QUANTIFYING THE EFFECT OF IMAGE COMPRESSION ON SUPERVISED LEARNING TASKS IN MICROSCOPY

E. Pomarico[1*], C. Schmidt[1], D. Nguyen[2], A. Planchette[2], A. Roux[1], S. Pagès[3], L. Batti[3], C. Clausen[4], T. Lasser[5], A. Radenovic[2], B. Sanguinetti[4], and J. Extermann[1]

[1]HEPIA/HES-SO, University of Applied Sciences of Western Switzerland, Switzerland

[2]Laboratoire de Biologie à l'Echelle Nanométrique, EPFL, Lausanne, Switzerland

[3]Wyss Center for Bio- and Neuroengineering, Geneva, Switzerland

[4]Dotphoton SA, Zeughausgasse 17, 6300 Zug, Switzerland

[5]Max-Planck Institute for Polymer Research, Mainz, Germany.

***E-mail: enrico.pomarico@hesge.ch**

The growth of throughput in microscopy has led to the widespread use of supervised learning (SL) models running on compressed imaging datasets for automated analysis. However, since lossy compression can produce unpredictable artifacts, quantifying the impact of data compression on SL tasks is of pivotal importance to assess their reliability, especially for clinical use. We propose an experimental method to evaluate the tolerability of image compression distortions in 2D and 3D cell segmentation SL tasks: predictions on compressed data are compared to the raw predictive uncertainty, which is numerically estimated from the raw noise statistics, measured through sensor calibration [1].

Predictions on segmentation parameters in phase-contrast (PC), as well as light-sheet microscopy and OPT, are altered by up to 15% and more than 10 standard deviations after 16-to-8 bits down-sampling or JPEG compression. In contrast, a recent lossless algorithm, offering up to 10:1 compression ratio, provides a prediction spread equivalent to that stemming from raw noise [2]. By setting a lower bound to the predictive uncertainty, our technique can be generalized to validate a variety of analysis pipelines in SL-assisted fields.

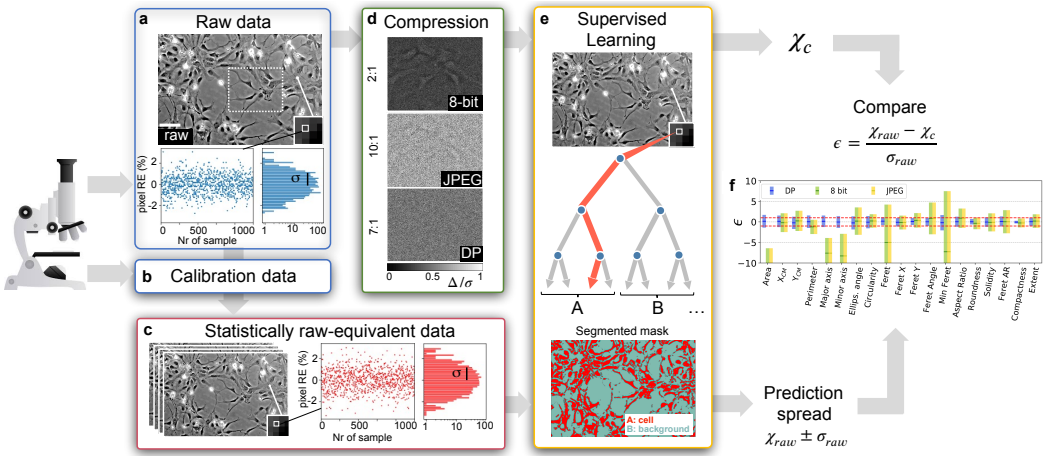

Figure 1: **a** Noise affecting raw data is shown by the pixel value statistics in a PC image of neural stem cells. **b, c** Via sensor calibration statistically raw-equivalent images are generated. **d** Compression is performed on raw data. **e** A trained SL model predicts a parameter value from compressed data χ_c and estimates from the synthetic data the standard deviation σ_{raw} associated to the raw parameter value χ_{raw}. **f** The standard score ϵ is calculated: if $|\epsilon| > 1$ predictions on the compressed image data exceed the statistical variability stemming from raw noise.

[1] E. Pomarico et al, under review, arXiv:2009.12570. [2] www.dotphoton.com