Mitotic index determination on live cells from label-free acquired quantitative phase images using a supervised autoencoder

SCHOLARONE™
Manuscripts

Mitotic index determination on live cells from label-free acquired quantitative phase images using a supervised autoencoder

Philippe Pognonec, Axel Gustovic, Zied Djabari, Thierry Pourcher and Michel Barlaud Fellow, IEEE

Abstract—This interdisciplinary work focuses on the interest of a new auto-encoder for supervised classification of live cell populations growing in a thermostated imaging station and acquired by a Quantitative Phase Imaging (QPI) camera. This type of camera produces interferograms that have to be processed to extract features derived from quantitative linear retardance and birefringence measurements. QPI is performed on living populations without any manipulation or treatment of the cells. We use the efficient new autoencoder classification method instead of the classical Douglas-Rachford method. Using this new supervised autoencoder, we show that the accuracy of the classification of the cells present in the mitotic phase of the cell cycle is very high using QPI features. This is a very important finding since we demonstrate that it is now possible to very precisely follow cell growth in a non-invasive manner, without any bias. No dye or any kind of markers are necessary for this live monitoring. Any studies requiring analysis of cell growth or cellular response to any treatment could benefit from this new approach by simply monitoring the proportion of cells entering mitosis in the studied cell population.

Index Terms—Mitotic index, non-invasive cell monitoring, Quantitative Phase Imaging (QPI), Supervised autoencoder.

✦

1 INTRODUCTION

Somatic cells constitute the vast majority of our body. These cells may have to multiply to allow our organism to maintain normal homeostasis. Growth occurs through a very tightly regulated process called the cell cycle. During this process, a cell ultimately divides into two identical cells within one or two days. Deregulation of the cell cycle may result in cell death [1] or in uncontrolled cell growth leading to cancer [2]. In addition, exposure to toxic compounds may also result in alteration to cell growth or cell death [3].

It is thus of utmost importance for biologists to monitor cell growth. This is usually done either with molecular tools on cell lysates, or via flow cytometry performed on isolated cells placed in suspension and processed with a FACS machine (Fluorescence-activated cell sorting) [4]. While both methods have advantages, they both interfere with the cells being monitored. Molecular analysis requires cell killing to obtain a workable lysate that is used to quantify molecular markers of the cell cycle. Flow cytometry, while being able under specific conditions to keep cells alive, will strongly interfere with the physiology of cells due to the extensive manipulations inherent to the FACS procedure [5].

Quantitative Phase Imaging (QPI) is a well established imaging technology that only requires short bursts of low intensity white light that are relatively harmless to living cells compared to fluorescence. QPI cameras produce phase images that are processed to extract biologically significant features. The optical phase differences measured after light propagates through biological samples are a direct reflection of the dry matter encountered along the optical path. Many reviews focus on this technique [6], [7], [8], [9], [10] [11],

[12], [13], [14]. The QPI strategy used here does not require temporally coherent illumination and can thus be used with conventional microscope sources, such as halogen lamps [15]. It accurately measures increases in the growing cell mass [16]. We previously showed that these and other features derived from QPI, associated with proximal machine learning methods, accurately recognize HeLa cells experimentally blocked in early S and M phases [17] following thymidine and nocodazole treatments, respectively. The S and M phases of the cell cycle correspond to the beginning of chromosome duplication (DNA synthesis) and to the onset of cell division (mitosis), respectively.

Stemming from this observation, we reasoned that we should be able to perform mathematical testing on normal (untreated) exponentially growing cell populations to characterize cells undergoing mitosis. This was performed with a supervised deep neural autoencoder. As presented in this work, our algorithms were able to identify cells with a very high accuracy from QPI-derived features. To biologically validate these computed results, we compared the phenotypes of the cells present in the predicted M class (Mitosis) to images of the same identical cells obtained with a regular CMOS camera in the presence of a permeant dye emitting light under fluorescent illumination. This dye specifically labels cellular nucleic acids (mainly DNA present in chromosomes). We could thus validate the coherence of the computed classification.

Deep neural networks have proven their effectiveness in bioinformatics for both classification and feature selection [18], [19]. Autoencoders were introduced into the field of neural networks decades ago and their most efficient application was dimensionality reduction [20], [21]. Stacked autoencoders were successfully used for denoising [22], and fault classification [23].

An autoencoder is a discriminative model which maps feature points to a high dimensional space to labels in a low dimensional latent space. Let us recall that variational autoencoders (VAEs) [24], [25] encourage the latent space to fit a prior distribution, like a Gaussian. These classical priors in the latent space are chosen for their computational simplicity rather than their compatibility with the latent structure and thus can lead to inaccurate latent low-dimensional representations of data.

Recently Barlaud and Guyard proposed a non-parametric supervised Autoencoder framework [26] where they relaxed the parametric distribution assumption in the latent space to learn a non-parametric data distribution of clusters. This supervised network encourages the latent space to fit a distribution learned with the clustering labels rather than a parametric prior distribution.

In this paper, we point out the following specific contributions:

- We estimate a new Mitosis reference cells using a hand made microscopy approach without any invasive chemical treatment.
- We use the efficient new autoencoder classification [26] method instead of the classical Douglas-Rachford method [17] [27].

The paper is organized as follows: Section 2 deals with live cell cycle monitoring. In section 3 we develop the autoencoder classification method. In section 4 experimental results are provided on real biological datasets with a discussion in section 5. Finally in section 6 we conclude the paper.

2 LIVE CELL CYCLE MONITORING

The cell cycle is a biological process ultimately resulting in cell division, yielding two daughter cells with exactly the same genetic content. This highly regulated process is of the utmost importance and is a direct reflection of cell growth. It consists in four phases:

- The mitotic phase per se (M-Phase), in which nuclei loose their membrane, nucleoli disappear and chromosomes are physically sorted to be equally split among two identical daughter cells. These new cells then enter:
- The G1 phase, in which the cell nucleus is reorganized following cell division. The entry into the G1 phase requires the passage of a checkpoint ("Spindle checkpoint") where cells verify that division occurs as planned. If not, cells cannot proceed to G1. The G1 phase is associated with the formation of nucleoli, sub-nuclear structures that produce ribosomal RNAs required for the synthesis of proteins from transcribed DNA, as well as for the normal physiology of the cell. DNA transcription into RNA restarts to support cellular functions. Following the G1 phase and after another checkpoint validating their well being, the cells proceed to:
- the S phase, in which DNA is replicated. This phase is associated with an increase in cell size and mass. After chromosomal duplication, cells enter:

- the G2 phase, in which there is a last checkpoint before entering the M-Phase again.

Monitoring the cell cycle allows direct evaluation of cell growth, but no method is currently available to do so without interfering with normal cell growth. Fluorescence-Activated Cell Sorting (FACS) is very powerful for monitoring cell cycling, but requires non-physiological cell handling and/or cell labeling with exogenously expressed fluorescent proteins [5]. We investigated whether high resolution quadri-wave lateral shearing interferometry, that allows quantitative linear retardance and birefringence measurements on biological samples could allow non-invasive monitoring of the cell cycle in populations. The QPI strategy, presented in [16], [28], only requires short bursts of low intensity white light which appears to have a negligible effect on cell physiology. The optical phase difference measured when photons exit a biological sample is a direct reflection of the dry mass present in the sample being analyzed. It has been shown for example that the mass increase associated with cellular cell cycle progression (2N toward 4N chromosomes) can be accurately determined [1]. In addition, the characteristic cell rounding associated with the late G2 phase is also easily detected by QPI [28]. Biologists partners [17] have investigated how reliable this approach would be to monitor a live cell population under the microscope. They first setup a controlled system in which HeLa cells were synchronized into the early S phase (double thymidine block) and late G2/mitosis (nocodazole block) [16]. Cells were classically labeled with Hoechst 33342 to visualize nuclei by fluorescence imaging with the dye. Multiple fields were then acquired in parallel with a regular CMOS camera (Zyla 5.5) to visualize the blue Hoechst fluorescence, and with a SID4Bio camera to acquire quantitative phase images. Typical images of cells obtained by Fluorescent labeling and Phase imaging are presented in Figure 1 and 2.
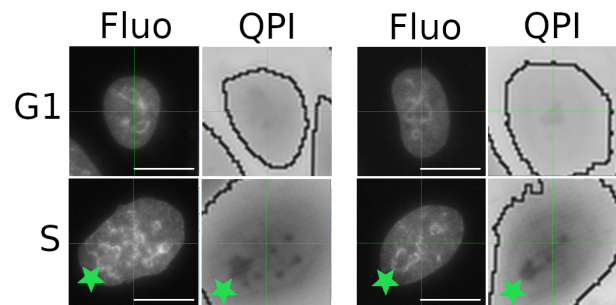

Fig. 1. "Fluorescence versus Phase imaging, G1 and S phases in HeLa Cells". G1 (top row): 2 representative cells before DNA synthesis are shown as seen by Fluorescence ("Fluo", left) and Phase ("QPI", right) imaging. S (bottom row) illustrates cells undergoing DNA synthesis. The cells are larger, and the nucleoli are more visible (pointed out with green stars) with Fluorescence labeling and in Phase imaging. White scale bar: 10μm.

As shown on Figure 2, the phenotype of the cells undergoing mitosis is very characteristic. The left hand fluorescent image displays individual chromosomes blocked before equatorial plate formation because of nocodazole treatment, while the right hand fluorescent image illustrates a typical equatorial plate (pointed out by

green stars). This work demonstrated the robustness of the approach. However, classification relied on training sets obtained from chemical treatments of the cells. While useful to arrest cells in their cycle, these treatments do not provide phenotypically normal cells.
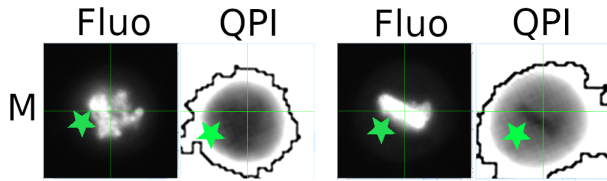

Fig. 2. "Fluorescence versus Phase imaging, M phase". Left pair: Cell blocked during mitosis following nocodazole treatment. The green stars point to chromosomes blocked in prophase, the first stage of mitosis. Chromosome visualization in Phase imaging is possible but difficult (contrast was enhanced) compared to Fluorescence imaging (where labeling is chromosome specific). However, the phase analysis of the image unambiguously indicates a bona fide mitotic cell. Right pair: Control cell in metaphase (second stage of mitosis), with densely packed chromosomes highly visible in Fluorescence imaging, equatorial plate also distinguishable in QPI (contrast was enhanced).

In this work, we focused on the classification of mitotic cells. These cells, or more precisely their percentage among the whole population, represent the so-called mitotic index [29]. It is a reliable reflection of the growth of a cell population. To avoid the bias of chemically arrested cells, training sets were obtained by supervised selection of cells undergoing mitosis. This strategy allowed us to monitor in a completely non-invasive manner the growth of a living cell population in real time by following the mitotic index. Since the cells are kept alive under a robotic imaging station allowing QPI time-lapse acquisitions, screening of the effect of pharmacological substances of interest becomes possible in the absence of any manipulation other than interferogram treatment and mathematical processing.

3 METHOD: NON-PARAMETRIC SUPERVISED AUTOENCODER FRAMEWORK

Let X be the dataset, as a $m \times d$ data matrix made of m line samples $x_1, \ldots, x_m$. Let $y_i = j, j \in [1...k]$ be the label, indicating that the sample x_i belongs to the j-th cluster. Projecting the data in the lower dimension latent space is crucial to be able to separate them accurately. In this paper we propose to use a deep neural network autoencoder framework.

Let us recall that the encoder (or discriminative part) of the autoencoder map features points of a high dimensional space to a low dimensional latent space in and that the decoder maps feature points of a low dimensional space to a high dimensional latent space.

Figure 3 depicts the main constituent blocks of our proposed approach. We have added to our autoencoder a "soft max" block to compute the classification loss.

Let Z, the latent space, $\widehat{X}$ the reconstructed data (Figure 3) and W the weights of the neural network.

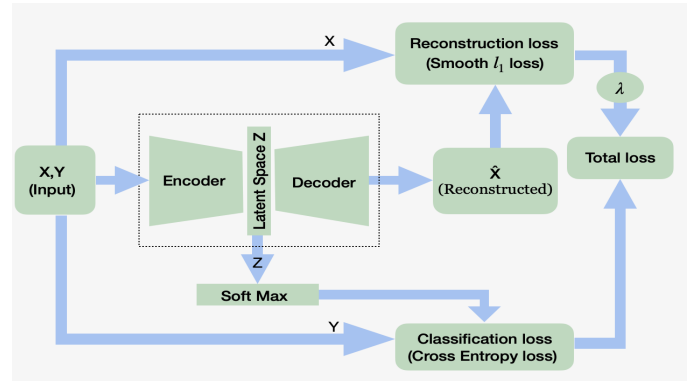The goal is to compute the weights W minimizing the

Fig. 3. Autoencoder framework

total loss which depends on both the classification loss and the reconstruction loss. Thus we propose to minimize the following criterion to compute the weights W of the autoencoder [26].

$$Loss(W) = \phi(Z, Y) + \lambda \psi(\widehat{X} - X) \text{ s.t. } \|W\|_1 \leq \eta. \quad (1)$$

Where $\phi(Z, Y)$ is the classification loss in the latent space and $\psi(\widehat{X} - X)$ is the reconstructed loss.

We also introduce a constrained regularization loss $\|W\|_1 \leq \eta$ for features selection [30], [31]. The parameter λ weights the classification loss and the reconstruction loss. We use the Cross Entropy Loss for the classification loss function. We use the robust Smooth ℓ_1 (Huber) Loss [32] as reconstruction loss function ψ. The size of the latent space is the number of clusters.

The authors of [33] proposed a double descent algorithm as follows: after training a network, set all weights smaller than some threshold to zero, rewind the rest of the weights to their initial configuration, and then retrain the network from this starting configuration but keeping the zero weights frozen (untrained). In order to minimize our global criterion 1, we use the modified double descent algorithm proposed by Barlaud and Guyard [34] where they replaced the thresholding by the ℓ_1 projection.

Algorithm 1 Projection on the ℓ_1 norm—$proj_{\ell_1}(V, \eta)$ is the projection on the ℓ_1-ball of radius η, $\nabla \phi(W, M_0)$ is the masked gradient with binary mask M_0, and f is the ADAM optimizer, γ is the learning rate

---

Input: $W*, \gamma, \eta$
for $n = 1, \ldots, N(epochs)$ **do**
    $V \leftarrow f(W, \gamma, \nabla \phi(W))$
end for
$W := proj_{\ell_1}((V, \eta)$
Output: W, M_0
Input: $W*$
for $n = 1, \ldots, N(epoch)$ **do**
    $W \leftarrow f(W, \gamma, \nabla \phi(W, M_0))$
end for
Output: W

In this paper, we used a linear fully connected network

(LFC) with an input layer of $d = 15$ neurons, one hidden layers (64 neurones) followed by a ReLU activation function and a latent layer of dimension k.

We have implemented our sparse learning method using a constraint approach in the PyTorch framework. The losses are averaged across observations for each mini-batch. We chose the ADAM optimizer [35], as the standard optimizer in PyTorch. We have used the default parameters of the Pytorch software or previous publications. However two parameters must be tuned. The parameter η is determined by the maximum accuracy after cross-validation. The "mini-batch" parameter is a compromise between the computational cost and accuracy. In our case the computational cost is not an issue and we have privileged the accuracy with a low value of "mini-batch". We used the Captum method [36] to compute the weights of features.

First we learn a neural network using the mitoses and G1/S reference cells and a part of the control dataset. Then we save the best learned network with a sense of accuracy. Finally we use this best learned network to predict the labels and probabilities on the control data.

4 EXPERIMENTAL RESULTS

4.1 Datasets

The following work was done on widely used HeLa cells, an immortalized cell line derived in 1953 from an epidermoid carcinoma of the cervix [37]. Hoechst 33342 labeling allows visualization of nuclei by fluorescence, and clearly detects in a proliferating population mitotic structures where chromosome condensation reaches a peak (see Figure 4, Panel A, arrows).

First, both visible (fluorescent labeling) and phase images were processed with software (CellProfiler for fluorescent images [38], and BioData (Phasics) for QPI) to extract multiple features stored as matrix (the X matrix defined in section 3), in which each line corresponds to a detected cell/nuclei and each column displays the values of a specific feature (diameter, phase value, ...) for all the cells.

In our experiment we have 15 features for QPI images: 8 features related to QPI Interferograms (Dry mass, Phase max, Phase min ...) and 7 features related to geometry of the cell (Surface, ...) [17]. Note that we obtain at least equivalent accuracy using 15 QPI features as when we used 65 fluorescent features [17]. Datasets were retrieved from cell segmented after interferogram processing. Each data set consists of one row per identified cell, and multiple columns corresponding to the multiple features calculated. These features have been described in [16].

Interferogram processing of the same field acquired via QPI produces the image shown in Panel B, where cells have been segmented. One can clearly see that the two mitotic cells visualized by nuclear staining correspond to cells with a high density signal, as pointed out by the green arrows. Double arrow labeled "2" points to a cell in which chromosomes have migrated to the just separated two daughter cells at the end of mitosis.

The data set were normalized as follows [39]: i) Log-transformation of the datas for the following main benefits: Reducing heteroscedasticity and thus the bias on regression, transforming multiplicative noise into additive noise and ii)
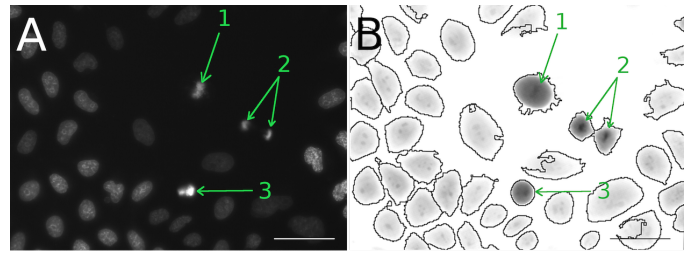

Fig. 4. "Fluorescence versus Phase imaging", Panel A: Fluorescence labeling of nuclei, Panel B: Phase imaging. The arrows point to cells in the M phase, clearly visualized with both fluorescence and phase imaging. The equatorial plate, corresponding to chromosome alignment during metaphase can be seen both with fluorescence and phase imaging (arrows 1 and 3). Double arrow "2" points to chromosomes migrating to the two daughter cells at the end of mitosis. Scale bar: 50μm.

Classical column zero mean.

In this paper, we provide experiments using three specific databases:

- The first dataset was presented in [17]. Briefly, HeLa cells were grown either in standard condition (Control cells, containing all phases of the cell cycle), or in the presence of chemicals to enriched the population in Mitosis phase (with nocodazole treatment) or in G1/early S phase (with double thymidine block). This dataset consist in 1163 G1/S reference cells and 9871 control cells (no chemical treatment).

- The second dataset was also built from HeLa cells. To avoid the bias resulting from mitotic nocodazole enrichment, we manually identified more than 150 cells entering mitosis and followed them with our imaging station, acquiring interferograms every 3 minutes until the end of mitosis, resulting in thousands of mitotic cell images. In parallel, we also acquired interferograms on the general HeLa population. This dataset consist in 1581 Mitotic reference cells and 84,411 control cells.

- In the third database, we use the 1581 Mitotic reference cells of the second database and 50,000 control cells from an acquisition performed on another day.

4.2 First experiment using chemical blocking

Initial results were published by the authors using a Douglas-Rachford classification method [17], [27]. Interestingly, we observed during this first study that the 15 QPI features used were at least as reliable as the 61 fluorescence features in term of classification accuracy. Mean accuracy Fluorescence = 89.24 and mean QPI accuracy = 91.86. This is why we now work exclusively with a QPI camera.

In the present paper, we use a non parametric supervised autoencoder [26]. The main benefits of the autoencoder is to improve the classification accuracy on the same dataset used in our original work (see Table 1) and to compute the latent space.

TABLE 1
Comparison of Accuracy of Mitosis and DNA detection using either
Douglas-Rachford or Autoencoder algorithm

	Mitosis	G1/S
Douglas Rachford Accuracy %	91.83	88.34
Autorencoder Accuracy %	98.17	95.8

TABLE 2
Number of detected Mitotic and G1/S cells

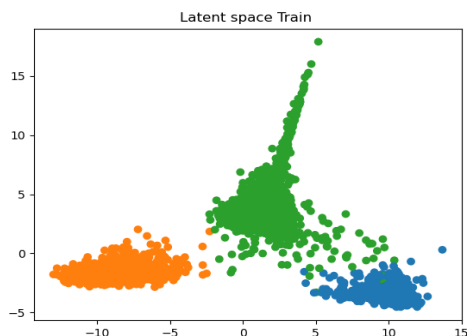| Training cells	600	650	700	750	800
Testing cells	10,426	10,394	10,305	10,320	10,303
Mitotic cells	422	332	350	342	332
Mitotic Index	4.04	3.19	3.39	3.31	3.12
G1/S Cells	86	102	70	67	67

Fig. 5. Latent space train on the first dataset using chemical datas selection: Mitosis (blue) and G1/S cells (orange).

specific point as shown on Figure 2. Indeed, nocodazole interferes with spindle formation which is required to bring individual chromosomes together to form the equatorial plate. The lack of kinetochores activates the spindle assembly checkpoint, and arrests the cycle in the prometaphase, that ultimately leads to apoptosis. Later stages of mitosis are thus not found in a nocodazole-blocked population: metaphase, anaphase, and telophase stages are missing. Thus training for mitosis on a nocodazole-blocked population will only identify early stages of mitosis. This is confirmed by the dotplot presented in Figure 7, using 2 QPI features (Cell surface and Phase average). Nocodazole arrested cells are in blue, and Control cells are in green. The plot shows both Control and Nocodazole treated cells and the Nocodazole cloud is located in the continuity of the small isolated cloud seen in Control cells. Phenotypic observation of these cells demonstrates that most of them are mitotic cells. Thus, it is clear that Nocodazole-treated cells do not exactly superimpose with Control mitotic cells. This is the reason for the lower than expected Mitotic index when using Nocodazole-treated cells as a training set.

Table 1 shows that the autoencoder outperforms Douglas-Rachford accuracy by 6.34% on the Mitosis, and by 7.46% on the G1/S cells.

Figure 5 of the train latent space in the training set demonstrates that Phase features allow the separation of the different populations according to their localization in the cell cycle. Interestingly, M phase cells (blue) form a characteristic sub-population, allowing for accurate calculation of the proportion of cells in mitosis, i.e. the determination of the proliferation status of the cell population monitored in Figure 6 of the test latent space .
This demonstrated that cells arrested in the M phase by nocodazole block or in the early S phase by double thymidine block, form homogeneous populations.

The Table 2 shows a lower than expected Mitotic index (3%). For these mitotic cells the expected index is usually close to 7% [40].
It should be kept in mind that the nocodazole-treated cells used for training arrests cells in the M phase at a very
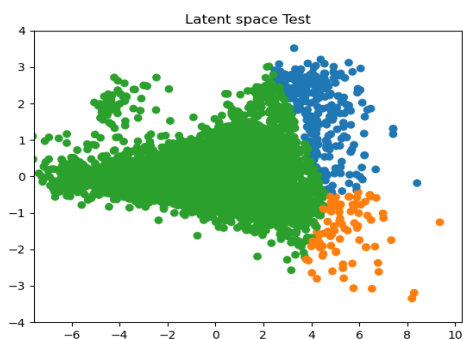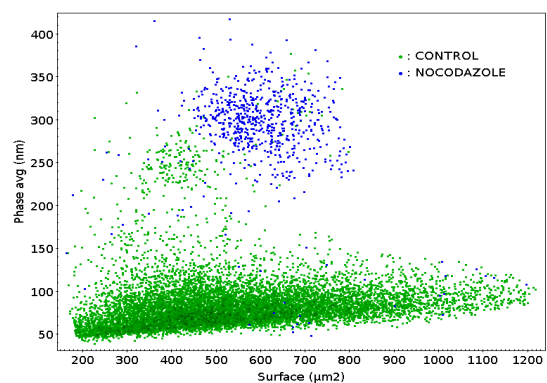

Fig. 7. "Mitotic versus Nocodazole-generated mitotic cells": Cell surface versus Phase average dotplot with both Control (green) and Nocodazole-treated cells (blue). The Nocodazole mitotic cloud is further out than the Control mitotic cloud.

As expected, some of the cells from the control population presented feature values similar to those of the 2 blocked populations, since there are cells in the early G1-S phase as well as cells in the M phase in the control population.
We thus proposed to visually select cells undergoing mitosis and to acquire their images every 3 minutes with time-lapse imaging using the QPI camera. This allowed us to build a robust bank of bona fide mitotic cells.
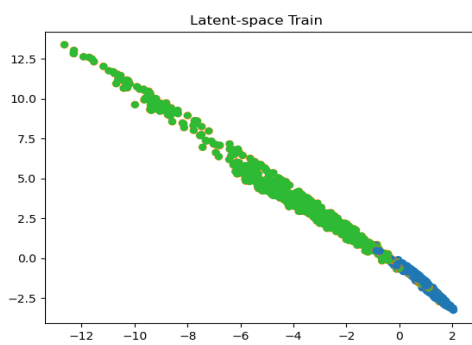

Fig. 6. Latent Space test using chemical datas selection: Mitosis (blue) and G1/S cells (orange)

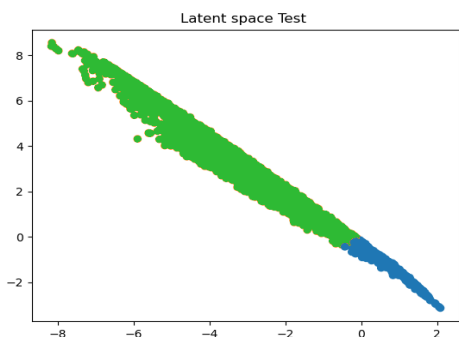Fig. 8. Latent Space on the training data set: Mitosis (blue)

Fig. 9. Latent space on the test set: Mitosis (blue)

4.3 Second experiment using biologist-supervised mitoses selection

Once this bank of mitotic cells was built, we followed the same strategy as described above using our autoencoder approach to identify mitotic cells. We decided to focus on mitotic cells for two reasons: i) mitoses, the percentage of which in a population is a https://www.overleaf.com/project/60032669b0517552d480edb5con marker of cell growth, are relatively quick events, representing roughly 5% of the cell cycle duration for this cell line, with very characteristic changes in phenotype; ii) the G1-S phase is a process that evolves slowly, lasting for more than 50% of the length of the cell cycle. In contrast to mitosis, there is no clear-cut change in phenotype, making it more difficult to unambiguously identify.

This experiment is illustrated by a pytorch code available at https://github.com/Gustoaxel/Cell-Cycle

4.3.1 Results on the second dataset

Figure 8 and 9 show that the distribution in the latent space of the training and test sets are similar and different to Gaussian distributions. Note that the mitoses cluster is more compact than the control cluster. The explanation for this observation is likely the fact that mitotic cells undergo a very abrupt and specific change of their morphology, while non-mitotic cells undergo a slow and progressive evolution of their phenotype.

Table 4 shows that the top 3 features are phase features which means that they are more efficient than geometric

TABLE 3
Accuracy of detection of Mitoses

Autoencoder	Mitosis
Accuracy %	98.6

TABLE 4
Selected features

Features	Weights
$Phase - avg(nm)$	0.87
$Phase - StdDev(nm)$	0.83
$Phase - median(nm)$	0.77
$Surface(micron2)$	0.40
$Optical - volume(micron3)$	0.32
$Phase - max(nm)$	0.30
$Surface - density(pg/micron2)$	0.25
$Dry - mass(pg)$	0.19

features for classification.

As expected, we found a significantly higher proportion of cells in mitosis compared to that found using the nocodazole-blocked population, since all stages of the M phase were present in the training set, as seen in Table 5 and Figure 10. More than 80,000 cells are plotted (Figure 10). The dense green zone close to the origins corresponds to small cellular debris.
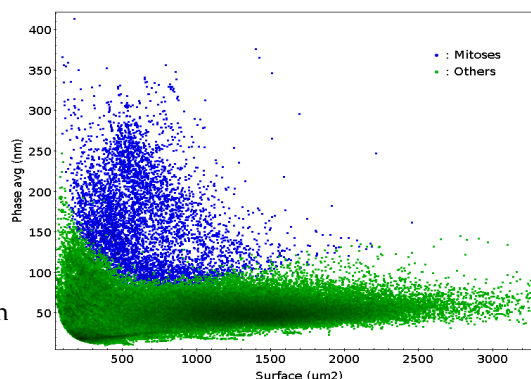

Fig. 10. "Mitotic cells in an exponentially growing population": Cell surface versus Phase average dotplot, showing both mitotic (blue) and non-mitotic cells.

TABLE 5
Number of detected mitotic cells of the second dataset

Testing cells	5,000	10,000	20,000	40,000	80,000
Mitotic cells	356	723	1,450	2,841	6,123
Mitotic index	7.1	7.2	7.3	7.1	7.7

TABLE 6
Number of detected mitotic cells on the third dataset

Testing cells	5,000	10,000	20,000	40,000	50,000
Mitotic cells	366	767	1,362	2,756	3,449
Mitotic index	7.3	7.7	6.8	6.9	6.9

4.3.2 Results on the third dataset

This mitotic bank, or sub-portions of it, yielded the same proportion of mitotic cells when used as training on different control populations. The averaged mitotic index was 7.3% in Table 5, in good agreement with the mitotic index calculated for exponentially growing HeLa cell populations by manual counting [40]. This demonstrates the robustness of the approach. Table 6 was obtained with the same training set, but on a dataset produced with the same cell line on another day. The averaged mitotic index obtained was 7.1%, very similar to that of Table 5.

This demonstrates that as long as the same exact acquisition settings are used (exposure, magnification, ...), the learned autoencoder can be used on multiple new acquisitions without any significant changes in the calculated mitotic index.

5 DISCUSSION

Exponentially growing cell populations grow because cells undergo mitosis. The mitotic index is defined as the percentage of cells undergoing mitosis in a cell population. There are many ways to determine the mitotic index, but only QPI allows non-invasive monitoring of the mitotic index. Using the strategy reported here, it is now feasible to screen chemical banks on live cell populations using high-throughput microscopy. The effect of these chemicals on the population growth/differentiation/cell death will be reflected by changes in their mitotic indexes. This should be of great interest for pharmacological studies to identify biologically active compounds.

This work demonstrates that Quantitative Phase Imaging is a very powerful and non-invasive way for biologists to monitor live cell populations with a relatively standard imaging station consisting of an automated microscope in a CO_2 thermostated and water-saturated incubation chamber. Classically, biologists interested in following cell populations under live conditions have had to rely on either "vital" labeling, as with the DNA intercalating agent Hoechst 33342, or exogenously expressed proteins tagged with a fluorescent domain. However, it is well known that "vital" dyes are quick to alter cell physiology, as seen with Hoechst 33342 mentioned above [41]. Similarly, while very useful, fluorescent fusion proteins result in many possible artifacts and toxicity, and thus cannot be considered physiological. [42]

The camera used in this study does not capture classical photonic images, but rather records interferograms of the live cells by measuring regular white light phase gradients created when going through biological samples. From these interferograms, several cell features are extracted. Since each cell type has its own phenotype, it is necessary to first build a mitotic reference data set for each cell type analyzed. Mathematical treatment of the cell features provides a very accurate supervised classification of existing sub-populations. This means that without any treatment of the cells, biologists are now able to calculate the proportion of cells going through mitosis, in a way reminiscent of the widely used FACS systems. In contrast to FACS analyses, the cells remain untouched in culture during the whole process, which can theoretically last as long as needed. In other words, all questions dealing with cell growth and phenotypic changes in cell populations exposed to any treatment of interest can now be reliably addressed in real time since normal cell physiology will not be affected by anything else but the tested products. While our approach should theoretically be usable with other QPI methodologies, we do not have access to these other systems. The only sure way to validate our concept to these other QPI methodologies would be to try it. Toxicity testing and drug development should be made more straightforward with this approach. In addition, screening costs will be kept as low as possible due to the absence of any treatment or manipulation other than computerized treatment of descriptors extracted from interferogram analyses, and mitotic index calculation with our new autoencoder algorithm.

Deep learning networks are of particular interest for classification analysis. Variational autoencoders are of increasing interest. Herein we show how a new supervised autoencoder can be of particular interest for cell monitoring.

6 CONCLUSION

In this paper we point out the following contributions:

- We shows that we now have a way to monitor live cell populations in a completely non-invasive manner thanks to the association of QPI technology with mathematical treatment of the QPI produced interferograms.

- As reported earlier [43], we confirm here that while convenient to produce training sets, chemical blockade of the cell cycle with nocodazole will only detect early stages of the M phase, which is too restrictive for the preparation of datasets representative of the whole mitotic process. Training on naturally occurring mitoses is thus necessary for the algorithm to accurately classify all mitotic events.

- Combining a hand made efficient mitosis reference and a new supervised autoencoder we obtained a very high accuracy classification of 98.6%. This result in an accurate Mitotic index estimation (7%) (more reliable as compared to chemical blocking) compared to state of the art results reported in the literature.

- A straightforward application, among others, will be live real-time high-throughput pharmacological screening of molecule banks. In addition to its easy implementation, this approach records changes in populations throughout the experiment, for example in response to changes in culture conditions. The proportion of mitoses in cell populations exposed to the tested chemicals will be a reliable reflection of any effect of the screened molecules on cellular proliferation.

REFERENCES

[1] L. L. Rubin, K. L. Philpott, and S. F. Brooks, "The cell cycle and cell death," *Current Biology*, vol. 3, no. 6, Jun. 1993.

[2] K. Vermeulen, D. R. Van Bockstaele, and Z. N. Berneman, "The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer: *Cell cycle regulation and deregulation*," *Cell Proliferation*, vol. 36, no. 3, pp. 131–149, Jun. 2003.

[3] Y. Zhang, "Cell toxicity mechanism and biomarker," *Clinical and Translational Medicine*, vol. 7, no. 1, Dec. 2018.

[4] A. Adan, G. Alizada, Y. Kiraz, Y. Baran, and A. Nalbant, "Flow cytometry: basic principles and applications," *Critical Reviews in Biotechnology*, vol. 37, no. 2, pp. 163–176, Feb. 2017.

[5] I. Andrä, H. Ulrich, S. Dürr, D. Soll, L. Henkel, C. Angerpointner, J. Ritter, S. Przibilla, H. Stadler, M. Effenberger, D. H. Busch, and M. Schiemann, "An Evaluation of T-Cell Functionality After Flow Cytometry Sorting Revealed p38 MAPK Activation," *Cytometry Part A*, vol. 97, no. 2, pp. 171–183, Feb. 2020.

[6] B. Kemper and G. von Bally, "Digital holographic microscopy for live cell applications and technical inspection," *Appl. Opt.*, vol. 47, no. 4, pp. A52–A61, 2008.

[7] M. K. Kim, "Principles and techniques of digital holographic microscopy," *Journal of Photonics for Energy*, p. 018005, Apr. 2010.

[8] K. Lee, K. Kim, J. Jung, J. Heo, S. Cho, S. Lee, G. Chang, Y. Jo, H. Park, and Y. Park, "Quantitative Phase Imaging Techniques for the Study of Cell Pathophysiology: From Principles to Applications," *Sensors*, vol. 13, no. 4, pp. 4170–4191, Mar. 2013.

[9] P. Marquet, C. Depeursinge, and P. J. Magistretti, "Review of quantitative phase-digital holographic microscopy: promising novel imaging technique to resolve neuronal network activity and identify cellular biomarkers of psychiatric disorders," *Neurophotonics*, vol. 1, no. 2, p. 020901, Sep. 2014.

[10] Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," *Nature Photonics*, vol. 12, no. 10, pp. 578–589, Oct. 2018.

[11] D. H. Park, C. M. Ho, Y. Chang, and H. Zhang, "Gradient-coherent strong regularization for deep neural networks," *arXiv preprint arXiv:1811.08056*, 2018.

[12] R. Kasprowicz, R. Suman, and P. O'Toole, "Characterising live cell behaviour: Traditional label-free and quantitative phase imaging approaches," *The International Journal of Biochemistry & Cell Biology*, vol. 84, pp. 89–95, Mar. 2017.

[13] T. A. Zangle and M. A. Teitell, "Live-cell mass profiling: an emerging approach in quantitative biophysics," *Nature Methods*, vol. 11, no. 12, pp. 1221–1228, Dec. 2014.

[14] G. Popescu, *Quantitative phase imaging of cells and tissues*, ser. McGraw-Hill biophotonics. New York: McGraw-Hill, 2011.

[15] P. Bon, G. Maucort, B. Wattellier, and S. Monneret, "Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells," *Optics Express*, vol. 17, no. 15, p. 13080, Jul. 2009.

[16] S. Aknoun, J. Savatier, P. Bon, F. Galland, L. Abdeladim, B. Wattellier, and S. Monneret, "Living cell dry mass measurement using quantitative phase imaging with quadriwave lateral shearing interferometry: an accuracy and sensitivity discussion," *Journal of Biomedical Optics*, vol. 20, no. 12, p. 126009, Dec. 2015.

[17] Pognonec, Philippe, Barlaud Michel, Wattellier Benoit, Pourcher Thierry, Zhou Yuxiang, Aknoun Sherazade, Yonnet Manuel, and Antonini Marc, "Non Invasive Live Cell Cycle Monitoring using Quantitative Phase Imaging and Proximal Machine Learning Methods," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. Cordoba, Spain: IEEE, Jun. 2019, pp. 483–488.

[18] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851–869, 07 2016.

[19] L. Zhang, C. Lv, Y. Jin, G. Cheng, Y. Fu, D. Yuan, Y. Tao, Y. Guo, X. Ni, and T. Shi, "Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma," *Frontiers in Genetics*, vol. 9, p. 477, 2018.

[20] G. Hinton and R. Zemel, "Autoencoders, minimum description length and helmholtz free energy." in *Advances in neural information processing systems*, 1994, pp. 3–10.

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016, vol. 1.

[22] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.

[23] Y. Wang, H. Yang, X. Yuan, Y. A. Shardt, C. Yang, and W. Gui, "Deep learning for fault-relevant feature extraction and fault classification with stacked supervised auto-encoder," *Journal of Process Control*, vol. 92, pp. 79–89, 2020.

[24] D. Kingma and M. Welling, "Auto-encoding variational bayes," *International Conference on Learning Representation*, 2014.

[25] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in Neural Information Processing Systems*, 2014, pp. 3581–3589.

[26] M. Barlaud and F. Guyard, "Learning a sparse generative nonparametric supervised autoencoder," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, TORONTO , Canada*, June 2021.

[27] M. Barlaud and M. Antonini, "Robust classification with feature selection using an application of the Douglas-Rachford splitting algorithm," *ESAIM: PROCEEDINGS AND SURVEYS*, 2021.

[28] S. Aknoun, M. Yonnet, Z. Djabari, F. Graslin, M. Taylor, T. Pourcher, B. Wattellier, and P. Pognonec, "Quantitative phase microscopy for non-invasive live cell population monitoring," *Scientific Reports*, vol. 11, no. 1, p. 4409, Feb. 2021.

[29] J. G. Hoffman, "Wright's Hypothesis: Its Relation to Volume Growth of Tissue Cells and Mitotic Index," *Science*, vol. 106, no. 2754, pp. 343–344, Oct. 1947.

[30] L. Condat, "Fast projection onto the simplex and the l1 ball," *Mathematical Programming Series A*, vol. 158, no. 1, pp. 575–585, 2016.

[31] G. Perez, M. Barlaud, L. Fillatre, and J.-C. Régin, "A filtered bucket-clustering method for projection onto the simplex and the ℓ_1-ball," *Mathematical Programming*, May 2019.

[32] P. J. Huber, "Robust statistics. 1981."

[33] H. Zhou, J. Lan, R. Liu, and J. Yosinski, "Deconstructing lottery tickets: Zeros, signs, and the supermask," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 3597–3607.

[34] M. Barlaud and F. Guyard, "Learning sparse deep neural networks using efficient structured projections on convex constraints for green ai," *International Conference on Pattern Recognition, Milan*, 2020.

[35] D. Kingma and J. Ba, "a method for stochastic optimization." *International Conference on Learning Representations*, pp. 1–13, 2015.

[36] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[37] W. F. Scherer, "Studies on the propagation in vitro of poliomyelitis viruses: iv. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix," *Journal of Experimental Medicine*, vol. 97, no. 5, pp. 695–710, May 1953.

[38] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini, "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biology*, vol. 7, no. 10, p. R100, 2006.

[39] R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf, "Centering, scaling, and transformations: improving the biological information content of metabolomics data," *BMC genomics*, vol. 7, 06 2006.

[40] S. Varia, D. Cheedu, M. Markey, K. Torres-Shafer, V. Battini, A. Bubulya, and P. Bubulya, "Alignment of Mitotic Chromosomes in Human Cells Involves SR-Like Splicing Factors Btf and TRAP150," *International Journal of Molecular Sciences*, vol. 18, no. 9, p. 1956, Sep. 2017.

[41] M. Purschke, N. Rubio, K. D. Held, and R. W. Redmond, "Phototoxicity of Hoechst 33342 in time-lapse fluorescence microscopy," *Photochemical & Photobiological Sciences*, vol. 9, no. 12, p. 1634, 2010.

[42] A. M. Ansari, A. K. Ahmed, A. E. Matsangos, F. Lay, L. J. Born, G. Marti, J. W. Harmon, and Z. Sun, "Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments," *Stem Cell Reviews*, vol. 12, pp. 553–559, 2016.

[43] N. Paweletz and U. Lang, "Fine structural studies of early mitotic stages in untreated and nocodazole-treated HeLa cells," *European Journal of Cell Biology*, vol. 47, no. 2, pp. 334–345, Dec. 1988.