ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

BrightNet: A Deep CNN for OLED-Based Point of Care Immunofluorescent Diagnostic Systems

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Samore A., Rusci M., Lazzaro D., Melpignano P., Benini L., Morigi S. (2020). BrightNet: A Deep CNN for OLED-Based Point of Care Immunofluorescent Diagnostic Systems. IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, 69(9), 6766-6775 [10.1109/TIM.2020.2973913].

Availability:
This version is available at: https://hdl.handle.net/11585/771945 since: 2021-03-28

Published:
DOI: http://doi.org/10.1109/TIM.2020.2973913

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

A. Samorè, M. Rusci, D. Lazzaro, P. Melpignano, L. Benini and S. Morigi, "BrightNet: A Deep CNN for OLED-Based Point of Care Immunofluorescent Diagnostic Systems," in *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6766-6775, Sept. 2020

The final published version is available online at ***https://dx.doi.org/10.1109/TIM.2020.2973913***

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

BrightNet: A Deep CNN for OLED based point of care immunofluorescent diagnostic systems

Andrea Samorè, Manuele Rusci, Damiana Lazzaro, Patrizia Melpignano, Luca Benini and Serena Morigi

***Abstract*—An automatic tool targeting low-cost, low-power point-of-care embedded systems is proposed for fluorescence diagnostic imaging. This allows for a quick and accurate diagnosis even when used by non-expert operators. To achieve this goal, an embedded system has been equipped with an end-to-end deep learning algorithm that does not require manual parameter tuning to perform a diagnosis. The proposed deep convolutional model, named BrightNet, is based on a single-shot detector neural network, modified to estimate the brightness of the detected fluorescent spots in a low-density protein or DNA microarray and finalize the diagnosis. Several optimization steps are presented to compress the inference model size, required for deployment into a portable resource-constrained device. The resulting inference time is about 66 [ms] on an i7 3770K desktop CPU and is estimated to be lower than 5 [s] on an ARM-Cortex M7 considering $1.1 \cdot 10^9$ multiply-accumulate operations. BrightNet has been successfully validated for the detection and discrimination of four different serotypes of the Dengue virus in a set of human samples, as well as for the diagnosis of West Nile virus in horse sera. When evaluated on the considered diagnostic tasks, BrightNet provides better average accuracy than a state of the art variational approach which requires operator intervention, with significant additional advantages of complete automation and quicker diagnosis.**

***Index Terms*—biomedical imaging, CNN, deep learning, brightness estimation, viral diagnosis**

I. Introduction

Point-of-care (POC) diagnostics offers several advantages with respect to tests currently performed by expert operators in hospitals or laboratories using benchtop equipment. POC systems are portable, inexpensive, and easy to use. Moreover, they are part of a clear trend towards automation in molecular diagnostics, that is expected to reduce costs while increasing efficiency and access to early diagnosis in regions where healthcare infrastructure is sparse [1].

In this work, an application-specific and advanced CNN is implemented in an OLED-based diagnostic portable device allowing for an efficient and reliable automatic diagnostic system. Applications in the field of epidemiological research could benefit from the use of the proposed Point-of-care Portable Diagnostic System (POC-PDS).

Andrea Samorè, Damiana Lazzaro and Serena Morigi are with the Department of Mathematics, University of Bologna, Bologna, Italy e-mail: (andrea.samore@unibo.it, damiana.lazzaro@unibo.it, serena.morigi@unibo.it).

Manuele Rusci and Luca Benini are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy e-mail: (manuele.rusci@unibo.it, luca.benini@unibo.it).

Patrizia Melpignano is with OR-EL d.o.o. Organska Elektronika, Kobarid, Slovenia e-mail: (office@or-el-doo.com).

Manuscript received April 19, 2005; revised August 26, 2015.

The proposed POC-PDS is based on immunofluorescence techniques which are currently applied in human medical diagnostics, where the detection of fluorescent spots in experiments of indirect immunofluorescence allows to determine the presence of antibodies against a certain pathogen in the biological fluid under analysis. In particular, biological probes can be labeled with fluorophores, and their weak optical signal can be detected using a charge-coupled device (CCD) camera after a suitable optical excitation. Several works demonstrated the utilization of an organic light-emitting device (OLED) source as an effective technology for the excitation of fluorophores, [2]–[4]. In our case, a specifically optimized and patented OLED has been used for this task [5], [6].

In [7], the authors proposed to replace the highly expensive and bulky CCD camera with a cheap and thin CMOS camera for fluorescence signal detection and quantification in an OLED-based device using low-density protein microarray biochips. This solution enables miniaturization and cost-reduction of the integrated fluorescence reader system paving the way to fast, cheap, and accurate measurement of low-density multiparametric protein or DNA microarray data, ideal to be used in hand-held portable point-of-care diagnostic systems. However, critical issues of the CMOS sensor, which presents a matrix of 24x24 pixels, 100 μm wide each, are its low spatial resolution and its relatively high signal noise. The use of large pixels, to improve the optical sensitivity, does not allow to achieve a high spatial resolution and, for this reason, the CMOS chip can be conveniently used only with low-density bio-probe matrices.

In this paper, the characteristics of a prototype architecture of a POC-PDS are outlined, focusing in particular on the design of the end-to-end diagnosis model. The POC-PDS consists of a heterogeneous system partitioned on two units: the acquisition unit (AU) and the diagnostic unit (DU). In the AU, the fluorophore in a suitable cartridge is excited by the OLED and then detected using a new CMOS sensor coupled to a miniaturized microscope camera lens. The AU is a low power embedded unit, which integrates all image acquisition components, and communicates with the DU, receiving commands or submitting images acquired for diagnostics. The DU processes the fluorescent images with a tailored deep CNN, named BrightNet, that allows for the identification of the kind of pathogen present in the biological sample and the quantification of its concentration, thus enabling automatic diagnosis.

Traditionally, this task is addressed by computing the brightness on each of the segmented and labeled spots through a postprocessing procedure. However, this approach requires

the manual intervention of expert operators to set various regularization and thresholding parameters for each image [7], and for this reason is not suitable for the considered application. Recently, computing pipelines based on deep learning methods have proved to be effective to learn nontrivial tasks and to provide completely automated solutions in a wide range of applications, from 3D object recognition [8] and material defect detection [9] to the solution of inverse problems [10]. Moreover, several deep learning approaches have been proposed to solve the segmentation problem, such as the U-Net [11] and Mask R-CNN [12], while other networks can efficiently solve the detection and localization problems in a single-shot like YOLO [13] and SSD [14]. In [15]–[17], the authors adapted single-shot object detector networks by adding an output to estimate the pose of the detected objects. Differently from previous works, BrightNet pursues three main tasks performed in a single shot: detection of the fluorescent spots in the image, localization of each spot in the cartridge, and finally estimation of the brightness level of each spot to classify it as pathologic or not. To the best of our knowledge, a computationally efficient end-to-end network that performs detection, localization and brightness estimation in a single step is not present in literature.

This work covers the design of the DU in the context of the development of a prototype of POC-PDS, and its validation on two different datasets, related to two different viruses: Dengue and West Nile. For the first time, the use of a CMOS image sensor, in combination with a suitable deep CNN-based analysis, has led to comparable results to a high sensitivity CCD camera and sophisticated operator-dependent variational diagnostic methods. The resultant POC-PDS is a very compact, low-cost, light i.e. portable, multi-parametric and quantitative Point-of-care device at a low price but with sensitivity better than standard ELISA [18]. Considering that the time required for sample preparation and image acquisition is about 30 minutes and the amount of sera required is only 3 μl, the OLED-based POC-PDS can be considered as a simple and reliable hand-held test for early pathology detection. In the specific case of Dengue diagnostics, a very accurate serotype recognition in the convalescent stage of the disease (complementary to PCR) has also been demonstrated.

Summarizing, the three main contributions of this work are the following:

- Specification of the architecture of a completely automatic low-cost point-of-care diagnostic system able to diagnose viral diseases in situations where trained medical personnel is not available;
- Development of an end-to-end neural network capable of detecting and estimating the brightness of spots in fluorescence low-resolution imaging;
- Optimization of the network model for deployment on a resource-constrained embedded system.

II. Point of Care Portable Diagnostic System

In the following, the design specifications of the proposed POC-PDS are provided, which feature both i) the hardware capabilities for data acquisition (AU) and ii) the computational power for the diagnostic task (DU). The input of the POC-PDS adopts an inexpensive disposable cartridge (DC) where a matrix of different antigens has been deposited (in drops of 1 μl each).

Figure 1 details the system architecture of the POC-PDS. The device is composed by the following main subsystems:

- **Acquisition Unit (AU)**. A low-power and low-cost imager provides the imagining capabilities for digitizing the samples under test. Moreover, an *OLED* source is used to illuminate the target during the image acquisition process. The AU is developed by OR-EL using a commercial 12 bit CMOS sensor. It is composed of the following units: a portable *Sample Preparation Unit (SPU)*, a *CMOS sensor* board, CMOS sensor driver and finally optical cube, used to interface the disposable slide to the CMOS sensor. The SPU combines the reaction with the biological fluid under analysis and a second reaction with a fluid containing secondary antibodies tagged with a fluorophore. The AU is connected to the DU where the images are analyzed.
- **Diagnostic Unit (DU)**. Besides the coordination of the whole system, a *Micro Controller Unit* runs locally the diagnostic tasks on the acquired images. Only the diagnosis outcome is transmitted out of the device through an *External Interface*. Moreover, an additional *Memory* component is needed for permanent and temporary storage. Given the wireless nature of the point of care device, energy is drawn from a on-board *Battery Pack*. To power each system component, a *Power Management* circuit handles DC-DC voltage conversions.

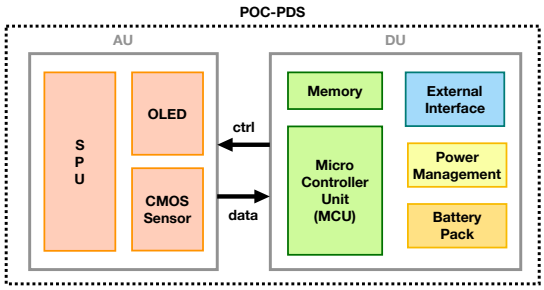


Fig. 1: Block diagram of the POC-PDS architecture.

III. Diagnostic Tasks

Two viral diseases are considered, as case studies, throughout this work: Dengue and West Nile. For both diseases, the diagnostic task is similar and involves both detection and brightness estimation of the fluorescence signal coming from four different spots reacting on a slide. Despite some similarities, there are also key differences in the diagnostic process, that are reflected in the dataset construction.

A. Dengue case study

For the Dengue diagnostic test, each of the four spots contains a different antigen, capable of reacting with one of the four different serotypes of the Dengue virus that can be found in human samples. Biological fluid is added to the slides, reacting with antigens and with a second fluid containing fluorophore-tagged secondary antibodies. The fluorophore is excited using an OLED and the resulting fluorescent signal is recorded by the CMOS camera [19].

Even if a single serotype is present in the blood sample, the diagnostic test is likely to suffer from cross-reactivity, thus more than one spot may emit a fluorescent signal. For this reason, in the Dengue diagnostic test, in addition to accurately estimating the fluorescence of each spot, it is very important to rank the fluorescence of the 4 fluorescent spots to discern actual infections from cross-reactivity.

The Dengue dataset is composed of 43 (24×24 pixels) images, acquired with the CMOS camera of the POC-PDS. To construct the dataset, each image is upscaled with bicubic interpolation to the input dimension for BrightNet: 300×300 pixels (see section IV for more details about the network architecture). Then, to assemble the ground truth labels used for training the DNN model, each ground truth box (GTbox) is manually selected by a trained operator through a graphical user interface by clicking the upper left corner of the ith box (x_{i1}, y_{i1}) and the lower right one (x_{i2}, y_{i2}). The brightness label b_i is then automatically computed by averaging the pixel values inside the box and normalizing it with respect to the dynamic range of the image thus to obtain $b_i \in [0, 1]$. Figure 2 illustrates the dataset labeling phase.

The resulting images are finally divided into three sets: training (25 images), validation (3 images) and testing (15 images). The testing set is the same as in [7] to allow for a direct comparison of results.

B. West Nile case study

For the West Nile diagnostic test, the four spots on the slide are replicates of the same antigen, as a single serotype of the West Nile virus exists. The objective is thus to estimate the brightness of each spot and average them to reduce noise and limit the contribution of possible artifacts in order to obtain a more accurate estimate. If the average fluorescent signal exceeds a threshold, the sample can be considered pathologic.

A set of images have been acquired with two different image sensors. The first detector, considered as the reference one, was the CCD Hamamatsu (ORCA C8484-03G02) while the second detector was a CMOS with 24×24 pixels of $100 \times 100 \quad \mu m^2$ area. Each image contains four fluorescent spots of 1 mm diameter deposited on a transparent substrate and illuminated with an OLED source as described in [7]. The fluorescent spots have been obtained by depositing four equal drops of 1 μl of the same dilution of an antibody tagged with a fluorophore. Then, several samples obtained by depositing different antibody concentrations have been prepared for the image acquisition and analysis. The antibody is a goat anti-horse IgG conjugated with the fluorophore AlexaFluor 430, using the conjugation procedure described in [19]. This kind of antibody has been selected for the detection of the West Nile Virus in the horse sera. In this experiment, the intensity of the fluorescence signal depends on the concentration of the conjugated antibodies in the deposited drops.

The West Nile dataset is composed of 35 (24×24 pixels) CMOS images and 35 (336×256 pixels) CCD images. The bounding box construction procedure is similar to the one previously described for the Dengue dataset and illustrated in Figure 2 but, in this case, the availability of CCD acquisitions is leveraged to generate ground truth intensity labels. The high-resolution CCD image corresponding to each CMOS acquisition is segmented using the variational approach described in [7] and then labeled to discern the four different spots. Ground truth brightness b_i of the ith spot is then computed by averaging the pixel values over the corresponding labeled mask and then by normalizing the result over the dynamic range of the image to obtain $b_i \in [0, 1]$.

C. Data Augmentation

In biomedical imaging, the collection of new data is often limited by cost and time constraints and data augmentation techniques are essential to train effectively the network. In this specific case, five additional images are generated from each image in the training and validation sub-datasets through rotations (90, 180 and 270 degrees), a horizontal and a vertical mirroring operation.

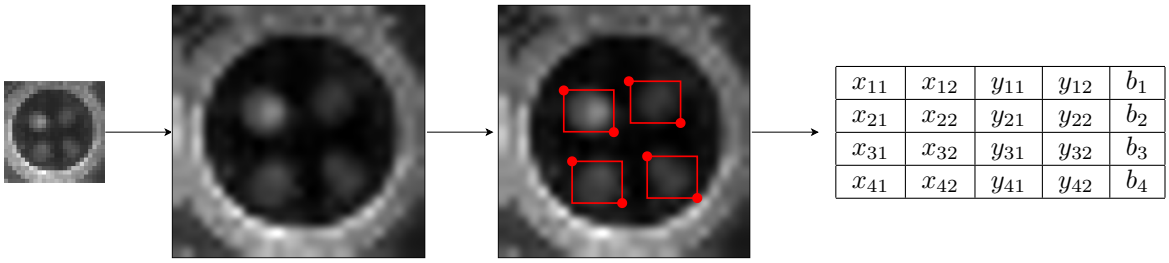
During training, additional stochastic augmentations are performed to improve learning. In particular, uniform brightness offsets, which follow a uniform distribution spanning $[-600, 600]$, are applied to each training image. It is important to note that uniform brightness offsets should not change the brightness value of the estimate as the labels are intensity values normalized over the dynamic range.

IV. BrightNet Architecture

BrightNet is an application-specific network capable of detecting and estimating the brightness of different fluorescent spots deposited on a slide. Since spots can appear everywhere on the slide, object detection is needed and the network is built starting from the Single-Shot multibox Detector (SSD) based on Mobilenet (SSD-MB) [14] [20] and is fine-tuned to optimize performance and reduce computational requirements for deployment in a point of care embedded diagnostic system.

Exploiting a pre-trained Mobilenet network is an essential design choice given the small size of the Dengue and West Nile datasets. Fine-tuning a network pre-trained on a very large and general dataset such as Imagenet [21] proved to increase performance even when the source and target applications (datasets) are significantly different, as is the case for general object classification and medical imaging [22].

A deep fine-tuning strategy is considered [22], where the pre-trained Mobilenet weights are loaded while the additional convolutional layers that perform object detection and brightness estimation are initialized according to the Xavier scheme [23]. All weights are then updated during training on the application-specific Dengue and West Nile datasets.



x_{11}	x_{12}	y_{11}	y_{12}	b_1
x_{21}	x_{22}	y_{21}	y_{22}	b_2
x_{31}	x_{32}	y_{31}	y_{32}	b_3
x_{41}	x_{42}	y_{41}	y_{42}	b_4

Fig. 2: Dataset construction (from left to right): 24×24 original CMOS image, 300×300 bicubic upsampled image, GTbox selection, extracted labels.

With respect to a standard implementation of SSD-MB, BrightNet has a different architecture and a customized loss function that allows for combined single-shot detection and brightness estimation. For a comparison with SSD-MB see section V-C.

As in SSD, a number K of fixed default boxes are generated, each represented by a coordinate vector that uniquely localizes a box inside a feature map. Unlike SSD, only a single class is meaningful - fluorescent spot - thus the class information is not made explicit.

During the training phase illustrated in Figure 3, a batch of images with the associated labels from the training set is fed to the network, that processes them through Mobilenet and a set of convolutional layers that produce the location (L), confidence (C) and brightness (B) feature maps. In particular, B is the tensor of elements $\bar{b}_i \in \mathbb{R}$ representing the brightness associated with the ith predicted box (Pbox), C is the tensor of elements $\bar{c}_i \in \mathbb{R}$ representing the confidence of the ith Pbox, while L is the tensor of elements $\bar{l}_i \in \mathbb{R}^4$ representing the location $\bar{x}_{i1}, \bar{x}_{i2}, \bar{y}_{i1}, \bar{y}_{i2}$ of the ith Pbox. These, together with the labels and the generated default boxes, are then passed to a multibox loss function which is used in the weight optimization. In the Multibox Loss block, a matching phase assigns the coefficients $x_{ij} \in \{0, 1\}$, where $x_{ij} = 1$ only if the ith Pbox is assigned to the jth GTbox by means of the best Jaccard overlap [14].

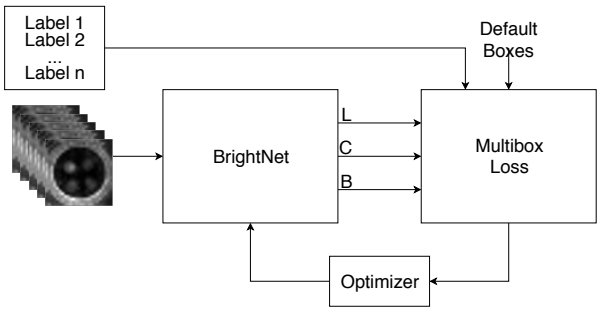


Fig. 3: Training Flow

To prevent overfitting, the loss on the validation dataset is monitored at each epoch during training. We employ the early stopping criterion described in [24], algorithm 7.1, where the maximum number of iterations is set to 120000, and the training is stopped if the validation loss does not improve for more than 5000 consecutive epochs (patience). The trained model for the inference phase is finally selected corresponding to the minimum validation loss.

At inference time the model is frozen (see Figure 4) and a test image is processed yielding the L, C and B feature maps that feed, together with the set of default boxes, a detector layer that performs the following two steps:

- I = threshold(softmax(C))
- I_{nms} = nms(L(I),softmax(C(I)),B(I))

where the first operator thresholds the confidence probabilities yielding a set of indices I, the second operator performs non-maximum suppression (nms) by selecting the Pboxes with maximum overlap greater than a given threshold in the subsets B(I) and L(I), and finally selects the first four with maximum confidence.

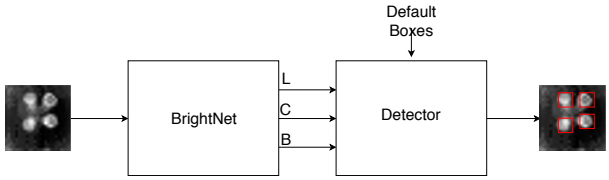


Fig. 4: Testing Flow

The Brightnet network architecture is illustrated in Figure 5.

The input image is initially processed by Mobilenet V1 depthwise separable convolutions. Two feature maps are extracted, the first one after layer 23 and the second one after layer 27. The extracted feature maps are then processed by convolutional layers that generate the predictions feature maps relative to the location of bounding boxes (L), their confidence (C) and their brightness estimate (B).

A. Loss function

Each PBox has a binary state, matched ($x_{ij} = 1$) or not matched ($x_{ij} = 0$) - with respect to the GTboxes - which is used in the loss function computation to finalize the cost to the more likely boxes. Given this information, the loss function generalizes the one proposed in [14] and is composed of a weighted sum of three terms: confidence loss, localization loss, and brightness loss:

$$Loss(x, c, l, \bar{l}, b, \bar{b}) = \alpha \cdot L_{cf}(x, c) + \beta \cdot L_{loc}(x, l, \bar{l}) + \gamma \cdot L_{br}(x, b, \bar{b}), \quad (1)$$

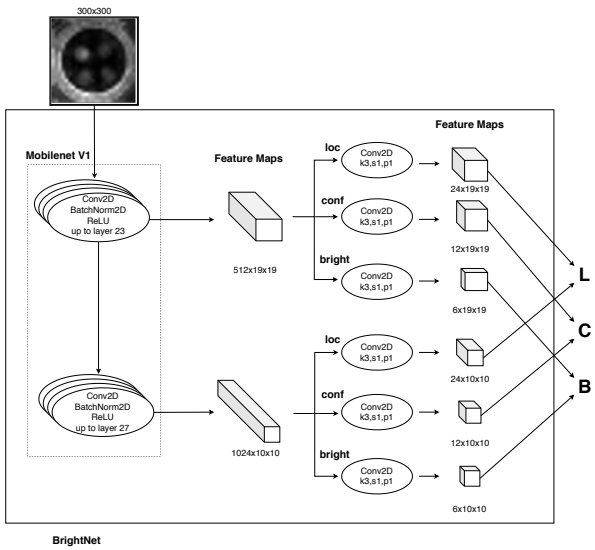


Fig. 5: BrightNet architecture: loc, conf and bright convolutions are performed with a 3x3 convolution kernel (k=3), stride 1 (s=1) and padding 1 (p=1).

where α, β and γ are parameters that balance the contribution of the confidence, location and brightness components of the loss function, respectively. The factors $\alpha = 1$, $\beta = 1$ and $\gamma = 7$ are empirically set.

Confidence loss is computed according to the formula:

$$L_{cf}(x,c) = -\left(\sum_{i=1}^{K} x_{ij} \cdot log(\hat{c}_i) + \sum_{i=1}^{K}(1 - \sum_j x_{ij}) \cdot log(1-\hat{c}_i)\right), \tag{2}$$

where $\hat{c}_i := softmax(c_i)$ and $\sum_j x_{ij} = 1$ if and only if Pbox i has been matched to a GTbox j. L_{cf} maximizes the confidences of the matched Pbox and minimizes the confidences of the remaining predictions.

Location loss measures the discrepancy between the predicted ($\bar{l}$) and the ground truth (l) box locations, and is described by the formula:

$$L_{loc}(x,l,\bar{l}) = \sum_{i=1}^{K} x_{ij} \sum_{k=1}^{4} \phi(l_i^{(k)} - \bar{l}_j^{(k)}), \tag{3}$$

where $\phi(\cdot)$ is the robust, differentiable Huber loss function, which is less sensitive to outliers in data than the squared error loss [25]. Considering the end-to-end nature of BrightNet, the detected bounding boxes do not need to be perfectly localized, as they are only used to distinguish spots in the input image and are not used for subsequent processing of the box contents. In this context, Huber loss facilitates the training by dampening the contribution of outliers.

Finally, brightness loss is determined by:

$$L_{br}(x,b,\bar{b}) = \sum_{i=1}^{K} x_{ij} \cdot (b_i - \bar{b}_i)^2. \tag{4}$$

For the brightness component of the loss function, the mean squared error loss is preferred over the Huber function. This weighs more the outliers and pushes the learning process to avoid them. It was experimentally verified that it allows for more reliable brightness estimation and ranking of the detected spots.

The role of the whole loss function is thus to measure how accurately our model is able to predict with high confidence the brightness of the matched Pboxes.

B. Performance Metrics

Two different metrics, M_1 and M_2, are computed to evaluate the performance of the network for the specific application of spot detection and brightness estimation for the diagnosis of Dengue and West Nile diseases.

Considering Dengue diagnosis, it is important to identify and correctly rank the four spots in order to pinpoint the brightest ones that are over a certain threshold and accurately diagnose the serotype(s), despite possible cross-reactivity of the antibodies present in the serum with the different antigens deposited in the four spots, representing the four Dengue serotypes. To compute M_1, both the intensity estimates generated by the network and the ground truth intensities, provided by a trained operator starting from the upsampled CMOS images, are ranked in decreasing order and then compared to yield a matching number between zero (no matching) and four (perfect matching).

In contrast, a single serotype of West Nile exists and in this case the four spots are replicates used to reduce uncertainty in the measurement. M_1 is thus not informative in this case.

The metric M_2 measures the difference between the intensity estimates by the network and the ground truth intensities. Let $\bar{s} = \frac{1}{4}\sum_{i=1}^{4} s_i$, be the average intensity value of the four spots in the sample image. Defining, for each spot, the variation in percentage from the mean value $\bar{s}$, as

$$p_i = \frac{(s_i - \bar{s}) \cdot 100}{\bar{s}}, \quad i = 1,2,3,4, \tag{5}$$

then the displacement between the estimated (e) and the ground truth (GT) corresponding concentrations in a reference sample, is measured by

$$M_2 = \frac{1}{4}\sum_{i=1}^{4} |p_{i,GT} - p_{i,e}|, \tag{6}$$

where $p_{i,e}$ and $p_{i,GT}$ are computed by (5), with s_i carried out by an automatic estimation method and by a human operator, respectively. In particular, the reference sample is, in case of the Dengue dataset, a CMOS sample, while for the West Nile dataset, a CCD sample.

Smaller values for M_2 are associated with a smaller difference and thus a better similarity between the two measurements. Thus M1 must be maximised, while M2 has to be minimised.

V. Results and Discussion

Numerical experiments to evaluate the object detection and brightness estimation performance of the proposed BrightNet CNN are illustrated in this section. The proposed network has been implemented using the Pytorch framework [26] and run

on an Intel i7-3770K with 24 GB of RAM and two NVIDIA Titan X GPUs.

To highlight the versatility of BrightNet, our evaluation has been conducted on the two datasets Dengue and West Nile, described in section III-A and III-B, respectively. Mobilenet parameters pre-trained on Imagenet are loaded while location, confidence and brightness layers are initialized using the Xavier scheme [23]. Training of the whole network is performed with the stochastic gradient-based optimizer ADAM, with the following default parameters: exponential decay rates $\beta_1 = 0.9$, $\beta_2 = 0.99$, gradient regularizer $\epsilon = 10^{-8}$ [27]. The learning rate has been suitably optimized (see Table II).

In this section, the results obtained by the BrightNet CNN are compared with those generated by the variational approach proposed in [7] on the Dengue dataset based on the suitable metrics introduced in section IV-B. Then, the versatility of the proposed network is illustrated by evaluating its performance on the detection of a different viral disease: West Nile. Finally, several steps of optimization needed for deployment on a resource-constrained embedded system are evaluated.

A. Evaluation on Dengue dataset

In this first experiment, the performance of the proposed deep learning diagnostic approach is compared with the variational solution presented in [7] on the Dengue dataset.

In Figure 6 the results are illustrated on a representative subset of test images. In particular, in the first row the original low-resolution images acquired by the CMOS camera are shown; in the second row the upscaled images input to the network are illustrated; while in the third and fourth rows the network outputs are reported in terms of bounding boxes superimposed on the input images together with the estimated brightness in the upper left corner (third row), and corresponding synthesized diagnostic results highlighting the estimated ranking of the four spots (fourth row). The chromatic scale represents with false colors the concentrations detected on each spot, for an easy and intuitive visual inspection of the results.

For what concerns the accuracy, the diagnosis results obtained by the variational method in [7], composed of a multistep computational pipeline involving variational super-resolution, manual cropping, image restoration, segmentation and quadrature, are compared with those of the proposed BrightNet approach, on the testing dataset.

Metrics M1 and M2 described in IV-B reveal a good performance of the proposed CNN (see Table I).

Method	M1	M2
BrightNet	78.3%	5.688
Variational	81.7%	7.941

TABLE I: Performance comparison with the variational approach [7]

These results highlight a significantly more accurate average brightness estimation (lower M2) provided by BrightNet when compared to the variational approach. In contrast, the lower M1 indicates a slightly worse ranking performance. In this regard, it was experimentally verified that the network presents a significant probability of misranking the spots only if the ground truth intensity difference between spots is smaller than 4% of the dynamic range of the image. This small difference in intensity is not significant from a diagnostic point of view, because it can be ascribed to some statistical variation rather normal in diagnostic assays.

Pursuing the goal of designing an automatic diagnostic system, a key aspect analyzed in this comparison procedure is the number of model parameters which involve a tuning, and the degree of human interaction which makes the successful application of the system inevitably operator-dependent. The performance of the variational method in [7] strongly relies on the selection of the space-variant regularization parameters involved in the model, while the end-to-end neural network is intrinsically tuned. Moreover, in [7] the pipeline procedure proposed requires human interaction to correctly localize the bounding boxes of the salient regions of interest in the fluorescence image to proceed with segmentation, and this makes the diagnostic result time consuming and significantly dependent on the skill level of the operator.

Finally, the diagnostic time, without considering sample preparation, has been significantly reduced from about 30 seconds (variational approach [7]) to less than a second (BrightNet).

B. Evaluation on West Nile dataset

A significant advantage of the presented network is its flexibility: it can be easily adapted to several serological diagnostic tests for the detection and discrimination of different viruses in a set of fluorescent images given a suitable dataset and a new training procedure.

To this purpose, the performance of BrightNet is analyzed on the West Nile dataset, demonstrating that the use of the proposed CNN-based low-cost point of care diagnostic system equipped with a CMOS camera is a reliable alternative to expensive state-of-the-art CCD-based systems.

In Figure 7 the results on a representative subset of the test images are illustrated. As in Figure 6, in the first row the original low-resolution images acquired by the CMOS camera are shown; in the second row the upscaled images input to the network are illustrated; while in the third and fourth rows the network outputs in terms of bounding boxes superimposed on the input image together with the estimated brightness in the upper left corner (third row), and corresponding synthesized diagnostic result highlighting the average spot (fourth row).

With 35 images available, 6-fold cross-validation is employed to be able to test all images. The average M2 metric for the whole dataset is 8.371. This result highlights the capability of the network to accurately map the brightness estimated from low-resolution CMOS images (inputs) to that of CCD images (ground truth). In comparison, the reference variational method was applied to the same task, yielding an M2 = 11.976, which is significantly worse than the one produced by BrightNet.

The West Nile dataset is composed of several different samples where spots have been obtained by depositing different dilutions of antibody tagged with a fluorophore. Moreover, different samples are acquired at different exposure times. In

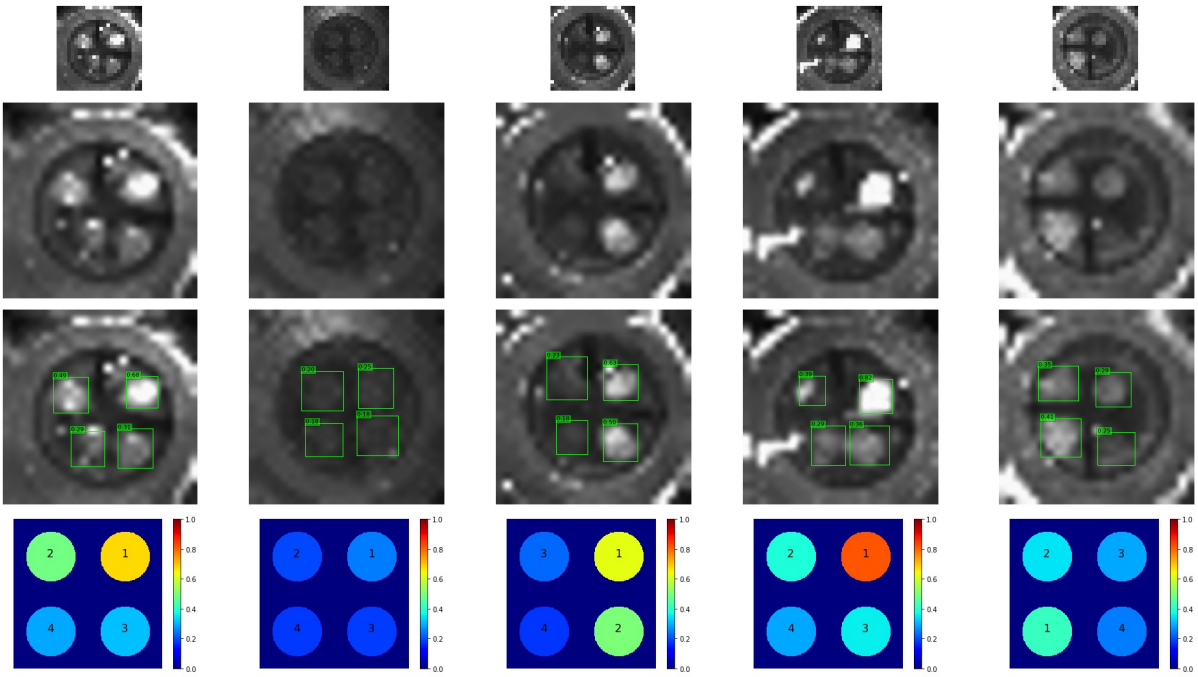


Fig. 6: Results over the Dengue dataset for five different testing images. First row: original low-resolution images. Second row: upscaled images input to the network. Third row: spot localization. Fourth row: estimated ranking of the spots.

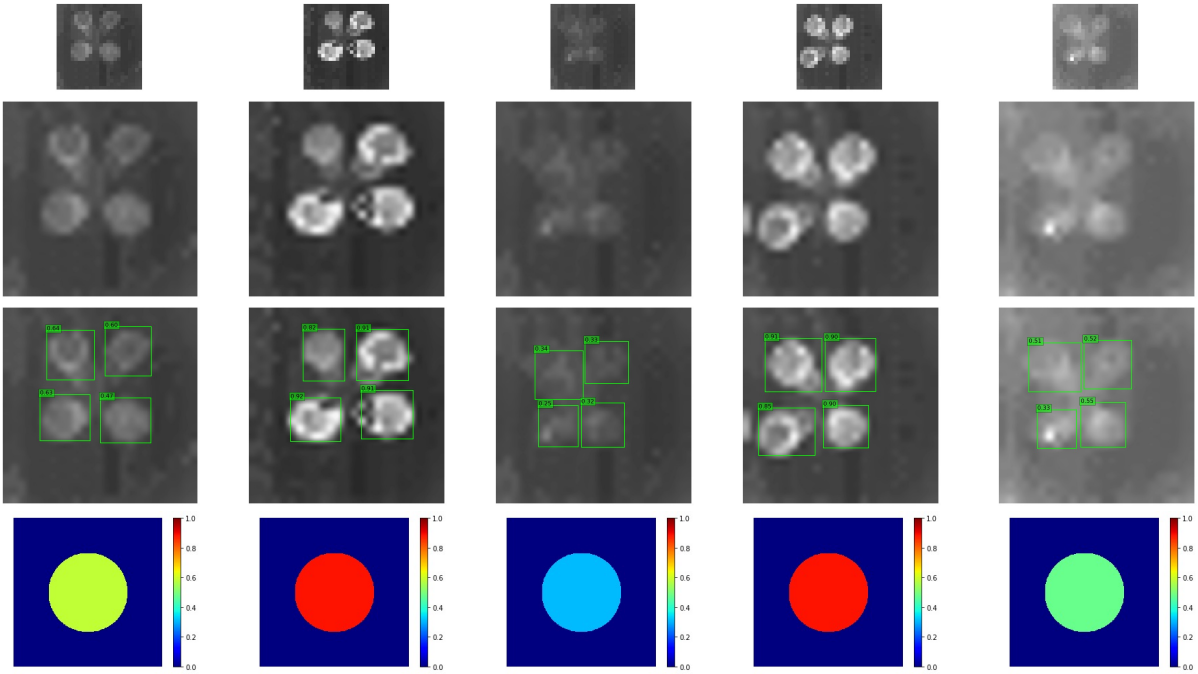


Fig. 7: Results over the West Nile dataset for five different testing images. First row: original low-resolution images. Second row: upscaled images input to the network. Third row: spot localization. Fourth row: estimated brightness of the average spot.

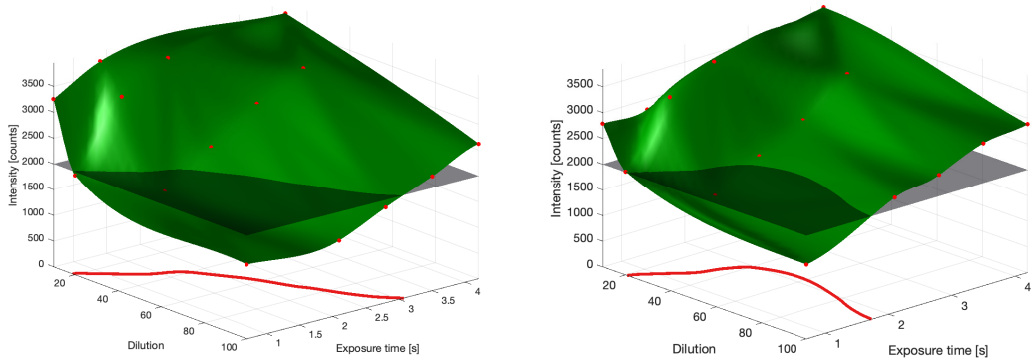


Fig. 8: Ground truth intensity surface (left) and estimated intensity surface (right) for various dilutions and exposure times.

order to adequately configure the sample acquisition procedure, it is possible to estimate a "safe region" defined by values of exposure times and dilutions which, according to the sensitivity of our system, allow to correctly detect the positivity to the West Nile virus. In fact, when the fluorescent signal exceeds a given threshold (2000 in Figure 8), the sample can be considered pathologic. Let this threshold be represented by the plane in Figure 8, and the intensities estimated by the proposed CNN be represented by the interpolated surface $z = f(x, y)$ among the points of coordinates (dilution, exposure time) in the plane x,y. The contour line on the plane x,y represents the boundary of the safe region. In Figure 8 (left) the surface is obtained by interpolating the intensities directly on the CCD images, while in Figure 8 (right) the intensities are estimated by BrightNet using CMOS input images. The safe region obtained with the CCD images is slightly smaller with respect to the one generated with BrightNet. The plots identify similar safe regions and highlight that our proposal allows for correct detection even using considerably lower exposure times.

C. CNN Optimization

The performance of the proposed network is analyzed as a function of the various steps performed to reduce the network size and computational requirements with the aim to deploy it on a low power embedded system. Dengue is considered as a case study as, for this dataset, both the M1 and M2 metrics are meaningful. The performance of the network in the various stages of optimization is summarized in Table II, both in terms of accuracy (M1, M2) and computational cost (Multiply-Accumulate operations - MACs, number of convolutional weights and CPU inference time on an i7 3770K processor). Only CPU times are listed, as GPU processing overheads obfuscate the difference in inference time for Mobilenet-based networks. These results have been obtained with the optimal learning rates reported in the second column of Table II.

Starting from a standard SSD implementation based on VGG, the network has been expanded to process also a brightness estimate of the detected objects. The resulting network (SSD-VGG-Bright) is fully functional and able to achieve satisfactory performance on most of the images in the tested datasets, but with a significant drawback: the high computational cost of VGG16. For this reason, a standard implementation of SSD based on Mobilenet V1 (SSD-MB) has been expanded to include a brightness branch, thus obtaining the SSD-MB-Bright model illustrated in Figure 9.

According to the M1 and M2 metrics computed on the Dengue dataset, a significant increase in performance with respect to SSD-VGG-Bright is obtained. It is possible to attribute this increase to the smaller size of Mobilenet V1 when compared to VGG16 and thus to less overfitting during the training procedure.

In addition, as the fluorescent spots size is fairly large and constant, multiscale detection is not critical for the considered application, so it is possible to trim all the extra layers that scale the feature maps (in blue in Figure 9), and the corresponding location, confidence and brightness layers that process their output, from SSD-MB-Bright. The shrunk network without the layers in the dashed box in Figure 9 is the proposed BrightNet, whose architecture is illustrated in Figure 5. Somewhat surprisingly, in this last step of optimization, the performance does not decrease but instead slightly improves. The reason is that, by removing feature map scaling, relevant prior information is included by suggesting to the network the correct size of spots that need to be detected (quite large in this case). This reduced network is significantly lighter than a full SSD-MB-Bright implementation, reducing the number of weights by 44% and the floating-point operations by 11%.

To further reduce the memory footprint and computational cost of the proposed network, experiments were performed by either reducing the width multiplier (BrightNet WM 0.25) or the resolution multiplier (BrightNet 120x120), as defined in the mobilenet paper [20]. Both multipliers trim the size of the feature maps and thus would reduce both the memory footprint and the computational cost of the network. Unfortunately, a significant decrease in performance with either optimization has been observed.

Recently, Mobilenet V2 has been released, featuring improvements over V1 in both computational cost and accuracy [28]. In this work, Mobilenet V1 was targeted as an example representative of a class of neural networks optimized for mobile applications. The modularity of BrightNet allows for

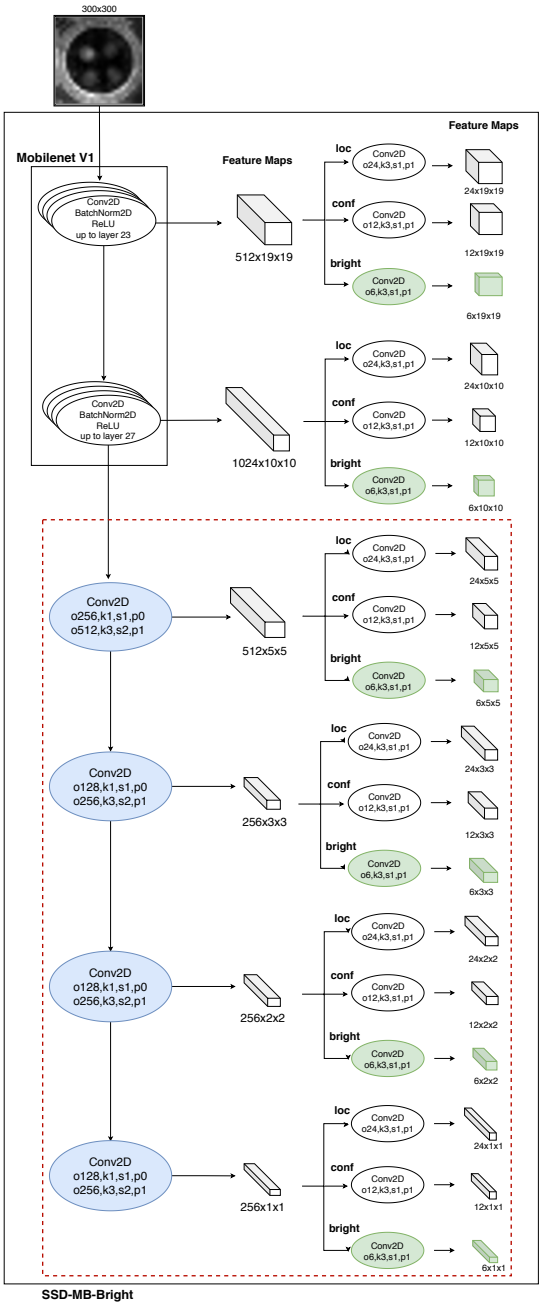


Fig. 9: SSD-MB-Bright network architecture with the brightness branches higlighted in green and the extra layers required for multiscale detection higlighted in blue. 'o' indicates the number of output channels for each convolution, 'k' is the kernel size, 's' is the stride and 'p' is the padding.

the substitution of the base feature extraction network, and the integration of new developments is likely to further improve performance and reduce computational cost. Future plans involve the evaluation of the performance of BrightNet based on Mobilenet V2.

BrightNet features $3.7 \cdot 10^6$ parameters and a computational cost of $1.1 \cdot 10^9$ MACs. Several embedded processing platforms for mobile applications fit these computational requirements and only necessitate of a few hundreds of milliseconds to run similar inference tasks [29]. However, given the requirements of our diagnostic application, even a microcontroller-class processor could be targeted for deployment, as an inference time of a few seconds is acceptable. Considering a high-end microcontroller platform, featuring a high-performance core ARM-Cortex M7 with a running frequency of 400MHz and an external RAM memory of a few megabytes, the estimated inference time is lower than 5 seconds with a computation efficiency of 0.6 MAC/cycle [30]. Nevertheless, the authors plan to investigate further network shrinking and model exploration in future works, to allow for fast inference on lower-end low-power devices.

VI. Conclusions

We presented an operator-independent diagnostic tool for the completely automated diagnosis of viral diseases with fluorescence imaging techniques on a POC-PDS. The ad-hoc, end-to-end deep learning model proved to be capable of quickly estimating the brightness of fluorescent spots on a microarray in a single step, with diagnostic performance comparable to state of the art variational methods that require parameter tuning by expert operators on an image by image basis. As such our automated approach is ideal for on-site diagnoses where the access to hospitals and laboratories is limited. BrightNet proved to be significantly faster than previous methods and versatile enough to diagnose different viral diseases, Dengue and West Nile, requiring only a fine-tuning with a short retraining procedure. This makes BrightNet potentially applicable to other problems that require brightness estimation in fluorescence imaging, such as the detection of contaminants in the food industry.

References

[1] V. Gubala, L. F. Harris, A. J. Ricco, M. X. Tan, and D. E. Williams, "Point of care diagnostics: Status and future," *Analytical Chemistry*, vol. 84, no. 2, pp. 487–515, 2012.

[2] A. Pais, A. Banerjee, D. Klotzkin, and I. Papautsky, "High-sensitivity, disposable lab-on-a-chip with thin-film organic electronics for fluorescence detection," *Lab on a Chip*, vol. 8, no. 5, pp. 794–800, 2008.

[3] B. Yao, G. Luo, L. Wang, Y. Gao, G. Lei, K. Ren, L. Chen, Y. Wang, Y. Hu, and Y. Qiu, "A microfluidic device using a green organic light emitting diode as an integrated excitation source," *Lab on a Chip*, vol. 5, no. 10, pp. 1041–1047, 2005.

[4] O. Hofmann, X. Wang, J. C. DeMello, D. D. Bradley, and A. J. DeMello, "Towards microalbuminuria determination on a disposable diagnostic microchip with integrated fluorescence detection based on thin-film organic light emitting diodes," *Lab on a Chip*, vol. 5, no. 8, pp. 863–868, 2005.

[5] A. Marcello, D. Sblattero, C. Cioarec, P. Maiuri, and P. Melpignano, "A deep-blue OLED-based biochip for protein microarray fluorescence detection," *Biosensors and Bioelectronics*, vol. 46, pp. 44–47, 2013.

Network	Learning Rate	M1	M2	MACs	# Weights	CPU time [s]
SSD-VGG-Bright	10^{-4}	0.65	9.78	$29.982 \cdot 10^9$	$23.877 \cdot 10^6$	0.504
SSD-MB-Bright	$5 \cdot 10^{-3}$	0.78	6.84	$1.229 \cdot 10^9$	$6.729 \cdot 10^6$	0.091
BrightNet	10^{-2}	**0.78**	**5.69**	$1.100 \cdot 10^9$	$3.787 \cdot 10^6$	0.066
BrightNet WM 0.25	10^{-3}	0.58	8.44	$121 \cdot 10^6$	$0.794 \cdot 10^6$	0.035
BrightNet 120x120	$5 \cdot 10^{-3}$	0.52	11.62	$211 \cdot 10^6$	$3.787 \cdot 10^6$	0.051

TABLE II: Performance at various stages of optimization, optimal learning rates have been experimentally determined.

[6] M. Manzano, F. Cecchini, M. Fontanot, L. Iacumin, G. Comi, and P. Melpignano, "OLED-based DNA biochip for Campylobacter spp. detection in poultry meat samples," *Biosensors and Bioelectronics*, vol. 66, pp. 271–276, 2015.

[7] D. Lazzaro, S. Morigi, P. Melpignano, E. Loli Piccolomini, and L. Benini, "Image enhancement variational methods for enabling strong cost reduction in OLED-based point-of-care immunofluorescent diagnostic systems," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 34.3, p. e2932, 2018.

[8] C. Ma, Y. Guo, Y. Lei, and W. An, "Binary Volumetric Convolutional Neural Networks for 3-D Object Recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 38–48, 2019.

[9] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep Architecture for High-Speed Railway Insulator Surface Defect Detection: Denoising Autoencoder With Multitask Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2679–2690, 2018.

[10] J. Lei, Q. Liu, and X. Wang, "Deep Learning-Based Inversion Method for Imaging Problems in Electrical Capacitance Tomography," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 9, pp. 2107–2118, 2018.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, pp. 234–241, 2015.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 779–788, 2016.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *Computer Vision – ECCV 2016*, pp. 21–37, 2016.

[15] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "SSD-6D : Making RGB-Based 3D Detection and 6D Pose Estimation Great Again," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1521–1529, 2017.

[16] P. Poirson, P. Ammirato, C.-y. Fu, W. Liu, J. Ko, A. C. Berg, U. N. C. Chapel, and H. George, "Fast Single Shot Detection and Pose Estimation," *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 676–684, 2016.

[17] B. Tekin and P. Fua, "Real-Time Seamless Single Shot 6D Object Pose Prediction," *arXiv*, 2018.

[18] E. Engvall and P. Perlmann, "Enzyme-linked immunosorbent assay, Elisa. 3. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes." *Journal of immunology (Baltimore, Md. : 1950)*, vol. 109, no. 1, pp. 129–35, 1972. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/4113792

[19] P. Melpignano, S. Morigi, E. Daniso, E. L. Piccolomini, and L. Benini, "Point-of Care OLED-based Multiparametric Biochip Exploiting Advanced Image Processing for the Dengue Serotype recognition," *CMBBE2018 Conference Proceedings, P.R. Fernanes and J.M. Tavares (Editors)*, pp. 1–14, 2018.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv*, 2017.

[21] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[22] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the 13th International Conference On Artificial Intelligence and Statistics*, vol. 9, pp. 249–256, 2010.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[25] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[26] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," *NIPS-W*, 2017.

[27] D. P. Kingma and J. L. Ba, "Adam: a Method for Stochastic Optimization," *ICLR 2015*, pp. 1–15, 2015.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[29] A. Ignatov, R. Timofte, W. Chou, K. Wang, M. Wu, T. Hartley, and L. Van Gool, "AI Benchmark: Running deep neural networks on android smartphones," *Computer Vision – ECCV 2018 Workshops*, 2018.

[30] STMicroelectronics, "STM32H7 Series - STMicroelectronics." [Online]. Available: https://www.st.com/en/microcontrollers-microprocessors/stm32h7-series.html