Friedrich Rieken Münke*, Luca Rettenberger, Anna Popova, and Markus Reischl

A Lightweight Framework for Semantic Segmentation of Biomedical Images

https://doi.org/10.1515/cdbme-2023-1048

Abstract: We introduce a lightweight framework for semantic segmentation that utilizes structured classifiers as an alternative to deep learning methods. Biomedical data is known for being scarce and difficult to label. However, this framework provides a lightweight, easy-to-apply, and fast-to-train approach that can be adapted to changes in image material though efficient retraining. Moreover, the framework is able to adapt to various input sizes making it robust against changes in resolution and is not tied to specialized hardware, which allows efficient application on standard laptops or desktops without GPUs. We benchmark two distinct models, a single structured classifier and an ensemble of structured classifiers, against a U-Net, evaluating overall performance and training speed. The framework is versatile and can be applied to multi-class semantic segmentation. Our study shows that the proposed framework can effectively compete with established deep learning methods on diverse datasets in terms of performance while reducing training time immensely.

Keywords: Semantic Segmentation, Structured Classifier, Structured Random Forest, Deep Learning, Machine Learning, Benchmark

1 Introduction

Computer vision has become increasingly important in a wide range of fields, from self-driving cars to medical imaging. One of the most fundamental tasks in computer vision is semantic segmentation, which involves labeling each pixel in an image to identify different objects and regions of interest. In particular, biomedical image segmentation is widely used to analyze experimental results or to support the diagnosis of diseases. While deep learning approaches, such as Convolutional Neural Networks (CNNs), show state-of-the-art performance on a variety of datasets, they are complex and computationally expensive [9]. These methods often require large amounts of annotated data to train, which can be a significant hurdle in biomedical applications where labeling is difficult and time-consuming [12]. In addition, biomedical data can be highly variable, with differences in color, resolution, annotator noise and other modalities between datasets. This variability makes it challenging for CNNs to generalize across different datasets, which may require retraining the model for each new dataset. To address these challenges, we propose a lightweight and thus fast to train framework for semantic segmentation based on structured random forests.

Structured random forests are a technique used in computer vision for assigning labels to each pixel in an image to identify regions of interest [3]. This approach is similar to a CNN, where a kernel defines the neighborhood for each pixel, and is used to select features per pixel for the final classification. Structured random forests have been used for various image segmentation tasks, including edge detection [3], brain tumor segmentation [15], and road crack segmentation [13]. It is possible to further increase performance by combining multiple structured classifiers [1].

Our objective is to use this method to enable biomedical researchers to analyze and interpret their image data, enabling faster and efficient evaluations of experiments and supporting diagnoses of diseases.

2 Method

The proposed framework integrates structured segmentation algorithms as modular components to enable the creation of efficient and powerful multi-class semantic segmentation models. We showcase the capabilities of the framework by presenting two models designed to compete with a U-Net in terms of performance while maintaining fast training times.

The first model (SC: Structured Classifier Model) consists of one structured classifier. The classifier processes the image at its original resolution classifying each pixel based on is surrounding pixels. This classifier comprises two components: a kernel and a regular classifier. The kernel determines the pixels that are taken into consideration for classification, while the regular classifier makes the final decision. The kernel gathers the pixel values surrounding a given pixel and organizes them

***Corresponding author: Friedrich Rieken Münke,** Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany, e-mail: friedrich.muenke@kit.edu
Luca Rettenberger, Markus Reischl, Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany
Anna Popova, Institute of Biological and Chemical Systems - Functional and Molecular Systems, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany

into a single feature vector. The position of each value in the feature vector corresponds to its relative position to the pixel. In our case, the kernel considers all pixels within a radius of 5 pixels. This model has a limited capacity since it is not able to learn features at a large scale. The second structured classifier model (ED: Encoder-Decoder Model) scales the previous model to enable the model to process larger context in an image. Instead of using a single classifier for segmentation, the model utilizes six structured classifiers as an ensemble, which are stacked on top of each other. The concept is depicted in Fig. 1.

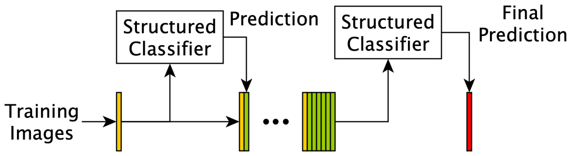


Fig. 1: The ED model is an ensemble of six classifiers, where each classifier provides it predictions for the next classifiers as features. Each classifier has its defined image scale and is only trained on one third of all images.

The predictions from all previous classifiers serve as input for the next one. The classifiers learn features of different scales by downscaling the image by a factor of two for the first three classifiers and then increasing its size to the original resolution for the last three classifiers. Each individual classifier is trained on one-third of the dataset, which is randomly sampled, to reduce training time and prevent overfitting. Both proposed models are able to process input images of any size during training and inference and support colored and gray-scale images. In our case, both models process gray-scale images by default to further boost their speed. We use an extra tree classifier [5] to classify the pixels for all models.

The models are developed to use the same data format as deep learning-based methods, ensuring compatibility with various image labeling methods and frameworks specifically designed for deep learning applications. Moreover, these models can be easily adjusted to process either gray-scale or color images.[1]

Our benchmark model is a U-Net [10] with a ResNet18 [6] backbone and the encoder pretrained on the Imagenet dataset [11]. To train the U-Net, we used common augmentation techniques such as rotation, flipping, cropping, blurring, salt-and-pepper noise, and brightening. Our proposed models are not using augmentations techniques to simplify their training procedure. We split the dataset into training and validation sets using a 90-10 split, and we selected the weights that performed best on the validation dataset as our final model. If there was no improvement after 100 epochs, we stopped the training. The U-Net was trained with an ADAM optimizer, a learning rate of 10^{-4} and the binary-cross-entropy loss. The U-Net processes colored images, due to its pretrained encoder. We score the models by training time in minutes and the segmentation performance in F1-Score (Dice Coefficient). Each model was trained three times to validate against effects of randomness and the average result is used as a representative value. The training time was measured on a MacBook Pro (2020) with an 2,3 GHz Quad-Core Intel Core i7 CPU and 16 GB Ram.

3 Datasets

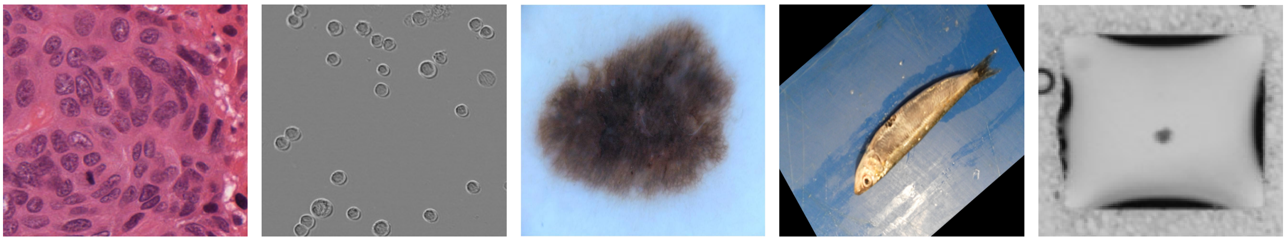
In Tab. 1, we introduce biomedical datasets as general benchmark for evaluating semantic segmentation algorithms. In Fig. 2 we present exemplary images and discuss the diverse challenges. Each dataset is prepared to present a binary segmentation task to the models separating between foreground and background.

Dataset	Train	Test	Domain
CryoNuSeg [7]	148	30	Nuclei
LIVECell [4]	2.640	176	Cell
Derma ISIC [2]	2.000	600	Skin Lesion
Seafood [14]	8.550	450	Seafood Types
Sphere [8]	201	38	Tumor Spheroids

Tab. 1: An overview over all benchmark datasets used to showcase the functionality of the proposed framework and the distribution of training and test images.

The CryoNuSeg [7] dataset is the first annotated dataset of Hematoxylin and Eosin (H&E)-stained images for nuclei segmentation from 10 different organs. Its diverse textures make it a challenging dataset that requires adaptation to the appearance of different organs. The LIVECell [4] dataset is a high-quality dataset of phase-contrast images for cell segmentation. Due to time constraints, we use a subset of the original dataset for training, where the training dataset consists of fixed-size crops (256px x 256px) while the test set has a different resolution (704px x 520px). Structural difficulty is common in other datasets and underscores the importance of input size flexibility. The Derma ISIC [2] dataset focuses on skin lesion analysis and melanoma detection, with nine diagnostic categories and varying image sizes in the training and test datasets. This

[1] For a more detailed explanation and the implementation, refer to https://github.com/FMuenke/structured_segmentation



(a) CryoNuSeg [7] **(b)** LIVECell [4] **(c)** Derma ISIC [2] **(d)** Seafood [14] **(e)** Sphere [8]

Fig. 2: Five examples of images from each benchmark dataset for semantic segmentation, showcasing the variability in object size and area, textures, and shape features present in the datasets.

dataset poses a difficult challenge due to its varying appearance and large differences in the melanoma scale in the image. The Seafood [14] dataset comprises nine seafood types, and its images are heavily augmented with rotation, resulting in the largest dataset. This dataset is distinct from others due to its unique combination of color and texture information. Finally, the Sphere [8] dataset consists of tumor spheroids captured in a high-throughput Droplet Microarray (DMA). This datasets poses a unique challenge since there is only one segmentation target and it often overlaps with the borders of the capturing setup.

4 Results

We evaluate the models on the introduced datasets. We measure the performance by computing the F1-Score on the test dataset, the results are shown in Tab. 2. The training time is measured in seconds and reported in Tab. 3.

Model	Performance				
	CryoNuSeg	LIVECell	ISIC	Seafood	Sphere
U-Net	59.08	81.97	82.51	92.81	92.52
SC	59.90	92.50	57.44	64.79	79.92
ED	60.77	93.04	69.62	84.55	90.24
ED-c	56.70	92.93	71.65	91.14	83.81

Tab. 2: Summary of the performance in F1-Score on all datasets. ED-c represents the ED model with color processing enabled.

On the CryoNuSeg dataset, both of our models achieved better results than the U-Net by approximately 1% and 2%. However, this dataset is challenging due to its complexity and diversity, resulting in lower performance across all algorithms. On the LIVECell dataset, both proposed models outperformed the U-Net by approximately 11%, and perform similarly to each other. Although the ED model has the capability to analyze more context, we believe it was not necessary in this case since the segmentation targets were small in size. The dataset features a unique characteristic of different resolutions between the test and training images. Our proposed models are able to easily accommodate inputs of varying sizes, resulting in no impact on their performance. In contrast, the U-Net was unable to adjust to this variation. On the ISIC dataset, the U-Net model exceeds the performance of the ED model by approximately 13% and the SC model by approximately 25%. The diversity of the ISIC dataset, which includes varying image capture setups and a range of melanoma types, is likely the reason for the notable differences in performance among the models. It's worth noting that the melanomas in the dataset also vary greatly in scale, which could be another contributing factor. On the seafood dataset, the U-Net outperformed our proposed models by approximately 8% and 28%. Since both proposed models were processing gray-scale images, crucial information was lost. On the Spheroid dataset, the U-Net again performed better, outperforming our proposed models by approximately 2% and 13%. Additionally, we are evaluating the model ED-c, which is able to process colored images. The results show a positive impact on the F1-Score for the Seafood (+6.59%) and ISIC (+2.03%) dataset, while the impact is small or negative for the datasets CryoNuSeg, LIVECell and Sphere.

We also evaluated the speed of training, the results are shown in Tab. 3. Both of our proposed models showed notably faster training times compared to the U-Net, with the simpler SC model being about 100 times faster and the ED model performing about 10 times as well on all evaluated datasets. During inference the U-Net is able to process an image in 113ms, while the SC model requires 924ms and the ED model takes 1699ms. While the inference is slower it is still able to process images with reasonable speed.

When comparing the performance of the SC and ED models, the ED model outperforms the SC model. Particularly, the ED model displays superior performance on the last three datasets, which contain larger objects or require a broader context to identify the primary target against background objects.

However, the difference in datasets with numerous smaller objects such as CryoNuSeg and LIVECell was not as pronounced. It is important to note that this improvement comes at the expense of training speed, as the SC model is 10 times faster than the ED model. Thus, we recommend using the SC model for segmenting many small objects, and the ED model when a larger image context is necessary.

Overall, our proposed models were able to outperform the U-Net on two out of five datasets and provide competitive performance on the Sphere and Seafood dataset. We also demonstrated that for complex datasets with important color features, the performance can be further improved by activating colored inputs. Additionally, we proof that our framework enables fast training cycles and is able to quickly adapt to new datasets.

Model	Training Time				
	CryoNuSeg	LIVECell	ISIC	Seafood	Sphere
U-Net	53.5	886.0	702.2	27665.0	71.4
SC	0.5	1.7	4.8	8.1	0.3
ED	5.2	26.4	23.9	79.6	1.3
ED-c	7.3	45.4	43.6	97.0	3.4

Tab. 3: Summary of the training times on all datasets in minutes. ED-c represents the ED model with color processing enabled.

5 Conclusion

Our paper presents a lightweight framework for semantic segmentation. Biomedical data is notoriously scarce and hard to label, making it challenging to train deep-learning models that can generalize to new datasets. Our framework overcomes this challenge by enabling researchers to quickly generate specialized models for each specific dataset. This allows the framework to be easily adapted to any changes in image material that may arise. Through our experiments, we evaluated two different models in terms of training speed and F1-Score performance. Our findings demonstrate that the framework can achieve impressive results on four out of five benchmark datasets with diverse characteristics, and that its fast training time, up to 100 times faster than the U-Net on a CPU, provides a feasible alternative to popular deep-learning frameworks in scenarios where GPU resources are limited or unavailable. We recommend the use of the ensemble of structured classifiers model when larger image context is needed and the single structured classifier model when segmenting many small objects.

Future works may involve enhancing the framework with image augmentation techniques, mechanisms for automatic detection of the need for colored image processing, and quantifying the minimum number of images required for good results to reduce unnecessary labeling effort.

References

[1] Amiri S, et al. Bayesian Network and Structured Random Forest Cooperative DeepLearning for Automatic Multi-label Brain Tumor Segmentation. In: 10th International Conference on Agents and Artificial Intelligence (ICAART 2018). 2018, 183–190.

[2] Codella NCF, et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC) 2018.

[3] Dollar P, et al. Structured Forests for Fast Edge Detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2013.

[4] Edlund C, et al. LIVECell—A large-scale dataset for label-free live cell segmentation. *Nature Methods* 2021;18.

[5] Geurts P, et al. Extremely randomized trees. *Machine Learning* 2006;63.

[6] He K, et al. Deep Residual Learning for Image Recognition. *arXiv* 2015;1512.03385.

[7] Mahbod A, et al. CryoNuSeg: A Dataset for Nuclei Instance Segmentation of Cryosectioned H&E-Stained Histological Images. *Computers in Biology and Medicine* 2021; 132:104349.

[8] Popova A, et al. Facile One Step Formation and Screening of Tumor Spheroids Using Droplet-Microarray Platform. *Small* 2019;15.

[9] Rettenberger L, et al. Self-Supervised Learning for Annotation Efficient Biomedical Image Segmentation. *IEEE Transactions on Biomedical Engineering* 2023;1:1–11.

[10] Ronneberger O, et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* 2015; abs/1505.04597.

[11] Russakovsky O, et al. ImageNet Large Scale Visual Recognition Challenge. *IJCV* 2015;115:211–252.

[12] Schilling M, et al. KaIDA: a modular tool for assisting image annotation in deep learning. *Journal of Integrative Bioinformatics* 2022;19:20220018.

[13] Shi Y, et al. Automatic Road Crack Detection Using Random Structured Forests. *IEEE Transactions on Intelligent Transportation Systems* 2016;17:3434–3445.

[14] Ulucan O, et al. A Large-Scale Dataset for Fish Segmentation and Classification. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). 2020, 1–5.

[15] Zhang J, et al. Segmentation of Perivascular Spaces Using Vascular Features and Structured Random Forest from 7T MR Image. In: Machine Learning in Medical Imaging. 2016, 61–68.