



OPEN

## Self-normalized density map (SNDM) for counting microbiological objects

Krzysztof M. Graczyk<sup>1,4</sup>, Jarosław Pawłowski<sup>2,3,4</sup>, Sylwia Majchrowska<sup>2,3</sup> & Tomasz Golan<sup>2</sup>

The statistical properties of the density map (DM) approach to counting microbiological objects on images are studied in detail. The DM is given by U<sup>2</sup>-Net. Two statistical methods for deep neural networks are utilized: the bootstrap and the Monte Carlo (MC) dropout. The detailed analysis of the uncertainties for the DM predictions leads to a deeper understanding of the DM model's deficiencies. Based on our investigation, we propose a self-normalization module in the network. The improved network model, called *Self-Normalized Density Map* (SNDM), can correct its output density map by itself to accurately predict the total number of objects in the image. The SNDM architecture outperforms the original model. Moreover, both statistical frameworks—bootstrap and MC dropout—have consistent statistical results for SNDM, which were not observed in the original model. The SNDM efficiency is comparable with the detector-base models, such as Faster and Cascade R-CNN detectors.

An inseparable part of every statistical data analysis is the discussion of the uncertainties that characterize the data and the model's predictions. The first type of uncertainty is essential when formulating a model, building the likelihood, etc., while the other type reflects the model's predictive abilities. Indeed, providing the model's predictions with uncertainties allows assessing how confident the model is. A detailed error analysis of the model's predictions enables one to investigate the quality of the model, particularly verifying whether the model over- or under-fits the data. The latter aspect is crucial if one makes predictions outside the data domain.

Deep learning (DL) methods<sup>1</sup> proved their excellence in many domains of life<sup>2</sup>. Indeed, DL systems are utilized in image recognition and classification<sup>3</sup> as well as speech recognition<sup>4</sup> problems, or even in the game GO<sup>5</sup>. DL systems help optimize complex computational systems<sup>6</sup>, extracting non-trivial information hidden in big data<sup>7</sup>. DL systems are used in high-risk applications, such as autonomous driving<sup>8</sup> or medical image analysis<sup>9–11</sup>. In high-risk applications, the knowledge of how certain the system's predictions are is decidedly of great importance.

There is no doubt that proper estimation of uncertainties of a DL system's prediction's is crucial to control the validity of its model and applicability. Therefore, many groups have recently been studying this problem—for a recent review, see Gawlikowski et al.<sup>12</sup>. As many model parameters define a DL system, optimizing it requires immense computational power. Hence the successful statistical approach to the DL system, one which takes into account various sources of uncertainties, should be simple to implement and, at the same time, should not decrease the efficiency of the resulting DL model.

The methods for estimating the uncertainties for (shallow) neural networks (NN) were developed before the origin of deep learning. One of the directions was to use the Bayesian statistics that offer a consistent approach to error analysis for NNs<sup>13</sup>. Unfortunately, the Bayesian tools are not readily applicable to DL systems. For instance, the Bayesian technique developed by MacKay<sup>14</sup> utilizes the Laplace approximation, which is simple in implementation and fast in execution for the shallow NN, but very difficult to implement for DL systems. Indeed, in MacKay's approach to predicting the model uncertainties, the hessian matrix must be inverted, which is impossible to perform in the case of a DL system due to too many model parameters. Other Bayesian methods utilize the Monte Carlo (MC) chain algorithms<sup>15</sup>. Again, using this type of approach for the DL is difficult due to the massive number of model parameters. Nevertheless, the Bayesian-inspired, as well as the non-Bayesian statistical techniques for DL, have been developed recently<sup>16,17</sup>. They enable DL users to perform error analysis. Many proposed approaches are based on either the MC dropout or the bootstrap technique. The MC dropout<sup>18</sup>

<sup>1</sup>Institute for Theoretical Physics, University of Wrocław, pl. Maxa Borna 9, 50-343 Wrocław, Poland. <sup>2</sup>NeuroSYS, Rybacka 7, 53-656 Wrocław, Poland. <sup>3</sup>Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology, Wybrzeże S. Wyspiańskiego 27, 50-372 Wrocław, Poland. <sup>4</sup>These authors contributed equally: Krzysztof M. Graczyk and Jarosław Pawłowski. ✉email: krzysztof.graczyk@uwr.edu.pl

is an example of the Bayesian-inspired approach. In contrast, the bootstrap technique is an excellent example of the non-Bayesian method<sup>19</sup>, which statisticians have successfully exploited for years. Presently, it is widely adapted for shallow and deep NN<sup>20</sup>.

This paper aims to consider one of the practical (microbiological) applications of DL systems—a problem of counting microorganisms on images of Petri dishes. Our goal is to perform the error analysis, which will allow us to understand the validity of the counting system and let us propose the significant model's improvements. Indeed, our main achievement is a proposal of the self-normalization mechanism, implemented as additional modules in the network architecture. A counting system with self-normalization modules works more efficiently, accurately, and in the broader domain than the vanilla model.

There are many different approaches to object counting on images. An obvious way to face this type of problem is to build a regression model<sup>21,22</sup>. The biggest advantage of this method is that it is enough to label each image with the number of objects in it. If a dataset includes more detailed annotation with a bounding box for every object in an image, it is feasible to leverage detectors for object counting<sup>23–25</sup>. Eventually, an autoencoder can be used to estimate the density map (DM) based on a given image, which can be later integrated to obtain the number of objects<sup>26–30</sup>. It requires objects to be labeled with the coordinates of their center. The latter two methods performed better on the dataset we worked on in our experiments. The main goal of this study is to perform a comprehensive statistical analysis for a deep learning solution. Thus, we decided to focus on the density map approach because the architecture of autoencoders is usually less complex than recent neural networks for object detection, and the methods introduced in this paper are more transparent.

Counting microbial colonies on Petri dishes is an essential step in a microbiological laboratory to evaluate the cleanliness of the samples. Traditionally, this task is done manually or semi-automatically using traditional computer vision methods<sup>31,32</sup>. However, recent studies prove that a DL-based methodology accelerates the process<sup>23,25,29,33–35</sup>. We shall utilize the DM method to predict the number of microbial colonies on a Petri dish. The DM system is defined by U<sup>2</sup>-Net DL model<sup>36</sup>. To test the DM models, we utilize the AGAR (Annotated Germs for Automated Recognition) dataset<sup>23</sup>, which includes several thousand pictures of Petri dishes with five species of microbial colonies grown.

The vanilla U<sup>2</sup>-Net network accurately predicts the number of microbes up to about 60 colonies but underestimates the counts above. It also overestimates its prediction for some samples with just a few colonies. We adapt two known statistical approaches to perform an error analysis of the U<sup>2</sup>-Net predictions. Namely, we implement the bootstrap and MC dropout frameworks, respectively. The analysis of the model's uncertainties indicates that the main problem with the performance of U<sup>2</sup>-Net is not the localization of microorganisms itself, but rather the normalization of density maps for a large range of possible outcomes (from 0 to 100 colonies on a Petri dish). Thus, we propose a modification of the U<sup>2</sup>-Net, called *Self-Normalized Density Map* (SNDM) architecture, which includes extra normalization layers significantly improving the system's performance in the previously failing domain. Moreover, the SNDM works as efficiently as detector-base approaches, such as Faster<sup>37</sup> and Cascade<sup>38</sup> R-CNN.

The paper is organized as follows: in second section the DM and U<sup>2</sup>-Net architecture is introduced and the AGAR dataset is described; third section reviews two statistical approaches for neural networks designed to estimate the uncertainties, namely, bootstrap and MC dropout; in fourth section we present our numerical results. The SNDM is introduced, and its performance is studied in detail. Conclusions are drawn in the last section.

## Density map and microbiological data

**U<sup>2</sup>-net as density map.** The DM, denoted by  $\mathcal{M}$ , transforms the given input image,  $\mathbf{x}$ , into a density map target

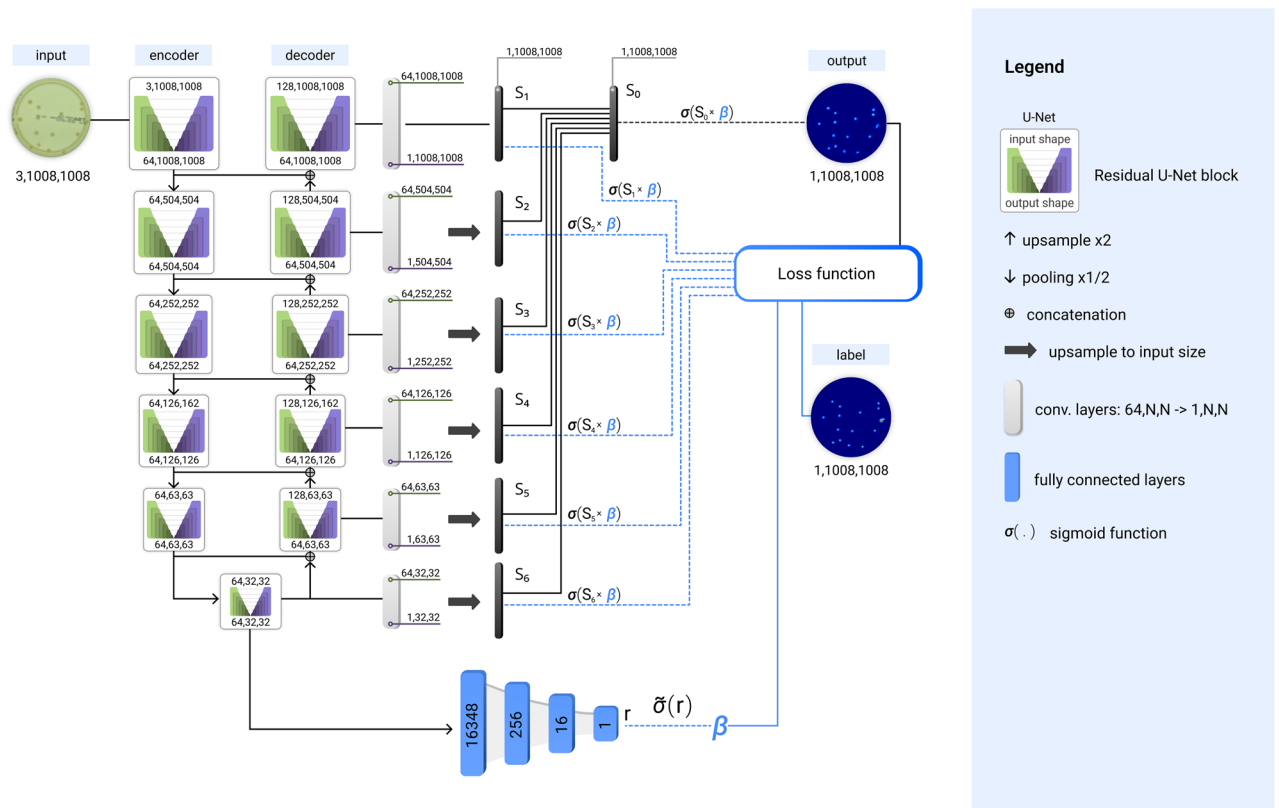
$$\mathbf{y} = \mathcal{M}(\mathbf{x}). \quad (1)$$

Note that  $\mathcal{M}$  is parametrized by weights and hyperparameters, which we intentionally omit in notation because they are not relevant to our discussion. In our approach, the total number of objects in the image is obtained by summing over the entries (the pixels) of the density map  $\mathbf{y}$ .

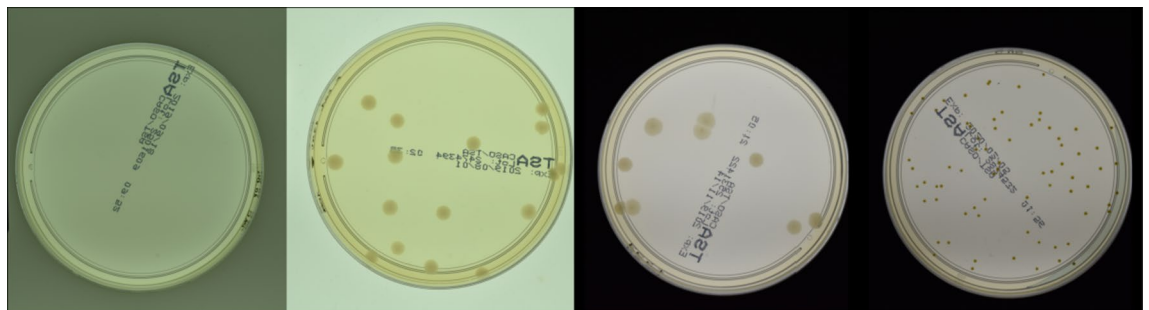
The convolutional neural network (CNN) has been designed to face problems such as image recognition, image classification<sup>39,40</sup>, and segmentation. Hence the CNN-based models are also proposed to transform the image into a density map. Among several types of CNNs, architecture U-Net-type<sup>41</sup> is the most successful in reproducing the density map of the objects<sup>30</sup>. In the present paper, we discuss the results of the analysis in which the U<sup>2</sup>-Net<sup>36</sup> is utilized.

The U<sup>2</sup>-Net has a form of a two-level nested U-shaped structure as presented in Fig. 1. The top-level is built of eleven U-Net blocks, connected by pooling (encoder branch) or upsampling layers (decoder branch), with additional concatenation connections. On the bottom level, each U-Net block itself also has an encoder-decoder structure, within additional residual connections in the U-Net block.

The analysis is performed on the AGAR dataset, including images of Petri dishes with bounding box annotations for each colony. To generate density maps, we assume that the center of objects lies in the center of the bounding box, and corresponding coordinates are added to an empty map in the form of single pixels. Next, the map is blurred with a Gaussian filter, which leads to a smooth density map normalized so that the sum over all pixels gives the total number of colonies. In that way we create the annotation map for each input image. The autoencoder is trained to transform a real image into a density map. Note that each density map is normalized so that the sum over all pixels gives the total number of objects. Hence the network predicting the density map for an input picture also provides information about the total number of objects.



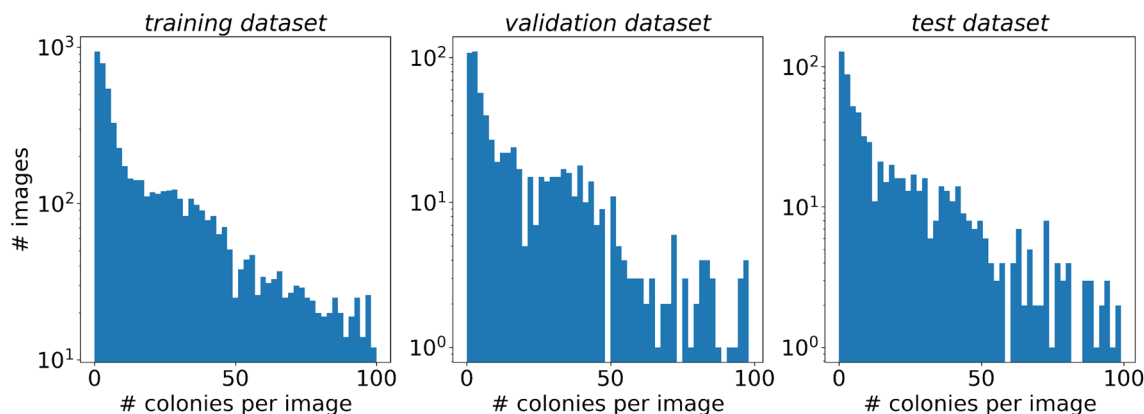
**Figure 1.** SNDM network: U<sup>2</sup>-Net extended by the normalization bypass-layers (blue) that output  $\beta$  normalization parameter. Note that  $\beta$  parameter multiplies last layers of the U<sup>2</sup>-Net that creates the output map.



**Figure 2.** Exemplary images of Petri dish from the *higher-resolution* subset of the AGAR dataset that was used during the numerical experiments. From the left: empty dish with no colonies, dish containing 16 colonies of *B. subtilis*, dish containing 9 colonies of *E. coli*, and dish containing 72 colonies of *S. aureus*. Samples differ in numbers of colonies, types of microbial species cultured, and image acquisition setups—detailed information can be found in Supplementary Material for Majchrowska et al.<sup>23</sup>.

**AGAR dataset.** As mentioned before, we utilize the AGAR dataset<sup>23</sup> to test the properties of the U<sup>2</sup>-Net in the MC dropout and bootstrap approaches. It is a huge microbiological database including 18 000 high-resolution images of Petri dishes with cultured colonies of 5 standard microbes taken under diverse lighting conditions. Professional microbiologists manually annotated microbial colonies by precisely labeling each colony using a *bounding box*. AGAR is diverse data collected to automate microbial colony counting. The data covers edge cases, including highly crowded plates with overlapping colonies, cases with extensive and tiny colonies, and cases where colonies are hardly visible, e.g., when located close to the dish edge. Typical plates captured in the various setups (three different illumination conditions) with microbial colonies of different shapes and sizes are shown in Fig. 2.

Note that AGAR mainly contains images with fewer than 50 colonies, about 85% of samples (see Fig. 3), or even empty dishes, which are essential for testing the models’ tendency to generate false-positive counts.



**Figure 3.** Histogram of the number of annotated microbial colonies per image in the training (left), validation (center), and test (right) subsets.

### Uncertainties in DM

Some methods of estimation of uncertainties in the predictions of the DM, for the U-Net architecture, have recently been discussed by Eaton-Rosen et al.<sup>42</sup>. Our paper focuses on two approaches: bootstrap and MC dropout adapted for U<sup>2</sup>-Net. We briefly review both of them in the following two subsections.

**Bootstrap method.** A bootstrap method for error estimation is one of the oldest and the most popular methods of statistics<sup>19</sup>. An extensive review can be found in the work of Hastie et al.<sup>43</sup>, whereas Tibshirani<sup>44</sup> discusses the application of the bootstrap technique in the neural network analyses. Two types of bootstrap approaches are distinguished, namely, pairs sampling and residual sampling algorithms. We adapt the first option.

The main idea of the bootstrap is to train the same model on some number of data subsets sampled (with replacement) from the original set. Then, as a result, one obtains the ensemble of models. In the inference mode, one makes predictions of all models in the ensemble. The bootstrap model prediction is then given by the mean over the model predictions in the ensemble. Additionally, the variance over the models' outcomes gives  $1\sigma$  uncertainty.

The procedure of taking the average over the models is robust to overfitting. In our applications, we consider  $B = 20$  bootstrap training subsets. The same network architecture is trained for each subset, and the best model is selected using the validation subset. If the noise of the data is random and not correlated (from sample to sample), then the total error obtained for the ensemble of models is  $B$  times smaller than the error of one model. For the detailed explanation, see Bishop<sup>45</sup>. Note that the bootstrap approach is not a Bayesian approach, but if it is combined with early stopping, it is interpreted as approximate inference<sup>46</sup>.

The known strengths of the bootstrap approach are the following:

- It is simple in implementation and relatively fast in execution, i.e., it is usually enough to run the analysis for  $B = 10$  to 20.
- The obtained model is given by the ensemble of networks of the same architecture but with a different configuration of network parameters.
- The predictions of the bootstrap model are similar to the Gaussian process (GP) in data range<sup>20</sup>. Indeed, the GP model is defined by some density distribution used to sample the model. The best model prediction is given by the average over the outcomes of the model samples.

The mean overall models give the response of the bootstrap model. Therefore the bootstrap model should not overfit the data. Indeed, a single fit obtained for a particular subset of the data might be overtrained, but the mean over the models is not. Hence the bootstrap model should be characterized by a good generalization ability.

*The main steps of the bootstrap approach:*

- consider the training data set  $\mathcal{D}_{train}$  with the total number of samples equals  $N$ ;
- obtain  $B$  subsets  $C_i \subset \mathcal{D}_{train}$  with an equal number of images.  $C_i$  is obtained by sampling from the data  $\mathcal{D}$  (with replacement)  $b$  images. Note that in our experiments  $B = 20$ , moreover  $b \approx 0.63 N$  as it is recommended in the literature<sup>43</sup>;
- consider network  $\mathcal{M}$  and train it on  $C_i$  data subset, as the result one gets the model  $\mathcal{M}_i^B \equiv \mathcal{M}(C_i)$ ;
- to make the predictions for given input  $\mathbf{x}$  compute the outcome of each network  $\mathcal{M}_i^B$  and take the mean, namely:

$$\overline{\mathcal{M}}(\mathbf{x}) = \frac{1}{B} \sum_{i=1}^B \mathcal{M}_i(\mathbf{x}), \quad (2)$$

- (v) to estimate the uncertainty for the prediction of the network calculate the variance:

$$\Delta^2 \mathcal{M}(\mathbf{x}) = \frac{1}{B} \sum_{i=1}^B (\overline{\mathcal{M}}(\mathbf{x}) - \mathcal{M}_i(\mathbf{x}))^2, \quad (3)$$

where  $\sqrt{\Delta^2 \mathcal{M}(\mathbf{x})}$  is interpreted as  $1\sigma$  uncertainty.

**MC dropout.** In 1995, Neal proved that a class of neural networks with a single hidden layer and an infinite number of units converge to the Gaussian process<sup>15</sup>. Moreover, Rasmussen and Williams<sup>47</sup> studied Gaussian process methods in supervised learning. In both cases, the Bayesian statistics stand a background for all derivations. It is believed that the GP approach allows accessing the model prediction uncertainties with reasonable accuracy.

Motivated by the Bayesian statistics as well as the success of GP techniques in the estimation of model uncertainties, Gal and Ghahramani studied the dropout technique<sup>18,48</sup>. They showed that a deep neural network with a dropout layer after each weight layer can be understood as the Gaussian process.

The main idea of the approach is to keep dropout layers active in the training and inference modes. It means that, for a given input  $\mathbf{x}$ , every prediction is computed with a slightly different configuration of active units. Therefore, to obtain the MC dropout prediction for a given neural network, one has to run it  $N$ -times. Then, the mean over the predictions gives the best value while the variance estimates the uncertainty of the predictions.

*The main steps of the MC dropout:*

- (i) Consider a network  $\mathcal{M}_{dropout}$ ;
- (ii) Train the  $\mathcal{M}_{dropout}$  on the data set  $\mathcal{D}_{train}$ . The training is run with regularization—we apply Adam algorithm with weight decay  $\mu = 0.0005$ . Note that during the training the dropout layers are active.
- (iii) To make a prediction for given input vector  $\mathbf{x}$ , run the network  $\mathcal{M}_{dropout}$ ,  $r = 20$  times, keeping dropout layers active. Then the dropout network prediction is given by the

$$\overline{\mathcal{M}}_{dropout}(\mathbf{x}) = \frac{1}{r} \sum_{i=1}^r \mathcal{M}_{dropout}^i(\mathbf{x}), \quad (4)$$

where  $\mathcal{M}_{dropout}^i(\mathbf{x})$  denotes the  $i$ -th run prediction of the network. The  $1\sigma$  uncertainty is given by the square root of the variance:

$$\Delta^2 \mathcal{M}_{dropout}(\mathbf{x}) = \frac{1}{r} \sum_{i=1}^r (\overline{\mathcal{M}}_{dropout}(\mathbf{x}) - \mathcal{M}_{dropout}^i(\mathbf{x}))^2. \quad (5)$$

## Numerical results

This section examines the properties of the U<sup>2</sup>-Net architecture as a density map predictor. The results of these studies allow us to propose the self-normalization mechanism in the U<sup>2</sup>-Net model. We justify the relevance of this approach in the following sections.

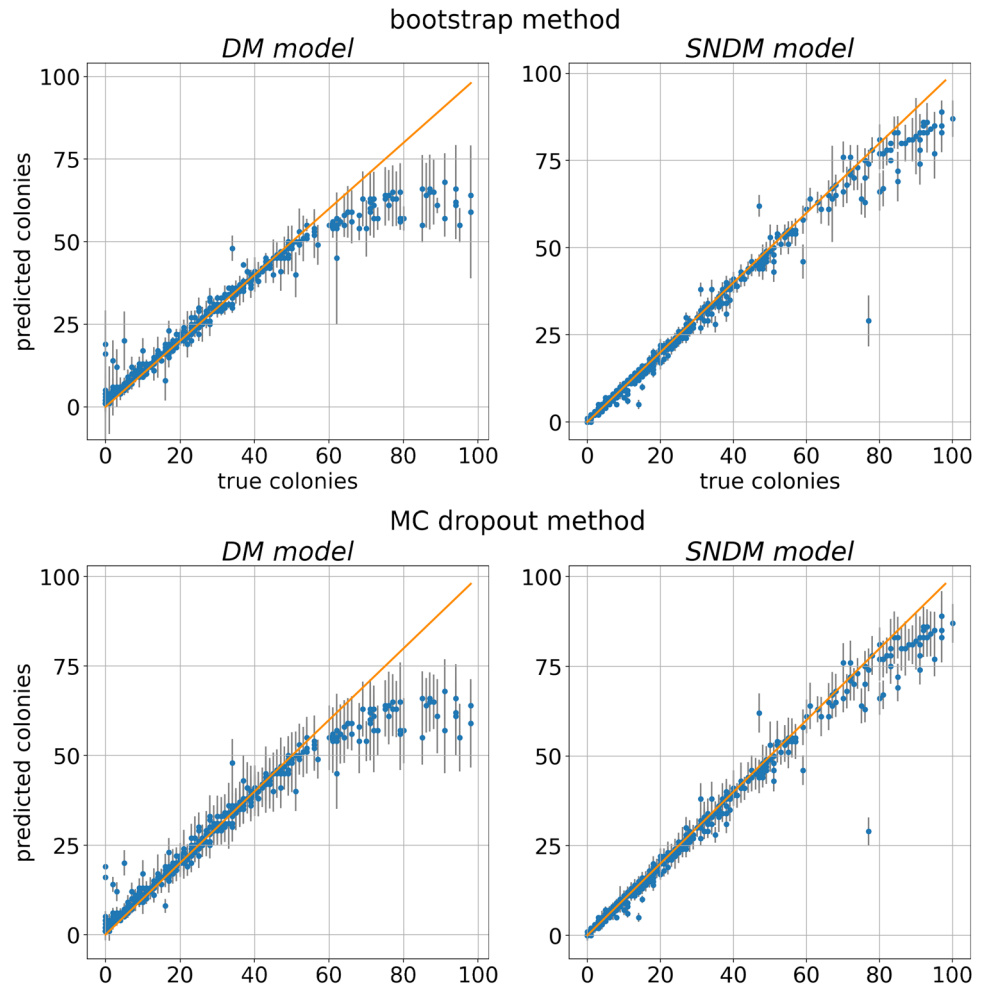
To study the U<sup>2</sup>-Net properties, we consider the *higher-resolution* subset of the AGAR dataset, represented by RGB images of Petri dishes with a resolution of about 4000 × 4000 pixels. We limit our discussion to images containing no more than 100 colonies. Because, from the microbiological point of view, the accuracy of automatic counting is less relevant for more populated samples, i.e. determining whether the sample is contaminated or not is the most important.

The analyzed data is divided into three subsets: training (5656 images), validation (706 images), and test (706 images) data subset. As introduced earlier, we consider two types of statistical approaches for estimating uncertainties in predictions of deep neural networks: the bootstrap and the MC dropout. In both methods, we exploit the same network architecture shown in Fig. 1. Moreover, the same training scheme is adapted for both models—the Adam optimization algorithm with *learning rate* of 0.0001, *weight decay* equals 0.0005, and the size of the minibatch of two. The training takes at least 100 epochs.

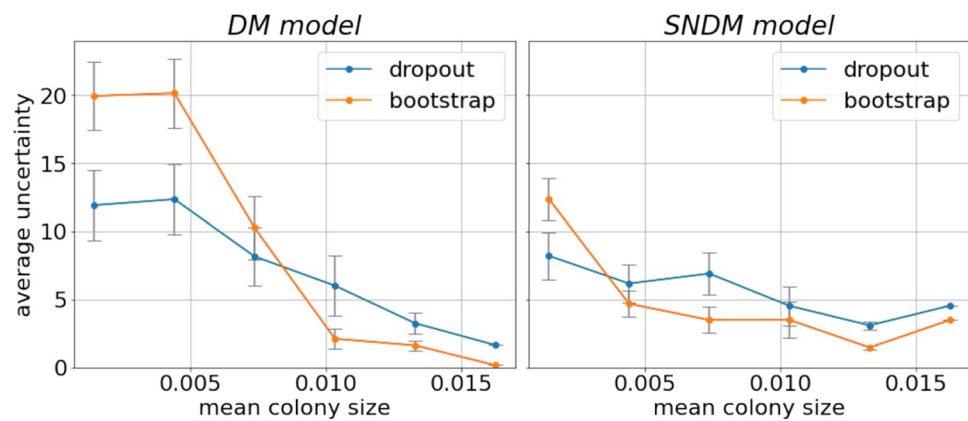
We ran the computations on NVIDIA TITAN RTX GPU with 24 GB of memory. The training time was about 40 h for MC dropout. The bootstrap experiments took longer, and the training lasted about four days, performed in parallel on five GPUs.

**Performance of vanilla U<sup>2</sup>-Net as a density map predictor.** To access the quality of both discussed models, we compare the predictions of the network versus the ground truth computed on the test data set; see the left column in Fig. 4. Each predicted number of colonies is given with  $1\sigma$  uncertainty. Note that both statistical models work well for the samples with the number of colonies smaller than 60. However, for images with the number of colonies lower than 10, the DM count is overestimated. Predictions with tiny uncertainties characterize the MC dropout for these cases. Consequently, the model disagrees with the data within the  $1\sigma$  confidence level. In contrast, the bootstrap model predictions have larger uncertainties. Hence, the model agrees with the ground truth in the  $1\sigma$  level. From that perspective, the bootstrap approach works better than the MC dropout.

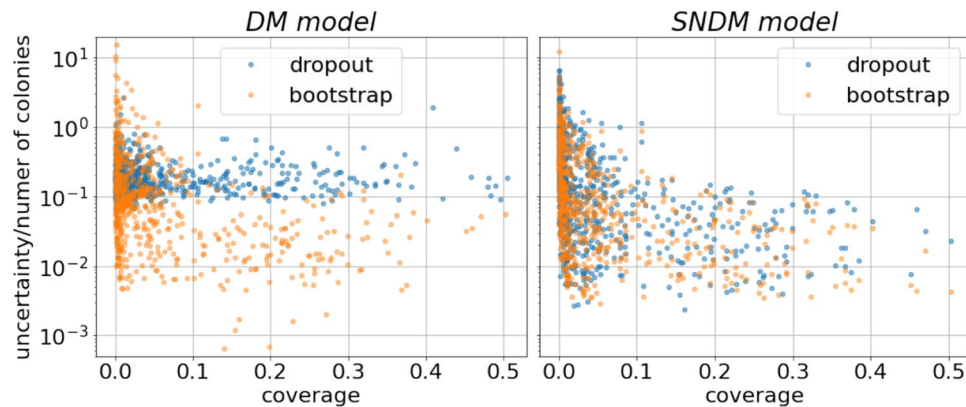
The detailed examination of our results shows that the uncertainty of the model's predictions depends on the mean size of colonies for a given dish. It is illustrated in Fig. 5 (left panel). Interestingly, for the images with



**Figure 4.** Dependence of the predicted number of colonies on ground truth for the test data set within the DM model (right), and self-normalized DM model (left). Note that for ideal detection every blue dot representing a single Petri dish image should lay on the  $y = x$  orange line. Uncertainties, denoted by black bars, in the bootstrap and MC dropout approaches are given in top and bottom rows, respectively.



**Figure 5.** Dependence of the counting uncertainty on mean colony size calculated for the test subset within both models. Uncertainty values are averaged on intervals of mean colonies size.



**Figure 6.** Dependence of the counting uncertainty per single colony on coverage factor calculated for the test subset using both models.

smaller colonies, the uncertainties for the bootstrap approach are larger than for the dropout model. However, when the size of the colonies increases, the effect is the opposite.

We noted that the uncertainties of the model's prediction (normalized to the number of colonies) weakly depend on the dish coverage (the fraction of the dish covered by colonies), as shown in Fig. 6 (left panel). It is interesting to note that the MC dropout uncertainties are systematically larger than for the bootstrap approach, see Table 1, where the average (over the test dataset) uncertainties are given.

In Fig. 7, we showed four panels. Each contains an input picture, a map of annotations, network prediction, and a map of uncertainties computed for each output pixel. We show the results for two dishes that contain 16 and 72 bacterial colonies, respectively. The DM prediction agrees with the ground truth in the first case (first panel). In contrast to the second case (third panel), the DM underestimates the count. However, in the latter case, the DM correctly reproduces spots representing the objects in the image, but the most significant uncertainties are consistently observed in the spot positions. Therefore, we conclude that the DM reconstructs the map of spots (objects) correctly but has a problem finding the correct normalization for the images with a larger number of colonies. Therefore, we propose a modification of the network architecture so that the network can correct by itself the normalization of the output.

**Self-normalized DM (SNDM).** To propose the normalization module in the network, we must explain how the loss function is constructed in the U<sup>2</sup>-Net model.

Let us denote by  $E_{U^2}$  the loss function U<sup>2</sup>-Net. It is a function of the data and model parameters. From the statistical perspective, it refers to the logarithm of the likelihood function. In practice,  $E_{U^2}$  has a quite complicated structure<sup>36</sup>:

$$E_{U^2} = \sum_{\text{images } M=0}^6 \mathcal{L}(P_M, P^G) \quad (6)$$

where  $P_M$  and  $P^G$  denote predicted density map by  $M$ -th U-Net block and the ground truth, respectively.

The  $\mathcal{L}_M$  is the standard binary cross entropy function:

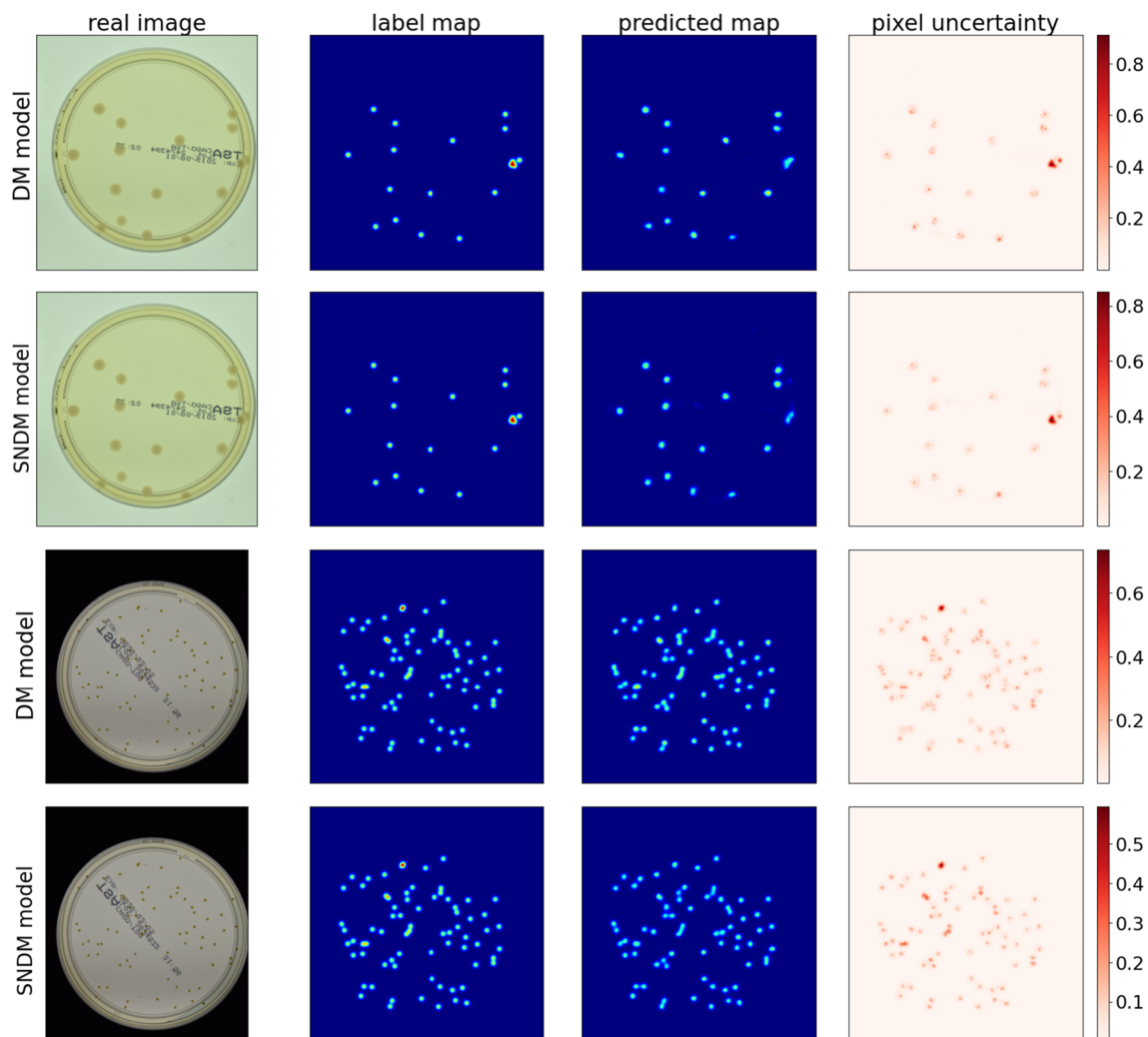
$$\mathcal{L}(P, P_G) = - \sum_{i,j}^{H,W} [P_{Gij} \ln P_{ij} + (1 - P_{Gij}) \ln(1 - P_{ij})], \quad (7)$$

where  $i, j$  denotes the pixel indexes in the figure of the width  $W$  and height  $H$ .

The U<sup>2</sup>-Net generates six output maps, denoted by  $S_{1,\dots,6}$ , from the encoder ( $S_6$ ) and five from the decoder branch ( $S_{1,\dots,5}$ ), respectively. Outputs from smaller blocks ( $S_{2,\dots,6}$ ) are next upsampled to fit the size of the output map. Note that  $S_1$  already has the proper  $1008 \times 1008$  size. After upsampling (we still denote them by  $S_i$ ), each of them gives the density map and  $P_i = \sigma(S_i)$ , where  $\sigma$  is the sigmoid activation function. All maps  $P_{1..6}$  contributes to the loss function. Moreover,  $S_{1..6}$  maps are fused with a concatenation operation followed by a  $1 \times 1$  convolution layer (giving  $S_0$ ). Then  $P_0 = \sigma(S_0)$  is the final density map output of the U<sup>2</sup>-Net. The  $P_0$  contributes also to the loss function.

Our idea is to allow the network to correct the normalization of output pixels in  $S_i$  maps. Since every output pixel ranges from 0 to 1, one can not simply re-scale the  $P_i$ -outputs by some normalization number. Therefore, we propose the rescaling of pixels normalization in  $S_i$  map. We use the property that sigmoid is a linear function on neighborhood, namely,  $\sigma(x) \approx \frac{1}{2} + \frac{x}{2} + \dots$ . Hence, the rescaling of the sigmoid argument changes the total normalization of the sigmoid almost linearly in this range. In practice, we rescale each  $S_i$  output by a factor  $\beta$ , so

$$P_i = \sigma(S_i) \rightarrow \sigma(\beta S_i). \quad (8)$$



**Figure 7.** Examples of density maps for two exemplary input images from the test subset (with 16 and 72 labelled colonies), together with their uncertainties per pixel: the DM model (rows 1st and 3rd), and the SNDM model (rows 2nd and 4th). In the first case DM model predicts  $17 \pm 0.99$ , while SNDM predicts  $16 \pm 0.44$  colonies. In the second case DM predicts  $63 \pm 3.32$ , and the SNDM predicts  $71 \pm 1.57$  colonies. Uncertainty maps calculated using MC dropout method.

The normalization value  $\beta$  might depart from one, and it can be a function of the distribution of colonies, the total coverage, etc. Therefore, we assume that  $\beta$  depends on the input image and is an additional network outcome. To limit the number of parameters that define  $\beta$  we connect it with the smallest  $U$ -Net block of  $U^2$ -Net, the encoder part, see Fig. 1. In the rest of the paper, the  $U^2$ -Net architecture with our modification will be called a *self-normalized density map* (SNDM).

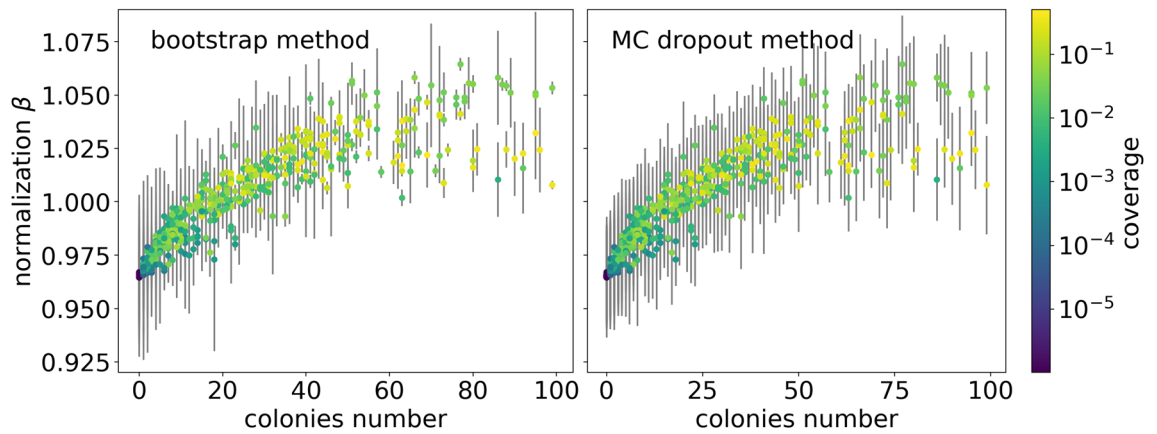
In principle, the system should work without the need for corrections,  $\beta = 1$ , hence we expect that  $\beta$  should take values around one. To keep the control  $\beta$ -dependence, we add to the loss term a penalty term:

$$E_{U^2} \rightarrow E_{U^2} + \sum_{\text{images}} \frac{1}{2} (1 - \beta)^2. \quad (9)$$

In general  $\beta$  can take any positive number, but, in practical applications, we assume that  $\beta = \tilde{\sigma}(\dots) = \lambda \sigma(\dots)$  to speed up the process of training. It guarantees that normalization ranges from 0 to 1.5. In our experiments, we tested a sigmoid with  $\lambda = 2$  and  $\tilde{\sigma} = \text{ReLU}$ . However, the most optimal results have been obtained for the sigmoid activation function with  $\lambda = 1.5$ .

In Fig. 4 (right column of panel), we plot the predictions of SNDM versus ground truth. We see that modifying  $U^2$ -Net architecture by introducing the self-normalization module significantly improves the network's predictions. Indeed, the SNDM works well for images containing less than 60 objects and samples with the number of objects larger than 60. Interestingly, in the improved model, the bootstrap and MC dropout predictions are in agreement. The estimated uncertainties for both approaches became similar, see Figs. 5 and 6 (right panel), as well as Table 1. Moreover, the bootstrap uncertainties for SNDM are reduced with respect to the original model.

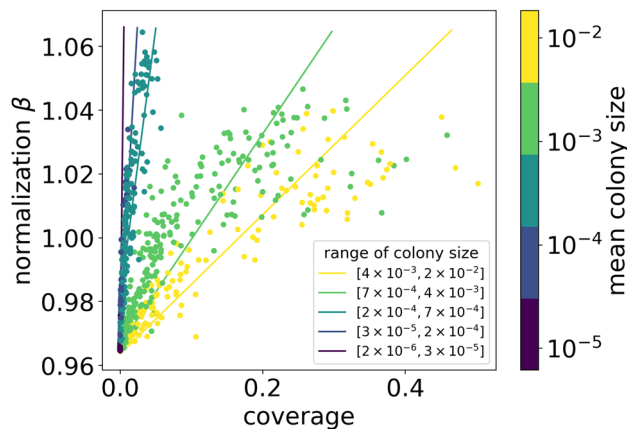




**Figure 8.** Dependence of the normalization parameter  $\beta$  on the number of colonies. Uncertainties of the  $\beta$  parameter calculated using bootstrap method (left) and MC dropout (right). Results for the test subset.

Dishes with:	Bootstrap		MC dropout	
	DM	SNDM	DM	SNDM
$\leq 50$ colonies	1.26	0.80	2.38	1.31
$> 50$ colonies	5.67	3.94	7.74	5.01
overall	1.76	1.19	2.99	1.78

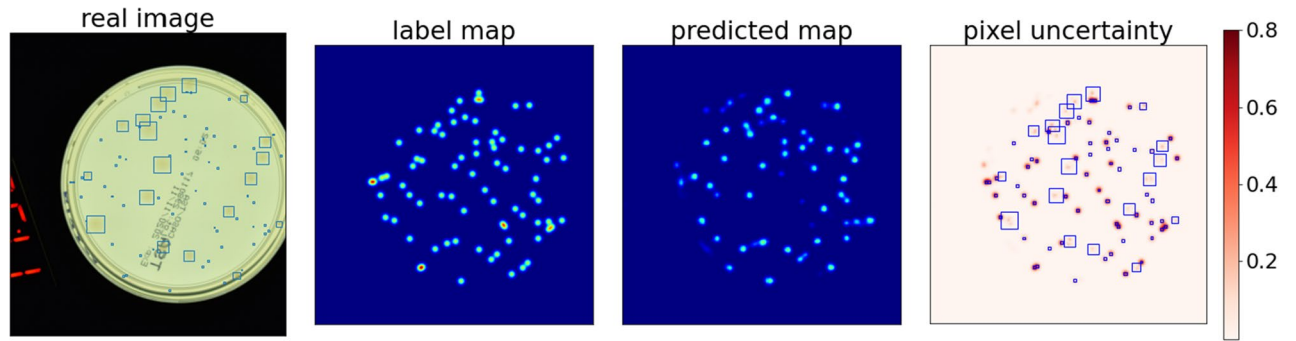
**Table 1.** Average uncertainties of predictions for the test dataset within the DM and the SNDM models in the bootstrap and MC dropout approaches (as presented in Fig. 4). Uncertainties averaged in a group of less (more) crowded dishes with  $\leq 50$  ( $> 50$ ) colonies and overall.



**Figure 9.** Dependence of  $\beta$  parameter on dish coverage, together with mean colony size indicated by the color of points. The points for different ranges of colony size (see the legend) cluster in subsets that show approximately linear dependence between  $\beta$  and coverage. Additionally, a linear function is fitted in each range, with a slope that decreases with colony size. Results obtained within MC dropout for the test subset.

The normalization parameter  $\beta$  depends on the features of the input image, such as the number of colonies. It is illustrated in Fig. 8. For both statistical approaches,  $\beta$  varies from 0.96 to 1.06 with the tendency to take lower values for less crowded images. On the other hand, for the images containing more than 50 objects,  $\beta$  varies less. It is also worth mentioning that  $\beta$  takes a wide range of values for a given number of colonies, which varies with the number of colonies, which translates to coverage and other features like mean colony size. It is visible in Fig. 9 where we plot the dependence of  $\beta$  on the dish coverage. When we group the samples by the mean colony size (indicated by different colors of points in Fig. 9), one can observe that  $\beta$  almost linearly depends on coverage. Moreover, the slope of the fitted lines increases gradually with the mean size of colonies.

Figure 7, shows density maps together with pixel uncertainties for two input pictures with 16 and 72 objects, respectively. The predictions for the SNDM are shown in the second and the fourth panel, respectively. We see



**Figure 10.** Density maps of one of the outliers from Fig. 4 predicted with the SNDM. The model predicts  $29 \pm 7.30$  colonies, while the sample contains 77 colonies of two types: 19 *P. aeruginosa* (big ones), and 58 *C. albicans* (small ones). The blue boxes represent all the colonies' ground truth position and size. Uncertainties calculated using the bootstrap method.

that SNDM outperforms the original U<sup>2</sup>-Net. Indeed, the SNDM accurately predicts the number of colonies for both images.

Eventually, we note that, overlapping objects is one of the difficulties the counter system must face. In Fig. 7, we see an example of colonies that overlap. In this case, the amplitude of density maps is naturally enlarged. But the interesting thing is that also uncertainties have higher values there—see the pixel uncertainty maps in the last column of Fig. 7. Therefore, we may conclude that these areas have a more substantial impact on counting uncertainties. However, if one compares DM and SNDM pixel uncertainty in the overlapping regions, then in the SNDM case, uncertainties are slightly reduced compared to DM.

The detailed inspection of the results presented in Fig. 4 shows that there are input figures for which the SNDM network overshoots. In Fig. 10 we plot the predicted map with uncertainties for such a case. We see that the network's main problem is the presence of colonies of different sizes on the same dish. It complicates the inference of the normalization parameter, which also depends on the size of the colonies, see in Fig. 9. The blue boxes in Fig. 10 represent ground truth labels showing the precise position and size of the colonies. It should be also noted that the pixel uncertainties for small colonies in the small box area are much larger than for big colonies located in bigger boxes.

We close the discussion of the numerical result by comparing the SNDM model predictions with the other popular counting systems. In one of our previous papers<sup>23,25</sup> we discussed some of them. For the comparison, we choose two of the most successful approaches (in our experiments) to count AGAR microbiological colonies, namely, Faster R-CNN detector<sup>37</sup> and Cascade R-CNN detector<sup>38</sup>. Note that, for both models, the algorithm first performs the detection of interesting objects and then counts them.

To quantitatively compare various models predictions, we consider a standard Mean Absolute Error:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |n_i - \tilde{n}_i|, \quad (10)$$

and less common in usage, symmetric Mean Absolute Percentage Error:

$$\text{sMAPE} = \frac{100\%}{N} \sum_{i=1}^N \frac{|n_i - \tilde{n}_i|}{|n_i + \tilde{n}_i|}, \quad (11)$$

where  $N$  is a number of all samples,  $n_i$  is true count of microbe colonies present on  $i$ -th image, and  $\tilde{n}_i$  is predicted number of colonies. Note that the sMAPE shows the relative error with respect to the overall number of colonies.

Both metrics are obtained for the *higher-resolution* subset of the AGAR dataset and the four models: Faster and Cascade detectors, as well as DM and SNDM in the MC dropout approach. The metrics are given in Table 2. We see that counting by detection gives better results than the standard DM model. But the SNDM gives comparable or even slightly better counting results than the detectors. Further discussion of other deep learning detectors, e.g., well-known YOLO, can be found in our previous paper<sup>24</sup>. Note that preparing annotations for the density map models are less expensive than for detector models. Moreover, the training of the detector systems is more resources-demanding and time-consuming than the density map models. Therefore, for the data for which the density map approaches work with the same or better efficiency as the detectors, the density map systems as SNDM are recommended.

## Summary

We have shown that the detailed analysis of the uncertainties in the deep neural network model leads to a deeper understanding of its limitations. It allows us to assess how uncertain the predictions are. More profound knowledge of the model allowed us to modify the DM model by introducing the Self-Normalization mechanism. The SNDM significantly improves the DM predictions. Indeed, before the modification, the model underestimated the number of colonies for highly populated dishes. In contrast, for some images with a small number of colonies, the

Model	Detection metrics	
	MAE	sMAPE
DM	5.74	12.58%
Self-normalized DM	3.65	4.05%
Faster R-CNN	4.75	5.32%
Cascade R-CNN	4.31	4.86 %

**Table 2.** Counting metrics, mean absolute error (MAE) (Eq. 10) and symmetric Mean Absolute Percentage Error (sMAPE) (Eq. 11), for the DM and SNDM model (in MC dropout) in comparison with other counting methods that use popular R-CNN object detectors: Faster<sup>37</sup> and Cascade<sup>38</sup>, which are discussed by Majchrowska *et al.*<sup>23</sup>.

DM model overestimated the counts. The SNDM network proposed by us can correctly count in both mentioned domains. Its performance is comparable with the other counting systems, such as Faster and Cascade R-CNN detectors. Eventually, we note that the bootstrap and the MC dropout predictions and the estimate of uncertainties in the SNDM are in statistical agreement, which was not observed in the original model.

### Data availability

The AGAR dataset used during the current study is available from <https://agar.neurosys.com/> on reasonable request.

Received: 17 March 2022; Accepted: 14 June 2022

Published online: 22 June 2022

### References

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 EP (2015).
2. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536. <https://doi.org/10.1038/323533a0> (1986).
3. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25 (Curran Associates, Inc., 2012).
4. Amodei, D. *et al.* Deep speech 2: End-to-end speech recognition in english and mandarin, *arXiv preprint arXiv:1512.02595* (2015).
5. Silver, D. *et al.* Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–489. <https://doi.org/10.1038/nature16961> (2016).
6. Graczyk, K. M. & Matyka, M. Predicting porosity, permeability, and tortuosity of porous media from images by deep learning. *Sci. Rep.* **10**, 21488. <https://doi.org/10.1038/s41598-020-78415-x> (2020).
7. Najafabadi, M. M. *et al.* Deep learning applications and challenges in big data analytics. *J. Big Data* **2**, 1. <https://doi.org/10.1186/s40537-014-0007-7> (2015).
8. Pek, C., Manzinger, S., Koschi, M. & Althoff, M. Using online verification to prevent autonomous vehicles from causing accidents. *Nat. Mach. Intell.* **2**, 518–528. <https://doi.org/10.1038/s42256-020-0225-y> (2020).
9. Nair, T., Precup, D., Arnold, D. L. & Arbel, T. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med. Image Anal.* **59**, 101557. <https://doi.org/10.1016/j.media.2019.101557> (2020).
10. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1> (2019).
11. Wang, G., Ye, J. C. & De Man, B. Deep learning for tomographic image reconstruction. *Nat. Mach. Intell.* **2**, 737–748. <https://doi.org/10.1038/s42256-020-00273-z> (2020).
12. Gawlikowski, J. *et al.* A survey of uncertainty in deep neural networks, *arXiv preprint, arXiv:2107.03342* (2021).
13. C. M., Bishop. *Neural Networks for Pattern Recognition* (Oxford University Press, 1995).
14. MacKay, D. *Bayesian Methods for Adaptive Models*. Ph.D. thesis, California Institute of Technology (1991).
15. Neal, R. M. *Bayesian learning for neural networks*. Ph.D. thesis, Graduate Department of Computer Science in University of Toronto (1995).
16. Caldeira, J. & Nord, B. *Deeply uncertain: Comparing methods of uncertainty quantification in deep learning algorithms*, *arXiv preprint, arXiv:2004.10710* (2020).
17. Ovadia, Y. *et al.* Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift, *arXiv preprint, arXiv:1906.02530* (2019).
18. Gal, Y. & Ghahramani, Z. *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*, *arXiv preprint, arXiv:1506.02142* (2015).
19. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**, 1–26. <https://doi.org/10.1214/aos/1176344552> (1979).
20. Osband, I., Blundell, C., Pritzel, A. & Roy, B. V. *Deep exploration via bootstrapped DQN*, *arXiv preprint, arXiv:1602.04621* (2016).
21. Chattopadhyay, P., Vedantam, R., Selvaraju, R. R., Batra, D. & Parikh, D. Counting everyday objects in everyday scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1135–1144 (2017).
22. Hoekendijk, J. *et al.* Counting using deep learning regression gives value to ecological surveys. *Sci. Rep.* **11**, 1–12 (2021).
23. Majchrowska, S. *et al.* AGAR a microbial colony dataset for deep learning detection. *arXiv preprint, arXiv:2108.01234* (2021).
24. Majchrowska, S. *et al.* Deep neural networks approach to microbial colony detection—a comparative analysis. *arXiv preprint, arXiv:2108.10103* (2021).
25. Pawłowski, J., Majchrowska, S. & Golan, T. Generation of microbial colonies dataset with deep learning style transfer. *Sci. Rep.* **12**, 5212. <https://doi.org/10.1038/s41598-022-09264-z> (2022).
26. Lempitsky, V. & Zisserman, A. Learning to count objects in images. *Advances in neural information processing systems* **23** (NIPS 2010).
27. Arteta, C., Lempitsky, V., Noble, J. A. & Zisserman, A. Interactive object counting. In *European conference on computer vision*, 504–518 (Springer, 2014).

28. Zhang, C., Li, H., Wang, X. & Yang, X. Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 833–841 (2015).
29. Jiang, N. & Yu, F. Multi-column network for cell counting. *OSA Continuum* **3**, 1834–1846. <https://doi.org/10.1364/OSAC.396603> (2020).
30. Xie, W., Noble, J. A. & Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **6**, 283–292. <https://doi.org/10.1080/21681163.2016.1149104> (2018).
31. Selinummi, J., Seppälä, J., Yli-Harja, O. & Puhakka, J. A. Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *Biotechniques* **39**, 859–863 (2005).
32. Chen, W.-B. & Zhang, C. An automated bacterial colony counting and classification system. *Inf. Syst. Front.* **11**, 349–368 (2009).
33. Beznik, T., Smyth, P., de Lannoy, G. & Lee, J. A. Deep learning to detect bacterial colonies for the production of vaccines. *arXiv preprint arXiv:2009.00926* (2020).
34. Zhang, Y., Jiang, H., Ye, T. & Juhas, M. Deep learning for imaging and detection of microorganisms. *Trends Microbiol.* **29**, 569–572. <https://doi.org/10.1016/j.tim.2021.01.006> (2021).
35. Wang, H. *et al.* Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light: Sci. Appl.* **9**, 1–17 (2020).
36. Qin, X. *et al.* U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recogn.* **106**, 107404. <https://doi.org/10.1016/j.patcog.2020.107404> (2020).
37. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural. Inf. Process. Syst.* **28**, 91–99 (2015).
38. Cai, Z. & Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6154–6162 (2018).
39. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
40. Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Fürnkranz, J. & Joachims, T. (eds.) Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814 (2010).
41. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation, *arXiv preprint, arXiv:1505.04597* (2015).
42. Eaton-Rosen, Z., Varsavsky, T., Ourselin, S. & Cardoso, M. J. As easy as 1, 2...4? uncertainty in counting tasks for medical imaging. In *Shen, D. et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 356–364 (Springer International Publishing, Cham, 2019).
43. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*. Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001).
44. Tibshirani, R. A comparison of some error estimates for neural network models. *Neural Comput.* **8**, 152–163 (1996).
45. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, New York, 2007).
46. Duvenaud, D., Maclaurin, D. & Adams, R. *Early stopping as nonparametric variational inference*, *Proceedings of Machine Learning Research* Vol. **51**, 1070–1077 (Cadiz, Spain, 2016).
47. Rasmussen, C. & Williams, C. *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning* (MIT Press, Cambridge, MA, 2006).
48. Gal, Y. & Ghahramani, Z. *Dropout as a bayesian approximation: Appendix*, *arXiv preprint, arXiv:1506.02157* (2015).

## Acknowledgements

Project “Development of a new method for detection and identifying bacterial colonies using artificial neural networks and machine learning algorithms” is co-financed from European Union funds under the European Regional Development Funds as part of the Smart Growth Operational Program. Project implemented as part of the National Centre for Research and Development: Fast Track (grant no. POIR.01.01.01-00-0040/18). K.M.G. has been partly supported by the program “Excellence initiative—research university” for the years 2020–2026 for the University of Wrocław, as well as by MOZART grant founded by Wrocław Academic Center.

## Author contributions

K.M.G. formulated the statistical framework for the density map and proposed a self-normalization mechanism. He wrote the manuscript and participated in the discussion of the results. At the initial stage, he performed the first numerical analyses not shown in the present paper. J.P. implemented the SNDM network model, performed all numerical analyses, and participated in studying the results. He worked on the manuscript’s text, and he is the author of all figures presented in the manuscript. S.M. proposed the usage of the U2-net model to count the objects. She participated in the initial discussions and took part in works on the manuscript’s text. T.G. took part in all discussions concerning the formulation of the model, planning the numerical experiments, analyses of the numerical results. He participated in the preparation of the manuscript text too.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to K.M.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022