Adaptive Illumination in Microscope with Visual Attention for Specimen Classification

Ravitashaw Bathla
Duke University
Durham, NC 27705
rb369@duke.edu

Abstract

The image acquisition process from a digital microscope can be automated by optimizing the physical layer with machine learning algorithms. In this paper, a recurrent visual attention mechanism, based on self-attention, is used to generate a latent space in conjugate with optimizing the illumination pattern for the classification task. The network decides the information available in the latent space to decide if the information is sufficient to make a classification decision. The network exits from the loop when a decision is made for the latent space. Several experiments are conducted in this work to understand how an adaptive LED illumination pattern for each sample provides better classification accuracy. The number of iterations per image sample for the network to make a decision (average length of the trajectory) is observed to correlate with the classification accuracy.

1. Introduction

With the advent of machine learning, there has been significant research in digital microscopy to improve the overall process of specimen capturing, understanding, and classification. Several preprocessing and postprocessing that are already available in digital microscopes that enhances the capturing of biological specimen. However, such microscopes are manual, and a rigorous number of manual adjustments is required for the acquisition of specimen, and often is time-consuming. Machine Learning techniques can automate mundane as well as complicated tasks can save time for the researchers and scientists. These digital microscopes can be automated using machine learning algorithms that can learn these manual adjustments to capture specimen samples that can be classified without much manual intervention. The specimen capture process under a microscope involves several adjustments like illumination pattern, the brightness of the light source, lens aperture, the transmittance of the illumination system. By employing the use of Machine Learning, the process of image capture from microscopes can be automated and be made more task-specific.

In this paper, a controlled study of learned illumination pattern is performed to classify the images of blood cells from thin smear slides used in the clinical diagnosis of malaria. To automate the end-to-end specimen classification process through a microscope, the illumination pattern is simulated as the physical layer. A deep neural network is trained, along with the optimization of the physical layer. This study is an extension of the work done by Chawre et al. (2020). Chawre et al. (2020) performed a study to classify the specimens using a closed feedback loop by balancing the image acquisition cost and classification accuracy. The study employed an exit strategy using a close feedback loop to decide if a new illumination pattern needs to be learned by the network for the classification of blood cells. The sequence of the number of steps (trajectory) that explores different LED illuminations has been done using variants of Recurrent Neural Network - LSTM (Hochreiter et al., 1997) and GRU(Cho et al., 2014).

The goal of this paper is to use the attention-based mechanism for accumulating image samples in the latent space for classification as used in transformers in natural language processing. There are primarily two experiments performed in this paper, first, the performance of fixed illumination (common) for all the images is compared to adaptive illumination for each image sample. Second, with the adaptive illumination classification accuracy is measured for a variable length of trajectory to observe if there is any performance gain for the length of the trajectory.

2. Related Work

There have been several pieces of research around the optimization of the physical layer along with training deep neural networks (DNN). Horstmeyer et al. (2017) employ the use of a convolutional neural network to determines an illumination pattern to highlight important sample features to classify the detected images in the postprocessing stage. Kellman R et al. (2019) work on coded-illumination patterns for a LED array microscope for phase reconstruction algorithm using DNNs on mouse fibroblast cells showed performance gain. Cooke et al. (2020) performed experiments on varying illumination patterns for different image samples suggesting an increase in performance for virtual fluorescence microscopy. Chawre et al. (2020) further developed on the study of Cooke et al. (2020) to perform adaptive illumination by using a reinforcement learning mechanism. The study involves rewarding if the classification decision is made in fewer attempts of LED illumination for a given image sample. The iterative policy per sample was designed using Long Short-Term Memory (LSTM) for generating a latent space that is encoded for classification.

Recent advances in iterative algorithms have opened several avenues of research in designing DNNs for imaging and borrowing of DNN architectures from natural language processing. The NLP based model has provided great success in natural language translation, understanding, and classification. In particular, Transformers (Vaswani et al., 2017), based upon self-attention have resulted in state of art in NLP. Inspired by NLP successes, multiple works combined CNN-like architectures with self-attention (Wang et al., 2018). Dosovitskiy et al. (2020) suggest eliminating the convolutional neural network and use a self-attention-based model for computer vision tasks. This study has shown improvement in performance by using a self-attention-based network over convolutional neural networks.

Recurrent Visual Attention was studied by Mnih et al. (2014) and showcased how networks can learn to self-select visual information most relevant for further processing. Chaware et al. (2020) further elaborated on the idea by applying the concept of visual attention to microscopy. This paper is an attempt to apply self-attention with visual attention on microscopy for sample classification.

3. Data

The malaria dataset of blood cells sample from thin smear slides is taken from Cooke et al (2019) which is an extension of experiments conducted in Horstmeyer et al. (2017) as mentioned in Chaware et al. (2020). The images are cropped and labeled to construct a binary classification task of classifying the blood cell sample as infected or not infected by the malaria parasite. Variable illumination was provided by 29 LEDs where each LED comprised of 3 spectral channels creating 87 uniquely illuminated images. The data was split into 80:20 percentage for training and testing. A blood sample is shown in figure 1 for a spectral channel from 10 LEDs illuminated. The data was available in the form of tensors of size 87x28x28.

4. Method

The network architecture consists of several components that are optimized in conjugation with each other, thus, providing a single automated stand-alone system for classification. In the network architecture (Figure 1), the illuminated image is encoded using a convolutional neural network and multi-layer perceptron. This encoded image serves as input to the transformer module. The transformer module consists of a self-attention-based encoder and decoder with a memory unit of the previous encoding. The decoded self-attention embedding is reduced to mean for each image and further clipped using hyperbolic tangent as suggested in Dosovitskiy et al. (2020). This output embedding is passed through the decision network to decide if there is enough information in the latent space to make a classification decision. The decision network module will continue to gather information in the latent space in the form of a closed-loop. On each iteration of the loop, an LED pattern is optimized using the output embedding, this is the physical layer component of the network. Once the network exits from the loop, a classification is made.

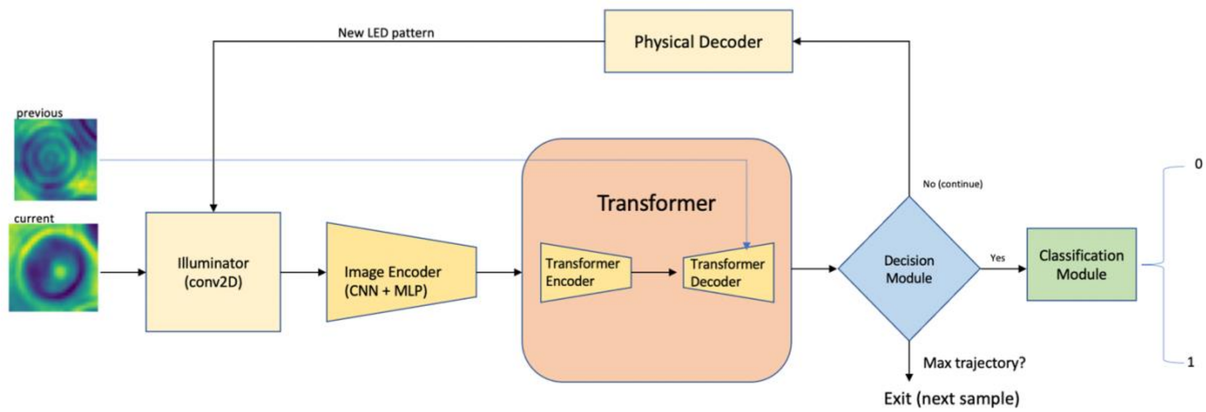


Figure 1: Network Architecture Design for Adaptive Illumination using Transformers

a. Loss

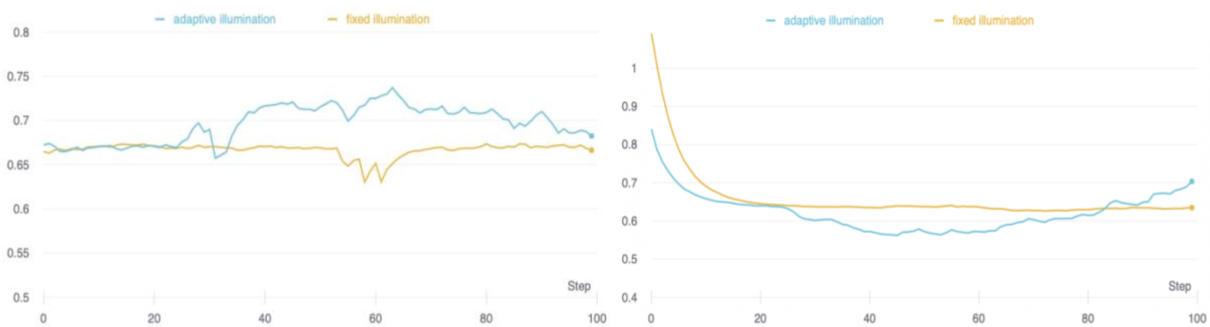
There are primarily two loss components that are combined to optimize the entire network. Both the loss are negative log-likelihoods (cross-entropy loss) for classification as well as the decision module. The loss in the decision module is scaled such that there is a penalty for time out – no decision was made for the specified trajectory length, as well as a penalty for incorrect classification.

b. Experiments

The first experiment is run for a single trajectory length with fixed illumination (common optimization of LED pattern for all samples) and adaptive illumination (optimization of LED pattern per sample). The purpose of this experiment is to see if there is any substantial gain in performance for optimization of LED pattern per sample. This also serves as the baseline model for performance measurement. An experiment is conducted to measure performance for a different number of trajectory lengths (common for all samples). The trajectory length that is tested in this experimental setup is in the range from 2 to 4. This experiment was conducted with a prime objective to see if there is an improvement in accuracy with different trajectory length. This experiment will optimize the complete network, so the result of this experiment will help in deciding the optimal trajectory length and validating the finding with the next experiment for adaptive trajectory per sample. The final experiment conducted is to test for adaptive trajectory length per sample. This experiment will provide insights about the beneficence of variable trajectory length per sample.

5. **Results**

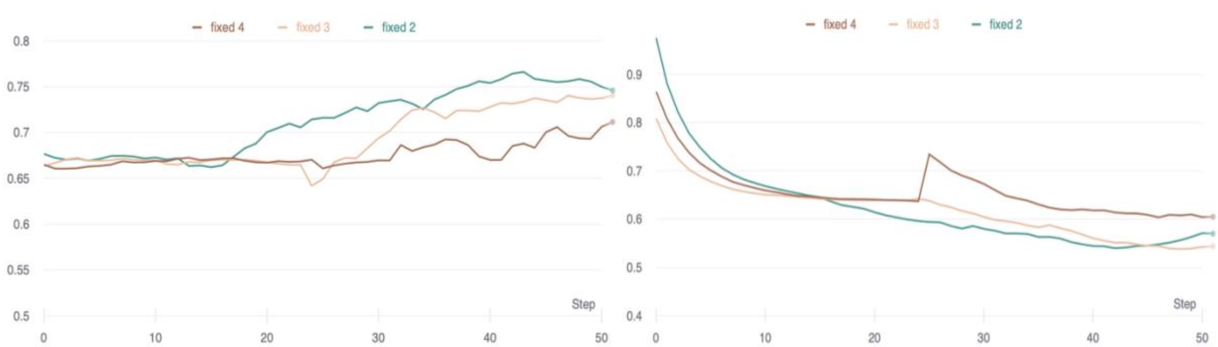
The fixed illumination experiment yielded an average accuracy of 67 percent and the adaptive illumination yielded an average accuracy of 71 percent. The experiment was run with a trajectory length of 1, therefore, for each sample, only a classification decision was made. Although the difference in accuracy is small, the losses across the two experiments are close. The network is overfitting on the data in the adaptive illumination experiment (Figure 2(b)). If early stopping is considered as the decision criteria, the best model accuracy is 74 percent for the adaptive illumination.



(a) (b)

Figure 2: Accuracy (a) and Loss (b) of Fixed and Adaptive Illumination using fixed trajectory of size 1 on malaria dataset

Next, the result of the controlled study on a different number of trajectories with adaptive illumination is shown in Figure 3. There is no major gain in performance with an increase in the trajectory length. The learning was stopped at 50 epochs as the model was overfitting beyond that point. Since the network was overfitting, these results might not be conclusive of this pattern. A different network architecture with less overfitting might produce the results otherwise. One of the image samples classified as infected by the network with an illumination pattern is shown in Figure 4.



(a) (b)

Figure 3: Accuracy (a) and Loss (b) of adaptive illumination using fixed trajectory of size 2, 3 and 4 on malaria dataset

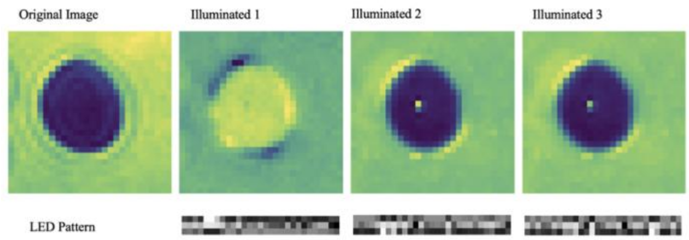


Figure 4: Illuminated Infected Malaria Cells for trajectory of length 3

Lastly, an experiment was performed to check the relationship between accuracy and trajectory length. The results of the experiment are shown in Figure 5. After several experiments with the trajectory length, the maximum trajectory length was set to 8, and the training was performed to compare the performance with the average trajectory length. It can be observed that the classification accuracy of the images has a linear relation with the trajectory length. A similar pattern, although, at a different scale is seen for both training and validation accuracies.

Figure 5: Average Trajectory Length with respect to Accuracy of training and validation on malaria dataset

6. Discussions

From the above experimental results, it can be concluded that there is a gain in performance on having an adaptive illumination for each sample over a common illumination pattern for the samples. Contrary to the belief, the increase in the fixed trajectory of variable length did not increase the performance as per the second experiment. This could likely be due to several reasons; the network can classify in one or two trajectory lengths, any optimization beyond this trajectory seems futile. It could be the limitation of the network architecture; a denser network may be able to learn better with an increased variable trajectory length. With the results of the third experiment on comparing adaptive trajectory length for each sample with accuracy, it can be inferred that there is a linear relationship between performance and average trajectory length. This is contrary to the results of the second experiment. Therefore, it can be argued that a denser network or further fine-tuning the current network will provide further insights about real-time performance for simulating adaptive trajectory and adaptive illumination for specimen classification.

7. References

Chaware, A., Cooke, C. L., Kim, K., & Horstmeyer, R. (2020, May). Towards an intelligent microscope: adaptively learned illumination for optimal sample classification. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 9284-9288). IEEE.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Cooke, C. L. (2019). Fourier Malaria. *Retrieved October 25, 2020 from https://www.kaggle.com/clvcooke/fourier-malaria*

Cooke, C. L., Kong, F., Chaware, A., Zhou, K. C., Kim, K., Xu, R., ... & Horstmeyer, R. (2020). Physics-enhanced machine learning for virtual fluorescence microscopy. *arXiv preprint arXiv:2004.04306*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Uszkoreit, J. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Horstmeyer, R., Chen, R. Y., Kappes, B., & Judkewitz, B. (2017). Convolutional neural networks that teach microscopes how to image. *arXiv preprint arXiv:1709.07223*

Kellman, M. R., Bostan, E., Repina, N. A., & Waller, L. (2019). Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, *5*(3), 344-353.

Kim, K., Konda, P. C., Cooke, C. L., Appel, R., & Horstmeyer, R. (2020). Multi-element microscope optimization by a learned sensing network with composite physical layers. *Optics Letters*, *45*(20), 5684-5687.

Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. In *Advances in neural information processing systems* (pp. 2204-2212)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794-7803).