

Physics-Enhanced Machine Learning for Virtual Fluorescence Microscopy

Colin L. Cooke^{*1}, Fanjie Kong¹, Amey Chaware¹, Kevin C. Zhou², Kanghyun Kim², Rong Xu¹, D. Michael Ando³, Samuel J. Yang³, Pavan Chandra Konda², and Roarke Horstmeyer^{1,2}

¹Duke University, Electrical and Computer Engineering Department

²Duke University, Biomedical Engineering Department

³Google Research, Applied Science Team

Abstract

This paper introduces a new method of data-driven microscope design for virtual fluorescence microscopy. We use a deep neural network (DNN) to effectively design optical patterns for specimen illumination that substantially improve upon the ability to infer fluorescence image information from unstained microscope images. To achieve this design, we include an illumination model within the DNN’s first layers that is jointly optimized during network training. We validated our method on two different experimental setups, with different magnifications and sample types, to show a consistent improvement in performance as compared to conventional microscope imaging methods. Additionally, to understand the importance of learned illumination on the inference task, we varied the number of illumination patterns being optimized (and thus the number of unique images captured) and analyzed how the structure of the patterns changed as their number increased. This work demonstrates the power of programmable optical elements at enabling better machine learning algorithm performance and at providing physical insight into next generation of machine-controlled imaging systems.

1. Introduction

The optical microscope remains a critical tool across a wide variety of disciplines, ranging from high-content screening in biological labs to quality control in factories. With the continued growth of automated software analysis tools, many microscope images are now rarely viewed directly in their raw format by humans, but are instead commonly processed first by a computer. Examples include the automatic classification of different cell types within large

cell cultures [1], segmentation of cancerous areas from thin pathology tissue sections [2], and, as focused upon here, the automatic creation of fluorescence images from bright-field data [3], which we term “virtual fluorescence” (an example of the virtual fluorescence process is shown in Figure 1).

Despite the continued advancement of image analysis software, microscope hardware has changed relatively little over the past several centuries. Most microscopes still consist of standard illumination units and objective lenses that are optimized for direct human inspection. Optical microscopes are constrained by several physical limitations, including a limited resolution, field-of-view, image contrast, and depth-of-field, which restrict the amount of information that can be captured within each image. Standard microscope design biases this limited information towards human analysis, potentially impacting the accuracy of automated analysis.

Here, we attempt to optimize the hardware of a new microscope design to improve the performance of a deep learning based image labelling task. To achieve this goal, we present a modified neural network which includes a physical model of our experimental microscope, which is optimized jointly with the deep neural network (DNN) during training. In this work, we limit our physical model to include only the spectral and angular properties of the microscope’s illumination, realized through a programmable LED array, but leave open the possibility of considering other important parameters (focus setting, lens design, detector properties) in future work. Our proposed system models the microscope illumination pattern as a set of linear weights that are directly integrated into the DNN, allowing the calculation of gradients through back-propagation and end-to-end optimization during supervised training. After training, the optimized “physical” weights can be interpreted as the distribution of LED brightnesses and colors to use in our experimental imaging setup, which transfers performance gains seen in training to a physical setup.

^{*}Colin Cooke is corresponding author: colin.cooke@duke.edu
Source code available at: github.com/clvcooke/virtual-fluorescence

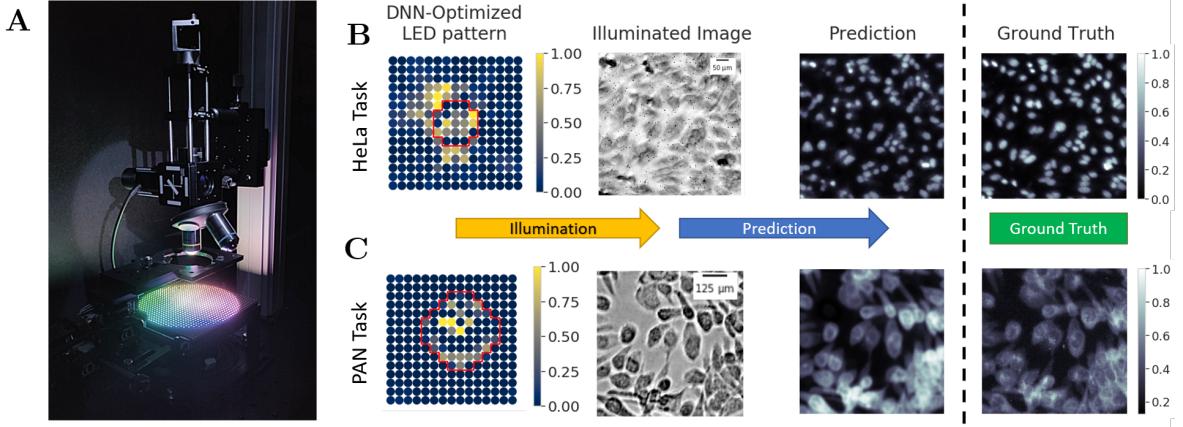


Figure 1: (A) Experimental setup: Microscope with both fluorescence and non-fluorescence imaging paths centered above a programmable LED array for training data capture. (B/C) Example fluorescence inference pipelines: The optimized LED pattern illuminates a sample to form an image. The neural network processes LED-illuminated images to produce an estimated fluorescence result. The ground truth is shown to the right of the inference result. Example B shows the *HeLa* task while Example C shows the *PAN* task. The red outline on the patterns indicates the bright/dark-field cutoff of the LED array. Where in bright-field illumination, both the scattered and unscattered light transmitted through the sample is directly observed through the lens, and in dark-field only scattered light (light which has been re-directed by the sample) is observed.

Here, our goal was to train a DNN to convert conventional transmission microscope imagery, captured with our learned illumination pattern, into a simultaneously captured fluorescence image that highlights specific features of interest via fluorescent markers. This goal of moving from unlabelled images to labelled ones synthetically is what we term “virtual fluorescence”, and has recently received interest as a promising means to avoid the need to fluorescently label specimens and to instead simply rely on deep learning to post-process standard image data[3, 4, 5].

We have two main goals within our work. The first is to understand the impact of illumination pattern design on task performance for virtual fluorescence image generation. While illumination dramatically alters the appearance of a microscopic specimen, the impact of the incident light’s spatial, angular and spectral properties on fluorescent-specific feature identification is challenging to directly establish. Second, we will explore how the number of designed illumination patterns used for image capture changes both in terms of pattern structure and overall task performance. To achieve this latter goal, we vary the number of patterns which are simultaneously optimized with the DNN. For each set we examine the interaction between patterns within each optimized illumination set, and how the patterns change as a function of set size. We find that optimized patterns not only yield higher performance than conventional alternatives, but the structure of the patterns themselves provides a certain degree of physical intuition between scattered bright-field light and fluorescent emission that can be used to improve future data collection

strategies.

Furthermore, our method represents a way of performing virtual fluorescence that is cheaper, more customizable, and potentially faster than existing alternatives. Previous approaches built upon a physical intuition to design their data capture system, such as Christiansen et al. [3] capturing a z-stack of images, or Cheng et al. [5] using pre-designed illumination patterns. In contrast, our method of data collection is entirely data-driven, where the trade-off between data collection speed and system performance is explicit. Additionally, since our hardware setup is inexpensive and uses no moving parts, it is easily deployable and accessible.

2. Related work

In recent years, convolutional neural networks (CNNs) have become commonplace for both medical and natural image processing [6]. Segmenting images to find specific cell types or sub-cellular features (e.g. cell nuclei), for example, is now a common biomedical image analysis task that CNNs excel at. The U-net structure [7], perhaps one of the most widely used CNNs within biomedical image analysis, has been applied across a wide variety of segmentation tasks [8] and makes efficient use of annotated data during training through its fully convolutional architecture.

As the use of neural networks continues to increase in popularity, many researchers are now also applying them to automatically analyze fluorescence imagery. Belthangady et al. [9] recently reviewed this increasing body of work and summarized it into two general categories: virtual labelling

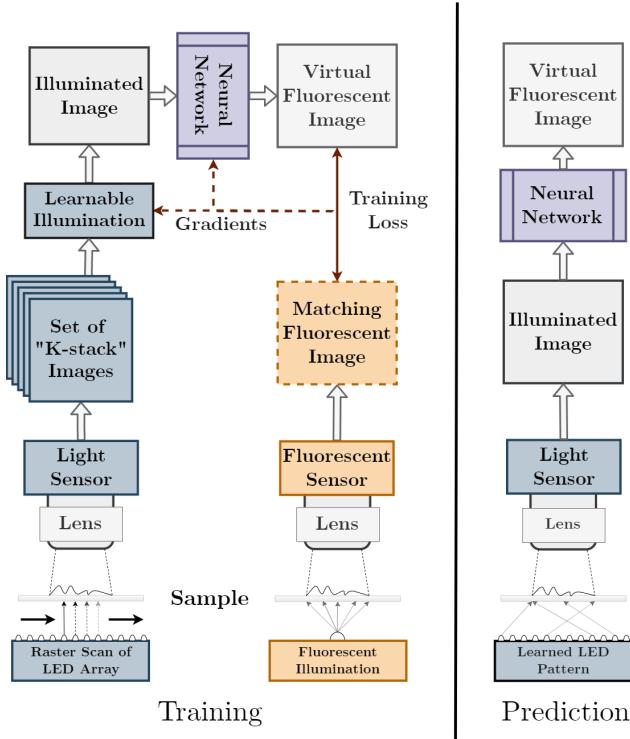


Figure 2: Diagram of training and inference pipelines for physics-enhanced fluorescence image inference. The left and right regions contain the setups used to capture data for network training, and to capture data in application for fluorescence image prediction, respectively. The fluorescence image and generated “k-stack” images are captured through the same objective lens. Loss is taken as the MSE between each predicted virtual fluorescence image and matching ground-truth fluorescence image, and then backpropagated to optimize the *Neural Network* and *Learnable Illumination* modules. For fluorescence image prediction, a *Learned LED Pattern* is used to illuminate the specimen. This *Learned LED Pattern* is configured using the optimized weights from the *Learnable Illumination* module determined during training. The same *Neural Network* is used from the training setup.

and fluorescence image enhancement.

Virtual labelling is the process of using unlabelled images to predict fluorescence images. While still a relatively new concept, several recent works demonstrate that DNNs can be quite effective at this task, which suggests that certain future experiments may forgo fluorescence imaging and staining entirely. Christiansen et al. [3] first demonstrated this concept by developing a CNN which predicted seven distinct fluorescence channels from a set of several dozen uniquely focused bright-field images. Since then numerous works have demonstrated success on similar prob-

lems, ranging from virtual fluorescence in 3D [4, 10] to reflectance microscopy [5]. Additionally the area of “virtual staining” [11, 12, 13] (inferring histologically stained images from unstained tissue) has recently shown promise and is in many ways similar to virtual fluorescence. These works all demonstrate that it is possible to infer information revealed by a fluorescent or histological stain from using unmodified unlabelled images, albeit at varying levels of accuracy.

Fluorescence image enhancement focuses on improving the quality of existing fluorescence images. Weigert et al. [14] developed a content-aware image restoration method powered by a CNN. This work showed that fluorescence image restoration was possible by predicting high resolution fluorescence images from ones which were undersampled. In an earlier work, Weigert et al. [15] used a CNN to perform isotropic reconstruction of 3D fluorescence data. Through these works, we can infer that data contained within fluorescence images is potentially redundant, and through the assumption of key underlying features, it is at times possible to enhance image quality via CNN post-processing.

However, in most studies, the focus is on post-processing data, rather than attempting to influence or improve the image acquisition process. While an early work used simple neural networks to effectively design components of optical systems [16], the first work (to the best of our knowledge) to examine hardware optimization in the context of CNNs was by Chakrabarti [17], who presented an optimal pixel level color-filter layout for color image reconstruction. A number of subsequent works have considered how to merge the optimization of various imaging hardware components into a differentiable optimization network [18, 19, 20, 21, 22, 23, 24, 25, 26]. However, few of these works have proposed the use of deep learning to optimize the image capture process for automated analysis.

Such an approach was recently considered by Muthumbi et al. [27], who suggested the use of a “physical layer” in a DNN to optimize an illumination pattern to improve the automated detection of the malaria infected blood cells. This same approach has been used to optimize illumination for other tasks, and has been extended to include the frequency response of an imaging system’s optics. Examples include optimizing the illumination and pupil function jointly for image classification ([28]), optimizing illumination to capture better images (phase contrast imaging [29, 30] and fourier ptychography [31, 32, 33]), and optimizing an imaging system’s optics for specific tasks (depth-of-field [34], depth detection [35], dynamic range [36], localization microscopy [22]).

In terms of using a microscope’s illumination to achieve new functionalities, prior work has clearly shown the benefits of applying programmable LED array illumination.

This includes variable bright-field and dark-field imaging [37], measurement of a specimen’s surface gradient [29, 30], and quantitative phase imaging [38] to name a few. Mathematically rigorous methods have also been used to combine variably-illuminated images to increase image resolution. Two prominent examples are Fourier ptychographic microscopy (FPM) [39, 31] and structured illumination microscopy (SIM) [40]. These works highlight the benefit of controlling illumination and provide ample evidence that it should be targeted for optimization.

3. Methods

3.1. Image Formation

In this work, we used a programmable LED array to control the spectral and angular qualities of our illumination by changing the brightness and color of each LED. We configured our setup such that the illumination from each LED can be modelled by a plane wave at the sample, with each LED contributing a wave which propagates at an angle corresponding to its position within the array.

When working with biological specimens which are primarily transparent, this multi-angle/multi-spectral illumination strategy yields diverse sample-specific image information that would be unavailable under standard on-axis illumination (illuminating with the center LED directly beneath the sample) [39]. However, this illumination strategy creates many degrees of freedom within an illumination pattern, making it challenging to directly determine which pattern is best for a specific task and specimen type. Here, we demonstrate a mechanism for jointly optimizing both the pattern that is used to illuminate the sample and the post-processing that maps the acquired image data into an accurate map of fluorescence emission.

To solve this joint optimization problem, we use a method called “Learned Sensing”, first introduced in Muthumbi et al. [27], which encapsulates the physical parameters of the imaging hardware (in our case, the LED brightness values) within a differentiable “physical layer”. We prepended the physical layer to a deep neural network (DNN) that has been designed to perform image labelling, making the entire process end-to-end trainable. In this work we further extend the initial concept of Learned Sensing to include multiple physical parameters that we can optimize simultaneously, to investigate the effect of capturing multiple unlabelled images per-sample (with each image taken under a different illumination pattern).

In this type of approach, the supervised training process optimizes both the physical parameters of our hardware (LED brightness and color values) and the parameters of our DNN. Therefore, while the DNN is being trained to perform virtual fluorescence (map unlabelled images to fluorescence maps) the illumination patterns are being op-

timized to aggregate an ideal set of unlabelled image data. After network training, we translate the parameters of our DNN physical layer (optimized linear weights) into a specific hardware configuration that can be realized experimentally (LED brightness and color values for multiple presented patterns).

The composition of the “physical layer” in this work is relatively straightforward. Since light from each LED is mutually incoherent, the image formed by an arbitrary illumination pattern is simply a weighted sum of the images formed by each LED individually. If we denote each individually illuminated image as I_n , its corresponding weight as w_n , and the final formed image as I' , then the image formation process can be expressed as:

$$I' = \sum_{n=1}^N w_n I_n \quad (1)$$

To perform our experiments, we collect this set of N raw individual images (I_n) for each specimen under study (we call this collection our k -stack), and use Eq. 1 to define how the parameters of our physical layer interact with the collected data to create the final image I' . To form multiple final images, we simply repeat the process within Eq. 1, using the same basis images (I_n for $n = 1, \dots, N$) in combination with differing weight vectors. To form M images for example, we learn a matrix $\mathbf{w} \in \mathcal{R}^{M \times N}$, where Eq. 1 is applied for every row m to create multiple formed images I_1, \dots, I_M . We provide a detailed derivation of Eq. 1 within the supplement.

3.2. Network Design

Our network consists of a modified U-Net architecture that is prepended with a physical layer that allows us to model Eq. 1 (Figure 3). The architecture and layer configuration was consistent across all reported experiments and is detailed in the supplement. During our multi-pattern experiments we constrained the weight matrix ($\mathbf{w} \in \mathcal{R}^{M \times N}$) within the physical layer to be positive only, allowing each row, m , of the matrix to be physically realizable with a single illumination pattern/captured image. However, when comparing our method to conventional illumination strategies we lifted this positivity constraint, since some of the competing strategies we considered require the subtraction of two images (such as differential phase contrast, DPC). Negative weights are easily realized experimentally by simply taking two images, one with the positive weights, and one with the negative, and taking the difference. Our rational for applying a positivity constraint during our multi-pattern experiment is that positive-only weights offer a larger degree of freedom for the optimizer when under a finite image constraint (one unconstrained pattern requires the same amount of images as two positive-only patterns)

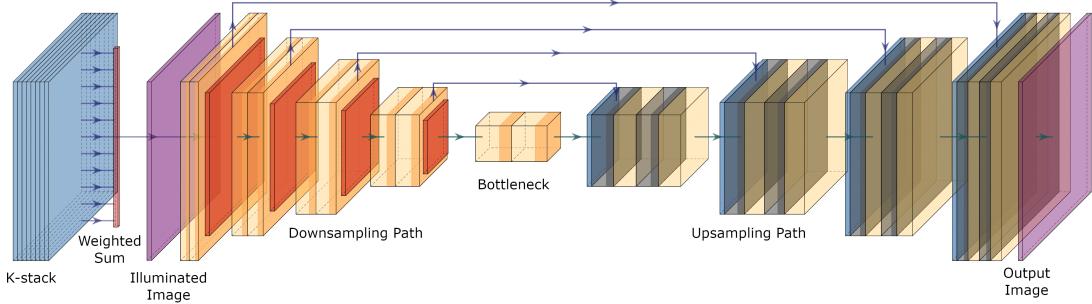


Figure 3: Neural network architecture used in this work. The “K-Stack” in blue represents the set of images of a specimen collected under illumination from each individual LED. The K-Stack is transformed into an illuminated image of the specimen, using a linear-weighting of each channel of the K-Stack (marked as “weighted sum” in red), after which the image is processed with a standard U-Net configured for regression.

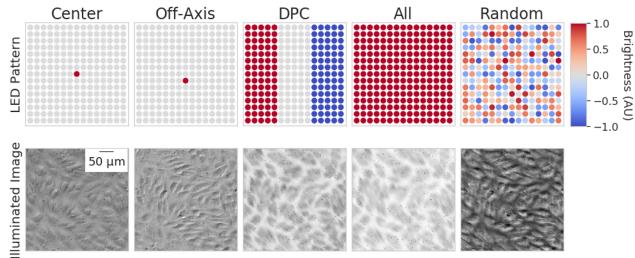


Figure 4: Visual comparison of standard LED patterns tested in this work, and resultant illuminated images.

and thus is a better gauge of how well additional information can be utilized.

Although the task of optimizing a task-specific illumination pattern has been studied before ([27, 28]), such previously developed models cannot execute image-to-image translation. Accordingly, our neural network here varies significantly in several important ways. First, we adopted a much deeper fully convolutional neural network (see Fig. 3). Second, our architecture uses *Skip Connections*, which enabling shorter paths for gradient calculation and preservation of high resolution feature maps generated early in the model. This had the added benefit of providing a more direct route between loss and the physical layer parameterization, without which we found successful convergence impossible. Finally, and most substantially, we modified the design of the physical layer to support *multi-pattern optimization*, which allows us to jointly optimize multiple sets of physical parameters in a controlled manner.

Although our underlying component LED images (slices of the *k-stack*) accurately capture the signal-to-noise ratio (SNR) of our imaging system, the “illuminated image” created by the physical layer will have an artificially inflated SNR (e.g., averaging N images produces a \sqrt{N} improvement in SNR). To compensate for this, we introduced a

Noise Layer, which adds dynamically generated Gaussian random noise to the data after the physical layer in both training and inference, and is modelled as: $I'' = \mathcal{N}(\mu = I', \sigma^2 = \psi \times |I'|)$. Where a hyperparameter ψ controls the scale of the noise in proportion to the pixel intensity. Note that the variance of the random noise is proportional to the image pixel intensity itself, and thus is consistent with a Poisson noise model.

Finally, we applied *L1* regularization to the weights of the physical layer, w . The regularization term was given a small weight proportional to the magnitude of the loss to drive weights to zero, if and only if they are not significantly contributing to the task performance. The addition of this type of penalty reduces variance across random seeds and aids in interpretation of the resulting LED patterns by suppressing elements which do not contribute to enhanced performance.

4. Experiments

In the included experiments, we used an inexpensive 15×15 programmable multi-color LED matrix (*Adafruit* product ID 607) to illuminate each specimen, where each addressable LED included three spectral channels (with center wavelengths of $\lambda = 480, 540$, and 632nm), representing a total of $15 \times 15 \times 3 = 675$ independent illumination sources. For each experiment, we constructed a *k-stack* of images by turning on each LED individually, capturing an image, and stacking the results into a single 3D matrix. To generate target labels, each sample was illuminated with a blue (470nm) fluorescent excitation source and captured using the same optical imaging setup used for *k-stack* generator. By using the same optical setup for each field-of-view (FOV), we ensured that sample positioning and lens distortion remained constant from source (*k-stack*) to target (fluorescent image).

After data capture, we processed each *k-*

stack/fluorescent image pair by splitting them into tiles. Each tile was 256×256 px for both experiments, and the tiling process was done such that there was no overlap between adjacent tiles. Our neural network model was trained to predict the fluorescent label from the unlabelled *k-stack* (represented by $256 \times 256 \times 675$ matrix). The training process jointly optimizes multiple illumination patterns, which control how the input images are formed, and the neural network which processes those images. We denote the illumination portion of our model as the “physical layer”, which produces M unique LED patterns, encapsulated within our weight matrix $\mathbf{w} \in \mathcal{R}^{M \times 675}$.

To compare our learned illumination method to standard, conventional, techniques we followed the same training process with a fixed \mathbf{w} , where the values of \mathbf{w} correspond to the LED brightness values of the target illumination pattern. We selected several standard patterns which are common in microscopy [41]; a qualitative descriptive of these patterns is provided below, and visualized for a sample FOV in Figure 4.

1. Center is spatially coherent illumination from the center LED and three colors
2. All is spatially incoherent illumination from all 225 LEDs and three colors
3. DPC stands for the differential phase contrast illumination method
4. Off-axis is from an LED located 4 mm off the optical axis, illuminating at approximately 3°
5. Random is a random set of LED brightness values

Generally speaking, both *DPC* and *Off-axis* tend to highlight edges within the sample, while *Center* and *All* provide more typical on-axis illumination. To provide a fair comparison between our *learned* method, and these conventional methods we optimized a weight vector which corresponds to a single illumination pattern, where $\mathbf{w} \in \mathcal{R}^{1 \times 675}$, however we allowed the values of \mathbf{w} to be negative, thus maintaining the same restrictions as the conventional methods.

To test our approach we captured multiple unique FOVs per biological specimen across two different fluorescently labelled sample types. While the general approach used across the two types of samples was consistent, some details, such as stain used, and magnification, differ. The details for each experiment are as follows.

HeLa Task: In our first experiment we used data captured in a prior experiment (see [42]), which imaged specimens of 90% confluent HeLa cells stained with a fluorescent nuclear label (DAPI). The employed microscope used a two-lens arrangement with a $f = 200$ mm tube lens (Thorlabs ITL200), and a $f = 50$ mm Nikon lens ($f/1.8D$ AF Nikkor). The setup had a collection numerical aperture (NA) of 0.085 and a magnification of 3.87x. The sample was imaged with a Prosilica GX6600 monochromatic camera (pixel size of $5.5\mu\text{m}$). The LED array was placed 80mm

beneath the sample and *k-stack* was generated by scanning through the inner 15×15 elements. We divided our full-FOV images into 442 non-overlapping 256×256 px tiles (with each tile having a 675 *k-stack* and matching fluorescent image). We randomly split these tiles into training, validation, and test sets containing 356, 48, and 48 tiles respectively.

Pan Task: For our second experiment we captured images of *Pan 16* pancreatic cancer cells stained with Cell-Mask Green plasma membrane stain (C37608). The sample was imaged using an Olympus PlanN 0.25NA , 10x, objective lens a Basler Ace (acA4024-29um) sensor. To capture the fluorescent label a set of filters were inserted into the optical setup (ThorLabs MDF-GFP excitation, emission, and dichroic filters), a photo of the setup used is displayed in Figure 1a. As with the previous experiment, a programmable LED array was placed beneath the sample of which the inner 15×15 elements were used. We captured 19 unique FOVs from the sample, which were split randomly into train, validation, and test sets (15, 2, and 2 FOVs respectively). Each image was broken into non-overlapping tiles, with the *k-stacks* and fluorescent labels having spatial dimensions of 256×256 px, with the train/validation/test sets having a total of 818, 108, and 108 tiles accordingly.

5. Results and Discussion

5.1. Performance

Across both experimental tasks we found that our physical layer’s learned patterns outperformed all tested alternatives. We evaluated the performance of our system both through root mean squared error (RMSE) and structural similarity index (SSIM [43]). Numerical results are provided in Table 1. This result demonstrates that under the same information restrictions (limited to illumination patterns of both negative and positive weights), learned illumination can lead to superior virtual fluorescence across multiple specimen types. To further demonstrate the superiority of our learned illumination strategy we note that even when under positive-only constraints ($\mathbf{w} \in [0, \infty)$) a single learnable illumination pattern out-performs conventional strategies which require two image to physically realize (*DPC* and *random*) for the HeLa task. While naturally there are additional illumination patterns which could be compared to our approach, we believe that our results as well as the mathematical framework described provide sufficient evidence to claim that a learnable illumination strategy is flexible enough to converge to any arbitrary pattern, given that the pattern is optimal for a specific task.

We next investigated the performance of our multi-pattern experiments. We varied the number of illumination patterns that are being optimized (the size of the weight matrix \mathbf{w}), ranging from a single pattern ($\mathbf{w} \in \mathcal{R}^{1 \times 675}$) to

Table 1: Performance across both tasks of conventional and learnable illumination strategies. The *Learned* configuration (bolded) offers the lowest average MSE and highest average SSIM for both tasks. All patterns are realized with an unconstrained representation of the illumination patterns, such that $\mathbf{w} \in (-\infty, \infty)$

Task	Center	All	DPC	Off-Axis	Random	Learned
HeLa-RMSE	.232 ± .0004	.226 ± .002	.144 ± .003	.232 ± .0004	.199 ± .002	.118 ± 0.001
HeLa-SSIM	.388 ± .001	.435 ± .006	.659 ± .006	.388 ± .001	.505 ± .075	.784 ± 0.009
PAN-RMSE	.144 ± .002	.178 ± .001	.126 ± .007	.144 ± .004	.152 ± .152	.122 ± 0.02
PAN-SSIM	.501 ± .001	.366 ± .001	.540 ± .007	.515 ± .007	.476 ± .020	.617 ± 0.006

Table 2: Performance across tasks of learnable illumination strategies using a constrained illumination pattern ($\mathbf{w} \in [0, \infty)$). Numerical results are reported as the mean and variance of three independent runs. The best results for each task/category are in bold, when equivalent multiple configurations are in bold.

Task	1	2	4	8	16	32	64
HeLa-RMSE	.119 ± .001	.092 ± .001	.084 ± .003	.079 ± .003	.075 ± .001	.074 ± .002	.071 ± 0.002
HeLa-SSIM	.778 ± .005	.866 ± .003	.881 ± .012	.897 ± .007	.904 ± .004	.908 ± .006	.914 ± 0.003
PAN-RMSE	.132 ± .011	.104 ± .004	.107 ± .017	.102 ± .013	.111 ± .005	.098 ± 0.006	.101 ± 0.015
PAN-SSIM	.568 ± .036	.639 ± .006	.641 ± .016	.654 ± 0.010	.635 ± .008	.653 ± 0.004	.645 ± 0.013

sixty-four patterns ($\mathbf{w} \in \mathcal{R}^{64 \times 675}$). The results are reported in Table 2 and example patterns are in Figure 6. Overall, we see that as more information is sampled from the *k-stack*, the system performance improves. However, there are diminishing returns, and it appears that the majority of the improvement is obtained in moving from 1 → 16 patterns.

Example predicted fluorescence maps for two unique FOVs for the HeLa and PAN tasks, using both the minimum number of patterns (1) and the maximum (64), are shown in Figure 5. From these predictions, we can see that the main difference between the results as the number of patterns increase is the detail (i.e., high spatial frequencies) in the generated fluorescence result. This is particularly evident for the 64 pattern configuration on the PAN task, where the image has crisper edges and a cleaner background. The shape of the generated fluorescence within the *HeLa* task prediction is also much closer to the ground truth for the 64 pattern configuration than the 1 pattern configuration.

While it is difficult to directly compare the trends in performance across tasks (due to the large difference in what is being labelled), the greater improvement observed within the *HeLa* task is nevertheless of interest. We hypothesize that this difference in improvement is at least in part due the *HeLa* setup’s use of an objective lens with a lower NA (i.e., lower maximum acceptance angle of light). Briefly, the objective lens NA characterizes the imaging system transfer function and can be conceptualized as a low-pass filter, with the spatial frequency cut-off proportional to the NA value. A lower NA corresponds to a lower frequency cut-off and a lower image resolution. By imaging at lower resolution, the impact of illumination optimization in transferring otherwise unobserved spatial frequency content to the imaging

plane leads to more effective *HeLa* task performance. This doesn’t mean that lenses with higher NA (such as the one used in our *PAN* experiment) prevent our technique from being useful (our results provide evidence to the contrary). However, it is likely that a different LED array (perhaps illuminating at higher angles, or through a condenser) would be more effective for optical setups with a high imaging NA.

5.2. Illumination Pattern Composition

In addition to improved performance, the DNN-optimized LED illumination patterns also offer several interesting physical insights. Illumination patterns for a select number of multi-pattern configurations are displayed within Figure 6. For example within the *HeLa* task, we see a mix of bright-field (within the red outline) and dark-field illumination, with some patterns being a mix, and others nearly exclusively bright-field. This likely means that the contrast provided by dark-field illumination (primarily sample scattering) is useful for task-specific biological structures of interest, but only when mixed with image data from the bright-field channel (primarily sample absorption).

Another interesting feature of our results is the diversity between patterns, or relative lack thereof. Using the same LED across patterns (within the same configuration) may seem to be yield redundant image data. Considering the presence of noise, however, it appears more sensible. Areas with weaker signal (which tend to be in the dark-field) exhibit higher noise and are thus less likely to be sampled in a single pattern/image. Oversampling can reduce the impact of noise to better highlight key features of interest, and it appears like the DNN identified this as an effective strategy.

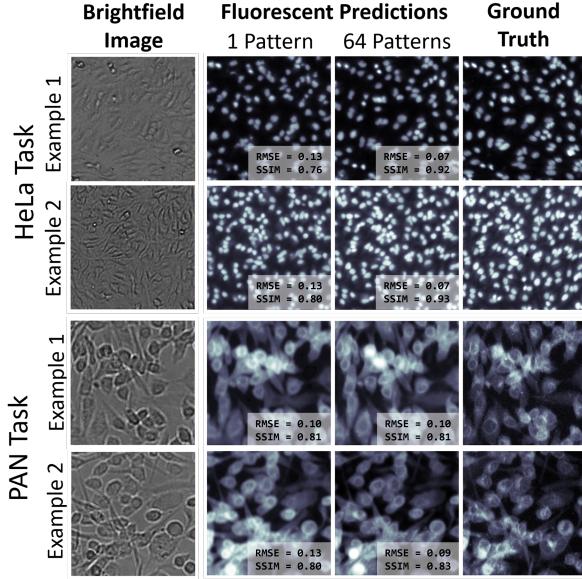


Figure 5: Example predictions across two distinct FOVs within the test set. The leftmost column is an example bright field image, where the two inner columns are example virtual fluorescence predictions for the 1-pattern and 64-pattern optimization cases. The rightmost column contains the true fluorescent image.

6. Conclusion

In summary, we have presented a novel method for jointly optimizing the illumination of an imaging system for deep learning-based fluorescence image inference. We focused our attention on microscope imaging with a programmable LED array, and determining an optimized set of brightnesses and colors to display on the array during image capture, which provides a simple and effective means to improve current setups. The two experiments we performed show that this technique is robust in its improvement, with the RMSE of fluorescence image reconstruction for each specimen type being minimized by the “learned” or jointly optimized approach. We believe our results support this kind of approach not only for virtual fluorescence, but other label-free imaging techniques (such as virtual staining). Furthermore, our experiments demonstrate that this technique scales well to optimize for multiple illumination patterns simultaneously. Our results showed that additional illumination patterns can be effectively utilized to improve task performance with diminishing returns.

While many physical components of an imaging system may be optimized in this manner to benefit inference performance, our results highlight the ease of use and simplicity of optimized illumination patterns. Requiring only an inexpensive LED array placed underneath the sample to experimentally realize, this particular strategy is generally

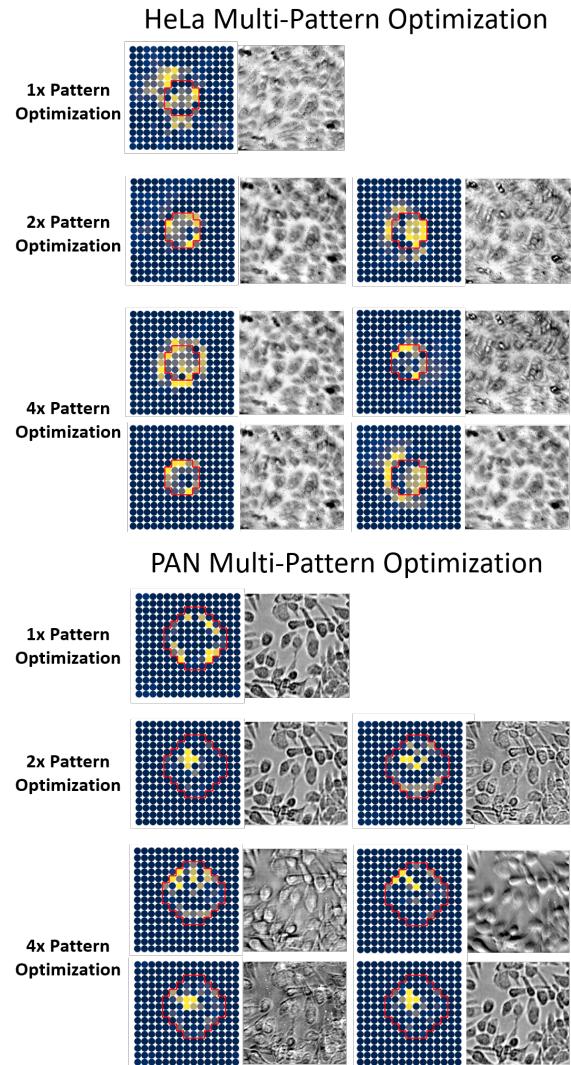


Figure 6: Selection of optimized illumination patterns and their corresponding formed images for both tasks, only a single spectral channel (green) is shown for brevity. The red outline within the pattern indicates the separation between the bright and dark-field areas of the LED array. As the number of patterns being optimized increases we observe a larger variance in pattern composition as well as a tendency to create bright/dark-field contrast.

less expensive than creating custom-fabricated optical elements. It is also dynamically controllable and can allow users to dial up specimen-dependent optimized illumination patterns, and explore the trade-off between data acquisition time (number of patterns) and desired performance. We hope our results continue to motivate the imaging and machine learning community to re-examine how they capture data and continue to develop understanding of the connection between data capture and data processing.

References

- [1] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, “Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments,” *PLoS computational biology*, vol. 12, no. 11, 2016. 1
- [2] J. Xu, C. Zhou, B. Lang, and Q. Liu, “Deep learning for histopathological image analysis: Towards computerized diagnosis on cancers,” in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pp. 73–95, Springer, 2017. 1
- [3] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O’Neil, K. Shah, A. K. Lee, et al., “In silico labeling: predicting fluorescent labels in unlabeled images,” *Cell*, vol. 173, no. 3, pp. 792–803, 2018. 1, 2, 3
- [4] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy,” *Nature methods*, vol. 15, no. 11, p. 917, 2018. 2, 3
- [5] S. Cheng, S. Fu, Y. M. Kim, W. Song, Y. Li, Y. Xue, J. Yi, and L. Tian, “Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy,” *Science advances*, vol. 7, no. 3, p. eabe0431, 2021. 2, 3
- [6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017. 2
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015. 2
- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015. 2
- [9] C. Belthangady and L. A. Royer, “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction,” *Nature methods*, pp. 1–11, 2019. 2
- [10] S.-M. Guo, L.-H. Yeh, J. Folkesson, I. E. Ivanov, A. P. Krishnan, M. G. Keefe, E. Hashemi, D. Shin, B. B. Chhun, N. H. Cho, et al., “Revealing architectural order with quantitative label-free imaging and deep learning,” *Elife*, vol. 9, p. e55502, 2020. 3
- [11] Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydin, J. E. Zuckerman, T. Chong, A. E. Sisk, et al., “Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning,” *Nature biomedical engineering*, vol. 3, no. 6, p. 466, 2019. 3
- [12] Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, and A. Ozcan, “Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue,” *Light: Science & Applications*, vol. 9, no. 1, pp. 1–13, 2020. 3
- [13] A. Rana, G. Yaune, A. Lowe, and P. Shah, “Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 828–834, IEEE, 2018. 3
- [14] M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broadbust, S. Culley, et al., “Content-aware image restoration: pushing the limits of fluorescence microscopy,” *Nature methods*, vol. 15, no. 12, pp. 1090–1097, 2018. 3
- [15] M. Weigert, L. Royer, F. Jug, and G. Myers, “Isotropic reconstruction of 3d fluorescence microscopy images using convolutional neural networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 126–134, Springer, 2017. 3
- [16] J. Macdonald, A. J. Breese, and N. L. Hanbury, “Optimization of a lens design using a neural network,” in *Applications of Artificial Neural Networks IV*, vol. 1965, pp. 431–442, International Society for Optics and Photonics, 1993. 3
- [17] A. Chakrabarti, “Learning sensor multiplexing design through back-propagation,” in *Advances in Neural Information Processing Systems*, pp. 3081–3089, 2016. 3
- [18] S. Diamond, V. Sitzmann, S. Boyd, G. Wetzstein, and F. Heide, “Dirty pixels: Optimizing image classification architectures for raw sensor data,” *arXiv preprint arXiv:1701.06487*, 2017. 3

- [19] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, “Reconnet: Non-iterative reconstruction of images from compressively sensed measurements,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 449–458, 2016. [3](#)
- [20] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging,” *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–13, 2018. [3](#)
- [21] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, “Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification,” *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018. [3](#)
- [22] E. Hershko, L. E. Weiss, T. Michaeli, and Y. Shechtman, “Multicolor localization microscopy and point-spread-function engineering by deep learning,” *Optics express*, vol. 27, no. 5, pp. 6158–6183, 2019. [3](#)
- [23] Y. Xue, S. Cheng, Y. Li, and L. Tian, “Reliable deep-learning-based phase imaging with uncertainty quantification,” *Optica*, vol. 6, no. 5, pp. 618–629, 2019. [3](#)
- [24] P. del Hougne, M. F. Imani, A. V. Diebold, R. Horstmeyer, and D. R. Smith, “Learned integrated sensing pipeline: Reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network,” *Advanced Science*, vol. 7, no. 3, p. 1901913, 2020. [3](#)
- [25] H. Sun, A. V. Dalca, and K. L. Bouman, “Learning a probabilistic strategy for computational imaging sensor selection,” *arXiv preprint arXiv:2003.10424*, 2020. [3](#)
- [26] G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” *Optica*, vol. 6, no. 8, pp. 921–943, 2019. [3](#)
- [27] A. Muthumbi, A. Chaware, K. Kim, K. C. Zhou, P. C. Konda, R. Chen, B. Judkewitz, A. Erdmann, B. Kappes, and R. Horstmeyer, “Learned sensing: jointly optimized microscope hardware for accurate image classification,” *Biomedical Optics Express*, vol. 10, no. 12, pp. 6351–6369, 2019. [3, 4, 5](#)
- [28] K. Kim, P. C. Konda, C. L. Cooke, R. Appel, and R. Horstmeyer, “Multi-element microscope optimization by a learned sensing network with composite physical layers,” *Optics Letters*, vol. 45, no. 20, pp. 5684–5687, 2020. [3, 5](#)
- [29] M. Kellman, E. Bostan, N. Repina, and L. Waller, “Physics-based learned design: Optimized coded-illumination for quantitative phase imaging,” *IEEE Transactions on Computational Imaging*, 2019. [3, 4](#)
- [30] B. Diederich, R. Wartmann, H. Schadwinkel, and R. Heintzmann, “Using machine-learning to optimize phase contrast in a low-cost cellphone microscope,” *PLoS one*, vol. 13, no. 3, p. e0192937, 2018. [3, 4](#)
- [31] M. Kellman, E. Bostan, M. Chen, and L. Waller, “Data-driven design for fourier ptychographic microscopy,” in *2019 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–8, IEEE, 2019. [3, 4](#)
- [32] S. Jiang, K. Guo, J. Liao, and G. Zheng, “Solving fourier ptychographic imaging problems via neural network modeling and tensorflow,” *Biomedical optics express*, vol. 9, no. 7, pp. 3306–3319, 2018. [3](#)
- [33] Y. F. Cheng, M. Strachan, Z. Weiss, M. Deb, D. Carone, and V. Ganapati, “Illumination pattern design with deep learning for single-shot fourier ptychographic microscopy,” *Optics express*, vol. 27, no. 2, pp. 644–656, 2019. [3](#)
- [34] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging,” *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 114, 2018. [3](#)
- [35] J. Chang and G. Wetzstein, “Deep optics for monocular depth estimation and 3d object detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10193–10202, 2019. [3](#)
- [36] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, “Deep optics for single-shot high-dynamic-range imaging,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1375–1385, 2020. [3](#)
- [37] G. Zheng, C. Kolner, and C. Yang, “Microscopy refocusing and dark-field imaging by using a simple led array,” *Optics letters*, vol. 36, no. 20, pp. 3987–3989, 2011. [4](#)
- [38] X. Ou, R. Horstmeyer, C. Yang, and G. Zheng, “Quantitative phase imaging via fourier ptychographic microscopy,” *Optics letters*, vol. 38, no. 22, pp. 4845–4848, 2013. [4](#)
- [39] G. Zheng, R. Horstmeyer, and C. Yang, “Wide-field, high-resolution fourier ptychographic microscopy,” *Nature photonics*, vol. 7, no. 9, p. 739, 2013. [4](#)

- [40] M. G. Gustafsson, “Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy,” *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, 2000. [4](#)
- [41] L. Tian and L. Waller, “Quantitative differential phase contrast imaging in an led array microscope,” *Optics express*, vol. 23, no. 9, pp. 11394–11403, 2015. [6](#)
- [42] J. Chung, J. Kim, X. Ou, R. Horstmeyer, and C. Yang, “Wide field-of-view fluorescence image deconvolution with aberration-estimation from fourier ptychography,” *Biomedical optics express*, vol. 7, no. 2, pp. 352–368, 2016. [6](#)
- [43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004. [6](#)