

# NCIS: Deep Color Gradient Maps Regression and Three-Class Pixel Classification for Enhanced Neuronal Cell Instance Segmentation in Nissl-Stained Histological Images

Valentina Vadori<sup>1</sup>, Antonella Peruffo<sup>2</sup>, Jean-Marie Graïc<sup>2</sup>, Livio Finos<sup>3</sup>, Livio Corain<sup>4</sup>, and Enrico Grisan<sup>1</sup>

<sup>1</sup> Dept. of Computer Science & Informatics, London South Bank University, UK  
{vvadori,egrisan}@lsbu.ac.uk

<sup>2</sup> Dept. of Comparative Biomedicine & Food Science, University of Padova, IT  
{antonella.peruffo,jeanmarie.graic}@unipd.it

<sup>3</sup> Dept. of Developmental Psychology and Socialisation, University of Padova, IT  
livio.finos@unipd.it

<sup>4</sup> Dept. of Management and Engineering, University of Padova, IT  
livio.corain@unipd.it

**Abstract.** Deep learning has proven to be more effective than other methods in medical image analysis, including the seemingly simple but challenging task of segmenting individual cells, an essential step for many biological studies. Comparative neuroanatomy studies are an example where the instance segmentation of neuronal cells is crucial for cytoarchitecture characterization. This paper presents an end-to-end framework to automatically segment single neuronal cells in Nissl-stained histological images of the brain, thus aiming to enable solid morphological and structural analyses for the investigation of changes in the brain cytoarchitecture. A U-Net-like architecture with an EfficientNet as the encoder and two decoding branches is exploited to regress four color gradient maps and classify pixels into contours between touching cells, cell bodies, or background. The decoding branches are connected through attention gates to share relevant features, and their outputs are combined to return the instance segmentation of the cells. The method was tested on images of the cerebral cortex and cerebellum, outperforming other recent deep-learning-based approaches for the instance segmentation of cells.

**Keywords:** Cell Segmentation · Histological Images · Neuroanatomy · Brain · Nissl Staining · Deep-Learning · U-Net · EfficientNet · Attention.

## 1 Introduction

Advancements in microscopy have made it possible to capture *Whole Slide Images* (WSIs) and obtain cellular-level details, revealing the intricate nature of the brain cytoarchitecture. This progress has opened up new avenues for conducting quantitative analysis of cell populations, their distribution, and morphology, which can help us answer a range of biological questions. Comparative

neuroanatomy studies examine differences in brain anatomy between groups distinguished by factors like sex, age, pathology, or species, investigating the connections between the brain’s structure and function [1,8,4]. A standard analysis pipeline involves the use of Nissl stain to label neuronal cells in tissue sections (histological slices) of brain specimens [6]. These sections are then fixed and digitized as WSIs for examination. WSIs are often characterized by their sheer size and complexity, making computerized methods necessary for their efficient and reproducible processing. Automatic cell instance segmentation plays a crucial role, as it allows to extract features at the single cell level.

In the field of digital pathology, numerous methods have been proposed to segment cells and nuclei and aid the detection and diagnosis of diseases. These methods mainly rely on a set of algorithms, including intensity thresholding, morphology operations, watershed transform, deformable models, clustering, graph-based approaches [16]. However, cell instance segmentation is very challenging due to the varying size, density, intensity and texture of cells in different anatomical regions, with additional artifacts that can easily influence the results. Recently, deep learning has shown remarkable progress in medical image analysis, and neural networks have been successfully applied for cell segmentation, achieving higher quality than traditional algorithms. In the last years, a considerable number of approaches have adopted a semantic segmentation formulation that employs a U-net-like convolutional neural network architecture [12]. These methods incorporate customized post-processing techniques, such as marker-controlled watershed algorithms and morphological operations, to separate cells instances. Some integrate the formulation with a regression task. Huaqian et. al [15] propose a framework with an EfficientNet as the U-Net encoder for ternary classification (contours between touching cells, cell bodies, and *background*, BG). Ultimate erosion and dynamic dilation reconstruction are used to determine the markers for watershed. StarDist [13] regresses a star-convex polygon for every pixel. CIA-Net [18] exploits two decoders, where each decoder segments either the nuclei or the contours. Hover-Net [7] uses a Preact-ResNet50 based encoder and three decoders for *foreground* (FG)/BG segmentation, cell type segmentation, and regression of horizontal and vertical distances of pixels from the cell centroid. Mesmer [9] considers classification into whole contours, cell interiors, or BG, and regression of the distance from the cell centroid.

Since in histological images the boundaries between cells that are in contact are often incomplete or ambiguous and they can appear between cells with differing characteristics and orientations, we recognized the pivotal role of correctly predicting these boundaries for accurate cell separation. Therefore, we have developed an approach that focuses on enhancing the prediction of contours. Specifically, we propose NCIS as an end-to-end framework to automatically segment individual neuronal cells in Nissl-stained histological images of the brain. NCIS employs an U-Net-like architecture, which synergistically combines solutions from [15,7,18] to classify pixels as contours between touching cells, cell body, or BG, and to regress four color gradient maps that represent distances of pixels from the cell centroid and that are post-processed to get a binary mask of

contours. Since cells are often slanted and there are configurations where smaller cells burrow into the concavities of larger ones, we hypothesized that the prediction of diagonal gradients together with horizontal and vertical ones could help to strengthen the approach. NCIS was created to examine the cytoarchitecture of brain regions in diverse animals, including cetaceans, primates, and ungulates. The primary objective is to conduct comparative neuroanatomy studies, with a particular emphasis on diseases that impair brain structure and functionality, such as neurodegeneration and neuroinflammation. We tested NCIS on images of the auditory cortex and cerebellum, outperforming other recent deep-learning-based approaches for the instance segmentation of cells.

## 2 Dataset

The data of this study are a set of 53 2048x2048 histological images extracted from Nissl-stained 40x magnification WSIs of the auditory cortex of *Tursiops truncatus*, also known as the bottlenose dolphin. Brain tissues were sampled from 20 specimens of different subjects (new-born, adult, old) stored in the Mediterranean Marine Mammals Tissue Bank (<http://www.marinemammals.eu>) at the University of Padova, a CITES recognized (IT020) research center and tissue bank. These specimens originated from stranded cetaceans with a decomposition and conservation code (DCC) of 1 and 2, which align with the guidelines for post-mortem investigation of cetaceans [10]. The images were divided into 3 subsets: 42 images for training, 5 for validation and 6 for testing.

To assess the generalizability of the proposed method on cerebral areas not seen during training, we considered an independent 2048x2048 Nissl-stained histological image of the cerebellum of the bovine [4]. In this case, the animal was treated according to the present European Community Council directive concerning animal welfare during the commercial slaughtering process and was constantly monitored under mandatory official veterinary medical care. Compared to the training set, the additional image is characterized by the presence of a higher density granular layer of touching or overlapping cells with small dimensions and predominantly circular shape and a thin Purkinje cell layer with relatively larger and sparse pear-shaped cells.

All images were annotated using QuPath [2] software, resulting in 24,044 annotated cells of the auditory cortex, and 3,706 of the cerebellum.

## 3 Method

The proposed NCIS framework for automatic instance segmentation of neuronal cells can be observed in Fig. 1. An image of arbitrary size is divided into patches of size 256x256 via a tiling step with 50% overlap between patches. Patches are individually fed to the deep learning model, NCIS-Net, whose architecture is shown in Fig. 2. The model outputs for each patch are combined via a untiling step to get 7 outputs with size equal to that of the original image. Finally, a series of post-processing steps is applied to generate the instance segmentation.

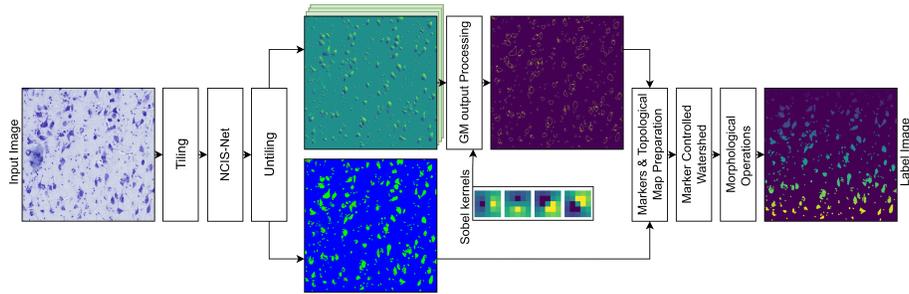


Fig. 1: An overview of the overall NCIS approach for neuronal cell instance segmentation in Nissl-stained whole slice images.

### 3.1 NCIS-Net

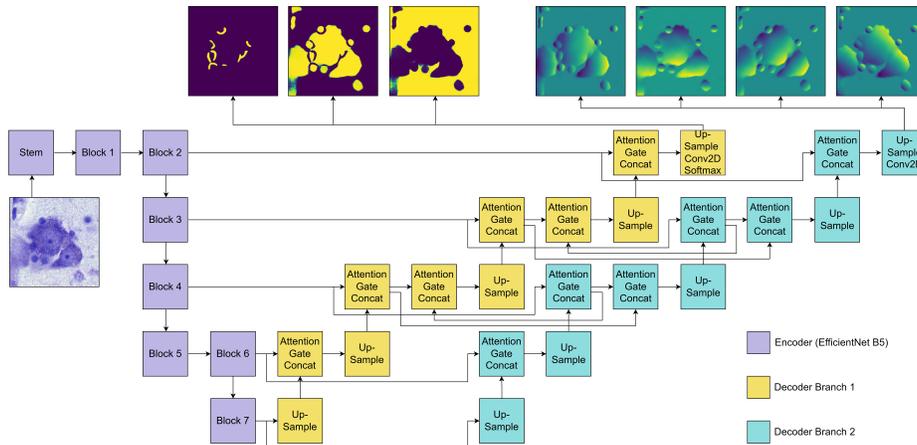


Fig. 2: An overview of the architecture of the proposed NCIS-Net.

**Architecture** The proposed NCIS-Net has a U-Net-like structure, as shown in Fig. 2. The down-sampling, feature extracting branch of the encoder is based upon a state-of-the-art network, *EfficientNet-B5*, whose building blocks are arrays of mobile inverted bottleneck *MBCConv* optimized with neural architecture search [14]. NCIS-Net is characterized by two decoding branches for a classification and regression task. The first branch performs three-class pixel classification. Its output is a three-channel image with probabilities for boundaries between touching cells, cell bodies, and BG, respectively. The union of the first two classes constitutes the FG. The second branch regresses four color gradient

maps. Its output is a four-channel image, where each channel represents a color gradient map with vertical, horizontal, and diagonal directions. As illustrated in Fig. 2, in a gradient map, each individual cell is represented by a color gradient where pixel values increment from the minimum of  $-1$  (blue in the figure) to the maximum of  $1$  (yellow in the figure) according to the prescribed direction (vertical from left to right, horizontal from top to bottom, diagonal from bottom left to top right and diagonal from top left to bottom right).  $0$  corresponds to the cell centroid. Four skip connections leveraging attention gates [11] connect the encoder branch to each decoder branch, promoting a focused exploitation of the features encoded at different resolutions. Skip connections with attention gates are also introduced between the two decoder branches to favor feature sharing.

**Loss function** NCIS-Net is trained through the backpropagation algorithm applied to a loss function  $\mathcal{L} = \mathcal{L}_{PC} + \mathcal{L}_{GR}$  that jointly optimizes the encoder, pixel classification ( $PC$ ) and gradient regression ( $GR$ ) branches, with

$$\mathcal{L}_{PC} = \lambda_1 \mathcal{C} + \lambda_2 \mathcal{D}_1 + \lambda_3 \mathcal{D}_2, \quad \mathcal{L}_{GR} = \lambda_4 \mathcal{M}_1 + \lambda_5 \mathcal{M}_2 \quad (1)$$

where  $\lambda_1 = \lambda_3 = \lambda_4 = \lambda_5 = 2, \lambda_2 = 1$  are weighting factors set via hyperparameter validation.  $\mathcal{C}$  is the categorical cross-entropy between the ground truth (GT) and  $\hat{Y}_{PC}$ , the output of the  $PC$  branch.  $\mathcal{D}_k$  is the Dice loss for class  $k$ :

$$\mathcal{D}_k = 1 - \frac{2 \sum_{i=1}^N (Y_{PC,i,k} \hat{Y}_{PC,i,k}) + \epsilon}{\sum_{i=1}^N Y_{PC,i,k} + \sum_{i=1}^N \hat{Y}_{PC,i,k} + \epsilon} \quad (2)$$

where  $N$  is the total number of pixels in the input image,  $\epsilon$  is a smoothness constant. Specifically,  $\mathcal{D}_1$  and  $\mathcal{D}_2$  in Eq. 1 are the Dice losses for contours and cell bodies, respectively.  $\mathcal{M}$  indicates a mean squared error loss.  $\mathcal{M}_1$  is the mean squared error loss between the GT ( $Y_{GR}$ ) and the predicted output  $\hat{Y}_{GR}$  of the  $GR$  branch, while  $\mathcal{M}_2$  is defined as:

$$\mathcal{M}_2 = \frac{1}{N \cdot J} \sum_{i=1}^N \sum_{j=1}^J (\nabla_j Y_{GR,i,j} - \nabla_j \hat{Y}_{GR,i,j})^2 \quad (3)$$

where  $J$  is the number of gradient maps (4 in our case) and  $\nabla$  is the derivative operator. Note that for each gradient map, finite derivatives are taken by convolving the map with  $5 \times 5$  Sobel kernels with orientation and direction matching that of the corresponding gradient map, as illustrated in Fig. 1.

**Training** Data batches are created from the training set of 2048x2048 images. Images are picked randomly with replacement and a series of random augmentations including rotations, flipping, deformations, intensity variations, and blurring, are applied to each selected image and corresponding three-class semantic and cell instance segmentation GT masks, when necessary. GT gradient maps are computed at this stage based on the transformed instance segmentation

GT masks. Random cropping is then applied to the transformed images to get 256x256 images to be fed to the network. By utilizing this approach, it is highly likely that the batches used during each epoch will not only have varying transformations but also distinct cropped original images, ultimately aiding in the prevention of overfitting. All models are trained with the TensorFlow 2 framework in two phases. In the first phase the encoder, pre-trained on the *Image-Net* dataset [5], is frozen, while in the second fine tuning phase the encoder is unfrozen, with the exception of batch normalization layers. Each phase continues for a maximum of 50 epochs with early stopping (patience of 8 epochs, 400 batches per epoch, batch size of 16). The validation set for early stopping is created from the validation set of 2048x2048 images by cropping them into 256x256 with no overlapping and no augmentations. We use the AMSGrad optimizer with clipnorm of 0.001 and a learning rate of  $10^{-3}$  in the first phase and  $10^{-5}$  in the second phase. The deep learning model architecture, training and inference code will be made available upon acceptance on GitHub.

### 3.2 Untiling and Post-Processing

Patches from the same image are blended together via interpolation with a second order spline window function that weights pixels when merging patches [3].

Within each of the color gradient map, pixels between different cells should have a meaningful difference. Therefore, Sobel kernels are utilized to get the following contour map:

$$C_{PC,i} = \max(\nabla_1 \hat{Y}_{GR,i,1}, \nabla_2 \hat{Y}_{GR,i,2}, \nabla_3 \hat{Y}_{GR,i,3}, \nabla_4 \hat{Y}_{GR,i,4}) \quad (4)$$

where  $\hat{Y}_{GR,i,j}, j = 1, \dots, 4$ , is normalized between 0 and 1.  $C_{PC,i}$  is thresholded on a per-image basis via the triangle method [17], so that ones in the thresholded binary version  $C_{PC,i,th}$  correspond to contours. A binary mask is defined with pixels set to 1 if they are most likely to belong to cell bodies based on the outputs of both decoders ( $\hat{Y}_{PC,i,2} > \hat{Y}_{PC,i,1}$  and  $\hat{Y}_{PC,i,2} > \hat{Y}_{PC,i,3}$  and  $C_{PC,i,th} = 0$ ). Connected components smaller than 80 pixels are removed. The mask is eroded with a disk-shaped structuring element of radius 4 to force the separation of touching cells that are only partially separated. The resulting connected components bigger than 3 pixels are used as markers for the marker-controlled watershed algorithm applied to the topological map given by the complement of  $\hat{Y}_{PC,i,2}$ . A foreground binary mask is obtained as  $\hat{Y}_{PC,i,1} + \hat{Y}_{PC,i,2} > \hat{Y}_{PC,i,3}$ . After holes filling and morphological opening to smoothen the cell boundaries, it is used to constrain the watershed labelling. Segmentations are refined via morphological closing and opening, ensuring that boundaries between cells are maintained.

## 4 Results

### 4.1 Evaluation Metrics

To evaluate the semantic segmentation performance on test images, we utilize the Dice Coefficient (Dice). Instance segmentation performance is evaluated according to the average precision with threshold 0.5 (AP@0.5). Predicted instances

Table 1: Segmentation performance of the proposed NCIS method and compared approaches. #P indicates the number of network trainable parameters.

Methods	#P(M)	Auditory Cortex		Cerebellum	
		Dice	AP@0.5	Dice	AP@0.5
Huaqian et. al [15]	40.15	0.915	0.782	0.708	0.329
Hover-Net [7]	44.98	0.912	0.780	0.792	0.230
Mesmer [9]	25.50	0.862	0.657	<b>0.801</b>	0.156
NCIS - <i>no attention</i>	58.07	0.920	0.810	<b>0.801</b>	<b>0.458</b>
NCIS	61.91	<b>0.925</b>	<b>0.814</b>	0.768	0.402

are compared to the GT and a match (*true positive*, TP) is established if a GT object exists whose intersection over union (IoU) is greater than 0.5. Unmatched objects are counted as *false positive* (FP) and unmatched GT objects are *false negatives* (FN). The average precision is then given by  $AP = TP / (TP + FP + FN)$ .

## 4.2 Experimental Results

We evaluated NCIS, Huaqian et. al [15], Hover-Net [7], and Mesmer [9] on images of the auditory cortex. The Dice coefficient for all methods (except Mesmer) is higher than 0.912, as shown in Table 1, highlighting that U-net-like architectures can produce reliable results for semantic segmentation regardless of the specific architecture. Huaqian et. al, based on three-class pixel classification, and Hover-Net, based on binary pixel classification and regression of two distances from the cell centroid, achieve similar performance. NCIS, which focuses on contours predictions through three-class pixel classification and regression of four distances (or color gradients) from the cell centroid, displays the best performance in terms of both semantic and instance segmentation accuracy. NCIS - *no attention*, the NCIS version with no attention gates, also performs well compared to the other methods but slightly worse than NCIS. Mesmer, which regresses the distance from the cell centroid to determine markers instead of contours during the post-processing, performs substantially worse.

When testing the methods on the cerebellum, an area not seen during training, we can observe lower performance overall, as expected. Interestingly, NCIS - *no attention* is the overall top performer, indicating that attention may be detrimental if the training set does not adequately represent the test set. For semantic segmentation, Mesmer is on par with NCIS - *no attention*, followed by Hover-Net. NCIS is the second-best for instance segmentation.

The qualitative outcomes for three example tiles are presented in Fig. 3. The segmentation of NCIS appears to be visually appealing, being smoother and more conforming to ground truth. Common segmentation errors include cell merging and inaccurate identification of artifacts as cells. For the cerebellum, all approaches struggle on the higher density granular layer, but the Purkinje cells, which closely match the cells encountered during training, are correctly isolated.

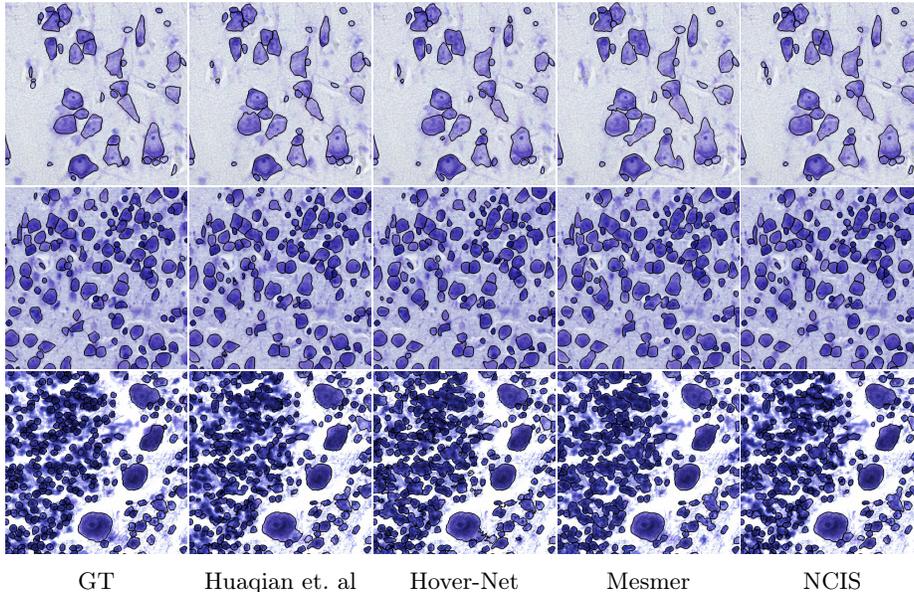


Fig. 3: Qualitative results for two sample tiles from the Auditory Cortex (top and center) and Cerebellum (bottom) datasets. GT is shown in the first column.

## 5 Discussion

Currently, there are limited techniques tailored for the segmentation of neuronal cells in Nissl-stained histological slices of the brain. To address this issue, we propose a new segmentation framework, called NCIS, which employs a dual-decoder U-Net architecture to enhance contour prediction by combining three-class pixel classification and regression of four color gradient maps. Our model outperforms existing state-of-the-art methods on images of the auditory cortex, demonstrating its ability to effectively deal with the challenges of neuronal cell segmentation (cells with variable shapes, intensity and texture, possibly touching or overlapping). If tested on an area of the brain not seen during training, the NCIS semantic segmentation accuracy is promising, but the instance segmentation performance indicates the need to enrich the training set. We believe that the number of NCIS-Net parameters could be reduced without compromising performance through minor architectural changes (e.g., summation instead of concatenation in skip connections).

NCIS could be particularly useful in processing histological WSIs of different species for comparative neuroanatomy studies, potentially contributing to the understanding of neurodegenerative and neuroinflammatory disorders.

## References

1. Amunts, K., Schleicher, A., Zilles, K.: Cytoarchitecture of the cerebral cortex—more than localization. *NeuroImage* **37**(4), 1061–1065 (2007)
2. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al.: Qupath: open source software for digital pathology image analysis. *Scientific Reports* **7**(1), 1–7 (2017)
3. Chevalier, G.: Make smooth predictions by blending image patches, such as for image segmentation (2017), <https://github.com/Vooban/Smoothly-Blend-Image-Patches>
4. Corain, L., Grisan, E., Grač, J.M., Carvajal-Schiaffino, R., Cozzi, B., Peruffo, A.: Multi-aspect testing and ranking inference to quantify dimorphism in the cytoarchitecture of cerebellum of male, female and intersex individuals: a model applied to bovine brains. *Brain Structure and Function* **225**(9), 2669–2688 (2020)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
6. García-Cabezas, M.Á., John, Y.J., Barbas, H., Zikopoulos, B.: Distinction of neurons, glia and endothelial cells in the cerebral cortex: an algorithm based on cytological features. *Frontiers in Neuroanatomy* **10**, 107 (2016)
7. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis* **58**, 101563 (2019)
8. Grač, J.M., Peruffo, A., Corain, L., Finos, L., Grisan, E., Cozzi, B.: The primary visual cortex of cetartiodactyls: organization, cytoarchitectonics and comparison with perissodactyls and primates. *Brain Structure and Function* **227**(4), 1195–1225 (2022)
9. Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C.C., McIntosh, B.J., Leow, K.X., Schwartz, M.S., et al.: Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology* **40**(4), 555–565 (2022)
10. IJsseldijk, L.L., Brownlow, A.C., Mazzariol, S.: Best practice on cetacean post mortem investigation and tissue sampling. *Jt. ACCOBAMS ASCOBANS Doc* pp. 1–73 (2019)
11. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. pp. 234–241. Springer (2015)
13. Schmidt, U., Weigert, M., Broaddus, C., Myers, G.: Cell detection with star-convex polygons. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II* 11. pp. 265–273. Springer (2018)
14. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*. pp. 6105–6114. PMLR (2019)

15. Wu, H., Souedet, N., Jan, C., Clouchoux, C., Delzescaux, T.: A general deep learning framework for neuron instance segmentation based on efficient unet and morphological post-processing. *Computers in Biology and Medicine* **150**, 106180 (2022)
16. Xing, F., Yang, L.: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Reviews in Biomedical Engineering* **9**, 234–263 (2016)
17. Zack, G.W., Rogers, W.E., Latt, S.A.: Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry* **25**(7), 741–753 (1977)
18. Zhou, Y., Onder, O.F., Dou, Q., Tsougenis, E., Chen, H., Heng, P.A.: Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. In: *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings* 26. pp. 682–693. Springer (2019)