Image Segmentation of Liver Stage Malaria Infection with Spatial Uncertainty Sampling

Ava P. Soleimany [1 2]   **Harini Suresh** [1]   **Jose Javier Gonzalez Ortiz** [1]   **Divya Shanmugam** [1]   **Nil Gural** [1]
John Guttag [1]   **Sangeeta N. Bhatia** [1 3]

Abstract

Global eradication of malaria depends on the development of drugs effective against the silent, yet obligate liver stage of the disease. The gold standard in drug development remains microscopic imaging of liver stage parasites in *in vitro* cell culture models. Image analysis presents a major bottleneck in this pipeline since the parasite has significant variability in size, shape, and density in these models. As with other highly variable datasets, traditional segmentation models have poor generalizability as they rely on hand-crafted features; thus, manual annotation of liver stage malaria images remains standard. To address this need, we develop a convolutional neural network architecture that utilizes spatial dropout sampling for parasite segmentation and epistemic uncertainty estimation in images of liver stage malaria. Our pipeline produces high-precision segmentations nearly identical to expert annotations, generalizes well on a diverse dataset of liver stage malaria parasites, and promotes independence between learned feature maps to model the uncertainty of generated predictions.

1. Introduction

Malaria remains a major global health scourge, with nearly half of the world's population remaining at risk (World Health Organization, 2018). *Plasmodium vivax* is the main barrier to malaria eradication because it harbors dormant forms in the liver, termed hypnozoites, which can reactivate weeks to years after the initia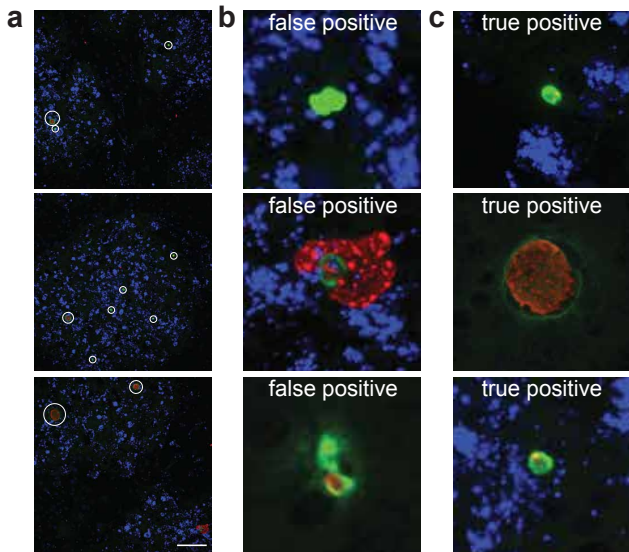l infection and cause relapsing disease. Thus, a malaria eradication campaign cannot be envisioned without eliminating the hypnozoite reservoir. However, there is only one clinically available drug that has anti-hypnozoite activity, which underlies the pressing need for continued drug screening and development (Wells et al., 2010; Alonso et al., 2011).

Current drug development approaches for liver stage antimalarials involve large-scale screening against *in vitro* cultures of liver cells that have been infected with malaria parasites (Gural et al., 2018; Antonova-Koch et al., 2018; Meister et al., 2011; Roth et al., 2018), with immunofluorescent microscopy as a primary readout. Though manual image segmentation for parasite identification and sizing remains standard, it is time and labor intensive, requiring significant technical expertise.

To address this need for automation in the anti-malarial screening pipeline, we present a convolutional neural network-based (CNN) architecture for automated segmentation of parasites in liver stage malaria infection, and develop a Bayesian deep learning approach for estimating uncertainty in image segmentations. We demonstrate accurate parasite detection and segmentation on a challenging liver stage *P. vivax* dataset, characterized by its variability in terms of parasite shape, density, and size. Finally, through dropout sampling of feature maps during training, we estimate the uncertainty of our segmentations and develop a generalizable algorithm for epistemic, or model, uncertainty (Kendall & Gal, 2017) metrics in image segmentation.

2. Related Work

Approaches to biomedical image segmentation fall into rule-based and network-based methods. A common rule-based approach is thresholding based on the intensity of fluorescent stains, which suffers in the face of noise and non-uniform stain intensities in the images (Phansalkar et al., 2011). A popular open-source software, CellProfiler uses user-defined features in conjunction with image processing algorithms to perform object identification for cell image analysis (Carpenter et al., 2006). CellProfiler can achieve high accuracies on many image processing problems, yet

[1]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA [2]Harvard University, Cambridge, Massachusetts, USA [3]Howard Hughes Medical Institute, Cambridge, Massachusetts, USA. Correspondence to: Ava P. Soleimany <asolei@mit.edu>.

Figure 1. **Dataset of liver stage *Plasmodium vivax* infection.** Representative images from dataset of *P. vivax* infected human hepatocyte cultures, eight days post infection. Green and red channels correspond to stains for parasite membrane and cytoplasmic proteins, respectively, and blue channel corresponds to nuclear stain. A: Representative data, where white circles indicate true positives. Scale bar is 100 μm. B: Representative examples of false positives in image patches. C: Representative examples of true positives in image patches, including the small forms or hypnozoites (top & bottom) and large forms or schizonts (middle).

is limited in its generalizability and robustness, as a new pipeline and a new feature set must be created for each distinct task.

CNNs have emerged as an attractive approach for the segmentation of biomedical images due to their strength on visual recognition tasks. For example, a convolutional network with a sliding-window setup to predict the class label of each pixel in an image has been presented (Ciresan et al., 2012), yet this approach suffers from computational inefficiency and requires significant spatial context to achieve high accuracy. Our approach addresses these limitations by using a convolutional U-Net (Ronneberger et al., 2015) for accurate, efficient image analysis of liver stage malaria infection. Furthermore, we develop a generalizable method for spatial uncertainty modeling (Amini et al., 2018) in image segmentation and provide concrete estimates of model uncertainty on our malaria dataset.

3. Methodology

Given a dataset of input image and binary segmentation pairs $(\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{Y}$ is a pixel-wise binary classification of the input image with positive pixel labels corresponding to the parasite class, we learn a functional mapping $\boldsymbol{f}$ parameterized by weights $\boldsymbol{W}$ such that the distance from its

output $\hat{\boldsymbol{Y}} = \boldsymbol{f}(\boldsymbol{X}; \boldsymbol{W})$ to the true labels $\boldsymbol{Y}$ is minimized.

3.1. Dataset

The liver stage *P. vivax* dataset consists of images of human hepatocytes cultured in the micropatterned co-culture format and imaged eight days following infection with *P. vivax* parasites. Cultures were fixed and stained for the parasite membrane protein UIS4 (green channel) and either a histone acetylation marker, H3K9ac, or a cytoplasmic parasite protein, BIP (red channel). DAPI (blue channel) staining was used to mark hepatocyte nuclei. On day eight, cultures comprise two parasite stages: the dormant hypnozoites which are small in size and uninucleated (Fig. 1c, top & bottom), and the maturing schizonts which are large and multinucleated (Fig. 1c, middle). Approximately 0.2-1% of hepatocytes are infected based on the typical infectivity of *P. vivax* in this culture system. All images were annotated by experienced researchers via manual identification and segmentation of the parasites present. The dataset was split into training, validation, and test cohorts, and 256×256 pixel RGB image patches were selected from each 1024×1024 images. Patches that contained the presence of the green UIS4 parasite membrane stain and the red BIP parasite protein stain above a threshold T were fed into the model for training. We further augmented the data by applying random rotations and symmetries. Approximately 5000 augmented image patches were used for training.

3.2. Model

To perform the segmentation task, we utilized a convolutional U-Net architecture (Ronneberger et al., 2015), which uses a autoencoder structure where the input image is compressed and then decompressed through successive convolutional layers, connected via skip-like connections (Fig. 2). The network was trained using the pixel-wise cross-entropy loss,

$$\mathcal{L}(y, \hat{y}) = -\sum_i y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (1)$$

where i is the index of a pixel in an image, y is the true label of a pixel, and $\hat{y} \in [0, 1]$ is the predicted label of the pixel. We performed one-fold cross validation using the validation data, and applied our trained model to a held-out test set to assess performance. All models were implemented in TensorFlow and trained on a NVIDIA Titan X GPU.

The performance of our U-Net model was benchmarked against the ground truth human annotation, a thresholding algorithm, and a regularized logistic regression model trained on class-balanced patches. We utilized a commonly used evaluation metric, the Average Precision Score (APS): $APS = \sum_n (R_n - R_{n-1})P_n$, where R_n and P_n are the recall and precision, respectively, at threshold n. Precision is
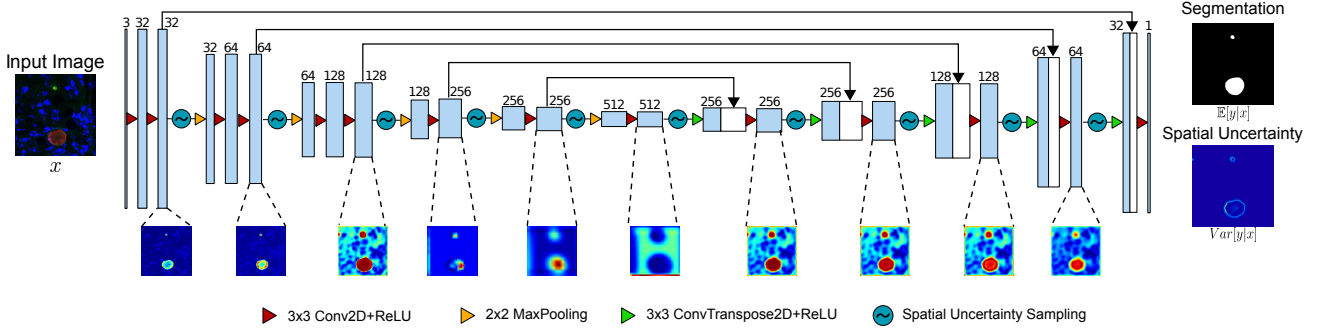
Figure 2. **Architecture for segmentation of images of liver stage malaria infection.** A convolutional U-Net successively compresses and decompresses the input image to output a pixel-wise classification. Convolutional blocks across the encoder and decoder portions of the network are connected by skip-like connections indicated in black arrows. Spatial uncertainty sampling via dropout is applied between successive convolutional blocks to estimate model uncertainty, and network outputs after convolutional blocks are visualized in inset.

defined as the percentage of segmented pixels that correctly correspond to a parasite, and recall is the percentage of true segmented pixels identified.

3.3. Uncertainty Estimation

To investigate the confidence of our model in generating parasite segmentations, we utilized a dropout based approach for estimating model uncertainty (Gal & Ghahramani, 2016a;b; Amini et al., 2018). It has been shown that applying dropout before every weight layer in a neural network is equivalent to approximating a probabilistic deep Gaussian process, and that this approximates the posterior distribution over the network weights, $q(\boldsymbol{W})$ (Gal & Ghahramani, 2015; 2016a). Given input image data $\boldsymbol{X}$ and a segmentation output $\boldsymbol{Y}$, we can use the dropout-based approximation of the posterior, to obtain an estimation for the predictive distribution $q(\boldsymbol{Y}|\boldsymbol{X})$, the likelihood of a segmentation given an input image. We have:

$$q(\boldsymbol{Y}|\boldsymbol{X}) = \int P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{W}) \, q(\boldsymbol{W}) \, d\boldsymbol{W} \qquad (2)$$

Given T stochastic forward passes through the network generated with dropout, $\{\boldsymbol{W_t}\}_{t=1}^{T}$, we can define the predictive mean as: $\boldsymbol{E}[\boldsymbol{Y}|\boldsymbol{X}] = \frac{1}{T} \sum_{t=0}^{T} \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{W_t})$. Thus, the predictive variance, which defines the model uncertainty, is:

$$\mathrm{Var}[\boldsymbol{Y}|\boldsymbol{X}] = \frac{1}{T} \sum_{t=0}^{T-1} \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{W_t})^2 - E[\boldsymbol{Y}|\boldsymbol{X}]^2. \quad (3)$$

Given that the U-Net architecture is purely convolutional, we utilize spatial Bernoulli dropout (Tompson et al., 2015; Amini et al., 2018) for the stochastic sampling. In this method, our uncertainty estimates result from sampling a Bernoulli random variable $\boldsymbol{Z}$ to drop entire feature maps in the network, where $\boldsymbol{z}^{(k,l)} \sim Bernoulli(p)$ corresponds to k-th feature map in the l-th layer and p is the probability that

all units in the feature map remain active (Gal & Ghahramani, 2015). This approach has previously been shown to be a special case of element-wise dropout and thus a valid way for estimating model uncertainty (Amini et al., 2018).

4. Experiments and Results

To assess the performance of our model on the segmentation of *P. vivax* parasites, we trained the U-Net architecture with pixel-wise cross entropy loss, evaluated our trained model on a held-out test dataset of *P. vivax* liver stage culture and assessed segmentation precision using the APS. As baselines, we evaluated the performance of the thresholding and logistic regression models on the test dataset.

Representative examples of the resulting segmentations are shown in Fig. 3a. We visually compared the segmentations returned by each of thresholding baseline, logistic regression, and the U-Net with cross entropy loss to images from the dataset with true positives highlighted, and found that the baseline models appear unable to distinguish parasites from other stains of similar brightness, or recognize that two regions of significantly different brightnesses could both be parasites. These results suggest that our U-Net architecture is both sensitive and specific in its detection and segmentation of parasites in these images. More quantitatively, we evaluated the precision of the U-Net and logistic regression models using the APS, and found that the U-Net trained with cross-entropy loss achieved a precision exceeding 98% on the held-out test set, while the logistic regression model achieved a precision of 80.7%.

In addition to the detection and segmentation of malaria parasites in hepatocyte cultures, statistics about the number and size of parasites present in an image are critical to both drug development and fundamental biology research. Using a simple clustering algorithm, we count and size parasites with high accuracy, identifying approximately 10%
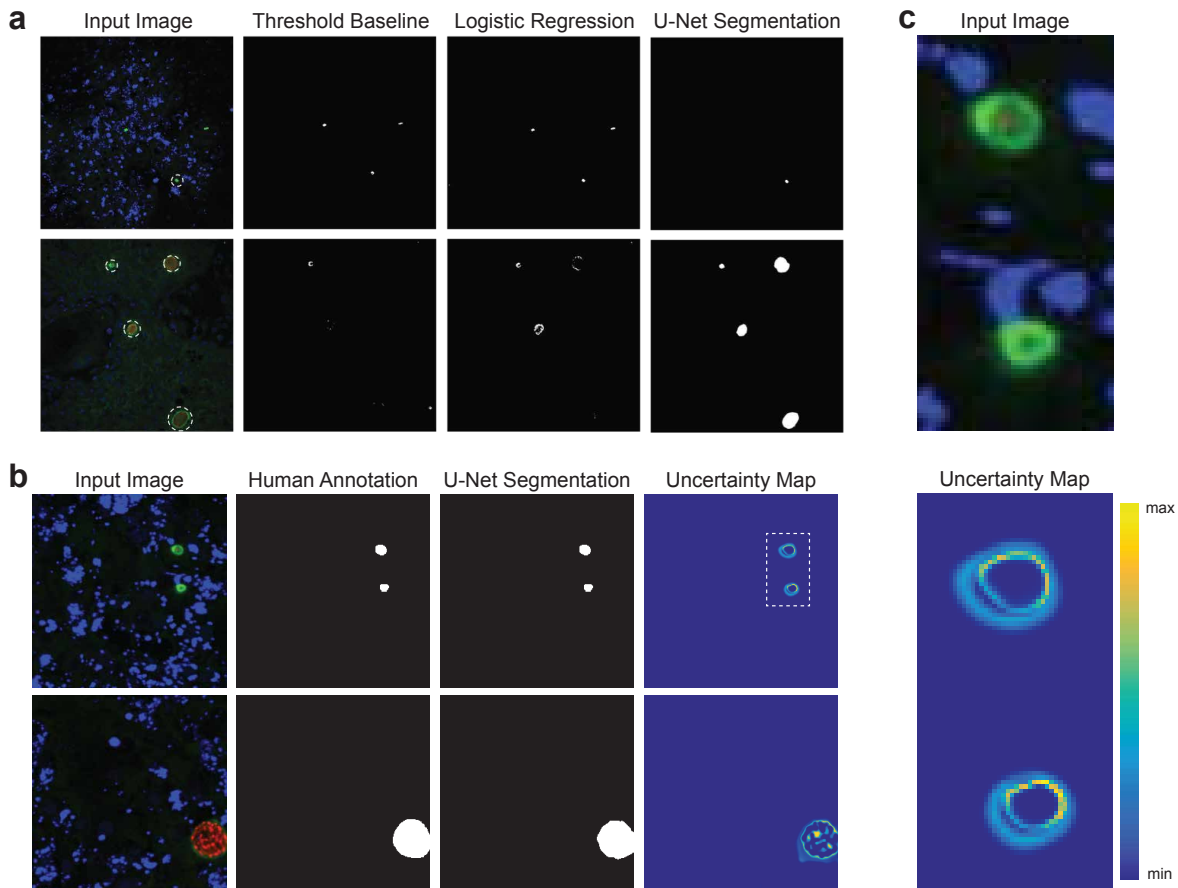
Figure 3. **Neural network segmentation performance and uncertainty estimation.** A: Results for the segmentation task on dataset images with true positives circled in white, comparing the segmentations returned by the threshold baseline, logistic regression, and U-Net models. B: Representative examples from dataset (left), ground truth human annotation (middle left), U-Net segmentations (middle right), and visualization of model uncertainty (right). C: Zoom-in of the boxed region in (b) comparing the input image (top) to estimations of model uncertainty over the region of interest (bottom).

for human annotation. Out of the 13.3% of the images where the number of parasites was incorrectly counted, 91.3% were within 1 of the true annotation.

Finally, we applied spatial dropout sampling to estimate and subsequently visualize model uncertainty, that is, the regions of predicted segmentation where we can think of the model as being less confident in its prediction (Fig. 3b, c). As shown in Fig. 3c, the U-Net model is more uncertain in its prediction around the edges and membranes of the segmented parasites, and more certain in the interior, cytoplasmic portion of the parasites, consistent with what we may expect in terms of a model's confidence in detection and segmentation of particular regions of an image. It is important to note that while the network's predicted *probability* for a given class may be high, the model may not necessarily be *confident* in that prediction. The uncertainty algorithm presented here may thus be used to capture these situations and to inform subsequent human annotation.

5. Conclusion

Our work supports the potential for neural network-based methods to automate previously time-intensive processes of the anti-malarial drug development pipeline. In particular, our U-Net architecture demonstrates robust segmentation of parasites during the liver stage of *P. vivax* malaria with near-human accuracy. Furthermore, we develop a novel method for uncertainty estimation in image segmentation tasks, utilizing spatial dropout sampling to visualize model uncertainty and evaluating our approach on the malaria dataset. We believe that our neural network approach provides an efficient, accurate method for automated image segmentation of liver stage malaria for applications in drug development and screening, and that our spatial dropout algorithm provides a generalizable method for robust estimation of model uncertainty in image segmentation tasks.

References

Alonso, P. L., Bassat, Q., Binka, F., Brewer, T., Chandra, R., Culpepper, J., Dinglasan, R., Duncan, K., Duparc, S., Fukuda, M., et al. A research agenda for malaria eradication: drugs. *PLoS medicine*, 8(1):e1000402, 2011.

Amini, A., Soleimany, A., Karaman, S., and Rus, D. Spatial uncertainty sampling for end-to-end control. *arXiv preprint arXiv:1805.04829*, 2018.

Antonova-Koch, Y., Meister, S., Abraham, M., Luth, M. R., Ottilie, S., Lukens, A. K., Sakata-Kato, T., Vanaerschot, M., Owen, E., Jado, J. C., et al. Open-source discovery of chemical leads for next-generation chemoprotective antimalarials. *Science*, 362(6419):eaat9446, 2018.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., et al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100, 2006.

Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pp. 2843–2851, 2012.

Gal, Y. and Ghahramani, Z. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015.

Gal, Y. and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059, 2016a.

Gal, Y. and Ghahramani, Z. A theoretically grounded application of dropout in recurrent neural networks. In *Advances in neural information processing systems*, pp. 1019–1027, 2016b.

Gural, N., Mancio-Silva, L., Miller, A. B., Galstian, A., Butty, V. L., Levine, S. S., Patrapuvich, R., Desai, S. P., Mikolajczak, S. A., Kappe, S. H., et al. In vitro culture, drug sensitivity, and transcriptome of plasmodium vivax hypnozoites. *Cell host & microbe*, 23(3):395–406, 2018.

Kendall, A. and Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pp. 5574–5584, 2017.

Meister, S., Plouffe, D. M., Kuhen, K. L., Bonamy, G. M., Wu, T., Barnes, S. W., Bopp, S. E., Borboa, R., Bright, A. T., Che, J., et al. Imaging of plasmodium liver stages to drive next-generation antimalarial drug discovery. *Science*, 334(6061):1372–1377, 2011.

Phansalkar, N., More, S., Sabale, A., and Joshi, M. Adaptive local thresholding for detection of nuclei in diversity stained cytology images. In *Communications and Signal Processing (ICCSP), 2011 International Conference on*, pp. 218–220. IEEE, 2011.

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

Roth, A., Maher, S. P., Conway, A. J., Ubalee, R., Chaumeau, V., Andolina, C., Kaba, S. A., Vantaux, A., Bakowski, M. A., Luque, R. T., et al. A comprehensive model for assessment of liver stage therapies targeting plasmodium vivax and plasmodium falciparum. *Nature communications*, 9(1):1837, 2018.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., and Bregler, C. Efficient object localization using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 648–656, 2015.

Wells, T. N., Burrows, J. N., and Baird, J. K. Targeting the hypnozoite reservoir of plasmodium vivax: the hidden obstacle to malaria elimination. *Trends in parasitology*, 26(3):145–151, 2010.

World Health Organization. *World malaria report 2018*. World Health Organization, 2018.