*: How marker position and segmentation performance are correlated?

We address questions *RQ1* and *RQ2* in this work. To answer them, we propose a strategy to evaluate distinct marker positions when using FLIM networks. First, we evaluate FLIM performance by computing the DICE score related to each marker. We also evaluate the resulting FLIM feature spaces. We proposed a random over-sampling strategy with five and thirty object and background markers, respectively. From that, we create a heatmap with segmentation performance color-coded in the marker position in the original image. With that, we can visualize both performance and marker position variation over each marked image. To investigate that, we consider three image datasets of different applications.

II. RELATED WORK

Feature Learning from Image Markers (FLIM) exploits the user knowledge to learn convolutional filters by learning those filters directly from user markers with a clustering algorithm [6]–[8]. The FLIM learning process is direct and does not require a backpropagation algorithm. It has been

employed in multiple 2d natural image classification tasks [6], [7] and for detecting COVID-19 from chest CT images [11]. For segmentation, FLIM was used to assist a graph-based segmentation method [8], to expand the segmentation in aerial images [14], and for medical image tasks [12], [13], [15].

The works also differ in terms of their methodology. For the image selection, some used the inspection of the training set, either visually [11]–[13], using a 2d projection tool [7] or interactively [16]. As for the annotation of the markers, most works used unlabeled markers. Regarding the filter estimation process from the markers, most studies employed clustering, and one study combined clustering with PCA [14].

FLIM’s previous work shows that it is possible to use a reduced number of weakly labeled images to learn a shallow feature extractor (1-3 layers) with a descriptive procedure while maintaining its performance compared to standard deep learning models. Even though FLIM offers competitive and compact networks, its performance is related to a series of user actions. Although some studies have used multiple users in their work, showing differences in performance between them [6], [13], and one study has shown the impact of marker size on normalization [12], further research is needed to explore how marker location affects FLIM’s performance and to compare those different markers.

III. METHOD

In this work, we investigate the problem of how markers in different positions impact filter learning with FLIM. Therefore, we focus our analysis on the scenario where there are only two markers, one for the object and one for the background, allowing the analysis to be more *controlled*. Therefore, by fixing the object marker and varying the background marker, we can measure the impact of each background position.

Figure 2 shows a schema of this evaluation. From an Image, we (1) sample the marker positions for the background (three in this schema), (2) and by varying the background marker, (3) we learn different FLIM models for each combination of background/object markers. (4) Each of these models will have a performance which could be (5) exploited visually in the original with a color range associated with the model’s performance. Therefore, our method can be divided into a few parts: (I) marker over-sampling, and (II) estimation of the importance of each marker. By running this method over multiple images and datasets, we can analyze the heatmaps to gain insights into marker locations for FLIM.

A. Marker over-sampling

To evaluate the position of markers for FLIM, we employ a random sampling strategy for sampling marker centers given an image I , its ground-truth image L , a desired number of background markers n_{bg} , and a marker size r . We chose to use random sampling because our goal is to analyze diversity marker positions in images without imposing any restrictions beyond background/object. This approach avoids the limitations of deterministic sampling methods [12], which can reduce the diversity of these markers.

1) *Sampling markers centers*: Our marker sampling is done by initially sampling a unique pixel that will be the center of an object marker, then over-sampling n_{bg} centers of background pixels. For this, we use a safety margin of 1.5 times the marker size r . Therefore, marker centers that are close to other centers or that are close to the image edges are rejected, and a new center selection is made respecting this safety margin (using $L2$ distance). This approach ensures diversity among marker centers without excluding markers that are close to each other.

2) *Drawing markers size*: Given a marker center $(x_c, y_c) \in \mathbb{R}^2$ and marker size r , we enlarge the center to its adjacency. Therefore, the marker set M of one image is composed of the centers (x_c, y_c) and their adjacency. For a given a marked pixel p_m with coordinates (x_c, y_c) , we use an adjacency relation A between two given pixels (p, q) , with $A : (p, q) || |p - q| \leq r$, therefore $A(p_m)$ is the adjacency of the marked pixel p_m according to the marker size r .

It is worth mentioning that in this work, there is no restriction on whether a pixel is on an edge of the object, that is, if there is a pixel q in the adjacency $A(p_m)$ from another label ($label(p_m) \neq label(q)$). This is because we are interested in evaluate markers on the edge of the objects, whether internal or external to the objects of interest.

B. Estimating markers importance

Once we obtained the set of background markers \mathcal{M}_{bg} and the set of object markers, composed of one marker per image, $\mathcal{M}_o = \{M_o\}$, we train a FLIM model f_i for each combination of object/background markers $\{M_o, M_{bg}^{(i)}\}$, $M_{bg}^{(i)} \in \mathcal{M}_{bg}$, then we evaluate this model on the validation set obtaining a metric $\alpha^{(i)}$. We then used this metric as the relevance of marker $M_{bg}^{(i)}$ and repeated the process for the remaining markers of \mathcal{M}_{bg} .

In order to get a visual feedback from the marker, we create a heatmap for each $\alpha^{(i)}$ related to the background marker $M_{bg}^{(i)}$. Instead of only setting every pixel $p = (x, y) \in M_{bg}^{(i)}$ with the value of $\alpha^{(i)}$, we used a function to have a spatial decay of $\alpha^{(i)}$, considering the marker’s adjacency relationship A . Therefore, for a heatmap image I_h , given a pixel q in the adjacency of the center of marker p , the value of $I(q)$ will follow $I(q) = 2 - ||p - q||^2 / r^2$. Finally, to a better representation we apply a colormap on $I(q)$ to get a colored image ranging from blue to reddish.

IV. EXPERIMENTAL SETUP

A. Datasets

We employ three segmentation datasets in this work:

- **Fish**: A subset of Fish segmentation dataset [17], composing 181 images of *Pomacentrus moluccensis*. The dataset was collected from live video, showing the animals in different positions and with differences in background.
- **Schisto**: An in-house dataset from microscopy imaging of parasite *Schistosoma mansoni* eggs. The dataset was manually labeled by clinical experts and we used only the images with eggs, resulting in 632 labeled images.
- **Refuge**: The refuge dataset, composed of 1200 annotated retinal fundus images [18]. The dataset consist of two

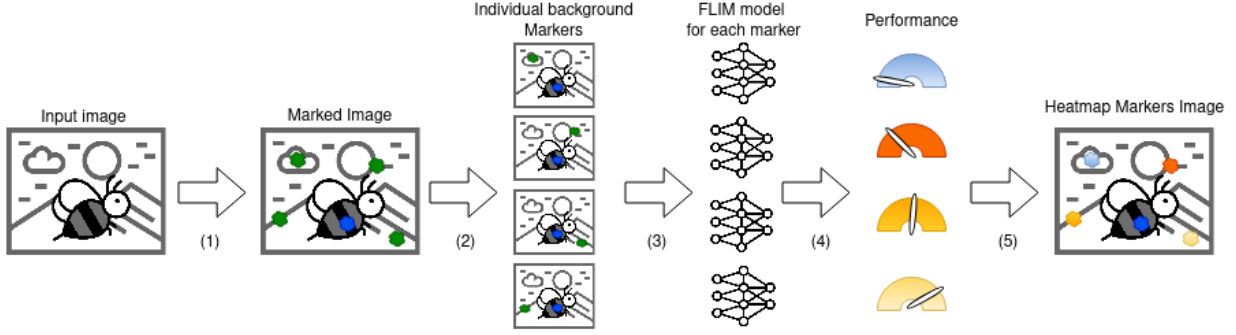


Fig. 2. Our evaluation method: Sampling markers position for image, obtain a FLIM model for each pair of object/background marker and evaluate the marker (in this case background) according to the models performance.

labels from, the optical disc/cup segmentation. In this work, we are focused in segmenting the optical disc, which is the outer part of the object.

For all datasets, we split the samples into a 50% holdout for testing. From the remaining 50%, we remove three images for marking, which we denote as ‘marked images’, and use the remaining images as a validation set.

B. Model

We employ a simple model f for evaluation, a 1-layer FLIM convolutional block followed by a decision layer (Figure 3). We trained the FLIM block with the marked image and using the pair of object/background markers, M_o and $M_{bg}^{(i)}$. For all datasets we used a convolutional FLIM block composed of 3×3 kernels and with output of $C = 16$ channels, followed by ReLU and an average pooling with 3×3 and $stride = 2$.

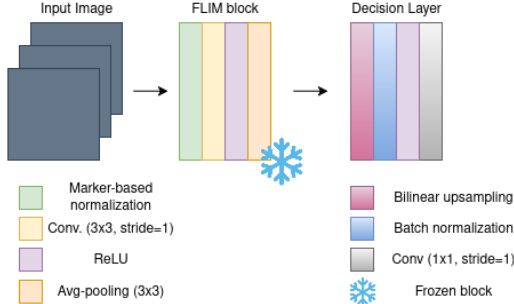


Fig. 3. Our evaluated model: A FLIM block followed by a decision layer.

We then interpolate and use a decision layer. As for the decision layer, we trained a 1×1 layer with batch normalization using only the marked image for tuning the decision layer weights. It is worth mentioning that we freeze the convolutional FLIM block, training only the decision layer.

We initialize the weights of the decision layer using Xavier. The weights were tuned using Adam optimizer with $lr = 0.1$ for 20 epochs, and using a mean of Cross entropy and Dice loss. We took the best weights for decision layer according to the validation set. The goal here is to evaluate the capacity of segmentation from a FLIM block trained with the pair M_o and $M_{bg}^{(i)}$. Also, we employ a single block (or layer) trained with FLIM because it is not very clear if the same markers position should be used for every layer (block) of FLIM.

C. Markers Parameters

We have two marker parameters, their size r and the number of background markers n_{bg} . For the Schisto dataset we used $r = 5$ and $n_{bg} = 30$, for the Fish we used $r = 3$ and $n_{bg} = 10$, and for Refuge we used $r = 4$ and $n_{bg} = 30$. Both parameters were adjusted based on the image’s spatial size and the size of objects within it. For example, Refuge and Schisto used images were 400×400 and had objects around 2 – 3% of that size. For Fish, we used images of 128×128 .

V. RESULTS

A. Correlation of marker position and segmentation performance

In order to evaluate the impact of different maker positions in the FLIM performance, we evaluate the feature space produced by FLIM and its segmentation performance using the DICE similarity.

Figure 4 presents (a) object (blue) and background (red) markers in the original image, (b) the 2d projection of image patches extracted from marker positions of (a), (c) the 2d projection feature space produced by FLIM over validation set, and DICE values from the validation set. Three different results are presented: a high, a medium, and a low DICE value. For computing the 2d projections, we use t-SNE [19] algorithm with perplexity of 5 for projection of (b) and 15 to (c), and other parameters as default. Worth to mention that projecting the all pixels of validation set would be unfeasible. Therefore, in (c) we took the center of image superpixels projected on their feature maps and project those *centers of features* using t-SNE. We used DISF [20] as superpixel method, with number of superpixels of 10, 25, 30 for *Fish*, *Schisto*, and *Refuge* dataset.

There are several aspects to analyze from this experiment. First, for Fish, it is possible to see how different marker positions impact DICE values: image 1 shows background marker in a texture region with the highest DICE value, while image 3 shows background marker in a region without any texture with the lowest DICE value. In image 2, even though the background marker is in a texture region, is a region with texture and color very similar to the object marker. Another aspect is that the defined markers (a), the 2d feature space produced by FLIM (c), and DICE values

(d) are **correlated**, thereby answering RQ2. Whenever there is a clear visual separation between all extracted centers of superpixels of object (orange) and background (blue) in the projection, the higher the DICE performance in segmentation (see image 1). The opposite is also seen: the little the visual separation between center of superpixels of object (orange) and background (blue) in the projection, the lower the DICE (see image 3). Additionally, one can notice that there is a clear separation of points of marked patches in (b), from object (orange points) and background (blue points) markers in all images, and for different performances of DICE, *i.e.*, there is no correlation between them.

Another important aspect is how the FLIM performance for segmentation can vary significantly depending on the marker. For example, for *Fish* dataset values of DICE are $[0.315, 0.846]$, for *Schisto* dataset values are $[0.331, 0.623]$, and for *Refuge* dataset values are $[0.000, 0.679]$.

B. Over-sampling of markers

Given that we have shown the correlation between marker positions and segmentation performance, we now intend to explore and compare *many different* marker positions for object and background, thereby answering RQ1. For that, we create a heatmap relating the different marker positions to their relevance. The performance of each background marker is computed individually and then placed in the same heatmap image for comparison. For example, in Figure 5, Position illustrates the distribution of random background markers across images. Also, we show their corresponding Heatmap. For Heatmap, a colormap from blue to red defines low to high DICE values, respectively. A single object marker is colored as solid red on the object. The diameter of each circle is the same size as the marker. In the following results, we present only the Heatmap due to page limit constraints.

Given that, we intend to evaluate different positions of background markers for different images. We present this evaluation in Figures 6, 7, and 8. For that, we randomly sample background markers for three different images (rows) and five splits of objects markers (columns). We present correspondent heatmaps and minimum and maximum DICE values for all marker combinations per image.

First, for the *Fish* dataset, Figure 6 shows a trend in which background markers are red (higher DICE values) at the corals and blue/cyan (lower DICE values) in other areas of the water. This trend is observed for all three images. Given that there is more texture and color patterns in the corals, FLIM can extract relevant information for segmentation from these regions. Another important observation is related to the lack of texture and high pixel intensity in corals. Notice how markers on plain white areas of the corals tend to lower metrics, shown as blue/cyan discs at the heatmaps, for example, in image 2.

Heatmaps for the *Schisto* dataset are presented in Figure 7. The same trend of the last dataset is noticed: background markers are reddish (higher DICE values) at textured regions compared to plain areas, which present more blue/cyan (lower DICE values). Images with more texture in the background,

such as images 1 and 3, show more cyan circles than blue as lower values as in image 2. Another interesting aspect is related to markers in the border of (i) the object and (ii) background markers. For (i), one can note that whenever there is an object marker on the border of the object –for example, in the first column of image 3– the higher the values of the background markers. For (ii), in the first column of image 1, background makers at the border of background textures (impurities) show better results when the object marker is in the middle of the object. Additionally, for image 1, in the second column, the marker at the border of the parasite yielded better overall metrics (0.523 vs 0.581, the maximum value).

Figure 8 presents results for *Refuge* dataset. It is possible to observe a trend where reddish background markers (higher DICE values) are present whenever an object marker is located on the border of the outer object (optic disc) rather than in the central region of the object (optic cup, the brighter area). For example, see the second column of image 1, and the first column of image 2: there is an object marker in the middle of the object, and all other background markers are blue/cyan (lower DICE values). Another interesting point is related to background marker positions. Background markers with higher DICE values are located near retinal vessels –especially regions with high contrast from vessels to background– and lower ones are in regions with little texture.

VI. CONCLUSION

In this paper, we propose to investigate the impact of different positions of markers on the performance of FLIM networks for segmentation. We proposed a strategy to oversample background markers, evaluate each marker individually, and display their performance in a single image as a heatmap, where marker positions and colors indicate the quality of FLIM performance when sampling this location as a marker. We evaluated our experiments in three different datasets, with five and thirty random samples for object and background markers, respectively. We also evaluated the feature space resulting from FLIM and its performance when using different marker positions.

Several key findings are as follows. First, we show how FLIM performance values may differ depending on the marker position. This indicates how FLIM performance is susceptible to the localization of markers in the original image. Second, both the position of the object and background makers also affect FLIM performance. Object markers on the border of the object can assist FLIM in capturing differences from the background. Background markers in the border of textures in the background can also improve the performance. Object and background markers within regions of similar texture/color can not guarantee a good performance. Third, using the projection of marked patches from object and background markers is not discriminative for choosing suitable markers for higher performance values.

Given these findings, we expect that the proposed method can be used as part of an FLIM experimental setup whenever a new dataset is used and finding suitable markers is needed. As

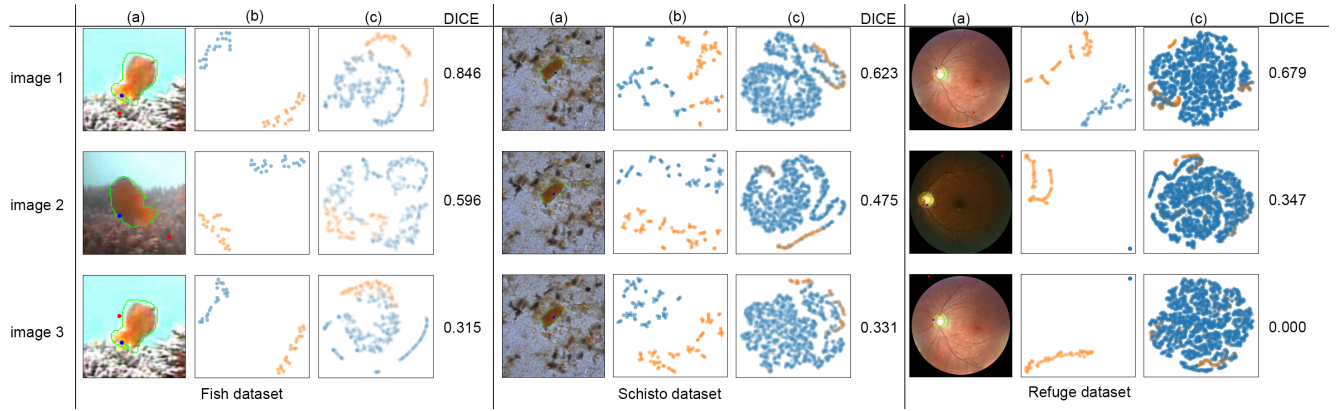


Fig. 4. Different markers for all datasets: *Fish*, *Schisto*, and *Refuge*. Each row is a different image/marker position. Column shows (a) the markers, (b) marked patches projection, (c) feature maps projection of superpixel centers, and DICE values.

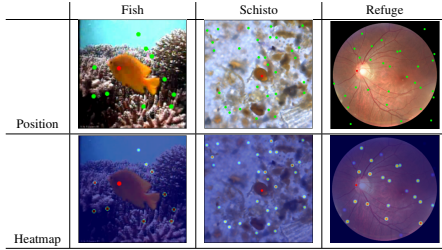


Fig. 5. Examples of marker and marker heatmap for all datasets. For Position, object markers are red and background markers are green. For Heatmap, a colormap from blue to red defines low to high DICE values, respectively. A single object marker is colored as solid red on the object.

future work, we plan to use our insights to recommend marker positions, investigate the combination of markers, evaluate additional datasets, and assess the classification task. We also intend to address other questions, for example, *how many markers/images to draw* and *if there are differences between object/background markers*, while comparing our results with other attention methods.

ACKNOWLEDGMENT

The authors thank FAPESP (2023/14427-8, and 2023/09210-0) and CNPq (304711/2023-3) for financial support.

REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [2] L. Yuan, Y. Chen, T. Wang, W. Yu, Y. Shi, Z.-H. Jiang, F. E. Tay, J. Feng, and S. Yan, "Tokens-to-token vit: Training vision transformers from scratch on imagenet," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 558–567.
- [3] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," *Advances in neural information processing systems*, vol. 27, 2014.
- [6] B. C. Benato, I. E. de Souza, F. L. Galvão, and A. X. Falcão, "Convolutional neural networks from image markers," *arXiv preprint arXiv:2012.12108*, 2020.
- [7] I. E. De Souza and A. X. Falcão, "Learning cnn filters from user-drawn image markers for coconut-tree image classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [8] I. E. de Souza, B. C. Benato, and A. X. Falcão, "Feature learning from image markers for object delineation," in *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2020, pp. 116–123.
- [9] J. A. Fails and D. R. Olsen Jr, "Interactive machine learning," in *Proceedings of the 8th international conference on Intelligent user interfaces*, 2003, pp. 39–45.
- [10] G. Ramos, C. Meek, P. Simard, J. Suh, and S. Ghorashi, "Interactive machine teaching: a human-centered approach to building machine-learned models," *Human-Computer Interaction*, vol. 35, no. 5-6, pp. 413–451, 2020.
- [11] A. M. Sousa, F. Reis, R. Zerbini, J. L. Comba, and A. X. Falcão, "Cnn filter learning from drawn markers for the detection of suggestive signs of covid-19 in ct images," in *EMBC*. IEEE, 2021, pp. 3169–3172.
- [12] L. Joao, M. Cerqueira, B. Benato, and A. Falcao, "Understanding marker-based normalization for flim networks," in *Proc. of VISIGRAPI - Volume 2: VISAPP*. INSTICC, 2024, pp. 612–623.
- [13] G. J. Soares, M. A. Cerqueira, S. J. F. Guimaraes, J. F. Gomes, and A. X. Falcão, "Adaptive decoders for flim-based salient object detection networks," in *Proc. of SIBGRAPI*, 2024, pp. 1–6.
- [14] I. E. de Souza, C. L. Cazarin, M. R. Veronez, L. Gonzaga, and A. X. Falcão, "User-guided data expansion modeling to train deep neural networks with little supervision," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [15] M. A. Cerqueira, F. Sprenger, B. C. Teixeira, and A. X. Falcão, "Building brain tumor segmentation networks with user-assisted filter estimation and selection," in *18th International Symposium on Medical Information Processing and Analysis*, vol. 12567. SPIE, 2023, pp. 202–211.
- [16] M. A. Cerqueira, F. Sprenger, B. C. A. Teixeira, S. J. F. Guimarães, and A. X. Falcão, "Interactive ground-truth-free image selection for flim segmentation encoders," in *2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2024, pp. 1–6.
- [17] B. J. Boom, P. X. Huang, C. Beyan, C. Spampinato, S. Palazzo, J. He, E. Beauxis-Aussalet, S.-I. Lin, H.-M. Chou, G. Nadarajan *et al.*, "Long-term underwater camera surveillance for monitoring and analysis of fish populations," in *International Workshop on VAIB, at ICPR*, 2012, pp. 1–4.
- [18] J. I. Orlando, H. Fu, J. B. Breda, K. Van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee *et al.*, "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical image analysis*, vol. 59, p. 101570, 2020.
- [19] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [20] F. C. Belém, S. J. F. Guimaraes, and A. X. Falcão, "Superpixel segmentation using dynamic and iterative spanning forest," *IEEE Signal Processing Letters*, vol. 27, pp. 1440–1444, 2020.


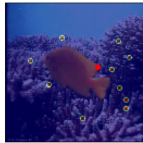
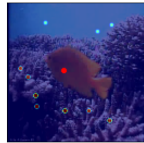
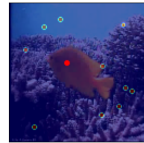
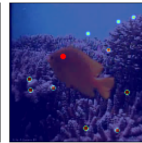

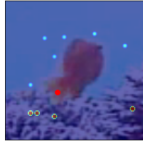
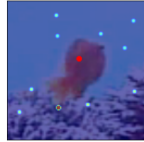
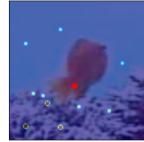
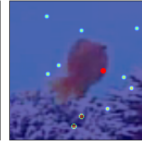





Image	Heatmap					min/max DICE
1						[0.369, 0.831]
2						[0.315, 0.846]
3						[0.396, 0.829]

Fig. 6. Heatmap for *Fish* dataset. Different images 1, 2, and 3 with different random samplings of object and background markers per image are presented per line. In each Heatmap, circles define marker position, and colors define low (blue/cyan) to high (red with green border) DICE values. A solid red circle on the fish is the only object marker. High DICE values relate to better segmentation performance. Minimum and maximum DICE values are presented per image.

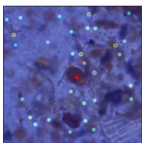
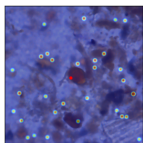
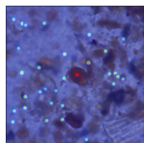
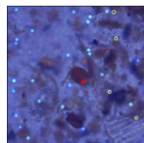
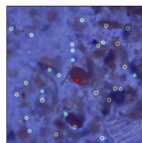
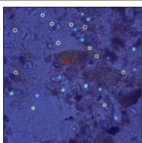
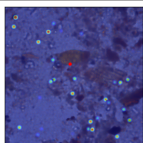
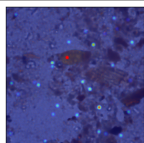
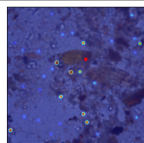
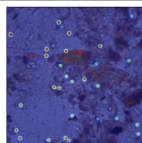
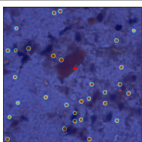
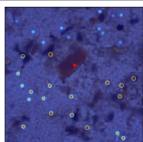
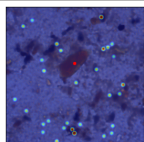
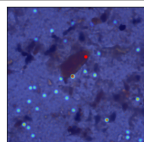
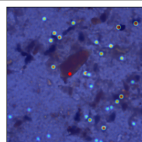
Image	Heatmap					min/max DICE
1						[0.186, 0.581]
2						[0.171, 0.613]
3						[0.305, 0.623]

Fig. 7. Heatmap for *Schisto* dataset. Different images 1, 2, and 3 with different random samplings of object and background markers per image are presented per line. In each Heatmap, circles define marker position on each image, and colors define low (blue/cyan) to high (red with green contour) DICE values. A solid red circle on the parasite is the only object marker. High DICE values relate to better segmentation performance. Minimum and maximum DICE values are presented per image.

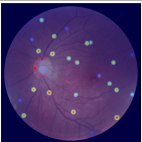
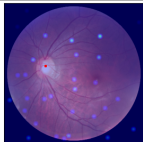
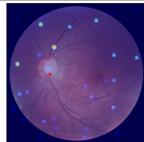
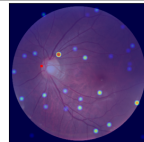
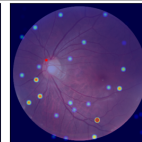
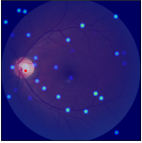
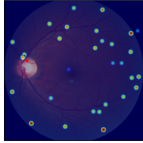
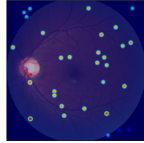
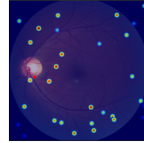
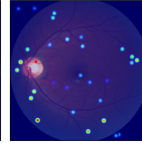
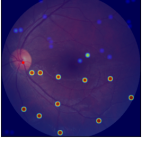
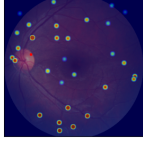
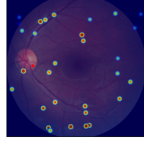
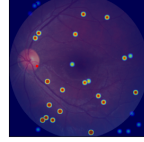
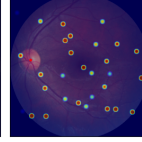
Image	Heatmap					min/max DICE
1						[0.000, 0.679]
2						[0.077, 0.543]
3						[0.131, 0.676]

Fig. 8. Heatmap for *Retinal* dataset. Different images 1, 2, and 3 with different random samplings of object and background markers per image are presented per line. In each Heatmap, circles define marker position, and colors define low (blue/cyan) to high (red with green border) DICE values. A solid red circle on the brighter region is the only object marker. High DICE values relate to better segmentation performance. Minimum and maximum DICE values are presented per image.