



**HAL**  
open science

# Cell Counting with Trainable h-Maxima and Connected Component Layers

Samy Blusseau, Xiaohu Liu, Santiago Velasco-Forero

► **To cite this version:**

Samy Blusseau, Xiaohu Liu, Santiago Velasco-Forero. Cell Counting with Trainable h-Maxima and Connected Component Layers. *Journal of Mathematical Imaging and Vision*, 2025, 67 (27). <hal-05033116>

**HAL Id: hal-05033116**

**<https://hal.science/hal-05033116v1>**

Submitted on 14 Apr 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Cell Counting with Trainable $h$ -Maxima and Connected Component Layers

Samy Blusseau<sup>1\*</sup>, Xiaohu Liu<sup>1,2</sup> and Santiago Velasco-Forero<sup>1</sup>

<sup>1</sup>Mines Paris, PSL University, Centre for Mathematical Morphology, Fontainebleau, 77300, France.

<sup>2</sup> Shanghai Jiao Tong University, Shanghai, China.

\*Corresponding author(s). E-mail(s): [samy.blusseau@minesparis.psl.eu](mailto:samy.blusseau@minesparis.psl.eu);  
Contributing authors: [liuxiaohu@sjtu.edu.cn](mailto:liuxiaohu@sjtu.edu.cn); [santiago.velasco@minesparis.psl.eu](mailto:santiago.velasco@minesparis.psl.eu);

## Abstract

Bright objects on a dark background, such as cells in microscopy images, can sometimes be modeled as maxima of sufficient dynamic, called  $h$ -maxima. Such a model could be sufficient to count these objects in images, provided we know the dynamic threshold that tells apart actual objects from irrelevant maxima. In a previous study we introduced a neural architecture that includes a morphological pipeline counting the number of  $h$ -maxima in an image, preceded by a classical CNN which estimates the dynamic  $h$  yielding the right number of objects. This was made possible by geodesic reconstruction layers, and a new module counting connected components. The resulting architecture could be trained end-to-end to count cells by minimizing a loss function focusing mainly on an intermediate image and very marginally on the counting errors. In the present paper, we extend that work with a detailed theoretical analysis of the derivability of such pipeline, and with more cell counting experiments on different datasets. Our analysis explains why it is difficult to train the proposed pipeline when penalizing counting errors, and ends up with a solution to that challenge. Our experiments show that our approach can obtain results comparable to the state of the art with much fewer parameters and an increased interpretability of the model.

**Keywords:** Cell counting, Mathematical morphology, Geometric deep learning

## 1 Introduction

Cell counting is a crucial step in biological experiments and medical diagnosis to provide both quantitative and qualitative information on cells. While the process can be tedious, low-efficiency and prone to subjective errors, especially when cell clustering occurs and cells show high variance in shapes and contrasts. Automation offers a quick and accurate estimation of cell quantity in a sample. In this paper, we focus on the cell counting task to quantify population on fluorescent images.

Previous studies [1–4] used deep learning with a Unet architecture or a similar one, which maps each input image to a density map, the integral of which yields the cell count. Another approach is to use an architecture for object detection (as YOLO[5] or Faster R-CNN [6]) and count the detected cells [5, 7].

These methods achieve satisfactory results, but use a large number of parameters even when the task seems quite simpler than crowd counting, as for example when cells, possibly diverse

in shapes and partially overlapping, still look like mere regional maxima over a dark background. Furthermore, density maps are ambiguous regarding what the model considers as a cell, especially in clusters, where the precise location of cells is often lost. This can be a limitation for users willing to check a posteriori the reliability of the count returned by the model, especially as the deep learning model is a black box and its decisions are hardly explainable. Images where cells appear as bright objects on a dark background are an interesting study case where a much simpler model, built on *a priori* knowledge, may achieve similar results as large deep networks but with higher interpretability, regarding for example geometrical and contrast criteria used to recognize cells.

In this paper, following and extending [8], we propose to model cells as regional maxima with sufficient dynamic and size. The dynamic of an extremum is, simply speaking, its depth (see Fig. 1), and well known morphological methods exist to select extrema with dynamic larger than a threshold  $h$ , called  $h$ -maxima or  $h$ -minima [9]. Accordingly, in [8] was presented a simple morphological pipeline consisting in counting the  $h$ -maxima of a size-filtered version of the input image, for a given dynamic  $h$ . Since the image resolution is constant in the dataset, as well as the cell size, this approach only depends on the choice of the dynamic parameter, which needs, however, to be adapted to each image in order to achieve accurate counting. Taking advantage of the recently introduced geodesic reconstruction layers [10, 11], which provide a neural implementation of the  $h$ -maxima computation, a pipeline has been designed in [8], where a small convolutional neural network (CNN) predicts the optimal dynamic  $h$  for the subsequent morphological layers. This pipeline is trained end-to-end by minimizing a joint loss where the main term focuses on an intermediate output image (the so-called  $h$ -transform of a filtered input image) and a second term, with much smaller weight, penalizes counting errors.

The goal of the present paper is to make the latter pipeline trainable by minimizing the counting errors only, because achieving so has several advantages. First, it saves us from pre-computing, for each training image, the intermediate results needed for other loss terms. Second, avoiding to

pre-compute these intermediate outputs allows to make trainable any preprocessing of the input images, as opposed to the need to freeze it when the intermediate outputs are pre-computed using a fixed preprocessing. Third, it alleviates a lot the burden of annotations, as it only requires a number of cells for each input image, which does not need a precise location of each cell (although the proposed method does provide the locations of the counted cells).

To that aim, in this article we investigate more in-depth the theoretical aspects of the pipeline proposed in [8], in particular its derivability, so as to find out how it can be trained with gradient descent when minimizing the counting errors only. The main novel contributions of this article with respect to Liu et al. [8] are the following:

1. A theoretical analysis of the aforementioned pipeline derivability, based on a detailed theoretical review of  $h$ -maxima.
2. A practical solution to overcome the derivability issues revealed by this analysis, allowing for the optimization of the pipeline by minimization of a loss penalizing the counting errors only.
3. New results on other cell image datasets (Cellpose, Fluorescent Neuronal Cells) with respect to [8].

The paper is structured as follows. In Section 2 the theoretical framework of  $h$ -maxima is presented, and a simple morphological algorithm using them to count cells, already introduced in [8], is recalled. In Section 3, we remind the deep learning architecture of [8], which includes this morphological algorithm, before studying its derivability and proposing a solution for training it with gradient descent even when minimizing the counting errors only. Experiments are conducted in Section 4, and their results analyzed in Section 5. Conclusions are drawn in Section 6.

## 2 Setting and Definitions

We assume images to be mappings of the discrete domain  $\Omega := ([0, W - 1] \times [0, H - 1]) \cap \mathbb{N}^2$ ,  $W, H \in \mathbb{N}^*$  to  $\mathbb{R}$ . We will denote by  $\mathcal{F}(\Omega, \mathbb{R})$  the set of such functions, and call *pixels* the elements of  $\Omega$ .

## 2.1 Image Poset

The set  $\mathcal{F}(\Omega, \mathbb{R})$  is a partially ordered set (or *poset*) for the partial order  $\leq_{\mathcal{F}}$  defined, for any  $f, g \in \mathcal{F}(\Omega, \mathbb{R})$ , by

$$f \leq_{\mathcal{F}} g \iff f(x) \leq g(x) \quad \forall x \in \Omega. \quad (1)$$

In the rest of the paper, we will simply write  $f \leq g$  when it is clear we are comparing functions.

The supremum and infimum of two functions  $f$  and  $g$  are denoted respectively by  $f \vee g$  and  $f \wedge g$  and also induced by the pointwise ordering, *i.e.*

$$\begin{aligned} \forall x \in \Omega, \quad (f \vee g)(x) &:= f(x) \vee g(x) \\ \text{and } (f \wedge g)(x) &:= f(x) \wedge g(x) \end{aligned} \quad (2)$$

where here  $\vee$  and  $\wedge$  also denote the supremum and infimum on real numbers,  $f(x) \vee g(x) = \sup\{f(x), g(x)\}$  and  $f(x) \wedge g(x) = \inf\{f(x), g(x)\}$ . The supremum and infimum of an arbitrary family of functions is defined likewise.

## 2.2 Connectivity

Throughout the paper, the notion of connectivity between pixels is crucial because the definitions of path and reconstruction operator, given below and extensively used in the paper, rely on it. It is determined by the unit discrete square  $B := \{x = (x_1, x_2) \in \mathbb{Z}^2, \max(|x_1|, |x_2|) \leq 1\}$ , classically referred to as the 8-connectivity *structuring element* in mathematical morphology.

Denoting by  $B_x := x + B = \{x + b, b \in B\}$  the set  $B$  translated at  $x \in \Omega$ , we say that  $y \in \Omega$  is a neighbor of  $x$ , noted  $y \in \mathcal{N}_x$ , if and only if  $y \in B_x \cap \Omega$ . Similarly, the neighbors of a set  $X \subset \Omega$  are defined as  $\mathcal{N}_X := \cup_{x \in X} \mathcal{N}_x$ .

The notion of neighborhood leads to classical definitions of graph theory, the vertices being here the pixels and the edges being the couples  $(x, y) \in \Omega^2$  such that  $y \in \mathcal{N}_x$ . A *path*  $\gamma$  from  $x$  to  $y$  of length  $n \in \mathbb{N}$ , is a tuple  $\gamma = (u_0, u_1, \dots, u_n) \in \Omega^{n+1}$  such that  $u_0 = x$ ,  $u_n = y$  and for any  $0 \leq k \leq n-1$ ,  $u_{k+1} \in \mathcal{N}_{u_k}$ . We will denote by  $\Gamma_{xy}$  the set of paths from  $x$  to  $y$ , and say that  $y$  is *connected* to  $x$  in a subset  $C \subseteq \Omega$  if there is a path from  $x$  to  $y$  included in  $C$ , *i.e.*  $\{\gamma \in \Gamma_{xy}, \gamma \subseteq C\} \neq \emptyset$ . Note that in our setting, since  $B$  is symmetric ( $b \in B \iff -b \in B$ ),  $y$  is connected to  $x$  if and only if  $x$  is connected to  $y$ ,

hence we shall simply say that  $x$  and  $y$  are *connected*. In particular, all pixels are connected in  $\Omega$ . Furthermore, the connectivity relation is also reflexive ( $x \in \mathcal{N}_x$  because  $0 \in B$ ) and transitive. It is therefore an equivalence relation, and the equivalence classes induced by it are the *connected components*. Finally, we say that a subset  $C \subseteq \Omega$  is connected if any two elements  $x$  and  $y \in C$  are connected.

## 2.3 Morphological Operators

We shall now recall the morphological operators which will be used in the rest of the paper. Note that these definitions are not the most general ones, but adapted to the structuring element  $B$  considered here, which is symmetric.

### 2.3.1 Dilation, Erosion, Opening and Closing by B

Given  $f \in \mathcal{F}(\Omega, \mathbb{R})$ , the *dilation of  $f$  by  $B$* , noted  $\delta_B(f)$ , is the element of  $\mathcal{F}(\Omega, \mathbb{R})$  defined by

$$\forall x \in \Omega, \quad \delta_B(f)(x) = \max_{b \in B} f(x+b) = \max_{y \in \mathcal{N}_x} f(y) \quad (3)$$

while the *erosion of  $f$  by  $B$* , noted  $\varepsilon_B(f)$ , is given by

$$\forall x \in \Omega, \quad \varepsilon_B(f)(x) = \min_{b \in B} f(x+b) = \min_{y \in \mathcal{N}_x} f(y). \quad (4)$$

Since  $0 \in B$ , here  $\delta_B$  is an extensive operator, *i.e.*  $\forall x \in \Omega, \delta_B(f)(x) \geq f(x)$  (and  $\varepsilon_B$  is anti-extensive,  $\varepsilon_B(f)(x) \leq f(x)$ ).

The opening and closing of  $f$  by  $B$ , noted  $\gamma_B(f)$  and  $\varphi_B(f)$  respectively, are the compositions of dilation and erosion, namely  $\gamma_B = \delta_B \circ \varepsilon_B$  and  $\varphi_B = \varepsilon_B \circ \delta_B$ .

### 2.3.2 Geodesic Operators

Given two functions  $f, g \in \mathcal{F}(\Omega, \mathbb{R})$  such that  $f \leq g$  the *geodesic dilation of  $f$  under  $g$* , noted  $\delta_g(f)$ , is defined by [9, 12]

$$\delta_g(f) := \delta_B(f) \wedge g \quad (5)$$

The geodesic dilation can be iterated and for  $p \geq 1$  we note  $\delta_g^{(p+1)}(f) := \delta_g(\delta_g^{(p)}(f))$ , with  $\delta_g^{(1)} = \delta_g$ . It is well known and easy to verify that this process converges to a fixed point after a finite number

$k$  of iterations (one can notice that,  $\delta_B$  being an extensive operator, the iterations of geodesic dilation produce an increasing sequence of values  $\delta_g^{(p)}(f)(x)$  at each pixel  $x$ ; furthermore, the possible values for each  $\delta_g^{(p)}(f)(x)$  are finite, as they come from those of  $f$  and  $g$  which both have  $W \times H$  pixels).

The *reconstruction by dilation* of  $g$  from  $f$ , noted  $\text{REC}(f, g)$  or  $\text{REC}_g(f)$  is then:

$$\text{REC}(f, g) := \text{REC}_g(f) := \delta_g^{(k)}(f) \quad (6)$$

where  $k$  is the smallest integer such that  $\delta_g^{(k)}(f) = \delta_g^{(k+1)}(f)$ .

Alternative formulations of the reconstruction by dilation exist, *e.g.* based on the level sets of  $f$  and  $g$  [13]. Here we mention one based on connectivity, which will be more convenient than (6) to prove the theoretical results of this paper. For any  $x \in \Omega$ , one can write

$$\text{REC}(f, g)(x) = \max_{y \in \Omega, \gamma \in \Gamma_{xy}} \left( f(y) \wedge \min_{z \in \gamma} g(z) \right). \quad (7)$$

To the best of our knowledge, this formulation has not been stated (hence not proved) in the morphological literature, but it can be deduced easily by induction on the length of paths using the sequential definition of (6), and the fact that  $\text{REC}(f_1 \vee f_2, g) = \text{REC}(f_1, g) \vee \text{REC}(f_2, g)$ .

## 2.4 Regional Maxima and $h$ -Maxima

In this section, besides recalling necessary definitions, we state several theoretical results which will not appear as new to experts in mathematical morphology. However, since we could not find them clearly stated and proved in the literature, we choose to present and prove them here.

### *Regional Maxima*

Given an image  $f \in \mathcal{F}(\Omega, \mathbb{R})$  we say that  $M \subset \Omega$ ,  $M \neq \Omega$ , is a *regional maximum* at level  $t \in \mathbb{R}$  if  $M$  is connected,  $f(x) = t$  for all  $x \in M$ , and  $f(y) < t$  for any  $y \in \mathcal{N}_M \setminus M$ . In words, a regional maximum is a plateau of  $f$  surrounded by lower values. Note that this definition imposes that a regional maximum  $M$  needs to be strictly

included in  $\Omega$ , hence  $\mathcal{N}_M \setminus M \neq \emptyset$  and  $\Omega$  itself cannot be a regional maximum. We will note  $\mathcal{M}(f)$  the set of regional maxima of  $f$  and  $\mathcal{PM}(f)$  the corresponding set of pixels, that is

$$\mathcal{PM}(f) := \bigcup_{M \in \mathcal{M}(f)} M.$$

Also, since  $f$  takes a unique value over a regional maximum  $M$ , we shall conveniently note  $f(M)$  this value (although  $M$  is not a pixel but a set of pixels). Importantly, because of the connectedness property, different regional maxima are necessarily disjoint.

Finally, we will denote by  $\text{RMAX}$  the characteristic function of regional maxima, *i.e.*

$$\text{RMAX}(f) := \mathbb{1}_{\mathcal{PM}(f)} = \sum_{M \in \mathcal{M}(f)} \mathbb{1}_M. \quad (8)$$

In the case where  $f$  takes values in a quantized set  $\{k \cdot \epsilon, 0 \leq k \leq L\}$  for some  $L \in \mathbb{N}^*$  and  $\epsilon > 0$ , then regional maxima can be computed thanks to the geodesical reconstruction of  $f$  from  $f - \epsilon$ . Indeed, in that case one can show [13] that the regional maxima of  $f$  are the connected components of

$$X_{\max}(f) = \{x \in \Omega, f(x) > \text{REC}_f(f - \epsilon)(x)\}, \quad (9)$$

the set of pixels where the reconstruction is different from (hence smaller than)  $f$ <sup>1</sup>. In that case,  $\mathcal{PM}(f) = X_{\max}(f)$  and therefore

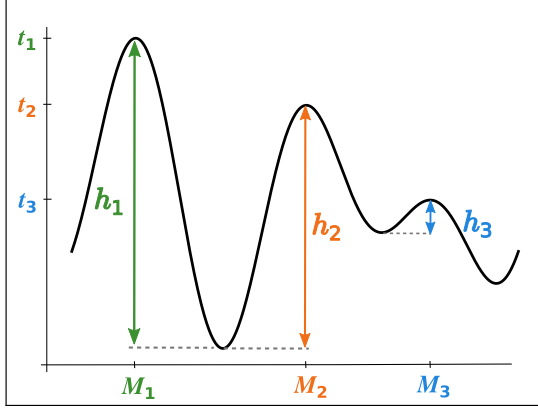
$$\text{RMAX}(f) = \mathbb{1}_{X_{\max}(f)}. \quad (10)$$

### *Dynamic*

The *dynamic* [14] of a regional maximum  $M$  of level  $t \in \mathbb{R}$  is the smallest height one needs to come down from  $M$  in order to reach a higher regional maximum  $M'$  of level  $t' > t$ , when there is one. Formally, in that case (refer to Section 6.3.5 in [9],

---

<sup>1</sup>This result can also be deduced from the "if" direction of Proposition 6, and a reasoning very similar to the one used to show the "only if" direction of the same proposition.



**Fig. 1** The dynamics of three maxima in a one-dimensional function. Note that in our 1D illustrations,  $\Omega$  is a discrete interval and the structuring element defining the neighborhood of each pixel is  $B = \{-1, 0, 1\}$ . Here each maximum  $M_i$  is reduced to one singleton.

and see Figure 1)

$$\text{DYNMAX}(M) := \min_{\substack{\gamma \in \Gamma_{MM'}, \\ f(M') > f(M)}} \left\{ \max_{z \in \gamma} (f(M) - f(z)) \right\} \quad (11)$$

where  $\Gamma_{MM'} := \{\gamma \in \Gamma_{xy}, x \in M, y \in M'\}$  and the variable  $M'$  spans all the regional maxima of  $f$  such that  $f(M') > f(M)$ . Note that, since  $f(M)$  does not depend on the path  $\gamma$ , the max term in (11) can be written  $f(M) - \min_{z \in \gamma} f(z)$ , and finally

$$\text{DYNMAX}(M) = f(M) - \max_{\substack{\gamma \in \Gamma_{MM'}, \\ f(M') > f(M)}} \min_{z \in \gamma} f(z). \quad (12)$$

By convention, the dynamic of the global maximum of  $f$  can be set to the difference between the maximum and the minimum values of  $f$ , denoted by

$$\Delta_{\Omega} f := \max_{\Omega} f - \min_{\Omega} f. \quad (13)$$

The dynamic is commonly used to distinguish between irrelevant maxima caused by noise and significant ones corresponding to underlying bright objects.

### *h-Maxima*

In order to define the *h-maxima* of  $f$ , we first need to introduce the *h-maxima transformation*. Given a parameter  $h \in [0, \Delta_{\Omega} f]$ , the  $\text{HMAX}_h$  image operator is the reconstruction by dilation of  $f$  from

$f - h$ , i.e

$$\text{HMAX}_h(f) = \text{REC}_f(f - h) \quad (14)$$

As illustrated in Figure 2 (a), this transformation suppresses what are called *h-domes* in [13].

Now, for a given  $h \in [0, \Delta_{\Omega} f]$ , the *h-maxima* of  $f$ , denoted by  $\mathcal{M}_h(f)$ , are defined as the regional maxima of  $\text{HMAX}_h(f)$ :

$$\mathcal{M}_h(f) := \mathcal{M}(\text{HMAX}_h(f)) = \mathcal{M}(\text{REC}_f(f - h)). \quad (15)$$

In particular,  $\mathcal{M}_0(f) = \mathcal{M}(f)$  and  $\mathcal{M}_{\Delta_{\Omega} f}(f) = \emptyset$ . Furthermore, as regional maxima of a function, distinct *h-maxima* are pairwise disjoint. Similarly to earlier, we will note  $\mathcal{PM}_h(f)$  the set of pixels belonging to *h-maxima*, that is

$$\mathcal{PM}_h(f) := \bigcup_{M \in \mathcal{M}_h(f)} M.$$

Figure 2 (b) shows in green the characteristic function of the *h-maxima* of  $f$ , noted  $\text{EMAX}_h$  for “extended maxima”:

$$\text{EMAX}_h(f) := \mathbf{1}_{\mathcal{PM}_h(f)} = \sum_{M \in \mathcal{M}_h(f)} \mathbf{1}_M. \quad (16)$$

Note that  $\text{EMAX}_h = \text{RMAX} \circ \text{HMAX}_h$ , and  $\text{EMAX}_0 = \text{RMAX}$ .

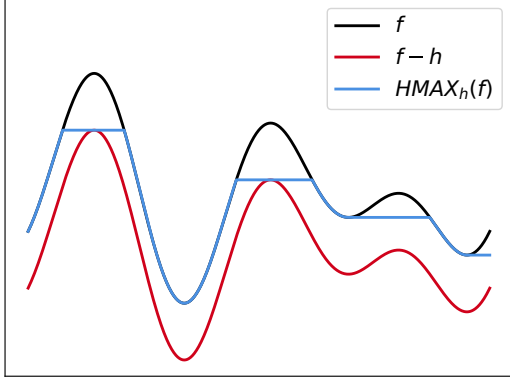
### *Links with the Dynamic*

An important property of *h-maxima* is that they contain the regional maxima of  $f$  of dynamic larger than  $h$ . These are given by the connected components where  $f$  is maximal inside the *h-maxima*. In turn, the regional maxima with dynamic larger than  $h$  determine the value of  $\text{HMAX}_h(f)$  over the *h-maxima* they belong to.

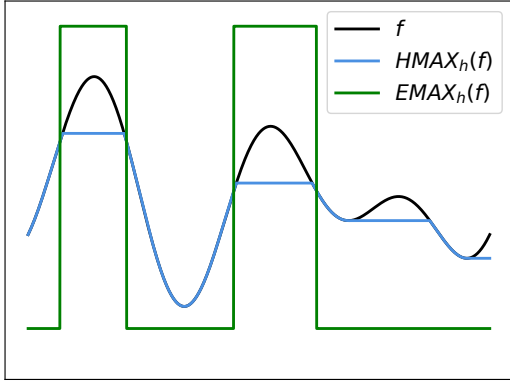
To put this more formally in Proposition 1, let us first introduce some notations. Given  $f \in \mathcal{F}(\Omega, \mathbb{R})$ , recall [15] that its upper level set of level  $t \in \mathbb{R}$ , which we will denote by  $\chi_t(f)$ , is defined by

$$\chi_t(f) := \{x \in \Omega, f(x) \geq t\}. \quad (17)$$

Also, given a non empty set  $M \subseteq \Omega$ , and  $M_0$  a connected subset of  $M$ , we denote by  $\text{CC}(M_0, M)$  the unique connected component of  $M$  containing  $M_0$ . Furthermore,  $\arg \max_M f$  denotes the set



(a)



(b)

**Fig. 2** Illustration of the computations of  $\text{HMAX}_h(f)$  and  $\text{EMAX}_h(f)$  for the same one-dimensional function  $f$  as in Figure 1, and for  $h_2 > h \geq h_3$ . (a)  $\text{HMAX}_h$  produces a regional maximum wherever there is a maximum with dynamic greater than  $h$  in  $f$ . (b)  $\text{EMAX}_h(f)$  is a binary function whose non-zero values indicate where the maxima of  $\text{HMAX}_h$ , hence the  $h$ -maxima of  $f$ , are. Note that in our 1D illustrations,  $\Omega$  is a discrete interval, represented by the  $x$  axis, and the structuring element defining the neighborhood of each pixel is  $B = \{-1, 0, 1\}$ .

$\{x \in M, f(x) = \max_{y \in M} f(y)\}$ . Then the following proposition holds.

**Proposition 1** Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$ ,  $h \geq 0$  and  $\lambda \in \mathbb{R}$ . Then

$$\begin{cases} M \in \mathcal{M}_h(f) \\ M_0 \text{ connected component of } \arg \max_M f \\ \lambda = \text{HMAX}_h(f)(M) \end{cases} \iff \begin{cases} M_0 \in \mathcal{M}(f), \text{DYNMAX}(M_0) > h, \\ \lambda = f(M_0) - h \\ M = \text{CC}(M_0, \chi_\lambda(f)). \end{cases} \quad (18)$$

According to Proposition 1, computing the  $h$ -maxima allows to detect bright objects of dynamic larger than  $h$ . This strategy is well known in the morphological literature, and is the one applied to count cells with the simple morphological algorithm described in Section 2.5.

Although this result has been at least partially stated in the literature [9], to the best of our knowledge no proof has been published, probably because it may appear completely obvious to experts, used to observe it on simple examples, like the one of Figure 3. We provide a proof in Appendix A.

### Larger $h$ , fewer $h$ -maxima

It is well known in the morphological literature, and easy to realize on simple examples, that the number of  $h$ -maxima is a decreasing function of  $h$ . This comes from a specific ordering relation between  $\mathcal{M}_h$  for different  $h$  values. Let us introduce the ordering relation  $\preceq$  on collections pairwise disjoint connected components of  $\Omega$  defined, for any two such collections  $\mathcal{M}$  and  $\mathcal{M}'$ , by

$$\mathcal{M} \preceq \mathcal{M}' \iff \forall M \in \mathcal{M}, \exists M' \in \mathcal{M}', M' \subseteq M. \quad (19)$$

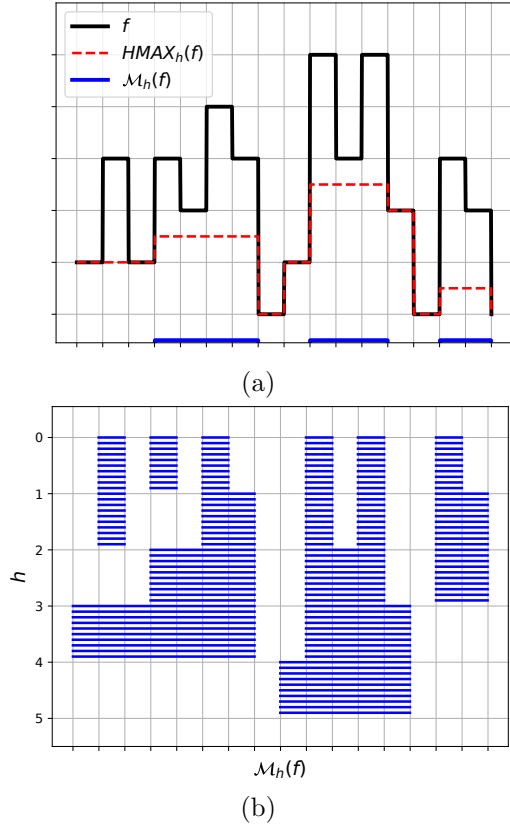
One can check that this is a partial ordering (reflexivity is ensured by the pairwise disjunction of two elements a collection). Then we have the following result.

**Proposition 2** For  $f \in \mathcal{F}(\Omega, \mathbb{R})$  and  $h_1, h_2 \geq 0$

$$h_1 \leq h_2 \Rightarrow \mathcal{M}_{h_2}(f) \preceq \mathcal{M}_{h_1}(f). \quad (20)$$

Recall that for  $h \geq 0$ , the elements of  $\mathcal{M}_h(f)$  are pairwise disjoint, as regional maxima of a function. Therefore, a  $h_1$ -maximum is included in at most one  $h_2$ -maximum. This is why  $\mathcal{M}_{h_2}(f) \preceq \mathcal{M}_{h_1}(f)$  implies that  $\mathcal{M}_{h_2}(f)$  has no more elements than  $\mathcal{M}_{h_1}(f)$ . Hence the number of  $h$ -maxima decreases with  $h$  from the total number of regional maxima, when  $h = 0$ , to zero, when  $h = \Delta_\Omega f$  (indeed one can see that  $\text{HMAX}_h(f)$  is just a constant function when  $h = \Delta_\Omega f$ , and has therefore no regional maximum by definition, i.e.  $\mathcal{M}_{\Delta_\Omega f} = \emptyset$ ). This is illustrated by Figure 3.

Note that Proposition 2 generalizes a part of Proposition 1, namely the one stating that



**Fig. 3** (a) A function  $f$ , the result of  $\text{HMAX}_h(f)$  for  $h = 2.5$  and the corresponding set  $\mathcal{M}_h(f)$  of  $h$ -maxima. The dynamics of the regional maxima of  $f$  are, from left to right: 2, 1, 4, 5, 5 and 3 respectively. (b) The evolution of  $h \mapsto \mathcal{M}_h(f)$  for  $h \in [0, \Delta_\Omega f]$ .

$\mathcal{M}_h(f) \subseteq \mathcal{M}_0(f) = \mathcal{M}(f)$  for  $h \geq 0$ . The proof of Proposition 2 is provided in Appendix B.

### Right Continuity of $\mu : h \mapsto \mathcal{M}_h(f)$

Let us denote by  $\mu$  the function mapping  $h \in [0, \Delta_\Omega f]$  to  $\mathcal{M}_h(f)$ , the set of  $h$ -maxima of  $f$ . With Proposition 2 in the previous paragraph, we saw that  $\mu$  is decreasing for the orderings  $\leq$  and  $\preceq$ . Now we will see that it is also piecewise constant and right continuous as illustrated by Figure 3. This will be useful in the study of the derivability of  $h \mapsto \text{HMAX}_h(f)$ , in Section 3.3.1.

**Proposition 3** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  non constant. Then there exists  $K \in \mathbb{N}$  and  $h_0 < h_1 < \dots < h_{K+1} \in [0, \Delta_\Omega f]$  with  $h_0 = 0$  and  $h_{K+1} = \Delta_\Omega f$ , such that*

$$\forall k \in \{0, \dots, K\}, \forall h \in [h_k, h_{k+1}), \mu(h) = \mu(h_k) = \mathcal{M}_{h_k}(f) \quad (21)$$

and for  $k \neq k'$ ,  $\mathcal{M}_{h_{k'}}(f) \neq \mathcal{M}_{h_k}(f)$ .

From now on we will note

$$\mathcal{H}_f := \{h_0, h_1, \dots, h_{K+1}\}$$

the finite subset of  $[0, \Delta_\Omega f]$  introduced by the proposition. Clearly, the elements of  $\mathcal{H}_f$  are the points of discontinuity of  $\mu$ , but since  $\mu(h_k) = \lim_{h \downarrow h_k} \mu(h)$ ,  $\mu$  is right continuous at these points and, therefore, everywhere. The proof of Proposition 3 is given in Appendix C.

## 2.5 Morphological Pipeline for Cell Counting

In this section we present a simple morphological algorithm for cell counting in images where individual cells are well approximated by regional extrema of significant dynamic<sup>2</sup>, like in Figure 4. We assume images to take quantized values in  $[0, V_{\max}] \cap \epsilon\mathbb{N}$  with  $V_{\max} > 0$  and  $0 < \epsilon < V_{\max}$ .

The algorithm takes as input an image  $f$ , as parameter a number  $h \geq 0$  and returns an estimated number  $n_c$  of cells after running the following steps:

1. Apply an alternate filter

$$\tilde{f} \leftarrow \varphi_B \circ \gamma_B(f)$$

where  $\varphi_B$  and  $\gamma_B$  are respectively the closing and opening by B, the unit square structuring element (see Section 2)

2. Compute the characteristic function of the  $h$ -maxima of  $\tilde{f}$ ,

$$\mathbf{I}_h \leftarrow \text{EMAX}_h(\tilde{f}) = \text{RMAX}(\text{HMAX}_h(\tilde{f})),$$

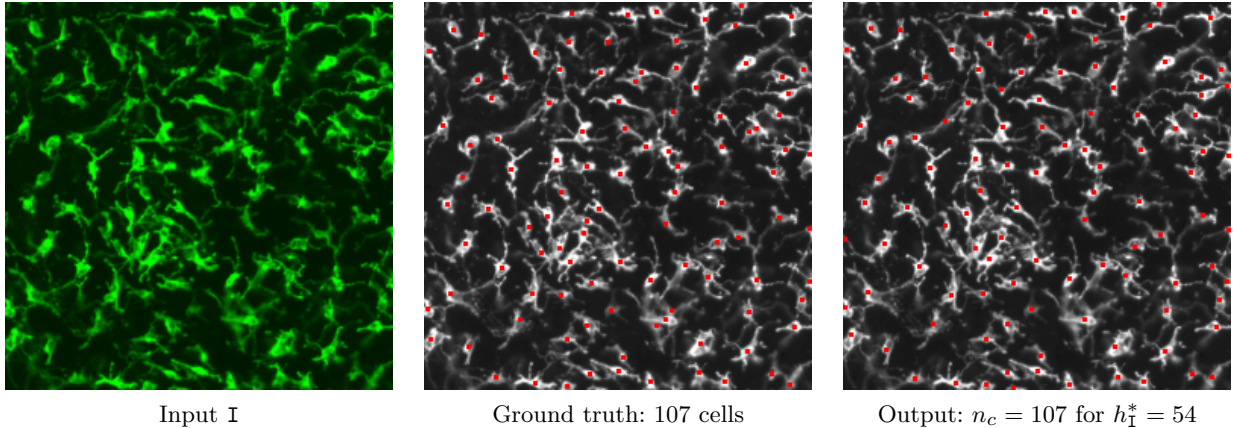
which is a binary image (see Section 2.4)

3. Apply to  $\mathbf{I}_h$  the operator CCC, which maps any binary image to its number of connected components:

$$n_c \leftarrow \text{CCC}(\mathbf{I}_h).$$

Let us briefly comment on the steps of this algorithm. The alternate filter is a very common one in morphological processing of images.

<sup>2</sup>This algorithm has already been presented in [8].



**Fig. 4** Example of cell counting by identifying cells to  $h$ -maxima, in the TRP1 melanocyte dataset (see Section 4 for the description of datasets). From left to right: input 8-bits RGB image (resized); green channel marked with ground truth location of the 107 cells (in red) ; detection results using the best  $h$  for the image:  $h_I^* = 54$ . Note that despite the perfect count, there are some false positives and false negatives (e.g. just below the top right hand corner).

The opening  $\gamma_B$  first removes brights structures smaller than the support of  $B$ , after what  $\varphi_B$  has the dual effect of filling in dark structures smaller than  $B$ . Hence the alternate filter denoises the image by suppressing small variations but preserving the shape of structures larger than  $B$ .

The second step of the algorithm detects, in the filtered image, the bright domes containing at least a regional maximum of dynamic strictly larger than  $h$ . Our hypothesis is that in such an image, and for a well chosen  $h$ , each cell contains one and only one such dome. This assumption is realistic given that cells do appear as bright shapes on a dark background, and that their brightness is fairly uniform over one image. Note that, in the present setting where  $f$  takes quantized values,  $\text{RMAX}$  can be computed with a geodesical reconstruction, following Eq. (9) and (10).

According to that hypothesis, counting the number  $n_c$  of connected components of the binary image  $\mathbf{I}_h$  yields an estimation of the number of cells in  $f$ . As we saw in Section 2, with Proposition 2, this number  $n_c$  is a decreasing step function of  $h$ , ranging from the total number of local maxima (for  $h = 0$ ) to zero (for  $h = \Delta_\Omega f$ , the dynamic of any global maximum). This is why we expect that for each input image, there is an interval of  $h$  values for which  $|\mathcal{M}_h(f)|$  is close to the right number of cells.

The goal of our method is to train end-to-end a deep learning architecture which first estimates

the right dynamic and then applies the above algorithm. This is the motivation of the next section.

### 3 Integration in a Deep Learning Architecture

The morphological pipeline described in the previous section depends on one crucial parameter, namely the minimum dynamic for a regional maximum to be considered a relevant object (in our case, a cell). Following [8], in this section we describe how this pipeline is implemented as a sequence of neural layers, and embedded in an architecture where it is preceded by a CNN which predicts an optimal dynamic for each image, in order to count the right number of cells at the output of the network. We also analyze the ability of such pipeline to be trained with gradient descent with respect to an appropriate loss, by studying the derivability of its components with respect to their parameters and input variables.

#### 3.1 Counting Connected Components (CCC) Layer

Counting connected components is an important step in many of the classical methods in data processing [16], especially in images and graphs. In the context of deep learning, this problem has been used to evaluate the generalization capacity of neural networks [17]. However, to be used

as a layer inside a network, it must be implemented so that gradient backpropagation can be computed in an auto-derivation software [18]. In this subsection we present an algorithm, already introduced in [8], meant to count the connected components of a binary image and relying mainly on the geodesic reconstruction defined in (6).

Let  $f \in \mathcal{F}(\Omega, \{0, 1\})$  be a binary image, and  $\mathcal{U}_\Omega \in \mathcal{F}(\Omega, [0, 1])$  an injective image, i.e. such that  $\forall x, y \in \Omega, x \neq y \Rightarrow \mathcal{U}_\Omega(x) \neq \mathcal{U}_\Omega(y)$ . Then it is clear, from Eq. (7), that reconstructing  $\mathcal{U}_\Omega \wedge f$  under  $f$  will label every connected component CC of  $f$  with the maximum value taken by  $\mathcal{U}_\Omega$  in CC. Since  $\mathcal{U}_\Omega$  is injective, this maximum is achieved only once in each connected component. Hence, setting to one each pixel  $x \in \Omega$  such that  $\text{REC}(\mathcal{U}_\Omega \wedge f, f)(x) = \mathcal{U}_\Omega(x)$  produces a binary image where exactly one pixel per connected component of  $f$  is lit. This is illustrated by Figure 5.

Therefore, we define

$$\mathcal{D}_f := \mathcal{B}(\mathcal{U}_\Omega, \text{REC}(\mathcal{U}_\Omega \wedge f, f)) \quad (22)$$

where the *binarization operator*  $\mathcal{B}$  for two functions  $f, g \in \mathcal{F}(\Omega, \mathbb{R})$  is:

$$\forall x \in \Omega, \quad \mathcal{B}(f, g)(x) := \begin{cases} 1 & \text{if } f(x) = g(x) \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

$\mathcal{D}_f$  denotes the final detection result, represented by locations of the isolated connected components.

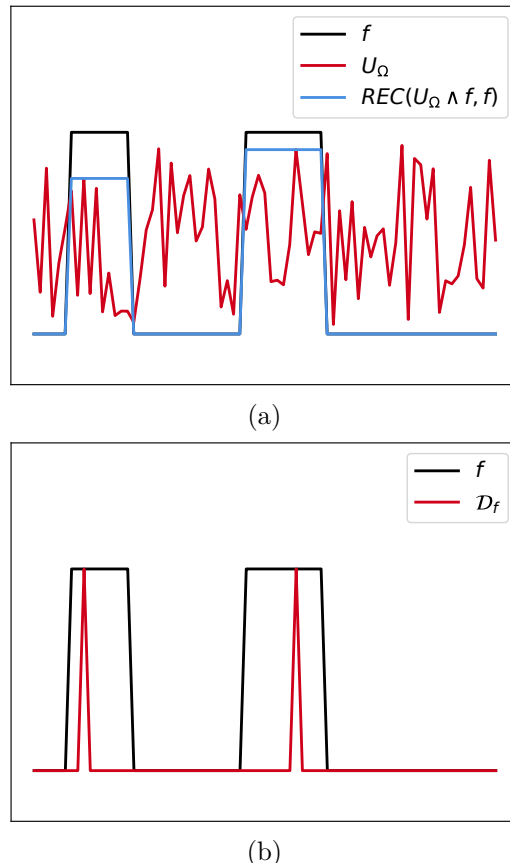
Accordingly, one can count the number of connected components on  $f$  by simply summing (22):

$$\text{CCC}(f) := \sum_{x \in \Omega} \mathcal{D}_f(x). \quad (24)$$

### 3.2 Loss Functions

The parameters of the proposed neural architecture, which are those of the CNN estimating  $h$ , are optimized by minimizing a loss function which can penalize different errors made by the network at different stages.

Since we are dealing with a counting task, a natural error to penalize is the counting error. Therefore we define the *counting loss* as the mean absolute error (MAE) between the estimated number of cells  $\widehat{\text{CCC}}_{\mathbf{I}} := \text{CCC}(\text{EMAX}_{\hat{h}_{\mathbf{I}}})$  and



**Fig. 5** Illustration of the CCC algorithm. (a)  $\mathcal{U}_\Omega$  coincides only once with  $\text{REC}(\mathcal{U}_\Omega \wedge f, f)$  inside each connected component, and never outside the connected components. (b) Hence  $\mathcal{D}_f$  has only two pixels activated, one per connected component of  $f$ .

the true number of cells  $\text{CCC}_{\mathbf{I}}^*$ , for an input image  $\mathbf{I}$ :

$$\mathcal{L}_{count}^{\mathbf{I}} := \left| \text{CCC}_{\mathbf{I}}^* - \widehat{\text{CCC}}_{\mathbf{I}} \right| = \left| \text{CCC}_{\mathbf{I}}^* - \text{CCC}(\text{EMAX}_{\hat{h}_{\mathbf{I}}}(\mathbf{I})) \right|. \quad (25)$$

In the case where we do know, for each training image  $\mathbf{I}$ , an optimal value  $h_{\mathbf{I}}^*$  for which the number of  $h$ -maxima is the closest to the number of cells, then we have two more possible errors to penalize. One is the error on the estimation of  $h$ , i.e.  $|h_{\mathbf{I}}^* - \hat{h}_{\mathbf{I}}|$ , and the other one is the *reconstruction loss*, defined as the mean squared error (MSE) between the predicted geodesic reconstruction image  $\text{HMAX}_{\hat{h}_{\mathbf{I}}}$  and the ground truth one

$\text{HMAX}_{h_I^*}$ :

$$\mathcal{L}_{rec}^I = \frac{1}{|\Omega|} \sum_{x \in \Omega} (\text{HMAX}_{h_I^*}(x) - \text{HMAX}_{\hat{h}_I}(x))^2. \quad (26)$$

Yet another possibility is the binary cross entropy between  $\text{EMAX}_{\hat{h}_I}$  and  $\text{EMAX}_{h_I^*}$ ,  $\text{BCE}(\text{EMAX}_{\hat{h}_I}, \text{EMAX}_{h_I^*})$ . Note that these loss terms are less and less constraining as we move from  $h$  to the output count, in the sense that

$$\begin{aligned} |h_I^* - \hat{h}_I| = 0 &\Rightarrow \mathcal{L}_{rec}^I = 0 \\ &\Rightarrow \text{BCE}(\text{EMAX}_{\hat{h}_I}, \text{EMAX}_{h_I^*}) = 0 \\ &\Rightarrow \mathcal{L}_{count}^I = 0 \end{aligned} \quad (27)$$

whereas none of the converse implications is true. In the present paper we focus on the counting and reconstruction losses, as previous experiments presented no improvement in including the other two terms.

Hence the final loss we consider is a joint loss of the form

$$\mathcal{L}^I = \alpha \mathcal{L}_{rec}^I + \beta \mathcal{L}_{count}^I \quad (28)$$

where  $\alpha \geq 0$  and  $\beta \geq 0$  control the importance of each of the terms, and need to be adjusted as hyperparameters (see Section 4).

We shall stress on the fact that, while both terms can help the optimization of the parameters, in practice the reconstruction loss is much more constraining as it requires the computation of an optimal  $h$  and its corresponding  $\text{HMAX}_h$  reconstruction, for each training image. This implies that we know, *before training*, the grayscale images that will be passed to the  $h$ -estimator CNN. In particular, this prevents from learning an optimal preprocessing of images before the estimation of  $h$ , *e.g.* a trainable filtering that would replace the proposed alternate filter  $\varphi_B \circ \gamma_B$ , and/or a trainable mapping from three-channel to one-channel images.

Therefore, training our pipeline by minimizing the counting loss only (*i.e.*,  $\alpha = 0$ ) is highly desirable. It is one of the targets of the present paper, as an improvement to [8].

### 3.3 Derivability of the Pipeline

In order to train the proposed pipeline with a gradient descent algorithm, we need to check that a

gradient can flow during backpropagation, allowing for the update of the CNN parameters. The alternate filter is non-trainable and applied to the input *before* the CNN, so it is not involved in the backpropagation. The CNN itself does not raise any particular problem compared to usual deep networks for images. The next step takes as input the  $h$  value estimated by the CNN, to compute the  $h$ -transformation  $\text{HMAX}_h(\tilde{f})$ , which should therefore be viewed as a function of  $h$  for a fixed input image  $\tilde{f}$ . The last step consists in counting the number of regional maxima of  $\text{HMAX}_h(\tilde{f})$ , *i.e.* applying  $\text{CCC} \circ \text{RMAX}$  to  $\text{HMAX}_h(\tilde{f})$ . Let us now analyze the derivability of these last two steps.

#### 3.3.1 Derivative of $h \mapsto \text{HMAX}_h(f)$

Here we consider a fixed  $f \in \mathcal{F}(\Omega, \mathbb{R})$  and see  $\text{HMAX}_h(f)$  as a function of  $h$ . Recall that we denote by  $\mathcal{M}_h(f)$  the set of  $h$ -maxima of  $f$  (see Section 2). Then we can state the following result.

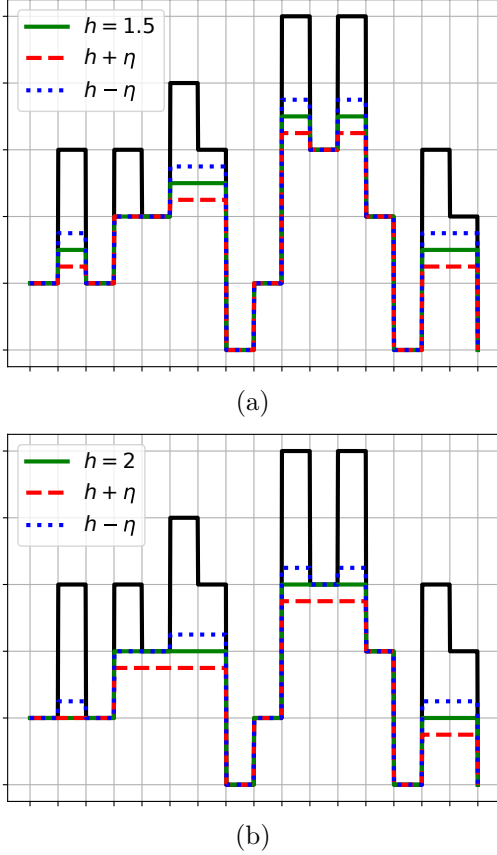
**Proposition 4** *Let  $h \geq 0$ , and assume there exists an interval  $I$  with  $0 \in I$ , such that for any  $\eta \in I$ ,  $\mathcal{M}_{h+\eta}(f) = \mathcal{M}_h(f)$ . Then we have*

$$\forall \eta \in I,$$

$$\text{HMAX}_{h+\eta}(f) = \text{HMAX}_h(f) - \eta \text{EMAX}_h(f). \quad (29)$$

The interpretation of Proposition 4 is the following. First, the assumption that  $\mathcal{M}_{h+\eta}(f) = \mathcal{M}_h(f)$  for  $\eta \in I$  just means that, for the considered  $f$  and value  $h$ , the set of  $h$ -maxima of  $f$  remains the same if we change  $h$  by a small enough (but non zero) quantity, in at least one direction, positive or negative. We will refer to this concept as the stability of  $h$ -maxima. Figure 6 shows an example where  $I$  is open, in (a), thus the stability holds in both directions, and an example where  $0 = \min I$ , in (b), where the stability holds only for non negative  $\eta$ .

Second, Equation (29) indicates that, under the condition of stability of  $h$ -maxima, a small change in  $h$  only changes the height of the  $h$ -maxima, and leaves the rest of  $\text{HMAX}_h(f)$  unchanged, as shown in Figure 6. The proof of Proposition 4 is provided in Appendix D.



**Fig. 6** Illustration of Proposition 4, showing the changes from  $\text{HMAX}_h(f)$  to  $\text{HMAX}_{h\pm\eta}(f)$  for a function  $f$  with integer values (black line), and two different  $h$ . In (a),  $f$  and  $h$  fulfill the stability condition of the proposition in both positive and negative directions, therefore a shift of  $\pm\eta$  occurs exactly at the  $h$ -maxima (regional maxima of the green function), confirming Eq. (29). In (b) however, the stability holds for nonnegative  $\eta$ , but the  $h$ -maxima vary even for the smallest negative  $\eta$ . Hence in the latter case the reconstructions vary outside the  $h$ -maxima and remain unchanged in some parts of these, contradicting (29).

Proposition 4 leads to the following consequence regarding the Fréchet derivability of  $h \mapsto \text{HMAX}_h(f)$ .

**Corollary 5** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  and  $\mathcal{H}_f = \{h_0, h_1, \dots, h_{K+1}\}$  be the finite subset of  $[0, \Delta_\Omega f]$  containing the points of discontinuity of  $h \mapsto \mathcal{M}_h(f)$ , as introduced by Proposition 3. Then the function  $h \mapsto \text{HMAX}_h(f)$  is differentiable at  $h^*$  if and only if  $h^* \notin \mathcal{H}_f$ , and in that case its derivative is*

$$\frac{\partial \text{HMAX}_{h^*}(f)}{\partial h} = -\text{EMAX}_{h^*}(f). \quad (30)$$

The non-derivability of  $h \mapsto \text{HMAX}_h(f)$  at any  $h \in \mathcal{H}_f$  appears clearly on the example of Figure 6 (b), where  $h^* = 2$ . In this case, on the one hand we do have, for  $0 \leq \eta < 1$ ,

$$\text{HMAX}_{h^*+\eta}(f) = \text{HMAX}_{h^*}(f) - \eta \text{EMAX}_{h^*}(f).$$

Therefore, if  $h \mapsto \text{HMAX}_h(f)$  was Fréchet differentiable at  $h^* = 2$ , its derivative would be  $\frac{\partial \text{HMAX}_{h^*}(f)}{\partial h} = -\text{EMAX}_{h^*}(f)$ . But on the other hand we have, for  $-1 < \eta < 0$ ,

$$\begin{aligned} \text{HMAX}_{h^*+\eta}(f) &= \text{HMAX}_{h^*}(f) - \eta \text{EMAX}_{h^*}(f) \\ &\quad - \eta \left( \text{EMAX}_{h^*+\eta}(f) - \text{EMAX}_{h^*}(f) \right) \end{aligned}$$

where  $\text{EMAX}_{h^*+\eta}(f) - \text{EMAX}_{h^*}(f) = \mathbf{1}_{\{1\}} - \mathbf{1}_E$ , with  $E = \{3, 4, 10\}$  if we index pixels from left to right and starting from zero. Since that quantity does not depend on  $\eta$ , it does not converge to the null function when  $\eta$  goes to zero from below, which contradicts the differentiability of  $h \mapsto \text{HMAX}_h(f)$  at  $h^* = 2$ . The general case is handled likewise in the proof of Corollary 5, in Appendix E.

It is clear from this example, and more formally from Proposition 1, that discontinuities in  $h \mapsto \mathcal{M}_h(f)$  can only be caused by discontinuities of the upper levelset function  $t \mapsto \chi_t(f)$ . For functions  $f$  taking quantized values (like in Figure 6), we have the guarantee that the level sets remain constant in between these values, and that therefore  $h \mapsto \text{HMAX}_h(f)$  is differentiable in the corresponding intervals.

### 3.3.2 Derivability of $\text{CCC} \circ \text{RMAX}$

As explained earlier, the operator  $\text{CCC} \circ \text{RMAX}$  consists in counting the regional maxima of a function (in our case,  $\text{HMAX}_h(f)$ ),  $\text{RMAX}$  returning the characteristic function of the regional maxima, and  $\text{CCC}$  the one counting the connected components of a binary image.

But, as seen in the previous paragraph, this number of  $h$ -maxima remains constant over intervals  $[h_k, h_{k+1})$ , and therefore the derivative of  $\text{CCC} \circ \text{RMAX}$  at  $\text{HMAX}_h(f)$ ,  $h \notin \mathcal{H}_f$  is zero.

On the other hand, when  $h \in \mathcal{H}_f$ , a discontinuity in the  $h$ -maxima occurs, causing a non-derivability of  $\text{CCC} \circ \text{RMAX}$  at  $\text{HMAX}_h(f)$  since, by Propositions 2 and 3, a change in  $\mathcal{M}_h(f)$  also

changes its cardinality. When relying on auto-differentiation with the algorithm described in Section 3.1, we rather observe exploding gradients.

Therefore, in order to avoid vanishing as well as exploding gradients, we propose to use explicit (approximations of) derivatives, as explained next.

### 3.4 Modified Backpropagation

Let us consider the setting where the training is done by minimizing the joint loss (28) with  $\beta > 0$ , that is to say including the counting loss term. As we have just shown in the previous section, numerical issues such as vanishing or exploding gradients are expected when relying on the auto-differentiation of (25), and therefore of  $h \mapsto |\mathcal{M}_h(f)| = \text{CCC} \circ \text{RMAX} \circ \text{HMAX}_h(f)$ . In our experiments, significant optimization difficulties have been observed, especially when the counting loss term was weighted similarly as the reconstruction term, that is  $\beta \approx \alpha$ . We overcame these difficulties by using derivatives of smooth approximations of the function  $h \mapsto |\mathcal{M}_h(f)|$ .

We propose two approximations, which were both tested in our experiments (see Section 4). The first one, more naive, is only based on the fact that  $|\mathcal{M}_h(f)|$  is a decreasing function of  $h$ . Without any further knowledge, we assume

$$|\mathcal{M}_h(f)| \approx -h + cst \quad (31)$$

which gives

$$\frac{\partial |\mathcal{M}_h(f)|}{\partial h} \approx -1. \quad (32)$$

We denote by (MB, -1) the class of models trained with this approximation, MB standing for Modified Backpropagation.

The second approximation, more sophisticated and realistic, assumes

$$|\mathcal{M}_h(f)| \approx \sum_{x \in \Omega} \mathcal{I}_N(x) \quad (33)$$

where, for some positive integer  $N$ ,

$$\mathcal{I}_N = \left( \frac{f - \text{HMAX}_h(f)}{h} \right)^N. \quad (34)$$

The justification for this approximation is given by Corollary 7, based on Proposition 6 here below.

We first introduce the notation

$$\overline{\mathcal{D}}_h(f) := \{M_0 \in \mathcal{M}(f), \text{DYNMAX}(M_0) \geq h\} \quad (35)$$

for the set of regional maxima of  $f$  with dynamic not smaller than  $h$ . We can now state the following proposition, proved in Appendix F.

**Proposition 6** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  non constant, and  $h \in (0, \Delta_\Omega f]$ . Then, for any  $x \in \Omega$ ,*

$$\begin{aligned} \text{HMAX}_h(f)(x) = f(x) - h &\iff \\ \exists M_0 \in \overline{\mathcal{D}}_h(f), x \in M_0. &\quad (36) \end{aligned}$$

Given that, additionally, we have by definition  $f \geq \text{HMAX}_h(f) \geq f - h$ , it follows that for any  $x \in \Omega$  and  $h > 0$ ,

$$0 \leq \frac{f(x) - \text{HMAX}_h(f)(x)}{h} \leq 1 \quad (37)$$

with equality to 1 if and only if  $x \in M_0$  for some  $M_0 \in \overline{\mathcal{D}}_h(f)$ . Hence, when  $N$  goes to infinity, we get

$$\lim_{N \rightarrow +\infty} \sum_{x \in \Omega} \mathcal{I}_N(x) = \sum_{M_0 \in \overline{\mathcal{D}}_h(f)} \mathbf{1}_{M_0}. \quad (38)$$

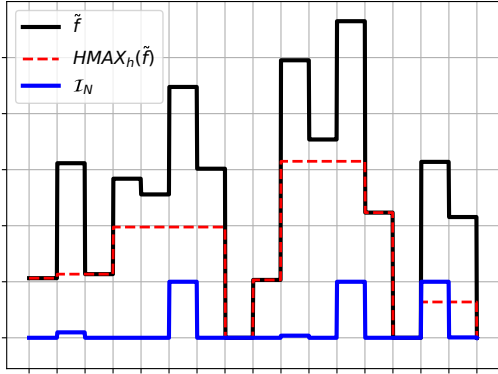
In words, when  $N$  is large enough, the sum in (33) approximates the number of pixels belonging to regional maxima of dynamic non smaller than  $h$ .

Now, note that for  $h \notin \mathcal{H}_f$ , as introduced by Proposition 3, there is no regional maximum  $M_0$  with dynamic exactly  $h$ . Indeed, if there was one, by Proposition 1 we would have  $M_0 \notin \mathcal{M}_h(f)$ , but  $M_0 \notin \mathcal{M}_{h-\epsilon}(f)$  for any  $\epsilon > 0$  (smaller than  $h$ ), so  $h$  would be a discontinuity point of  $h \mapsto |\mathcal{M}_h(f)|$ , i.e in  $\mathcal{H}_f$ . Therefore, for such  $h$  we have  $\overline{\mathcal{D}}_h(f) = \mathcal{D}_h(f)$ , where we recall that

$$\mathcal{D}_h(f) := \{M_0 \in \mathcal{M}(f), \text{DYNMAX}(M_0) > h\}$$

is the set of regional maxima of  $f$  with dynamic larger than  $h$ , already defined in (C8). Furthermore, Proposition 1 tells us that  $M_0 \in \mathcal{D}_h(f)$  if and only if  $M_0$  is a connected component of  $\arg \max_M f$  for some  $M \in \mathcal{M}_h(f)$ . Thus, for  $h \notin \mathcal{H}_f$ , we get

$$\sum_{M_0 \in \overline{\mathcal{D}}_h(f)} \mathbf{1}_{M_0} = \sum_{M \in \mathcal{M}_h(f)} \mathbf{1}_{\arg \max_M f}. \quad (39)$$



**Fig. 7** Illustration of Corollary 7. The function  $\tilde{f}$  is defined by adding some Gaussian noise (zero mean, 0.2 standard deviation) to the function  $f$  of Figures 3 and 6. Here  $h = 2.5 \notin \mathcal{H}_{\tilde{f}}$  and  $N = 10$ . The value to be approximated is  $|\mathcal{M}_h(f)| = 3 = |\mathcal{M}_h(\tilde{f})|$ , and we get  $\sum_{x \in \Omega} \mathcal{I}_N \approx 3.14$ , where  $\mathcal{I}_N$  is computed for  $\tilde{f}$ , which is injective.

As a consequence, if we further assume that in every  $h$ -maximum  $M$ , the set  $\arg \max_M f$  contains only one pixel, then when  $N$  goes to infinity the sum in (33) does provide the exact number of  $h$ -maxima. We have just proved the following corollary.

**Corollary 7** Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  be injective,  $h \in [0, \Delta_{\Omega} f] \setminus \mathcal{H}_f$ , and  $(\mathcal{I}_N)_{N \geq 1}$  the sequence of images defined by (34). Then

$$\lim_{N \rightarrow +\infty} \sum_{x \in \Omega} \mathcal{I}_N(x) = |\mathcal{M}_h(f)|. \quad (40)$$

Note that the approximation is quite accurate as soon as we add some small random noise to  $f$ , preventing any two pixels to have the same value, and for large enough  $N$ , as illustrated by Figures 7 and 8.

This leads to the following approximation of the derivative

$$\frac{\partial |\mathcal{M}_h(f)|}{\partial h} \approx \sum_{x \in \Omega} \mathcal{I}'_N(x) \quad (41)$$

with

$$\mathcal{I}'_N = -N \cdot \frac{\rho(f, h)^{N-1}}{h^N} \left( \frac{\rho(f, h)}{h} - \text{EMAX}_h(f) \right) \quad (42)$$

and  $\rho(f, h) = f - \text{HMAX}_h(f)$ . In the experimental section we will refer to this approximation as (MB, N).

## 4 Experiments

### 4.1 Data

To evaluate the benefits of our proposition, three databases are used. It is important to note, that the three types of images considered, have as common factor that the cells of interest are bright objects on dark backgrounds, for which using the dynamic for cell counting makes sense. In the three considered datasets, the input image is transformed to a single input channel, and then rescaled at a resolution of  $256 \times 256$  pixels. The image range was between 0 and  $V_{\max} = 255$ , quantized with a float precision  $\epsilon$ . Hence the previously described method applies, taking binary images with values in  $\{0, 255\}$ . Applying the morphological pipeline described in Section 2.5 for all the integer values  $h$  in  $[0, \Delta_{\Omega} f]$ , allowed to determine, for each training image  $I$ , an optimal dynamic  $h_{\text{T}}^*$ , defined as the lower bound of the interval of  $h$  values minimizing the difference between the true number of cells and the estimated one.

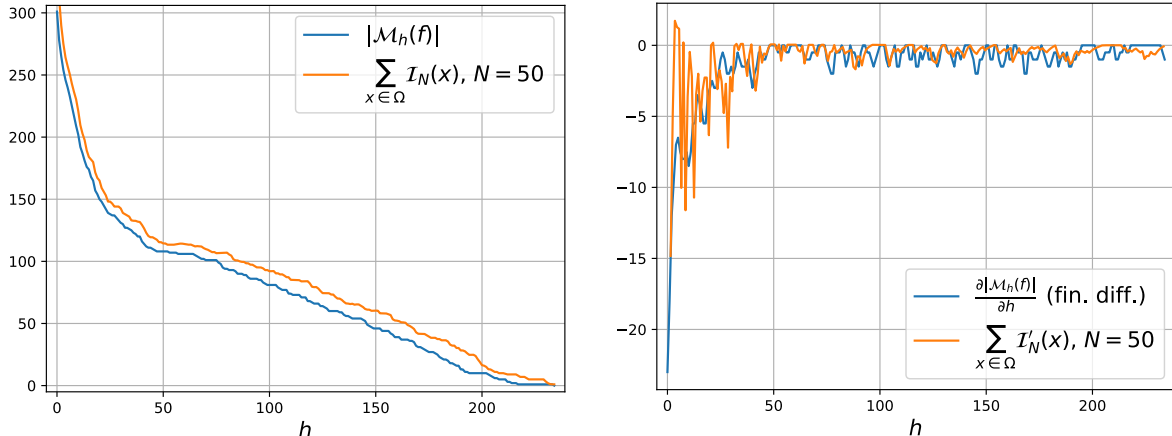
#### 4.1.1 TRP1 Dataset

The TRP1 dataset [3] we used contains two sets, called *set1* and *set2*, of 76 fluorescent melanocytes RGB images each, with resolution  $1024 \times 1024$  pixels, which showed a high variability in density and shapes<sup>3</sup>. They come along with manual annotations of the cell coordinates. The images were acquired from *in vitro* pigmented reconstructed skin samples submitted to TRP1 fluorescent labeling [19]. Figures 4 and 14 show examples of images and manual annotations.

#### 4.1.2 Fluorescent Neuronal Cells Dataset

The Fluorescent Neuronal Cells dataset [4] comprises 283 fluorescence microscopy images of neuron slices from the mice brain, with a resolution of  $1600 \times 1200$  pixels. These are accompanied by the corresponding ground truth cell location labels.

<sup>3</sup>The dataset is available at <https://bit.ly/melanocytesTRP1>.



**Fig. 8** Illustration of Equations (33) and (41), where  $f$  is the preprocessed green channel of the cell image of Fig. 4. **Left:** The actual number of  $h$ -maxima is plotted in blue as a function of  $h$ , whereas its approximation  $\sum_{x \in \Omega} \mathcal{I}_N(x)$  is plotted in

orange. **Right:** In blue we plot an estimation of  $\frac{\partial |\mathcal{M}_h(f)|}{\partial h}$  obtained by finite differences  $0.5 \times (|\mathcal{M}_{h+1}(f)| - |\mathcal{M}_{h-1}(f)|)$ . The orange curve is the approximation  $\sum_{x \in \Omega} \mathcal{I}'_N(x)$ .  $\mathcal{I}_N$  and  $\mathcal{I}'_N$  are computed for  $N = 50$  and the perturbed injective function  $\tilde{f}$  is obtained by adding Gaussian noise to  $f$  with standard deviation  $\sigma = 0.2h$ .

The dataset presents a number of challenges for the cell counting task. Firstly, the cell shapes are variable, the contrast of the cells relative to the background is variable, and the backgrounds themselves exhibit some variation in brightness. Example of input image are shown in Fig. 15. In addition, the images contain biological structures that are not neuronal cells, and the number of cells contained in the images is highly variable, ranging from none to dozens of cells, with instances of overlapping clasping between cells. These factors combine to present significant challenges for accurate cell counting.

#### 4.1.3 Subset of Cellpose Dataset

Originally used for generalist segmentation model training and evaluation, the Cellpose dataset [20] comprises a diverse array of microscopic cell images sourced from multiple avenues, including previous research and online searches. Furthermore, the dataset includes a proportion of non-cell images, which serve to enhance the model’s generalization capability. The dataset comprises a total of 616 images, of which 69 have been reserved for testing purposes. Each image is accompanied by ground truth segmentation labels.

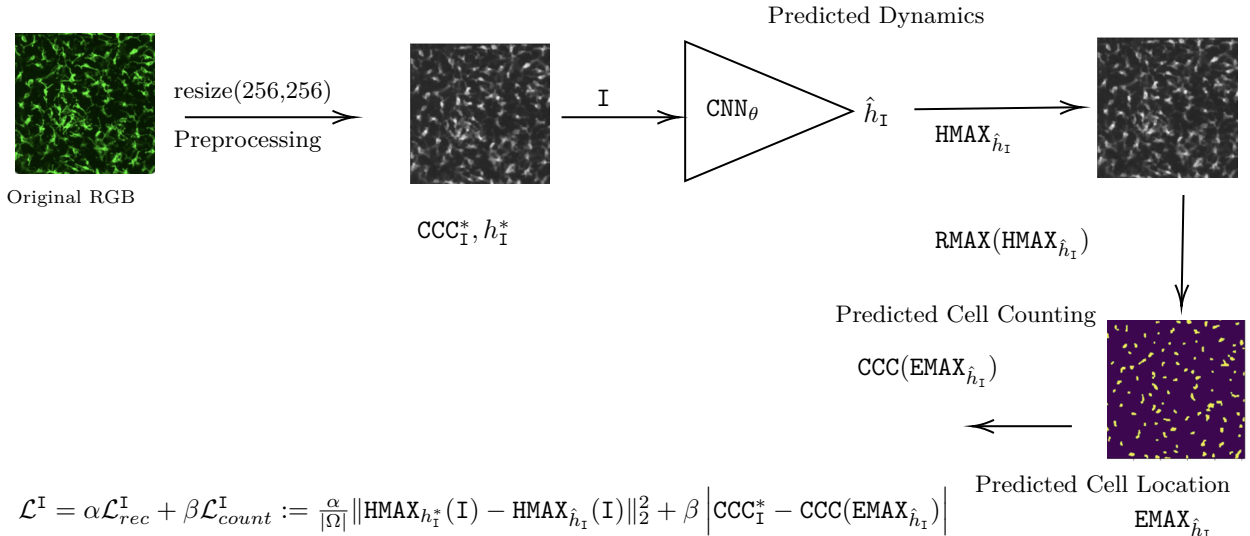
As our method is constrained to images featuring bright objects in darker backgrounds, we manually selected a subset of the entire dataset,

comprising images that align with this hypothesis. The selected fluorescent microscopic images typically feature nuclei that have been stained with DAPI and proteins that have been labeled with a fluorescent substance in the cytoplasm, the two stains are represented in different color channels. The examples of the selected cell images can be illustrated in left column in Fig. 16. Finally, 213 data samples are selected for training, validation and test.

## 4.2 Overall Architecture

The proposed neural architecture<sup>4</sup> is illustrated in Figure 9. Morphological layers apply an alternate filter (opening followed by closing) by a  $3 \times 3$  square structuring element, to the resized green channel, which yields the first preprocessed image  $I$ . This image is passed as input to a CNN, which estimates the best dynamic  $\hat{h}_I = \text{CNN}_\theta(I)$ , where  $\theta$  denotes the parameters of the CNN. This parameter is then used to compute the  $\hat{h}_I$ -maxima of the filtered image, by first feeding it to the HMAX reconstruction layer, followed by the RMAX one to get the binary image  $\text{EMAX}_{\hat{h}_I}$ . Finally, the number of connected components of the latter is obtained at the output of the proposed connected components

<sup>4</sup>Code available at [https://github.com/samyblusseau/trainable\\_hmaxima](https://github.com/samyblusseau/trainable_hmaxima).



**Fig. 9** Proposed end-to-end differentiable pipeline using a CNN and an HMAX layer to cell counting. The network is trained by minimizing a loss function that depends on 1) the difference between the geodesic reconstruction with the dynamic  $\hat{h}$  estimated by the network, and the best possible one; and 2) the difference between the number of cells and the number of regional maxima of  $\text{HMAX}_{\hat{h}}$ .

counting layer,  $\text{CCC}(\text{EMAX}_{\hat{h}_I})$ . The CNN architecture is summarized in Fig. 10. In particular, each convolutional layer<sup>5</sup> is followed by a Batch Normalization layer with ReLU as activation function. Following [11], global average pooling layer is used after the second Maxpooling layer instead of stacked flatten and fully connected layers, which has the advantage of further reducing the number of parameters of the network. What’s more, extracting global feature information is coherent with the objective of our task.

Regarding the theoretical complexity of this architecture, for the chosen fixed size of kernels and structuring elements, the CNN part of the pipeline has linear complexity in the number  $n$  of input pixels, because convolutions do. A dilation also has linear complexity but the geodesical reconstruction might require as many as  $\sqrt{n}$  dilations, each one followed by a pixelwise minimum and an image comparison test after each dilation. Therefore with our implementation (which does not use special data structures to reduce complexity) the morphological part has a much higher complexity of  $O(n^{3/2})$  in the worst case, which determines the complexity of the whole pipeline.

<sup>5</sup>Convolutional layers are implemented as a double convolution using the same kernel size for both convolutions, with a Batch Normalization layer in between.

Despite this asymptotic result, it is worth noting that the actual number of operations performed by the CNN is about  $5000 \times n = 5000 \times 256^2$ , and the morphological part performs two geodesical reconstructions, resulting in at most about  $20\sqrt{n} \times n \approx 5000 \times 256^2$  operations as well. So one could expect the CNN part to take about the same time as the morphological one during inference. In practice however, we observe that the morphological part takes from 13 to 29 times longer than the CNN on average, depending on the dataset (although all input images are resized at the fixed input size  $256 \times 256$  pixels). This could be explained by the fact that, within the machine learning framework we use (Tensorflow), convolutions are optimized way better than our custom geodesical reconstruction, which relies on a while loop and an image comparison after each geodesical dilation.

### 4.3 Training Protocol

**TRP1 dataset:** To train the architecture presented earlier, hence fit the parameters of the CNN, we randomly split *set1*, containing 76 images, into 60 training images and 16 validation images. We only worked with the green channel of image images. The parameters of the CNN were adjusted

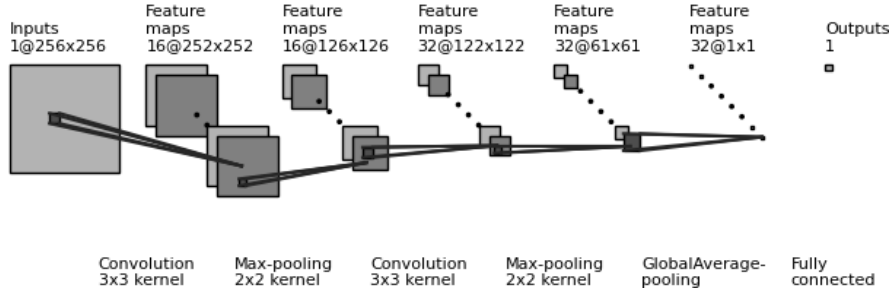


Fig. 10 CNN architecture used in the proposed pipeline.

through gradient descent on the training set, and the best model weights were chosen through evaluation metrics on the validation set. The whole *set2*, which also contains 76 images, was used for the final test only.

**Fluorescent Neuronal Cells dataset:** The same test set used in [4] is employed in our experiments, comprising 70 selected images. The remaining data is randomly split into a training set (80%) and a validation set (20%). For each fluorescent image, we converted it to grayscale using the linear luminance formula defined as  $0.2126R + 0.7152G + 0.0722B$ , where  $R$ ,  $G$ , and  $B$  correspond to the red, green, and blue channels, respectively. The image was then resized to a resolution of  $256 \times 256$  for input to the CNN.

**Subset of Cellpose dataset:** In order to keep consistency with the original Cellpose dataset, among the 213 selected data samples, we keep the ones that originally in the test split of Cellpose dataset as test set, the rest ones are randomly split into train and validation set with ratio 8:2, resulting in 158 samples for training, 39 for validation and 16 for test. For each input image, the maximum value for the RGB channels for each pixel is extracted in order to convert the input image into a single channel. Prior to inputting the CNN, the same resizing process is applied to the resolution of  $256 \times 256$ .

### 4.3.1 Data Augmentation

Convolutional operations are translation invariant, but not rotation or flip invariant. Moreover, for the image samples in the dataset, the flipped or rotated version still looks like a valid sample. Thus it is reasonable to use flip and rotation as data augmentation methods in the training stage. More precisely, we used both vertical and

horizontal flip transformations with probability  $1/2$ ; a rotation with the same probability and the rotation angle was uniformly selected from  $\{90^\circ, 180^\circ, 270^\circ\}$ . The true values of the number of cells, and the optimal value of dynamics, are not changed by this augmentation.

### 4.3.2 Optimization

The joint loss function was minimized with stochastic gradient descent. More precisely, we used the RMSProp optimizer [21], with initial learning rate of  $10^{-3}$ . Moreover, we used a batch size of 15 images for TRP1 and Cellpose, 17 for Fluorescent Neuronal Cells, and trained for a maximal number of 1600 epochs with early stopping as soon as there was no improvement on the validation loss for 300 epochs. Only the model weights with lowest validation error were saved and used for final evaluation on the test set.

### 4.3.3 Hyperparameters

**Loss weights  $(\alpha, \beta)$ :** In [8], experiments with the joint loss (28) had been carried out but without the modified backpropagation introduced in Section 3.4. Five-folds cross validation experiments on *set1*, with 11 ratios  $\beta/\alpha$  tested (ranging from  $10^{-4}$  to  $10^2$ , including zero<sup>6</sup>) had indicated that best results were expected for  $\alpha = 1, \beta = 0.001$ , that is to say for very little weight given to the counting loss. We now have an explanation for this, as we showed in Section 3.3 that derivating the reconstruction loss with respect to  $h$  is not an issue in general, but derivating the counting loss is. On the contrary, with the modified backpropagation, we observed no significant effect of

<sup>6</sup>Tested values:  $\alpha = 1$  and  $\beta \in \{0\} \cup 10^{\{-4, -3, -2, -1, 0, 1, 2\}} \cup 5 \times 10^{\{-4, -3, -2\}}$ .

the balance between  $\alpha$  and  $\beta$ : ( $\alpha = 1, \beta = 0.001$ ) gave similar performances on the validation sets as  $\alpha = \beta = 0.5$  and as ( $\alpha = 0, \beta = 1$ ). Therefore, we chose to set  $\alpha = \beta = 0.5$ .

**Asymptotic value  $N$ :** For the experiments where the modified backpropagation was defined following (41) and (42), the value of  $N$  was chosen by testing on training images the trade-off between the quality approximation of the number of  $h$ -maxima given by (33), and its smoothness. Indeed, a large  $N$  gives a better approximation of  $h \mapsto |\mathcal{M}_h(f)|$ , but also makes this approximation converge to a step function with zero derivatives almost everywhere, which is what we want to avoid. This study led us to choose  $N = 50$ , therefore the class of models trained with this modified backpropagation is denoted (MB, N=50).

**Random noise to compute  $\mathcal{I}'_N$ :** Recall that the approximation (41) of  $\frac{\partial |\mathcal{M}_h(f)|}{\partial h}$  relies on Corollary 7, which holds for injective images. To obtain almost surely an injective version  $\tilde{f}$  of an image  $f$ , possibly non injective, we add some random noise to the latter. The level of noise must be small enough not to change (too much) the number of  $h$ -maxima, but large enough to ensure a sufficient difference between the maximum value and the second largest value within a  $h$ -maximum, hence ensuring a quicker convergence in (40). Again, trials on the training data led to the choice of adding Gaussian noise with  $\sigma_h = 0.2h$  as standard deviation, where  $h$  is the value at which we want to approximate  $\frac{\partial |\mathcal{M}_h(f)|}{\partial h}$ .

#### 4.4 Evaluation Metrics

The following metrics have been used to evaluate our proposed pipeline. The *average relative error*, which averages the per-image relative counting error:

$$\mathcal{A}_{err}(\mathcal{S}) = \frac{1}{\mathcal{N}} \sum_{\mathbf{I} \in \mathcal{S}} \frac{|\text{CCC}_{\mathbf{I}}^* - \widehat{\text{CCC}}_{\mathbf{I}}|}{\text{CCC}_{\mathbf{I}}^*} \quad (43)$$

where  $\mathcal{S}$  denotes the test set with sample number  $\mathcal{N}$ . Considering the variability in cell numbers of the considered datasets, we also monitor the *total relative error* [2, 3], which sees the whole dataset as one skin sample, but avoids compensations of errors by taking an absolute difference in each

image:

$$\mathcal{T}_{err}(\mathcal{S}) = \frac{\sum_{\mathbf{I} \in \mathcal{S}} |\text{CCC}_{\mathbf{I}}^* - \widehat{\text{CCC}}_{\mathbf{I}}|}{\sum_{\mathbf{I} \in \mathcal{S}} \text{CCC}_{\mathbf{I}}^*}. \quad (44)$$

Additionally, for the comparison purpose, we use mean absolute error (MAE) and mean percentage error (MPE), recommended in [4]:

$$\text{MAE}(\mathcal{S}) = \frac{1}{\mathcal{N}} \sum_{\mathbf{I} \in \mathcal{S}} |\text{CCC}_{\mathbf{I}}^* - \widehat{\text{CCC}}_{\mathbf{I}}|, \quad (45)$$

$$\text{MPE}(\mathcal{S}) = \frac{1}{\mathcal{N}} \sum_{\mathbf{I} \in \mathcal{S}} \frac{\text{CCC}_{\mathbf{I}}^* - \widehat{\text{CCC}}_{\mathbf{I}}}{\text{CCC}_{\mathbf{I}}^*} \quad (46)$$

It is worth noting that finding the right number of cells does not imply correct detections, as false positives can compensate false negatives, like illustrated in Figure 4. However, we have deliberately limited our analysis to counting errors, as counting is the primary objective. All counting methods benefit from this type of compensation, which is only possible if the method is not overly biased toward over or underestimation.

## 5 Results

The results of our approach on the three test sets are summarized in Table 1, and compared to the state of the art for the TRP1 and the Fluorescent Neural Cells datasets. Since Cellpose is traditionally a segmentation benchmark, no published counting method was available (*i.e.* a method which does not rely on the ground truth masks of cells, but at most on their coordinates). For this reason, we followed the method described in [3] and trained a Unet to predict density maps, as a baseline for this dataset. More details for Unet architecture and training can be found in Appendix G.

Each of our models was trained with one of the two different modified backpropagations introduced in Section 3.4, namely (MB,-1) and (MB, N=50), and either with the joint loss ( $\alpha = \beta = 0.5$ ) or the counting loss only ( $\alpha = 0$ ). For each of these alternatives, ten models were trained following the protocol described in Section 4.3. We provide the average metric on the test sets  $\pm$  the standard deviation over the ten training instances. More detailed results of the best performing models are plotted in Figures 11, 12 and 13, as well as

some examples of their predictions in Figures 14, 15 and 16.

Based on these experiments, the following observations can be made. Firstly, our models achieve competitive results compared to the state of the art, sometimes surpassing it, with the advantage of being much lighter (50 to 100 times fewer parameters) and more interpretable: what is detected as a cell is a  $h$ -maximum, therefore one knows that missed or spurious detections are either due to a poor estimation of  $h$ , or to a deviation from our assumption of a one to one correspondence between cells and  $h$ -maxima. The robustness of our method also appears in Figures 11-13, where one can see that accurate count estimations can be achieved despite less accurate estimations of the optimal  $h$ . This is because the number of  $h$ -maxima is a step function of  $h$ , therefore the optimal number is reached for an interval of  $h$  values, usually wider as  $h$  is large. Furthermore, values outside that interval but close to it often give close to optimal count.

Secondly, no striking difference in performance appears between models trained with one or the other proposed modified backpropagation, nor between models trained with joint loss or counting loss only. It is remarkable in particular that training models with either the joint loss or the counting loss only, without the need for ground truth cell coordinates, converges well and consistently across different training instances. It is also interesting to note that the most naive approximation  $\frac{\partial |\mathcal{M}_h(f)|}{\partial h} = -1$ , implemented in the (MB, -1) models, seems to work as well, overall, as the more sophisticated and accurate one, (MB, N=50). This can be explained by the fact that the actual derivative is not far from  $-1$  in the range of  $h$  values of interest, usually in the middle part of  $[0, \Delta_\Omega(f)]$ . Besides its simplicity of implementation, the interest of (MB,-1) lies in a significantly reduced computation time during backpropagation (about 20%) compared to (MB,N=50).

Thirdly, the limitations of our approach are well illustrated by the examples of Figures 14-16. We can see that assuming a one-to-one correspondence between cells and  $h$ -maxima, for a single  $h$  value per image, leads to some over detections when non-cell objects happen to match this criterion, and missed cells especially when clusters

of cells appear as one large maximum and cannot be distinguished at the considered resolution and after the applied preprocessing (see in particular the examples from the Fluorescent Neural Cells dataset in Figure 15). We shall also comment on the limitation of the Unet trained on Cellpose as a baseline. It is worth noting that its architecture and training configurations were inspired by the study [3], designed for TRP1, and was not optimized for the Cellpose dataset. We decided to keep it for comparison, to show that the task is not straightforward.

## 6 Conclusion

In this paper, we have built on the recently introduced end-to-end trainable pipeline [8] where a CNN predicts the dynamic threshold for  $h$ -maxima to be recognized as melanocytes, and a layer counts the connected components of the  $h$ -maxima images, yielding an estimated number of cells. As reported in [8], when a counting error penalization term is included in the loss function, training with auto-differentiation often fails to converge and satisfactory results could only be achieved with a very small weight on the counting term with respect to the geodesical reconstruction term.

As an extension to [8], in this paper we further provide a detailed theoretical framework of  $h$ -maxima, on which is built the analysis of the derivability of the pipeline, explaining the issues related to the minimization of the counting loss. In order to overcome these issues, this work proposes and tests a training paradigm employing explicitly defined backpropagation rules, which show to work well even when training our pipeline with counting loss only. With respect to [8], we demonstrated the effectiveness and generalizability of the proposed pipeline on two other fluorescent cell datasets.

Training with a counting loss only allows our method to bypass the need for pre-defined preprocessing steps, paving the way for future research avenues such as the development of trainable morphological layers for adaptive cell size and shape learning, or convolutional layers for learned preprocessing. Moreover, future endeavors could also involve the prediction of *local*  $h$  values for smaller patches and the implementation

Dataset	Method	#Param	$\mathcal{A}_{err}(\%)$	$\mathcal{T}_{err}(\%)$	MAE	MPE(%)
TRP1 [3]	Lazard <i>et al</i> [3]	1,760,000	<b>9.28</b>	<b>8.72</b>	-	-
	Joint loss (MB,-1)	16,675	12.6 $\pm$ 0.6	10.4 $\pm$ 0.5	<b>5.82</b> $\pm$ 0.3	<b>-5.47</b> $\pm$ 0.8
	Joint loss (MB,N=50)	16,675	13.6 $\pm$ 1.2	11.1 $\pm$ 0.8	6.21 $\pm$ 0.45	-7.15 $\pm$ 1.8
	Count. loss (MB, -1)	16,675	13.7 $\pm$ 1.3	11.0 $\pm$ 1.1	6.2 $\pm$ 0.6	-8.60 $\pm$ 1.8
Fluorescent Neuronal Cells [4]	Morelli <i>et al</i> [4]	888,977	-	-	3.09	<b>-5.13</b>
	Joint loss (MB, -1)	16,675	34.4 $\pm$ 2.6	28.6 $\pm$ 0.4	2.89 $\pm$ 0.04	-9.1 $\pm$ 5.6
	Joint loss (MB, N=50)	16,675	33.0 $\pm$ 2.1	28.1 $\pm$ 0.7	2.84 $\pm$ 0.07	-6.05 $\pm$ 3.5
	Count. loss (MB, -1)	16,675	<b>31.7</b> $\pm$ 1.2	28.1 $\pm$ 0.7	2.84 $\pm$ 0.08	7.12 $\pm$ 1.6
Cellpose [20]	Count. loss (MB, N=50)	16,675	32.1 $\pm$ 1.2	<b>25.3</b> $\pm$ 0.8	<b>2.56</b> $\pm$ 0.08	-7.17 $\pm$ 3.3
	Unet [22]	7,852,033	12.1 $\pm$ 2.1	11.8 $\pm$ 1.9	6.31 $\pm$ 1.0	11.2 $\pm$ 2.5
	Joint loss (MB, -1)	16,675	<b>6.98</b> $\pm$ 0.7	8.09 $\pm$ 0.6	4.34 $\pm$ 0.33	<b>0.25</b> $\pm$ 1.4
	Joint loss (MB, N=50)	16,675	7.01 $\pm$ 0.93	7.97 $\pm$ 0.82	4.28 $\pm$ 0.44	1.35 $\pm$ 1.34
Cellpose [20]	Count. loss (MB, -1)	16,675	7.10 $\pm$ 1.2	<b>7.47</b> $\pm$ 1.7	<b>4.01</b> $\pm$ 0.89	0.94 $\pm$ 1.2
	Count. loss (MB, N=50)	16,675	8.87 $\pm$ 1.2	10.3 $\pm$ 1.3	5.52 $\pm$ 0.7	4.75 $\pm$ 2.3

**Table 1** Quantitative evaluation of the proposed method on the three datasets, and comparison to the state of the art (for TRP1 and Fluorescent Neuronal Cells) or a baseline deep network (for Cellpose). The metrics for our models are provided for the different optimization approaches described in earlier sections, namely: with joint loss ( $\alpha = \beta = 0.5$ ) or counting loss only, and with modified backpropagation (MB, -1) or (MB, N=50). For each of these alternatives, ten models were trained following the protocol described in Section 4.3. We provide the average metric on the test sets  $\pm$  the standard deviation over the ten training instances. The best metrics are shown in bold.

of contrast enhancement for blurred regions. Furthermore, as proposed in [23], the graph approach based on hierarchical clustering and gradient descent, combined with cell counting loss, also represents a potential avenue for dynamic analysis and cell counting task.

We posit that the synergy between neural networks and trainable morphological layers opens the door to a broader range of applications, such as counting astronomical objects in astronomical surveys. However, a deeper reflection is necessary to understand which are the best ways to initialize, optimize and regularize morphological layers [24, 25].

**Acknowledgements.** This work was granted access to the HPC resources of IDRIS under the allocation 2024-AD011012212R3 made by GENCI. Many thanks to Beatriz Marcotegui for helpful discussions about Proposition 4.

## Appendix A Proof of Proposition 1

Let us first state and prove the following lemma.

**Lemma 8** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$ ,  $h \geq 0$  and  $M \in \mathcal{M}_h(f)$ . Then*

$$\forall x \in \mathcal{N}_M \setminus M, \quad \text{HMAX}_h(f)(x) = f(x) < \text{HMAX}_h(f)(M) \quad (\text{A1})$$

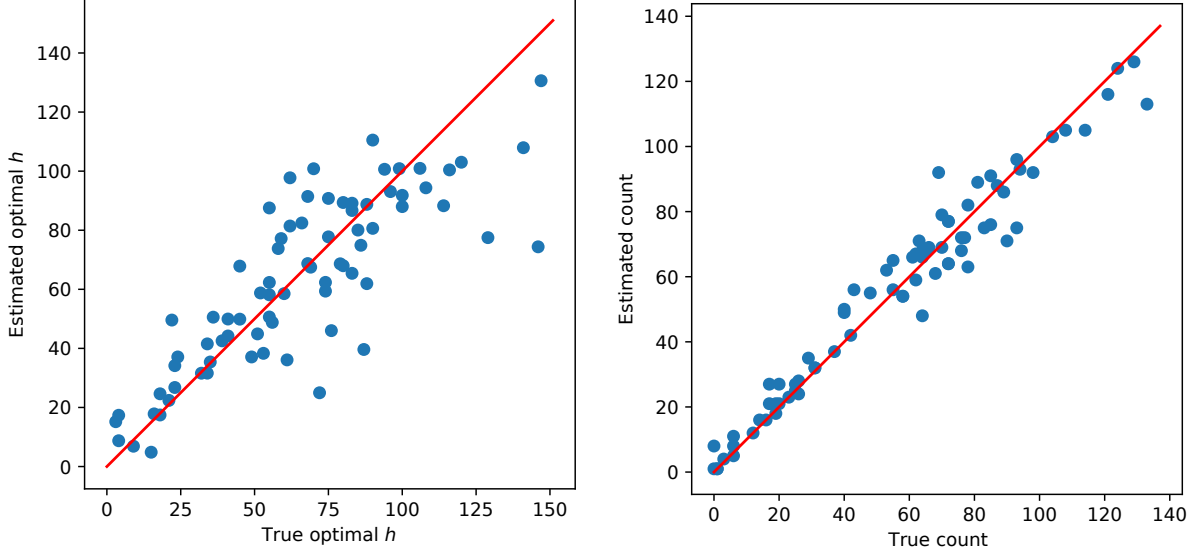
and

$$\text{HMAX}_h(f)(M) = \max_M f - h. \quad (\text{A2})$$

*Proof of Lemma 8* Since  $M$  is a regional maximum of  $\text{HMAX}_h(f)$ , by definition  $\text{HMAX}_h(f)$  is constant over  $M$  and, noting  $\lambda := \text{HMAX}_h(f)(M)$ , we have that for any  $x \in \mathcal{N}_M \setminus M$ ,  $\text{HMAX}_h(f)(x) < \lambda$ . Let us show that  $f(x) < \lambda$  as well. Recall that by definition of the geodesical reconstruction (6),

$$\begin{aligned} \text{HMAX}_h(f)(x) &= \delta_B(\text{HMAX}_h(f))(x) \wedge f(x) \\ &= \max_{y \in \mathcal{N}_x} \{\text{HMAX}_h(f)(y)\} \wedge f(x). \end{aligned} \quad (\text{A3})$$

But since  $x \in \mathcal{N}_M$ , there exists  $y \in \mathcal{N}_x$  such that  $y \in M$  and therefore  $\text{HMAX}_h(f)(y) = \lambda$ , so  $\delta_B(\text{HMAX}_h(f))(x) \geq \lambda$ . Hence, if  $f(x) \geq \lambda$ , then by (A3) we get  $\text{HMAX}_h(f)(x) \geq \lambda$ , which is false, as we just saw. So  $f(x) < \lambda$ . As we have just shown that



**Fig. 11** Results on the test set of the TRP1 dataset, for the best of our models on this dataset, trained with the joint loss and modified backpropagation (MB, -1). **Left:** predicted  $\hat{h}$  compared to the ground truth optimal  $h^*$ . **Right:** predicted versus true cell numbers. Each dot represents an individual image. In both plots, the red line represents the identity function. See Figure 14 for examples of predictions by this model.

$\delta_B(\text{HMAX}_h(f))(x) \geq \lambda$  and  $f(x) < \lambda$ , by (A3) we also get that  $\text{HMAX}_h(f)(x) = f(x)$ , which finishes to prove (A1).

Now let us prove (A2). Applying (7) to  $\text{HMAX}_h(f) = \text{REC}(f - h, f)$  we get,

$$\text{HMAX}_h(f)(x) = \max_{y \in \Omega, \gamma \in \Gamma_{xy}} \left\{ (f(y) - h) \wedge \min_{z \in \gamma} f(z) \right\}. \quad (\text{A4})$$

We denote by  $Q(y, \gamma, f, h) := (f(y) - h) \wedge \min_{\gamma} f$  and split Eq. (A4) as

$$\begin{aligned} \text{HMAX}_h(f)(x) &= \max \left\{ \max_{\substack{y \in \Omega, \gamma \in \Gamma_{xy} \\ \gamma \subseteq M}} Q(y, \gamma, f, h), \right. \\ &\quad \left. \max_{\substack{y \in \Omega, \gamma \in \Gamma_{xy} \\ \gamma \not\subseteq M}} Q(y, \gamma, f, h) \right\} \\ &= \max \left\{ \max_{\substack{y \in M, \gamma \in \Gamma_{xy} \\ \gamma \subseteq M}} Q(y, \gamma, f, h), \right. \quad (\text{A5}) \\ &\quad \left. \max_{\substack{y \in \Omega, \gamma \in \Gamma_{xy} \\ \gamma \not\subseteq M}} Q(y, \gamma, f, h) \right\} \\ &= \max(q_1(x, f, h), q_2(x, f, h)) \end{aligned}$$

with  $q_1(x, f, h) := \max_{\substack{y \in M, \gamma \in \Gamma_{xy} \\ \gamma \subseteq M}} Q(y, \gamma, f, h)$  and

$$q_2(x, f, h) := \max_{\substack{y \in \Omega, \gamma \in \Gamma_{xy} \\ \gamma \not\subseteq M}} Q(y, \gamma, f, h).$$

If  $y \in M$ , then  $\text{HMAX}_h(f)(y) = \lambda$  and  $\text{HMAX}_h(f)(y) \geq f(y) - h$  by definition of  $\text{REC}(f - h, f)$ . Furthermore, for any  $\gamma \in \Gamma_{xy}$  such that  $\gamma \subseteq M$ , and any  $z \in \gamma$ , we have  $f(z) \geq \text{HMAX}_h(f)(z) = \lambda$ , so  $\min_{\gamma} f \geq \lambda$ . As a consequence,  $(f(y) - h) \wedge \min_{\gamma} f = f(y) - h$  and finally  $q_1(x, f, h) = \max_M f - h$ .

Now, for any  $y \in \Omega$  and  $\gamma \in \Gamma_{xy}$  such that  $\gamma \not\subseteq M$ , there exists  $z \in \mathcal{N}_M \setminus M$  such that  $z \in \gamma$ , for which  $\min_{\gamma} f < \lambda$  by (A1), and consequently  $f(y) - h \wedge \min_{\gamma} f < \lambda$ . Therefore  $q_2(x, f, h) < \lambda$  and since  $\text{HMAX}_h(f)(x) = \lambda$ , we get

$$\text{HMAX}_h(f)(x) = q_1(x, f, h) = \max_M f - h$$

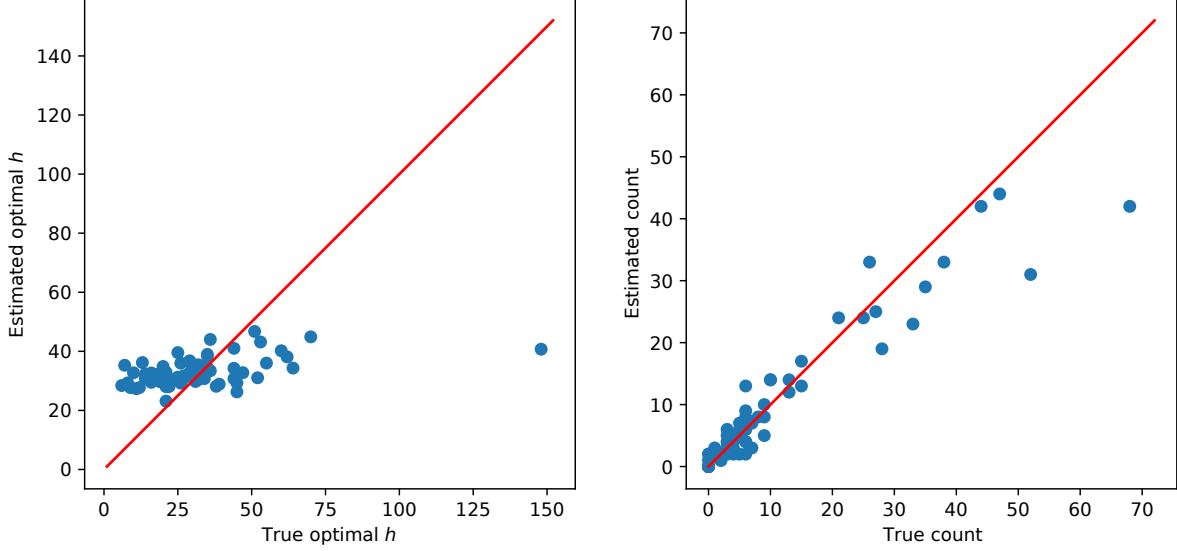
which concludes the proof of Lemma 8.  $\square$

Now let us prove Proposition 1.

*Proof of Proposition 1 “Only if” direction  $\Rightarrow$*

Let us start by the “only if” direction. Let  $h \geq 0$ ,  $M \in \mathcal{M}_h(f)$ ,  $\lambda = \text{HMAX}_h(f)(M)$  and  $M_0$  be a connected component of  $\arg \max_{M_h} f$ .

If  $M_0$  was not a regional maximum of  $f$ , then there would exist  $x \in \mathcal{N}_{M_0} \setminus M$  such that  $f(x) \geq f(M_0)$ ,



**Fig. 12** Results on the test set of the Fluorescent Neuronal Cells dataset, for the best of our models on this dataset, trained with the counting loss only and modified backpropagation (MB, N=50). **Left:** predicted  $\hat{h}$  compared to the ground truth optimal  $h^*$ . **Right:** predicted versus true cell numbers. Each dot represents an individual image. In both plots, the red line represents the identity function. See Figure 15 for examples of predictions by this model.

and therefore  $\text{HMAX}_h(f)(x) \geq f(x) - h \geq f(M_0) - h = \max_M f - h$ . Now, by Lemma 8,  $\text{HMAX}_h(f)(M) = f(M_0) - h$  and, since  $x \in \mathcal{N}_{M_0} \setminus M \subseteq \mathcal{N}_M \setminus M$ ,  $\text{HMAX}_h(f)(x) < \lambda$ . Therefore,  $M_0 \in \mathcal{M}(f)$  and  $\lambda = f(M_0) - h$ .

Now, let us lower bound the dynamic of  $M_0$  using Eq. (12). For any  $M'$  such that  $f(M') > f(M_0)$ , then  $M' \cap M = \emptyset$  since  $M_0 \subseteq \arg \max_M f$ . Therefore, any  $\gamma \in \Gamma_{M_0 M'}$  contains a  $z$  such that  $z \in \mathcal{N}_M \setminus M$ , implying  $f(z) < \text{HMAX}_h(f)(M) = f(M_0) - h$ , by (A1) of Lemma 8. As a consequence,  $\min_{z \in \gamma} f(z) < f(M_0) - h$ , and this is true for any  $M'$  such that  $f(M') > f(M_0)$  and any  $\gamma \in \Gamma_{M_0 M'}$ , which implies that

$$\max_{\substack{\gamma \in \Gamma_{M M'} \\ f(M') > f(M)}} \min_{z \in \gamma} f(z) < f(M_0) - h.$$

So by Eq. (12),  $\text{DYNMAX}(M_0) > h$ .

It remains to show that  $M = \text{CC}(M_0, \chi_\lambda(f))$ , where  $\lambda = f(M_0) - h$ . We may first remark that

$$\text{CC}(M_0, \chi_\lambda(f)) = \{x \in \Omega, \exists \gamma \in \Gamma_{x M_0}, f(z) \geq \lambda, \forall z \in \gamma\}. \quad (\text{A6})$$

Let us show that  $M \subseteq \text{CC}(M_0, \chi_\lambda(f))$ . We already have that, for any  $x \in M$ ,  $\text{HMAX}_h(f)(x) = \lambda = f(M_0) - h$  and by definition  $\text{HMAX}_h(f)(x) \leq f(x)$ . Since  $M$  is connected and  $M_0 \subseteq M$ , for any  $x \in M$  there is a path  $\gamma \in \Gamma_{x M_0}$  included in  $M$ , therefore  $f(z) \geq \lambda$  for any  $z \in \gamma$ . Hence, by (A6),  $M \subseteq \text{CC}(M_0, \chi_\lambda(f))$ .

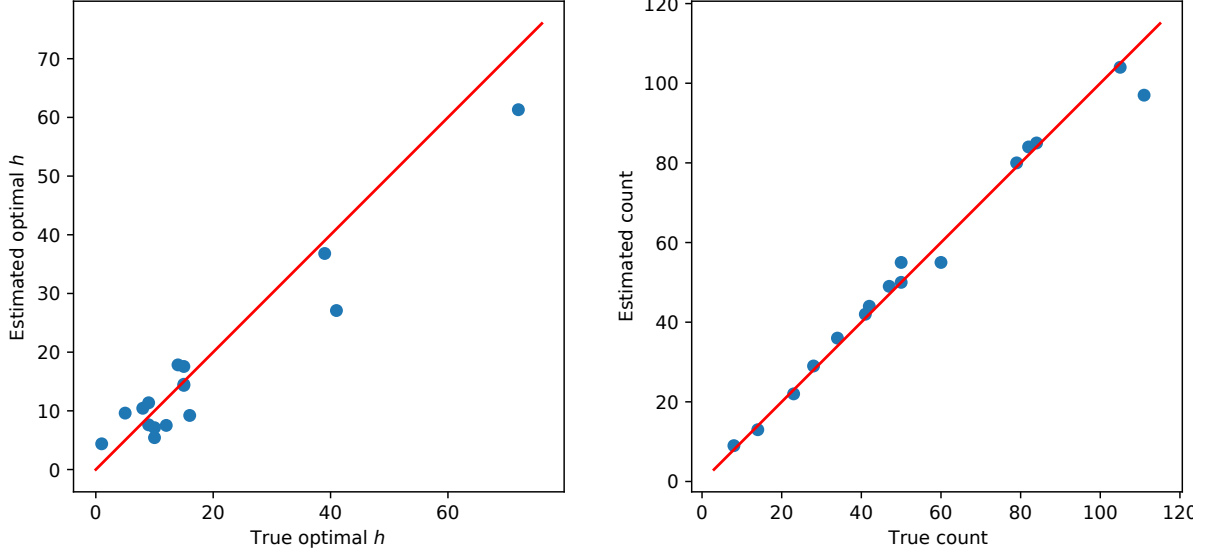
Let us now show that  $\text{CC}(M_0, \chi_\lambda(f)) \subseteq M$ . Let  $x \in \text{CC}(M_0, \chi_\lambda(f))$ . Then there is a path  $\gamma \in \Gamma_{x M_0}$  such that  $f(z) \geq \lambda$  for any  $z \in \gamma$ . If  $x \notin M$ , then there is  $z \in \gamma$  such that  $z \in \mathcal{N}_M \setminus M$ , and therefore  $f(z) < \lambda$  by (A1) of Lemma 8, which contradicts  $x \in \text{CC}(M_0, \chi_\lambda(f))$ . So  $x \in M$ , and we have the second inclusion, which leads to  $M = \text{CC}(M_0, \chi_\lambda(f))$  and finishes the proof of the “only if” direction.

“If” direction  $\Leftarrow$

For the “if” direction, let  $M_0 \in \mathcal{M}(f)$  be a regional maximum of  $f$ , with dynamic strictly larger than  $h \geq 0$ , let  $\lambda = f(M_0) - h$ , and let  $M = \text{CC}(M_0, \chi_\lambda(f))$ . Then  $M \neq \Omega$ , otherwise we would have  $f(x) \geq f(M_0) - h$  for any  $x \in \Omega$  and, by Eq. (12), we would get  $\text{DYNMAX}(M_0) \leq h$ , which is a contradiction.

Therefore we will show that  $M$  is a regional maximum of  $\text{HMAX}_h(f)$ , i.e.  $M \in \mathcal{M}_h(f)$ , by proving that  $\text{HMAX}_h(f)$  is constant over  $M$ , equal to  $\lambda$ , and that  $\text{HMAX}_h(f)(z) < \lambda$  for any  $z \in \mathcal{N}_M \setminus M$ .

Starting with the latter, let  $z \in \mathcal{N}_M \setminus M$ . Then  $f(z) < \lambda$ , since otherwise there would be  $x \in \mathcal{N}_z$  such that  $x \in M$ , and a path  $\gamma$  from  $x$  to  $M_0$  such that  $f(\gamma) \geq \lambda$ , by definition of  $M$ ; by appending  $z$  to  $\gamma$  we would have a path  $\gamma'$  from  $z$  to  $M_0$ , such that  $f(\gamma') \geq \lambda$ , and therefore  $z$  would be in  $M$ , which contradicts  $z \in \mathcal{N}_M \setminus M$ . Furthermore,  $\text{HMAX}_h(f)(z) \leq f(z)$  by definition, so  $\text{HMAX}_h(f)(z) < \lambda$ .



**Fig. 13** Results on the test set of Cellpose, for the best of our models on this dataset, trained with the counting loss only and modified backpropagation (MB, -1). **Left:** predicted  $\hat{h}$  compared to the ground truth optimal  $h^*$ . **Right:** predicted versus true cell numbers. Each dot represents an individual image. In both plots, the red line represents the identity function. See Figure 16 for examples of predictions by this model.

Now, to show that  $\text{HMAX}_h(f)(x) = \lambda$  for all  $x \in M$ , we will prove that  $f(M_0) = \max_M f$ , and then apply Eq. (7) to show that  $\text{HMAX}_h(f)(x) \geq \lambda$  and conclude.

Suppose the set  $\{x \in M, f(x) > f(M_0)\}$  is non-empty, and let  $M'_0$  a connected component of that set. Then  $M'_0$  is a regional maximum of  $f$ , otherwise we would have a  $z \in \mathcal{N}_M \setminus M$  such that  $f(z) \geq f(M'_0) > f(M_0) \geq \lambda$  and by the same reasoning as before, this would lead to  $z \in M$ , which is a contradiction. Therefore we can use Eq. (12) to show that the dynamic of  $M_0$  is not larger than  $h$ . Indeed, since  $M$  is connected, there is a path  $\gamma \in \Gamma_{M_0 M'_0}$  included in  $M$ , therefore  $\min_\gamma f \geq \lambda$ . Hence we have found a regional maximum  $M'_0$  such that  $f(M'_0) > f(M_0)$  and a path  $\gamma \in \Gamma_{M_0 M'_0}$  such that  $\min_\gamma f \geq f(M_0) - h$ . By Eq. (12), this implies that  $\text{DYNMAX}(M_0) \leq f(M_0) - (f(M_0) - h) = h$ , which contradicts our initial hypothesis. Therefore,  $f(M_0) = \max_M f$ , which also means that  $M_0$  is a connected component of  $\arg \max_M f$ , since  $M_0 \in \mathcal{M}(f)$ .

Now, from Equation (7) it appears that  $\text{HMAX}_h(f)(x) \geq \lambda = f(M_0) - h$  for  $x \in M$ . Indeed, applying (7) to  $\text{REC}(f - h, f)$ , we get Eq. (A4). Since  $M$  is connected and  $M_0 \subseteq M$ , for any  $y \in M_0$ , there exists  $\gamma \in \Gamma_{xy}$  included in  $M$ , for which therefore  $\min_\gamma f \geq f(M_0) - h = f(y) - h$ , and hence  $(f(y) - h) \wedge \min_\gamma f = f(M_0) - h$ . Therefore, taking the

maximum over  $y \in \Omega$  and over paths in  $\Gamma_{xy}$  yields  $\text{HMAX}_h(f)(x) \geq f(M_0) - h$  for  $x \in M$ .

To finally show that  $\text{HMAX}_h(f)(x) = f(M_0) - h$  for  $x \in M$ , we denote by  $Q(y, \gamma, f, h) := f(y) - h \wedge \min_\gamma f$ , like in the proof of Lemma 8, and split Eq. (A4) the same way, giving (A5).

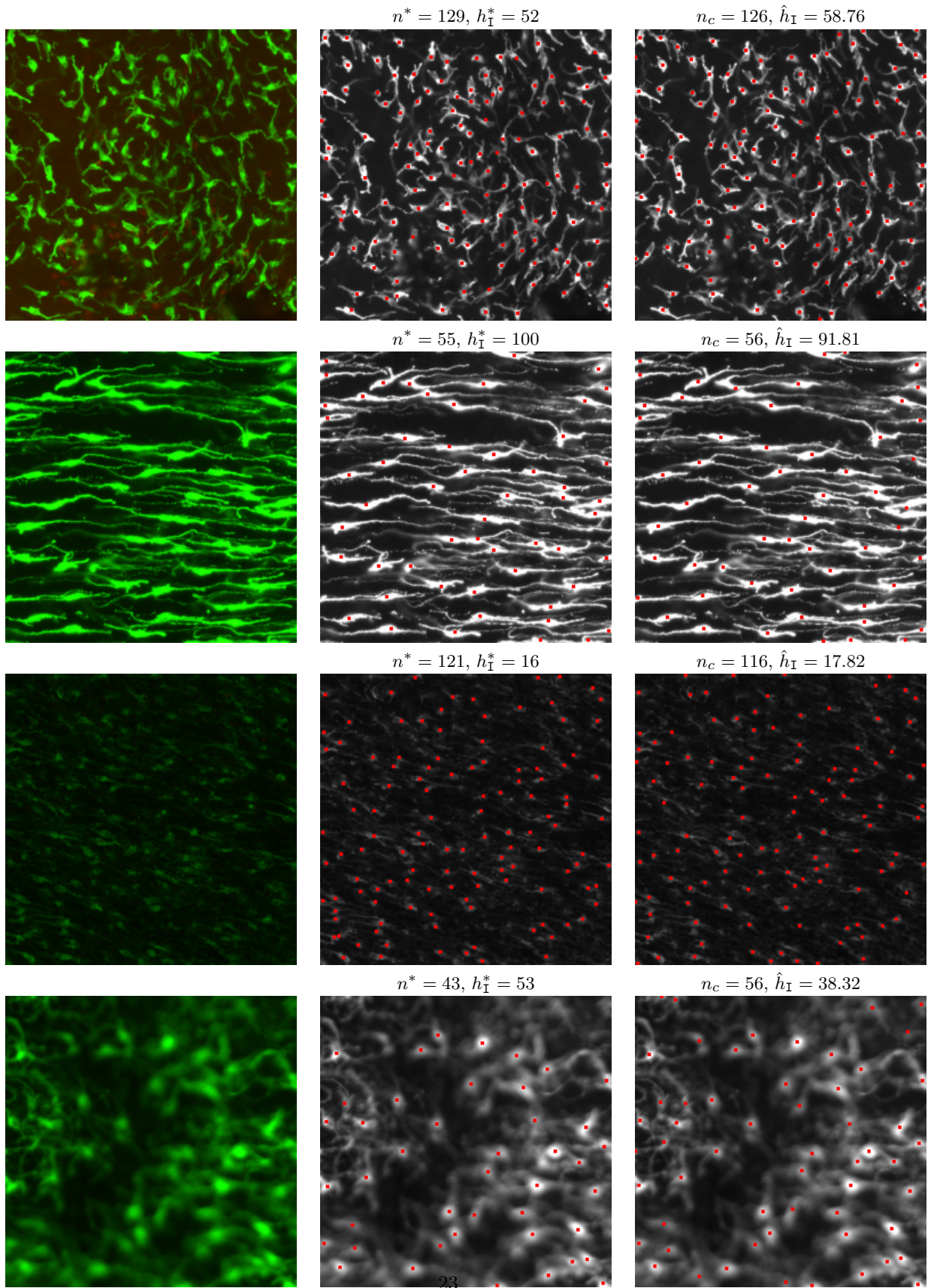
If  $x \in M$ , for any  $y \in \Omega$  and  $\gamma \in \Gamma_{xy}$  such that  $\gamma \not\subseteq M$ , there exists  $z \in \mathcal{N}_M \setminus M$  such that  $z \in \gamma$ , for which  $\min_\gamma f < f(M_0) - h$ , and consequently  $(f(y) - h) \wedge \min_\gamma f < f(M_0) - h$ . Therefore  $q_2(x, f, h) < f(M_0) - h$  and since  $\text{HMAX}_h(f)(x) \geq f(M_0) - h$ , we get  $\text{HMAX}_h(f)(x) = q_1(x, f, h)$ , *i.e.*

$$\text{HMAX}_h(f)(x) = \max_{\substack{y \in M, \gamma \in \Gamma_{xy} \\ \gamma \subseteq M}} \left\{ (f(y) - h) \wedge \min_\gamma f \right\}.$$

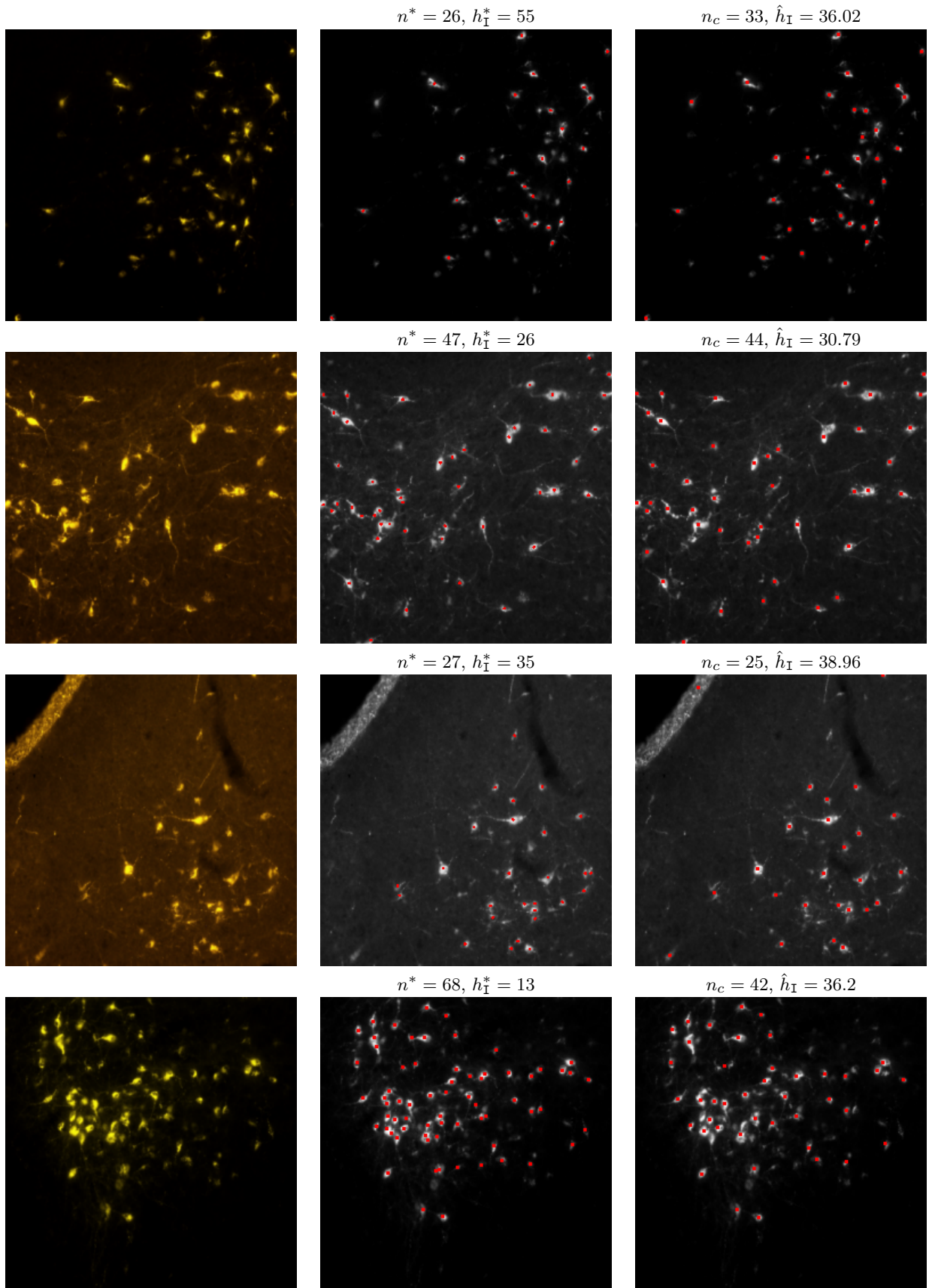
But for any  $y \in M$ ,  $f(y) - h \leq \max_M f - h = f(M_0) - h$ , and for any  $\gamma \in \Gamma_{xy}$  such that  $\gamma \subseteq M$ ,  $\min_\gamma f \geq f(M_0) - h$  by definition of  $M$ . Therefore  $(f(y) - h) \wedge \min_\gamma f = f(y) - h$  and finally

$$\text{HMAX}_h(f)(x) = \max_{y \in M} f(y) - h = \max_M f - h = f(M_0) - h.$$

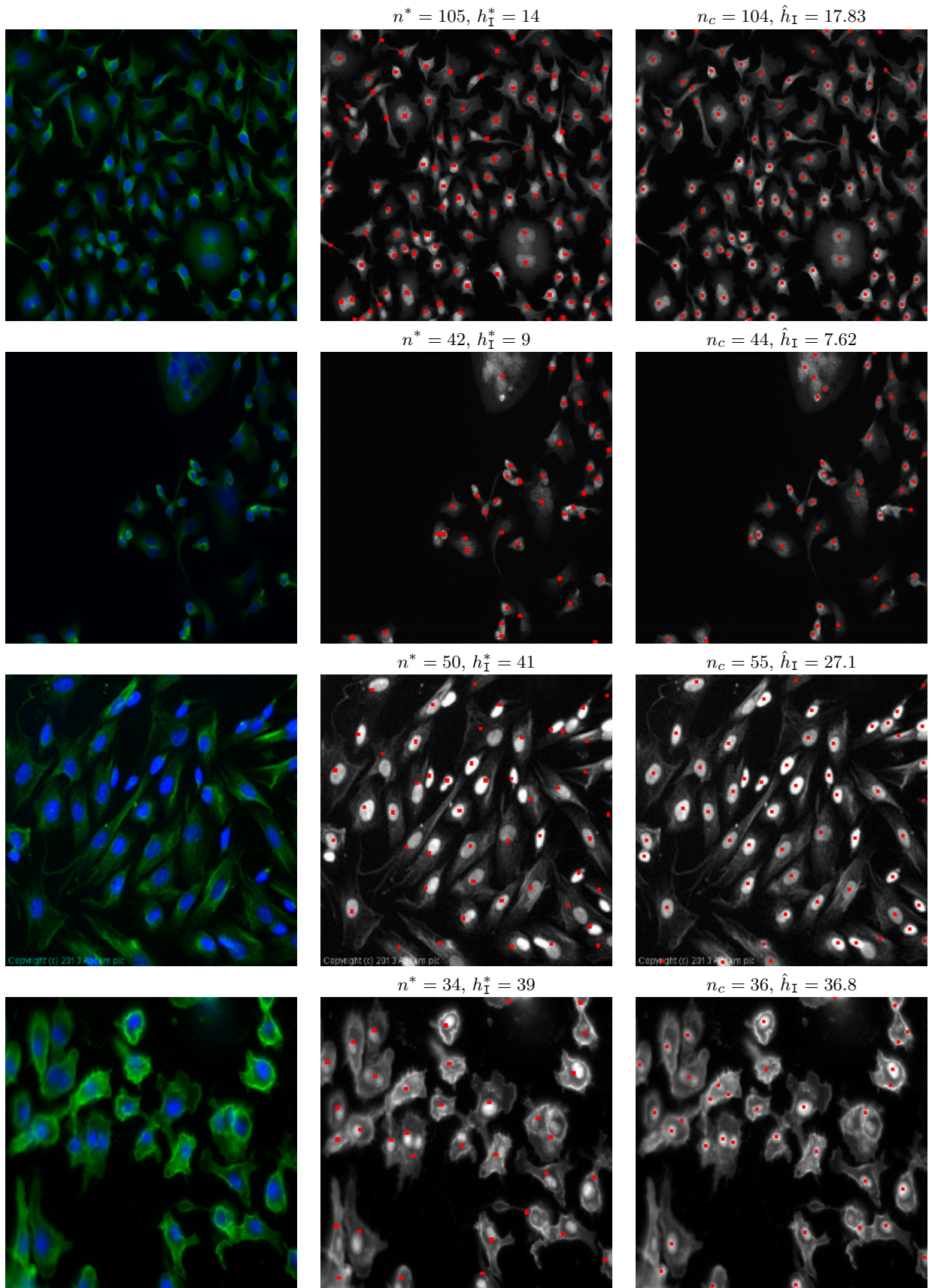
We have shown that for  $x \in M$   $\text{HMAX}_h(f)(x) = f(M_0) - h$  after having shown that  $\text{HMAX}_h(f)(z) < f(M_0) - h$  for  $z \in \mathcal{N}_M \setminus M$ . Since  $M$  is connected,  $M \in \mathcal{M}_h(f)$ , which concludes the “if” direction of Proposition 1.  $\square$



**Fig. 14** Examples of predictions on the test set of the TRP1 dataset, for the best of our models on this dataset, trained with the joint loss and modified backpropagation (MB, -1). From left to right: Input resized RGB image, green channel image with ground truth cell locations, true number of cells  $n^*$  and optimal dynamic  $h_I^*$ ; green channel image with predicted cell locations and their estimated number  $n_c$  obtained for the estimated dynamic  $\hat{h}_I$ .



**Fig. 15** Examples of predictions on the test set of the Fluorescent Neuronal Cells dataset, for the best of our models on this dataset, trained with the counting loss only and the modified backpropagation (MB,  $N=50$ ). From left to right: Input resized RGB image, green channel image with ground truth cell locations, true number of cells  $n^*$  and optimal dynamic  $h_I^*$ ; grayscale image with predicted cell locations and their estimated number  $n_c$  obtained for the estimated dynamic  $\hat{h}_I$ .



**Fig. 16** Examples of predictions on the test set of Cellpose, for the best of our models on this dataset, trained with the counting loss only and the modified backpropagation (MB, -1). From left to right: Input resized RGB image, maximum of three channels image with ground truth cell locations, true number of cells  $n^*$  and optimal dynamic  $h_I^*$ ; grayscale image with predicted cell locations and their estimated number  $n_c$  obtained for the estimated dynamic  $\hat{h}_I$ .

## Appendix B Proof of Proposition 2

Let  $0 \leq h_1 \leq h_2$ . Then if  $M_2 \in \mathcal{M}_{h_2}(f)$ , by Proposition 1, there is some regional maximum  $M_0$  of dynamic larger than  $h_2$ , such that  $M_2 = \text{CC}(M_0, \chi_{f(M_0)-h_2}(f))$ . But then  $M_0$  is also of dynamic larger than  $h_1$ , therefore  $M_1 = \text{CC}(M_0, \chi_{f(M_0)-h_1}(f))$  is in  $\mathcal{M}_{h_1}(f)$ . Furthermore,  $\chi_{f(M_0)-h_1}(f) \subseteq \chi_{f(M_0)-h_2}(f)$  because  $f(M_0) - h_2 \leq f(M_0) - h_1$ , hence  $\text{CC}(M_0, \chi_{f(M_0)-h_1}(f)) \subseteq \text{CC}(M_0, \chi_{f(M_0)-h_2}(f))$ , i.e.  $M_1 \subseteq M_2$ . This shows that  $0 \leq h_1 \leq h_2 \Rightarrow \mathcal{M}_{h_2}(f) \preceq \mathcal{M}_{h_1}(f)$ .

## Appendix C Proof of Proposition 3

We first state the right continuity of  $\mu$  in the following lemma, which will be used next.

**Lemma 9** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  non constant,  $h_\infty \in [0, \Delta_\Omega f]$  and  $(h_n)_{n \in \mathbb{N}} \in [0, \Delta_\Omega f]^{\mathbb{N}}$  a sequence such that  $h_n \downarrow h_\infty$ . Then*

$$\lim_{n \rightarrow +\infty} \mu(h_n) = \mu(h_\infty). \quad (\text{C7})$$

*Proof of Lemma 9* For  $h \in [0, \Delta_\Omega f]$ , let us denote by

$$\mathcal{D}_h(f) := \{M_0 \in \mathcal{M}(f), \text{DYNMAX}(M_0) > h\} \quad (\text{C8})$$

the set of regional maxima of  $f$  with dynamic larger than  $h$ , and for the considered  $h_\infty$ , we introduce

$$h_{\min} = \min\{\text{DYNMAX}(M_0), M_0 \in \mathcal{D}_{h_\infty}(f)\}. \quad (\text{C9})$$

Then  $h_{\min} > h_\infty$ , therefore there exists  $N_1 \in \mathbb{N}$ , such that  $n \geq N_1 \Rightarrow h_n \in [h_\infty, h_{\min})$ . As a consequence, one can check that for any  $n \geq N_1$ ,

$$\mathcal{D}_{h_n}(f) = \mathcal{D}_{h_\infty}(f).$$

Furthermore, by Proposition 1, we get that for any  $h \in [0, \Delta_\Omega f]$

$$\mathcal{M}_h(f) = \{\text{CC}(M_0, \chi_{f(M_0)-h}(f)), M_0 \in \mathcal{D}_h(f)\}. \quad (\text{C10})$$

Therefore, for any  $n \geq N_1$ ,

$$\mathcal{M}_{h_n}(f) = \{\text{CC}(M_0, \chi_{f(M_0)-h_n}(f)), M_0 \in \mathcal{D}_{h_\infty}(f)\}. \quad (\text{C11})$$

Now, it is well known that  $t \mapsto \chi_t(f)$  is left continuous i.e.  $\lim_{t \uparrow t_0} \chi_t(f) = \chi_{t_0}(f)$ . What is more, here each  $\chi_t(f)$  is a discrete set (as subset of  $\Omega$ ), hence the limit

is actually reached for  $t$  sufficiently close to  $t_0$ . This implies that there exists  $N_2 \in \mathbb{N}$  such that

$$n \geq N_2 \Rightarrow \chi_{f(M_0)-h_n}(f) = \chi_{f(M_0)-h_\infty}(f). \quad (\text{C12})$$

So finally, by (C11) and (C12), we get that for any  $n \geq \max(N_1, N_2)$ ,  $\mu(h_n) = \mathcal{M}_{h_n}(f) = \mathcal{M}_{h_\infty}(f) = \mu(h_\infty)$ , which concludes the proof.  $\square$

Now we can prove Proposition 3.

*Proof of Proposition 3* We shall notice first that, since  $\mu$  maps  $[0, \Delta_\Omega f]$  to the set of collections of subsets of  $\Omega$ , and since  $\Omega$  is a finite set,  $\mu$  can only take a finite set of values over  $[0, \Delta_\Omega f]$ , including  $\mu(0) = \mathcal{M}(f)$  and  $\mu(\Delta_\Omega f) = \emptyset$ . Hence, there exists  $K \in \mathbb{N}$  and  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_{K+1} \in \mathcal{P}(\mathcal{P}(\Omega))$  pairwise distinct, such that

$$\mu([0, \Delta_\Omega f]) = \{\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_{K+1}\}.$$

For every  $k \in \{0, \dots, K+1\}$  we note

$$\mathcal{I}_k := \mu^{-1}(\{\mathcal{C}_k\}) = \{h \in [0, \Delta_\Omega f], \mu(h) = \mathcal{C}_k\}. \quad (\text{C13})$$

Now, we can easily establish that the  $\mathcal{I}_k$ s are pairwise disjoint intervals.

Indeed, since  $k \neq k' \Rightarrow \mathcal{C}_k \neq \mathcal{C}_{k'}$ , we also have  $k \neq k' \Rightarrow \mathcal{I}_k \cap \mathcal{I}_{k'} = \emptyset$ , otherwise one  $h$  value could be mapped to two distinct set of  $h$ -maxima.

The fact that each  $\mathcal{I}_k$  is an interval is given by Proposition 2: if  $h \leq h' \in \mathcal{I}_k$  and  $h \leq h'' \leq h'$ , then

$$\mathcal{C}_k = \mu(h') \preceq \mu(h'') \preceq \mu(h) = \mathcal{C}_k$$

therefore  $\mu(h'') = \mathcal{C}_k$ , i.e.  $h'' \in \mathcal{I}_k$ .

Hence, without loss of generality, we can assume the  $\mathcal{I}_k$ s are indexed from left to right, i.e.  $k < k' \iff y < y' \forall y \in \mathcal{I}_k, y' \in \mathcal{I}_{k'}$ , which implies that

$$\mathcal{C}_0 = \mathcal{M}(f) \succ \mathcal{C}_1 \succ \dots \succ \mathcal{C}_{K+1} = \mathcal{C}_{\Delta_\Omega f}(f) = \emptyset.$$

Note that, as leftmost and rightmost intervals of a partition of  $[0, \Delta_\Omega f]$ ,  $\mathcal{I}_0$  must include 0 and  $\mathcal{I}_{K+1}$  must include  $\Delta_\Omega f$ . Since both intervals are included in  $[0, \Delta_\Omega f]$ , we get  $0 = \inf \mathcal{I}_0$  and  $\Delta_\Omega f = \sup \mathcal{I}_{K+1}$ .

Next, for every  $k \in \{0, \dots, K+1\}$  we note

$$h_k := \inf \mathcal{I}_k.$$

So  $h_0 = 0$  and, since the  $\mathcal{I}_k$ s form a partition of  $[0, \Delta_\Omega f]$ , it is straightforward that for  $k \leq K$  we also have  $h_{k+1} = \sup \mathcal{I}_k$ , so  $h_{K+1} = \Delta_\Omega f$ . Furthermore, by Lemma 9 we have that, for each  $k \in \{0, \dots, K+1\}$ ,  $\mu(h_k) = \mathcal{C}_k$ , hence  $h_k \in \mathcal{I}_k$  and  $\mathcal{C}_k = \mathcal{M}_{h_k}(f)$ . Since the  $\mathcal{I}_k$ s are pairwise disjoint, necessarily  $h_{k+1} \notin \mathcal{I}_k$  for  $k \leq K$ , so that for  $k \in \{0, \dots, K\}$

$$\mathcal{I}_k = [h_k, h_{k+1}).$$

This concludes the proof of Proposition 3.  $\square$

## Appendix D Proof of Proposition 4

To prove Proposition 4, it is sufficient to use Eq. (A2) of Lemma 8, and the following lemma.

**Lemma 10** *Let  $f \in \mathcal{F}(\Omega, \mathbb{R})$  and  $h \geq 0$ . Then*

$$\forall x \in \bigcap_{M \in \mathcal{M}_h(f)} M^c, \quad \text{HMAX}_h(f)(x) = \max_{\substack{M \in \mathcal{M}_h(f) \\ \gamma \in \Gamma_{xM}}} \min_{z \in \gamma} f(z). \quad (\text{D14})$$

Indeed, Eq. (A2) of Lemma 8 and Eq. (D14) yields

$$\begin{aligned} \forall x \in \Omega, \\ \text{HMAX}_h(f)(x) &= \sum_{M \in \mathcal{M}_h(f)} \left( \max_M f - h \right) \mathbb{1}_M(x) \\ &+ \left( \max_{\substack{M \in \mathcal{M}_h(f) \\ \gamma \in \Gamma_{xM}}} \min_{z \in \gamma} f(z) \right) \times \prod_{M \in \mathcal{M}_h(f)} \mathbb{1}_{M^c}(x). \end{aligned} \quad (\text{D15})$$

Therefore, if we assume the stability property of Proposition 4 that there exists an interval  $I$  with  $0 \in I$  such that, for any  $\eta \in I$ ,  $\mathcal{M}_{h+\eta}(f) = \mathcal{M}_h(f)$ , then we get, for any  $\eta \in I$  and  $x \in \Omega$ ,

$$\begin{aligned} \text{HMAX}_{h+\eta}(f)(x) &= \sum_{M \in \mathcal{M}_h(f)} \left( \max_M f - (h + \eta) \right) \mathbb{1}_M(x) \\ &+ \left( \max_{\substack{M \in \mathcal{M}_h(f) \\ \gamma \in \Gamma_{xM}}} \min_{z \in \gamma} f(z) \right) \times \prod_{M \in \mathcal{M}_h(f)} \mathbb{1}_{M^c}(x) \\ &= \text{HMAX}_h(f)(x) - \eta \sum_{M \in \mathcal{M}_h(f)} \mathbb{1}_M(x) \\ &= \text{HMAX}_h(f)(x) - \eta \text{EMAX}_h(f)(x) \end{aligned} \quad (\text{D16})$$

which is the result stated by Proposition 4.

Hence let us prove Lemma 10.

*Proof of Lemma 10* To alleviate notations, let us first denote  $f_h := \text{HMAX}_h(f)$ . Then, Equation (D14) can be proved by showing that

$$\forall x \in \Omega, \quad f_h(x) = \max_{\substack{M \in \mathcal{M}_h(f) \\ \gamma \in \Gamma_{xM}}} \left( f_h(M) \wedge \min_{z \in \gamma} f_h(z) \right). \quad (\text{D17})$$

Indeed, by Eq. (A1), for any  $x \in \bigcap_{M \in \mathcal{M}_h(f)} M^c$ , any  $M \in \mathcal{M}_h(f)$  and any  $\gamma \in \Gamma_{xM}$ , we have  $\min_{z \in \gamma} f_h(z) < f_h(M)$ . Therefore, assuming (D17) yields that for any  $x \in \bigcap_{M \in \mathcal{M}_h(f)} M^c$ ,  $f_h(x) = \max_{\substack{M \in \mathcal{M}_h(f) \\ \gamma \in \Gamma_{xM}}} \min_{z \in \gamma} f_h(z)$ , which is Eq. (D14).

To show (D17), we use an important property of the reconstruction by dilation: the fact that any function  $g \in \mathcal{F}(\Omega, \mathbb{R})$  is its own reconstruction by dilation from its regional maxima [13], i.e

$$g = \text{REC}(g \times \mathbb{1}_{\cup M \in \mathcal{M}(g)}, g). \quad (\text{D18})$$

Hence

$$f_h = \text{REC}(f_h \times \mathbb{1}_{\cup M \in \mathcal{M}(f_h)}, f_h) = \text{REC}(f_h \times \mathbb{1}_{\cup M \in \mathcal{M}_h(f)}, f_h).$$

Furthermore, since the elements of  $\mathcal{M}_h(f)$  are disjoint subsets of  $\Omega$ , one can write

$$f_h \times \mathbb{1}_{\cup M \in \mathcal{M}_h(f)} = \bigvee_{M \in \mathcal{M}_h(f)} f_h \times \mathbb{1}_M \quad (\text{D19})$$

and therefore  $f_h = \bigvee_{M \in \mathcal{M}_h(f)} \text{REC}(f_h \times \mathbb{1}_M, f_h)$ , i.e

$$\forall x \in \Omega, \quad f_h(x) = \max_{M \in \mathcal{M}_h(f)} \text{REC}(f_h \times \mathbb{1}_M, f_h)(x). \quad (\text{D20})$$

Now, by Eq. (7),

$$\text{REC}(f_h \times \mathbb{1}_M, f_h)(x) = \max_{\substack{y \in \Omega, \\ \gamma \in \Gamma_{xy}}} \left( f_h(y) \times \mathbb{1}_M(y) \wedge \min_{z \in \gamma} f_h(z) \right)$$

with  $f_h(y) \times \mathbb{1}_M(y) = 0 \leq \min_{z \in \gamma} f_h(z)$  for any  $y \in \bigcap_{M \in \mathcal{M}_h(f)} M^c$  and  $\gamma \in \Gamma_{xy}$ , therefore the terms

corresponding to such  $y$  do not contribute to the max in the equation above. Hence

$$\text{REC}(f_h \times \mathbb{1}_M, f_h)(x) = \max_{\substack{y \in M, \\ \gamma \in \Gamma_{xy}}} \left( f_h(M) \wedge \min_{z \in \gamma} f_h(z) \right). \quad (\text{D21})$$

Finally, combining (D20) and (D21) gives (D17). This concludes the proof of Lemma 10 and Proposition 4.  $\square$

## Appendix E Proof of Corollary 5

**“If” direction**  $\leftarrow$

Let  $h^* \in [0, \Delta_\Omega f]$ . If  $h^* \notin \mathcal{H}_f$ , then  $h^* \in (h_k, h_{k+1})$  for some  $k \in \{0, \dots, K\}$ . Therefore there exists an *open* interval  $I$  containing 0 such that for any  $\eta \in I$ ,  $\mathcal{M}_{h^*+\eta}(f) = \mathcal{M}_{h^*}(f) = \mathcal{M}_{h_k}(f)$ . Then Proposition 4 applies and Eq. (29) holds for all  $\eta \in I$ , which implies that  $h \mapsto$

$\text{HMAX}_h(f)$  is differentiable at  $h^*$  and its derivative is  $-\text{EMAX}_{h^*}(f)$ .

**“Only if” direction**  $\Rightarrow$

Now let us show that if  $h^* \in \mathcal{H}_f$  then  $h \mapsto \text{HMAX}_h(f)$  is *not* Fréchet differentiable at  $h^*$ .

This is already clear for  $h^* = h_0 = 0$  and  $h^* = h_{K+1} = \Delta_\Omega f$ , since by definition a function can only be Fréchet differentiable at points lying in the interior of its domain.

To prove it for the other elements of  $\mathcal{H}_f$ , we will first point out that

$$\begin{aligned} \forall k \in \{0, \dots, K\}, \forall \eta \in [0, h_{k+1} - h_k] \\ \text{HMAX}_{h_k+\eta}(f) = \text{HMAX}_{h_k}(f) - \eta \text{EMAX}_{h_k}(f). \end{aligned} \quad (\text{E22})$$

Indeed, for  $0 \leq k \leq K$  and any  $\eta \in [0, h_{k+1} - h_k)$ , we have  $\mathcal{M}_{h_k+\eta}(f) = \mathcal{M}_{h_k}(f)$ , therefore by Proposition 4, Eq. (29) holds for any  $\eta \in [0, h_{k+1} - h_k)$ . But since  $h \mapsto \text{HMAX}_h(f)$  is continuous (by continuity of the dilation and pointwise minimum), it also holds for  $\eta = h_{k+1} - h_k$ , yielding (E22).

Let us now assume  $h^* = h_k$  for some  $k \in \{1, \dots, K\}$ .

Then, on the one hand, (E22) holds, *i.e.* for non-negative  $\eta$ . Therefore, if  $h \mapsto \text{HMAX}_h(f)$  was differentiable at  $h_k$ , the derivative at that point would be  $-\text{EMAX}_{h_k}(f)$ .

On the other hand, for any  $\eta \in [h_{k-1} - h_k, 0)$ , we have  $\eta' := h_k - h_{k-1} + \eta \in [0, h_k - h_{k-1})$ , so that

$$\begin{aligned} \text{HMAX}_{h_k+\eta}(f) &= \text{HMAX}_{h_{k-1}+\eta'}(f) \\ &= \text{HMAX}_{h_{k-1}}(f) - \eta' \text{EMAX}_{h_{k-1}}(f) \\ &= \text{HMAX}_{h_{k-1}}(f) - \eta \text{EMAX}_{h_{k-1}}(f) \\ &\quad - (h_k - h_{k-1}) \text{EMAX}_{h_{k-1}}(f) \end{aligned}$$

But (E22) applied to  $k-1$  and  $\eta = h_k - h_{k-1}$  yields

$$-(h_k - h_{k-1}) \text{EMAX}_{h_{k-1}}(f) = \text{HMAX}_{h_k}(f) - \text{HMAX}_{h_{k-1}}(f)$$

therefore we finally get, for  $\eta \in [h_{k-1} - h_k, 0)$ ,

$$\begin{aligned} \text{HMAX}_{h_k+\eta}(f) &= \text{HMAX}_{h_k}(f) - \eta \text{EMAX}_{h_{k-1}}(f) \\ &= \text{HMAX}_{h_k}(f) - \eta \text{EMAX}_{h_k}(f) \\ &\quad + \eta (\text{EMAX}_{h_k}(f) - \text{EMAX}_{h_{k-1}}(f)). \end{aligned} \quad (\text{E23})$$

Since  $\text{EMAX}_{h_k}(f) - \text{EMAX}_{h_{k-1}}$  does not depend on  $\eta$  and is not null (as  $\mathcal{M}_{h_k} \neq \mathcal{M}_{h_{k-1}}$ ),  $\eta (\text{EMAX}_{h_k}(f) - \text{EMAX}_{h_{k-1}}(f))$  is not neglectable with respect to  $\eta$ ,

and therefore  $-\text{EMAX}_{h_k}(f)$  cannot be the derivative of  $h \mapsto \text{HMAX}_h(f)$  at  $h_k$ , which shows that the latter function is not Fréchet differentiable at  $h_k$ .

## Appendix F Proof of Proposition 6

**“If direction”**  $\Leftarrow$

Assume  $x \in M_0$  for a set  $M_0 \in \mathcal{M}(f)$  such that  $\text{DYNMAX}(M_0) \geq h > 0$ . Then, for any  $\epsilon > 0$ ,  $\text{DYNMAX}(M_0) > h - \epsilon$  and by Proposition 1 we get

$$\text{HMAX}_{h-\epsilon}(f)(x) = f(x) - h + \epsilon.$$

Since  $h \mapsto \text{HMAX}_h(f)$  is continuous, by letting  $\epsilon$  go to zero, it comes

$$\text{HMAX}_h(f)(x) = f(x) - h.$$

**“Only if direction”**  $\Rightarrow$

Let us assume  $\text{HMAX}_h(f)(x) = f(x) - h$  for some  $x \in \Omega$ , and denote

$$M_0 := \text{CC}(\{x\}, \chi_{f(x)}(f))$$

the connected component of  $x$  in the upper levelset  $\chi_{f(x)}(f)$ . We will show that  $M_0 \in \mathcal{M}(f)$ .

Let us first prove that  $f$  is constant over  $M_0$ , and  $f(M_0) = f(x)$ . We already have that  $f(y) \geq f(x)$  for any  $y \in M_0$  by definition of  $\chi_{f(x)}(f)$ , and it remains to show that  $f(y) \leq f(x)$ . Suppose there is  $y_0 \in M_0$  such that  $f(y_0) > f(x)$ . Then there exists a path  $\gamma \in \Gamma_{x y_0}$  such that  $\gamma \subseteq M_0$  because  $M_0$  is connected. Hence  $\min_{z \in \gamma} f(z) = f(x)$  because  $x \in \gamma$  and  $f(z) \geq f(x)$  for each  $z \in \gamma$ . Now, either  $f(y_0) - h > f(x)$  or  $f(y_0) - h \leq f(x)$ .

**Case 1,  $f(y_0) - h > f(x)$ :** In that case

$$(f(y_0) - h) \wedge \min_{z \in \gamma} f(z) = f(x).$$

Since by Eq. (7),

$$\text{HMAX}_h(f)(x) = \max_{\substack{y \in \Omega, \\ \gamma \in \Gamma_{xy}}} \left\{ (f(y) - h) \wedge \min_{z \in \gamma} f(z) \right\}$$

we get

$$\text{HMAX}_h(f)(x) \geq f(x) > f(x) - h$$

which is a contradiction of the initial assumption.  
**Case 2,  $f(y_0) - h \leq f(x)$ :** By the same reasoning, in that case

$$(f(y_0) - h) \wedge \min_{z \in \gamma} f(z) = f(y_0) - h,$$

so that

$$\text{HMAX}_h(f)(x) \geq f(y_0) - h > f(x) - h$$

which is again a contradiction.

We have just shown that  $f(y) \leq f(x)$ , and therefore  $f(y) = f(x)$ , for every  $y \in M_0$ .

Next, notice that  $M_0 \neq \Omega$ , otherwise  $f$  would be constant. Furthermore, for any  $y \in \mathcal{N}_{M_0} \setminus M_0$ ,  $f(y) < f(x)$  by definition. Therefore,  $M_0$  is a regional maximum of  $f$ .

Finally, let us show that  $\text{DYNMAX}(M_0) \geq h$ .

If  $M_0$  is a global maximum, then  $\text{DYNMAX}(M_0) = \Delta_\Omega f \geq h$ .

If not, by Eq. (12),

$$\begin{aligned} \text{DYNMAX}(M_0) < h &\Rightarrow f(M_0) - h < \max_{\substack{\gamma \in \Gamma_{M_0 M'}, \\ f(M') > f(M)}} \min_{z \in \gamma} f(z) \\ &= \max_{\substack{\gamma \in \Gamma_{x M'}, \\ f(M') > f(M)}} \min_{z \in \gamma} f(z) \quad (\text{F24}) \end{aligned}$$

where the last equality is due to  $f(M_0) = f(x)$ , which implies that for any path  $\gamma \in \Gamma_{M_0 M'}$  there is a path  $\gamma' \in \Gamma_{x M'}$  such that  $\min_{z \in \gamma} f(z) = \min_{z \in \gamma'} f(z)$ .

Therefore, if  $\text{DYNMAX}(M_0) < h$  there exists  $M' \in \mathcal{M}(f)$  such that  $f(M') > f(M_0) = f(x)$ ,  $y' \in M'$  and  $\gamma \in \Gamma_{x y'}$  such that  $\min_{z \in \gamma} f(z) > f(M_0) - h = f(x) - h$ . Furthermore,  $f(y') - h = f(M') - h > f(M_0) - h = f(x) - h$  too, so eventually

$$(f(y') - h) \wedge \min_{z \in \gamma} f(z) > f(x) - h.$$

Again, by Eq. (7), we get

$$\text{HMAX}_h(f)(x) \geq (f(y') - h) \wedge \min_{z \in \gamma} f(z) > f(x) - h$$

which is a contradiction.

Therefore,  $\text{DYNMAX}(M_0) \geq h$ , which concludes the proof of Proposition 6.

## Appendix G Unet baseline details

### G.1 Unet architecture

The Unet architecture employed in Section 5 is a standard Unet [22] modified in accordance with [3]. However, it employs a depth of four instead of three, as in [3], for better performance. More precisely, the initial convolutional filter count is 32, augmented to 64, 128, 256, comprising 512 filters for the bottleneck convolutional layer.

### G.2 Unet training protocol

In order to conduct Unet training, we followed the methodology proposed by [3] by applying an isotropic Gaussian filter with  $\sigma = 6$  to the normalized point ground truth cell locations map. Subsequently, the resulting Gaussian map was multiplied by 100, facilitating the training of the model and serving as the supervised training label. The final cell counting number can be obtained by summing the aforementioned Gaussian map, divided by 100. The loss function employed for Unet training is the  $L1$  loss. The remaining training protocol settings, data augmentation settings and optimisation settings are selected in accordance with the descriptions provided in Section 4.3 (Subset of Cellpose dataset), Section 4.3.1 and Section 4.3.2, respectively.

## References

- [1] Kong, Y., Li, H., Ren, Y., Genchev, G.Z., Wang, X., Zhao, H., Xie, Z., Lu, H.: Automated yeast cells segmentation and counting using a parallel U-Net based two-stage framework. *OSA Continuum* **3**(4), 982–992 (2020)
- [2] He, S., Minn, K.T., Solnica-Krezel, L., Anastasio, M.A., Li, H.: Deeply-supervised density regression for automatic cell counting in microscopy images. *Medical Image Analysis* **68**, 101892 (2021)
- [3] Lazard, T., Blusseau, S., Velasco-Forero, S., Decencière, É., Flouret, V., Cohen, C., Baldewick, T.: Applying deep learning to melanocyte counting on fluorescent TRP1 labelled images of in vitro skin model. *Image Analysis & Stereology* (2022)

- [4] Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., Zoccoli, A.: Automating cell counting in fluorescent microscopy through deep learning with c-resunet. *Scientific Reports* **11**(1), 22920 (2021)
- [5] Alam, M.M., Islam, M.T.: Machine learning approach of automatic identification and counting of blood cells. *Healthcare technology letters* **6**(4), 103–108 (2019)
- [6] Xia, T., Jiang, R., Fu, Y.Q., Jin, N.: Automated blood cell detection and counting via deep learning for microfluidic point-of-care medical devices. In: *IOP Conference Series: Materials Science and Engineering*, vol. 646, p. 012048 (2019). IOP Publishing
- [7] Shakarami, A., Menhaj, M.B., Mahdavi-Hormat, A., Tarrah, H.: A fast and yet efficient yolov3 for blood cell detection. *Biomedical Signal Processing and Control* **66**, 102495 (2021)
- [8] Liu, X., Blusseau, S., Velasco-Forero, S.: Counting melanocytes with trainable h-maxima and connected component layers. In: *International Conference on Discrete Geometry and Mathematical Morphology*, pp. 417–430 (2024). Springer
- [9] Soille, P.: *Morphological Image Analysis: Principles and Applications* vol. 2. Springer, Berlin (1999)
- [10] Velasco-Forero, S., Pagès, R., Angulo, J.: Learnable empirical mode decomposition based on mathematical morphology. *SIAM Journal on Imaging Sciences* **15**(1), 23–44 (2022)
- [11] Velasco-Forero, S., Rhim, A., Angulo, J.: Fixed point layers for geodesic morphological operations. In: *British Machine Vision Conference*, pp. 1–11 (2022)
- [12] Najman, L., Talbot, H. (eds.): *Mathematical Morphology: from Theory to Applications*. John Wiley & Sons, Oxford (2013)
- [13] Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE transactions on image processing* **2**(2), 176–201 (1993)
- [14] Grimaud, M.: New measure of contrast: the dynamics. In: *Image Algebra and Morphological Image Processing III*, vol. 1769, pp. 292–305 (1992). SPIE
- [15] Ballester, C., Caselles, V., Monasse, P.: The tree of shapes of an image. *ESAIM: Control, Optimisation and Calculus of Variations* **9**, 1–18 (2003)
- [16] Tarjan, R.: Depth-first search and linear graph algorithms. *SIAM journal on computing* **1**(2), 146–160 (1972)
- [17] Guan, S., Loew, M.: Understanding the ability of deep neural networks to count connected components in images. In: *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1–7 (2020). IEEE
- [18] Rall, L.B.: *Automatic Differentiation: Techniques and Applications*. Springer, Berlin, Heidelberg (1981)
- [19] Duval, C., Cohen, C., Chagnoleau, C., Flouret, V., Bourreau, E., Bernerd, F.: Key regulatory role of dermal fibroblasts in pigmentation as demonstrated using a reconstructed skin model: impact of photo-aging. *PLoS One* **9**(12), 114182 (2014)
- [20] Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* **18**(1), 100–106 (2021)
- [21] Tieleman, T., Hinton, G., *et al.*: Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* **4**(2), 26–31 (2012)
- [22] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical

image segmentation. In: Medical Image Computing and Computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer

- [23] Chierchia, G., Perret, B.: Ultrametric fitting by gradient descent. *Advances in neural information processing systems* **32** (2019)
- [24] Charisopoulos, V., Maragos, P.: Morphological perceptrons: geometry and training algorithms. In: ISMM 2017, pp. 3–15 (2017). Springer
- [25] Blusseau, S.: Training morphological neural networks with gradient descent: some theoretical insights. In: DGMM Proceedings. Springer, Berlin, Heidelberg (2024)