scientific **data**

Check for updates

DATA DESCRIPTOR

Fluorescent Neuronal Cells v2: multi-task, multi-format annotations for deep learning in microscopy

Luca Clissa[1,2 ✉], Antonio Macaluso[3], Roberto Morelli[2], Alessandra Occhinegro[4], Emiliana Piscitiello[4], Ludovico Taddei[4], Marco Luppi[4], Roberto Amici[4], Matteo Cerri[4], Timna Hitrec[4], Lorenzo Rinaldi[1,2] & Antonio Zoccoli[1,2]

Fluorescent Neuronal Cells v2 is a collection of fluorescence microscopy images and the corresponding ground-truth annotations, designed to foster innovative research in the domains of Life Sciences and Deep Learning. This dataset encompasses three image collections wherein rodent neuronal cell nuclei and cytoplasm are stained with diverse markers to highlight their anatomical or functional characteristics. Specifically, we release 1874 high-resolution images alongside 750 corresponding ground-truth annotations for several learning tasks, including semantic segmentation, object detection and counting. The contribution is two-fold. First, thanks to the variety of annotations and their accessible formats, we anticipate our work will facilitate methodological advancements in computer vision approaches for segmentation, detection, feature extraction, unsupervised and self-supervised learning, transfer learning, and related areas. Second, by enabling extensive exploration and benchmarking, we hope Fluorescent Neuronal Cells v2 will catalyze breakthroughs in fluorescence microscopy analysis and promote cutting-edge discoveries in life sciences.

Background & Summary

Fluorescence microscopy is a pivotal imaging technique in life-science experiments, allowing researchers to study biological structures or processes with remarkable precision. It employs fluorescent dyes or proteins that emit light at specific wavelengths depending on the illuminating wavelength they absorb. Exploiting this phenomenon, specific molecules can be tagged (*stained*) with fluorescent markers, and observed through a microscope by filtering only their emitted light, thus providing valuable insights into their localization, activity, and interactions. Based on this principle, several microscopy technologies are available, with modern solutions able to acquire 3D volumetric images containing rich neuromorphological information. Nevertheless, more traditional alternatives like epifluorescence microscopy and 2D-slice imaging are still very popular due to their suitability for a broad spectrum of applications, simplicity in design and operation, fast acquisition and cost-effective setup.

A major bottleneck in the adoption of these techniques is the lack of fully automated pipelines for the analysis of the resulting data, that instead often necessitates manual recognition and/or counting of the neuronal structures of interest[1–3]. For instance, in the study of torpor mechanisms, researchers depend on laborious hand-crafted operations to identify neuronal networks associated with this process[4]. This manual aspect typically delays the analyses, also introducing potential errors due to limitations of human operators. Moreover, the similarity between structures of interest and the background often leads to challenges in distinguishing and accurately recognizing biological compounds, resulting in inherent arbitrariness and interpretation bias.

For these reasons, there is a growing interest in automating the recognition and counting of tagged elements in fluorescence microscopy[5–9]. Deep learning approaches have demonstrated great promise in various

[1]National Institute of Nuclear Physics, Bologna, Italy. [2]University of Bologna, Department of Physics and Astronomy, Bologna, Italy. [3]German Research Center for Artificial Intelligence (DFKI), Agents and Simulated Reality Department, Saarbruecken, Germany. [4]University of Bologna, Department of Biomedical and Neuromotor Sciences, Bologna, Italy. ✉e-mail: luca.clissa2@unibo.it
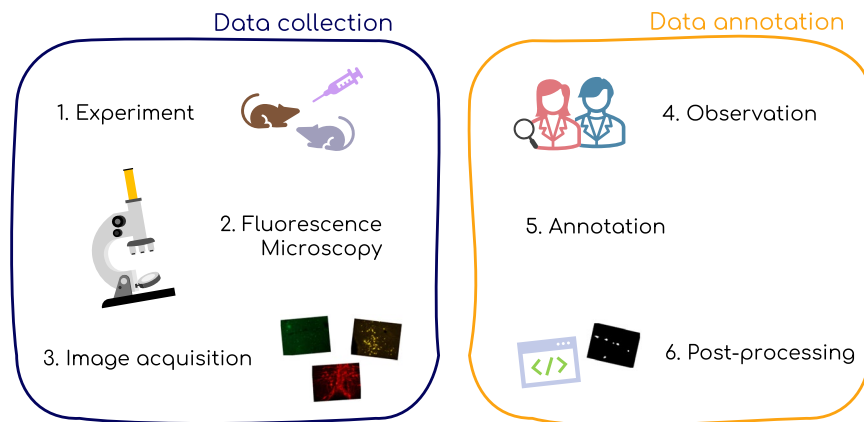
Fig. 1 Study design. The study was designed in two phases: *data collection*, where high-resolution pictures of rodent brain slices were acquired; and *data annotation*, where expert researchers collected annotations needed for supervised learning approaches.

object recognition tasks. However, their performance can deteriorate when applied to data from domains significantly different from those adopted for pre-training (*domain shift*[10,11]). Furthermore, the effectiveness of these approaches typically relies heavily on the availability of well-annotated data[12], which is often scarce and limited in the fluorescence microscopy domain.

To mitigate these issues, we present the Fluorescent Neuronal Cells v2 (FNC) dataset[13]. This archive features 3 data collections, for a total of 1874 high-resolution images of rodents brain slices capturing a diverse range of neuronal structures and staining patterns. To facilitate research in this field, we also provide 750 annotations in various formats, tailored to popular supervised learning tasks such as semantic segmentation, object detection, and counting. Apart from serving as an additional benchmark for testing model generalization in microscopy applications, the FNC dataset opens up several research opportunities. Firstly, the heterogeneity of biological structures and their visual characteristics enable testing the generalization of trained models, and validating transfer learning and domain adaptation methods[14,15]. Also, the availability of multiple annotation types allows the exploration of different learning paradigms, ranging from supervised and unsupervised approaches to self-/weakly-supervised techniques. Moreover, the specific challenges of our data well suit investigations into methodological advancements, e.g., assessing the effectiveness of different annotation formats and uncertainty estimation.

The design of the data collection process involved two distinct stages. Firstly, data collection was conducted following standardized experimental protocols. Specifically, controlled experimental conditions were applied to the animals, whose brains were sliced and processed by a classical immunofluorescence protocol to stain various neuronal substructures. Subsequently, a fluorescence microscopy was employed to capture high-resolution images of the areas of interest. Secondly, domain experts performed data annotation providing ground-truth labels necessary for supervised learning.

Despite the presence of open-source fluorescence microscopy datasets, several issues hinder their utilization for training deep learning models. Firstly, these collections typically lack accompanying ground-truth annotations, thus precluding the adoption of supervised learning techniques. Secondly, labelled datasets often include just a few dozens of images[16,17], that can be restrictive considering the data-intensive nature of deep learning models. Also, the moderate resolution of images in open datasets[18,19] hampers the effectiveness of resorting to crops as an alternative to whole images for augmenting sample size. Thirdly, most existing datasets predominantly include a single marker type[16,17], thus lacking diversity and limiting robust model training. Alongside these aspects, public datasets typically provide label types as dot-annotations or bounding boxes[16–19], which prevents their extension to fine-grained segmentation tasks. Additionally, the data accessibility is sometimes restricted due to the use of domain-specific formats[11], which complicates integration into deep learning frameworks and wide dissemination.

In response to these challenges, we present a large archive comprising high-resolution fluorescent microscopy images, encompassing different markers and cell types. Furthermore, the data are shared as easily accessible *PNG* files, and the corresponding annotations are provided in various types and formats, enabling the exploration of different learning approaches and tasks, thereby significantly expanding the scope of potential applications.

Methods

The FNC dataset[13] compiles images acquired from multiple studies and experimental conditions, while maintaining a consistent structure in the acquisition pipeline. Minor modifications were made to accommodate the specific requirements of each study and adapt to the current experimental circumstances and equipment. The data collection process consisted of two distinct and independent stages: *image acquisition* and *data annotation*. This section provides a comprehensive description of the data acquisition design, including the dedicated measures implemented for each image collection (refer to Fig. 1 for a visual summary).
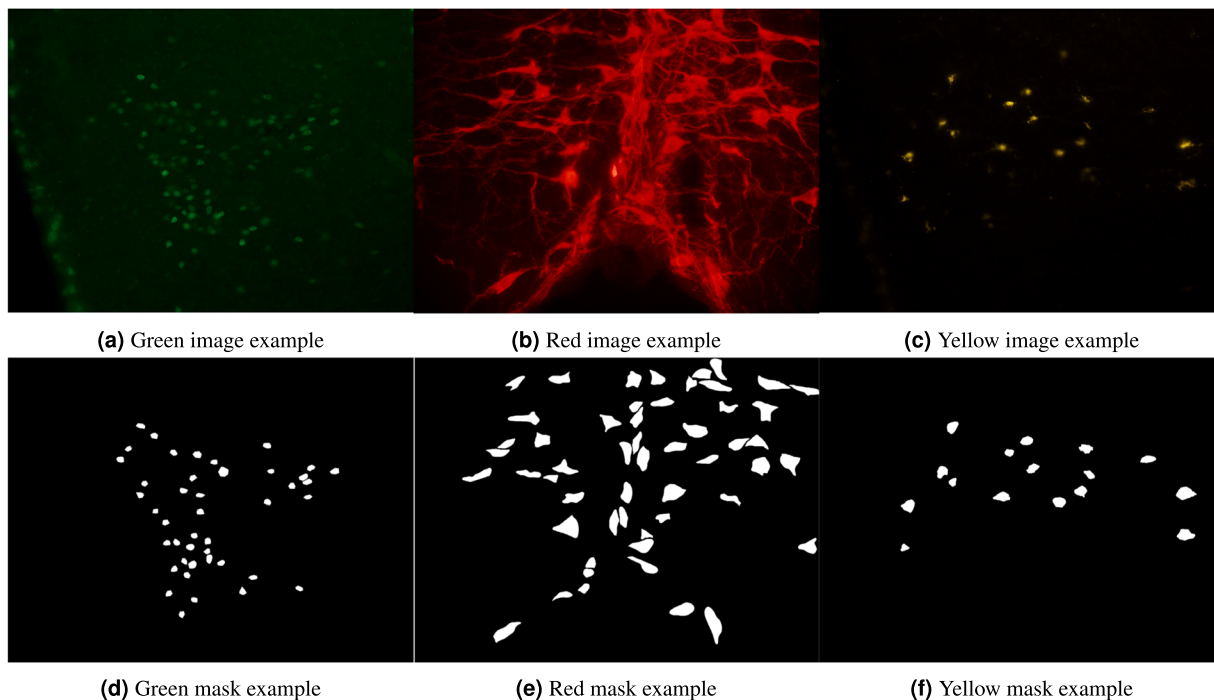
(a) Green image example   **(b)** Red image example   **(c)** Yellow image example

(d) Green mask example   **(e)** Red mask example   **(f)** Yellow mask example

Fig. 2 Data preview. The figures show examples of fluorescence micrscopy pictures (**a–c**) and the corresponding ground-truth binary masks (**d–f**).

Image acquisition.   In the image acquisition phase, a total of 68 rodents were subjected to controlled experimental conditions to study torpor and thermoregulatory mechanisms. At the end of the experimental session, the animals were deeply anaesthetized and transcardially perfused with 4% formaldehyde[4]. This process allowed for the tagging of several neuronal substructures located within the nucleus or cytoplasm of the neurons. All the experiments were conducted following approval by the National Health Authority (decree: No.141/2018 - PR/AEDB0.8.EXT.4), in accordance with the DL 26/2014 and the European Union Directive 2010/63/EU, and under the supervision of the Central Veterinary Service of the University of Bologna. All efforts were made to minimize the number of animals used and their pain and distress.

Rodents brains were then sectioned into $35\,\mu m$ thick tissue slices, with sampling conducted at regular intervals ($105\,\mu m$ for mice and of $210\,\mu m$ for rats) to avoid redundant data and ensure comprehensive coverage while maintaining manageable data size. Brain slices were finally stained for distinct markers following a standard immunofluorescence protocol[4]. Only some areas of interest were observed, namely the Raphe Pallidus (RPa), Dorsomedial Hypothalamus (DMH), Lateral Hypotalamus (LH), and Ventrolateral Periaqueductal Gray (VLPAG). These specific brain regions were chosen based on their relevance to the study of torpor mechanisms. The resulting specimens were observed by means of a fluorescence microscope equipped with a high-resolution camera. A specific wavelength of excitation light was selected for each collection based on the excitation wavelength of the chosen marker, resulting in pictures acquired with the application of green, yellow/orange or red filters. For simplicity, the image collections are named according to their prevalent hue. The original images were acquired as either *TIF* or *JPG* files depending on the camera default settings. To ensure traceability, a file naming convention was adopted to indicate their respective sample origins: `<animal_id>_S<sample_id>C<column_id>R<row_id>_<brain_area>_<zoom>_<collection_id>`. During the analysis phase, the raw data were converted to uncompressed *PNG* format, taking care to preserve the extensive set of associated metadata. This conversion aimed to enhance accessibility and facilitate broader utilization of the data, allowing for inspection and manipulation without the need for specialized software. Consequently, the FNC archive includes both these derived images and the original raw images, which are retained for data recovery and reproducibility purposes.

Green and yellow collections.   The images within these collections were obtained during the same experiment[4], in which brain sections from C57BL/6 J mice were stained with two markers to highlight specific substructures present in the neurons' nucleus and cytoplasm. The resulting brain slices were then observed using a Nikon Eclipse 80i microscope, equipped with a Nikon Digital Sight DS-Vi1 color camera, at a magnification of 200x.

More specifically, the **green** collection corresponds to cFOS staining (cf. Figure 2d). This staining method was employed to emphasize the nuclei of active neuronal cells[20], enabling the topographic analysis of brain areas that exhibit neuronal activity under specific experimental conditions. This approach is widely employed to identify neuronal cells responsible for regulating specific physiological phenomena.

In contrast, the **yellow** collection (cf. Figure 2c) utilized staining for the b-subunit of Cholera Toxin (CTb). This monosynaptic retrograde neuronal tracer migrates within the soma and axons of neuronal cells projecting
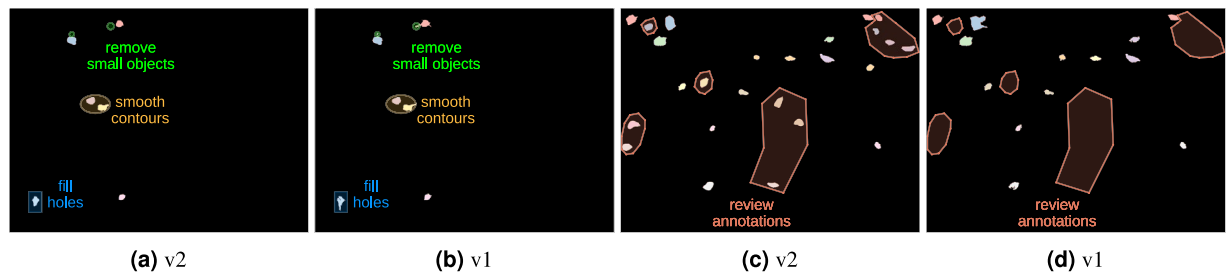
(a) v2   (b) v1   (c) v2   (d) v1

Fig. 3 Yellow masks v1 review. (**a,c**) illustrate how binary masks were reviewed compared to version 1, respectively (**b,d**). Improvements include: small objects removal, contour smoothing, holes filling and more consistent labelling.

to the brain area where CTb was previously injected during *in vivo* experiments[21]. Consequently, this staining technique facilitates the identification of morphological connections between different brain regions.

Red collection.   The **red** collection comprises images obtained from multiple unpublished experiments, concerning specimens of both mice and rats (cf. Figure 2b). Despite sharing the same experimental setup as green and yellow collections, this time the brain tissues were stained for various elements to phenotypically characterize the cells involved in the neural circuits underlying the physiological phenomena of torpor and thermoregulation. Specifically, slices were stained for orexin, tryptophan hydroxylase, and tyrosine hydroxylase. In this case, image acquisition was conducted using both the aforementioned Nikon Eclipse 80i microscope and an ausJENA JENAVAL microscope, equipped with a Nikon Coolpix E4500 color camera, at a magnification of 250x. For further details, please refer to the accompanying metadata for each image.

Data annotation.   The data annotation process was carried out by multiple proficient experimenters according to a fixed annotation protocol (see Annotations protocol.pdf file in the data archive), with multiple revision rounds to ensure data quality and minimize operator bias. We adopted the Visual Geometry Group Visual Image Annotator (VIA) annotation tool[22,23], which employs a web interface for image visualization and allows for the overlaying of annotations in different forms. In our study, the tagging process involved creating polygon contours, and the resulting annotations were exported into *CSV* format. To generate the binary masks required for training, the polygon contours were transformed using programming libraries such as OpenCV and scikit-image. For the yellow collection, we utilized the binary masks available from version 1[24] as pre-annotations. Specifically, we employed erosions and dilations techniques to address fragmented contours resulting from semi-automatic labeling based on thresholding. Furthermore, we applied methods to fill small holes within segmented objects, and removed spurious objects that went overlooked in the previous annotations or were erroneously added by prior processing. Subsequently, these pre-annotations were refined manually using VIA, enhancing their accuracy and ensuring better consistency across the annotations (see Fig. 3). In contrast, the green and red collections were annotated from scratch.

Upon completion of the labeling process, the polygon contours exported from VIA were also converted into multiple annotation types and formats. This conversion aims at facilitating accessibility for a wide range of users and promoting the exploration of various learning problems related to our data. For a more comprehensive understanding of the available formats and annotation types, please refer to the Section *Data Records*.

Data Records

The FNC dataset[13] is a collection of 1874 high-resolution fluorescent microscopy pictures, 750 of which also have their corresponding ground-truth segmentation masks, while the remaining 1124 are unlabelled. It is hosted on AMS Acta, the open access repository managed by the University of Bologna. The data are organized into three standalone image collections, named for simplicity *green*, *yellow*, and *red*, each available under the corresponding folder (see Fig. 4c). The collections share a common layout to facilitate easy access and analysis (see Fig. 4a).

To aid users in navigating the archive, the metadata_v2.xlsx file provides a comprehensive overview of the FNC data collection. It includes high-level metadata for each image, such as the corresponding animal, acquisition details, data partition, and annotation information.

Image collection folder structure.   The `trainval` and `test` folders contain all labelled images for each collection. These data partitions were obtained through a random 75%/25% split and are recommended as a suggested configuration to ensure reproducibility and comparability in future studies. The remaining images were collected under the `unlabelled` folder.

Inside each data partition folder, the `images` folder contains fluorescence microscopy images in *PNG* format. All the images are accompanied by a rich set of metadata, stored both in their *EXIF* tags and as a separate *TXT* file under the `metadata` folder. The `ground_truths` folder contains annotations in various formats commonly used within the machine learning community (see Fig. 4b).

Annotation types and formats.   The FNC collection[13] provides annotations of multiple *types*, encoded in several standard *formats*. In the masks folder, we find the binary masks typically used for segmentation tasks (cf. Figure 2d,f). The correspondence between the masks and the respective images can be established based
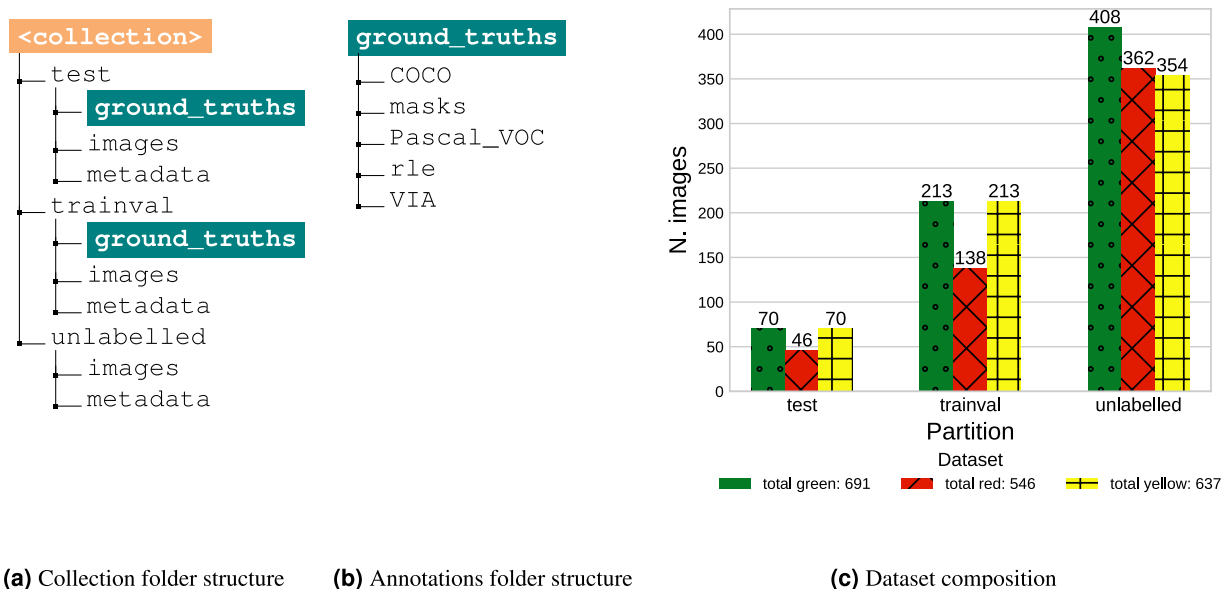
```
<collection>
  ├── test
  │     ├── ground_truths
  │     ├── images
  │     └── metadata
  ├── trainval
  │     ├── ground_truths
  │     ├── images
  │     └── metadata
  └── unlabelled
        ├── images
        └── metadata
```

```
ground_truths
  ├── COCO
  ├── masks
  ├── Pascal_VOC
  ├── rle
  └── VIA
```

(a) Collection folder structure   **(b)** Annotations folder structure   **(c)** Dataset composition

Fig. 4 FNC dataset structure. (**a**) shows the structure of each image collection folder, while (**b**) gives more details on the organization of the annotations directory. (**c**) summarizes the composition of each image collection, with the amounts of training, testing and unlabelled images.

on the filenames. The other folders store a light-weight encoding of the binary masks, enriched with additional annotation types/formats.

The `rle` directory contains *Running Length Encoding (RLE)* of the binary masks, stored as *pickle* files. This encoding is a compressed representation that can effectively save disk space while preserving the complete segmentation information. It is particularly convenient for high-resolution images like those present in our dataset.

The other directories provide several annotation types, and they are named after their annotation format. **Polygon** annotations are available in each of the `VIA`, `COCO` and `Pascal_VOC` directories, in the form of *json* or *xml* files. COCO[25] and Pascal_VOC[26] formats also features **bounding boxes** and **dot** annotations for object detection tasks, and **count** labels for object counting.

Fluorescent neuronal Cells v1 comparison. The FNC v2 archive introduced by this work features an extension and re-elaboration of FNC v1[6,24] content. In particular, the *Red* and *Green* collections present entirely novel and unpublished data, while the *Yellow* collection in v2 is derived from v1. Specifically, the 283 fluorescent microscopy images contained in v1 are reported in v2 inside the `trainval` and `test` partitions, with numbered filenames instead of the original ones. For an exact matching, the users may refer to the *original_name* and *image_name* columns in metadata_v2.xlsx for v1 and v2 filenames, respectively. As far as the annotations, *Yellow* v1 binary masks were used as pre-annotations for version 2 to improve their quality and consistency, as described in Subsection Data annotation. The other formats were then derived from the results of the previous re-elaboration.

Technical Validation

In order to demonstrate the potential for successful model training and analysis using the provided dataset, we conducted three types of checks to ensure the accuracy and quality of the annotations.

Firstly, polygon annotations were obtained by experienced researchers. During this phase, the annotations underwent multiple rounds of double-checking to ensure that the polygons did not have intersecting edges and that they accurately represented the objects when transformed into binary masks.

Secondly, we leveraged domain knowledge to validate the annotations. Precisely, we tested the binary masks against our expectations regarding the sizes and shapes of the biological structures involved. This validation process relies on a quantitative evaluation concerning objects' area and diameter, complemented by a visual scrutiny of the masks to ensure they align with the expected shapes and exhibit smooth contours. Table 1 reports summary statistics for the distribution of key features at the image and object levels, that can be leveraged for technical validation. For instance, the annotated objects display an average area of nearly 75, 247, and 133 μm^2 for green, red, and yellow cells, respectively. These values align with the expected size of the biological structures represented in each image collection. Additionally, the analysis of Feret and equivalent diameters provides an understanding of the typical form of the stained objects. In particular, the Feret diameter[27] can be interpreted as a measure of the maximum extension of an object, whereas the equivalent diameter represents the diameter the object would have if it were a perfect circle with the same area. Thus, comparing these two metrics can offer insight into the objects' shape regularity. For green cells, the values for the two measurements are relatively close (roughly 12 VS 10 μm), suggesting that these cells are broadly circular or oval in shape. A similar conclusion can be drawn for the yellow stains, albeit with slightly more variability (approximately 17 and 13 μm), indicating generally regular shapes with occasional deviations. In the case of red objects, instead, the comparison is

collection	signal %			cell count			area (μm²)			Feret diameter (μm)			equivalent diameter (μm)		
	green	red	yellow	green	red	yellow	green	red	yellow	green	red	yellow	green	red	yellow
count	283	184	283	4606ᵃ	4486ᵃ	2659ᵃ	4600	4483	2621	4600	4483	2621	4600	4483	2621
mean	0.64	2.61	0.65	28.65	34.96	25.76	74.78	246.52	132.55	12.20	25.68	17.24	9.64	17.13	12.63
std	0.57	1.74	0.80	17.49	17.90	19.88	23.39	130.24	66.02	1.92	8.86	4.86	1.48	4.50	3.04
min	0	0	0	0	0	0	12.41	24.81	17.62	5.48	7.08	6.09	3.97	5.62	4.74
10%	0.10	0.37	0	9	14	4	48.74	106.72	66.16	9.85	15.67	11.93	7.88	11.66	9.18
25%	0.25	1.34	0.13	15	21	9	58.58	144.04	88.02	10.88	18.64	13.90	8.64	13.54	10.59
50%	0.45	2.34	0.40	25	31	19	71.67	218.80	117.75	12.07	24.38	16.51	9.55	16.69	12.24
75%	0.92	3.53	0.82	39.75	50	44	87.33	320.75	161.17	13.35	30.99	19.76	10.54	20.21	14.33
90%	1.38	4.89	1.49	56	58	57	104.66	422.64	218.27	14.70	37.64	23.59	11.54	23.20	16.67
max	3.36	8.40	5.54	69	90	72	227.04	842.66	548.20	21.21	74.39	41.66	17	32.76	26.42

Table 1. Summary statistics of key features' distribution for each image collection. The top portion highlights global indicators, while the bottom one reports given percentiles of each distribution. ᵃThe difference compared to counts in following columns comes from the contribution of empty images. These amount to 6, 3 and 38 images for green, red and yellow collections, respectively.

metrics	segmentation			detection			counting		
	F_1 Score	Precision	Recall	F_1 Score	Precision	Recall	MAE	MedAE	MPE (%)
Green	0.69	0.79	0.62	0.77	0.86	0.69	1.20	1.00	9%
Red	0.28	0.33	0.25	0.60	0.69	0.53	7.13	7.00	42%
Yellow	0.65	0.67	0.63	0.77	0.78	0.75	1.54	1.00	18%

Table 2. Performance metrics by learning task. The *segmentation* portion refers to TP, FP, and FN computed based on objects overlapping (IoU). For *detection* metrics, predicted and true cells are associated based on their centers' distance. *Counting* metrics simply consider the difference between predicted and true objects.

markedly different. This time we observe a Feret diameter around 26 μm against an equivalent diameter of 17, which suggests that these stains are typically elongated in one direction rather than concentrated around a center of mass. All these observations are also corroborated when visually inspecting annotated cells, which confirms prior expectations about objects size and shapes based on the nature of the marked structures.

Learning. Thirdly, we conducted a sample training phase for each image collection using a **cell ResUnet** architecture[6], specifically designed for this type of application. Specifically, we trained a network from scratch for each collection using a Dice Loss[28] and the Adam optimizer[29]. The initial learning rate was set based on the "learning rate test"[30] implemented by *fastai*'s[31] lr_find() method. The training phase continued for 100 epochs with cyclical learning rates[32,33], and the best model was selected based on the best validation dice coefficient. For all technical details please refer to the GitHub repository (see Section Code availability). This training phase aims to verify the effectiveness of the data in facilitating the learning of beneficial cell features. Additionally, the intent is to highlight the relevant metrics for result evaluations. In particular, we suggest performance should be assessed differently depending on the end goal of future analyses.

For **segmentation** tasks, we provided an implementation where matching of actual and predicted neurons – i.e., the calculation of True Positives (TP), False Positives (FP) and False Negatives (FN) – is done based on their overlap, quantified as *Intersection-over-Union (IoU)*[34]. This approach not only ensures a 1-1 correspondence of true and predicted objects but also assesses how closely the predictions reconstruct the shape of ground-truth cells. Building on top of this definition, standard metrics such as *precision, recall* and F_1 *score* can be computed as measures of global performance.

Detection tasks, on the other hand, would benefit from a looser matching criterion, comparing predicted and true objects' centers instead of overlaps. This approach prioritizes recognition over precise shape reconstruction.

Finally, for **counting** tasks we suggest common regression metrics such as *Mean Absolute Error (MAE), Median Absolute Error (MedAE)* and *Mean Percentage Error (MPE)*.

Table 2 shows the results of the sample training for each image collection. These results are not intended to be a comprehensive exploration of the model's capabilities on FNC data[13], but rather to showcase some characteristics of various evaluation methods. Nonetheless, they may serve as a baseline for future studies.

Despite no optimization of the training pipeline, the initial results are mainly satisfactory (except for red segmentation), confirming the technical robustness of the data collection process. Going into more details, we observe a marked discrepancy between segmentation and detection metrics. As expected, the F_1 scores based on the distance between true and predicted centers of mass are significantly higher than the corresponding segmentation indicators. Moreover, the discrepancy is greater for image collections where the objects have more irregular shapes (green < yellow < red). This is a consequence of the more inclusive matching criterion used for detection tasks.

In terms of counting, performance is already very satisfactory. However, these metrics may not fully represent the model's performance as good results could arise due to a balancing effect between true positives and false negatives. Interestingly, despite low absolute errors, the percentage error is relatively high due to the impact of errors in images with few or no cells. To address this issue, we adopt the following formula for MPE computation: $\text{MPE} = \frac{(\text{predicted} - \text{true})}{\max(\text{true}, 1)}$. In this way, the fraction is not over-inflated when there are no cells in the original mask.

Usage Notes

The Fluorescent Neuronal Cells collection[13] is available both as a comprehensive archive and as individual image collections for specific research requirements. This enables users to download the data efficiently and selectively, based on their specific needs. The code provided is based on the Python and PyTorch frameworks, offering a robust foundation for analysis and modeling. However, thanks to the popularity of the annotation formats and the use of PNG images, users can easily employ their preferred deep learning framework.

Peculiar traits.   In all image collections, the visual representation is characterized by the prevalence of two distinct color tones, which result from the deliberate selection of a specific wavelength. One tone appears darker, indicating areas where light has been filtered out, while the other tone is brighter and more intense, emitted by the fluorophore corresponding to the color of each collection (see Fig. 2a–c). As a result, the images can generally be depicted using variations of a single color. Consequently, a 1-D representation may be sufficient, or an alternative color space other than RGB could provide more informative and less redundant data.

Notice, however, that the specific colors employed in our studies were dictated not by any inherent or functional property of the stained biological structures, but rather by their accessibility at the time of the experiments. Therefore, it would be a misinterpretation to associate specific colors to particular neuronal substructures. In fact, these colors serve only as contrasting elements to discern the stained foreground objects from the background. Consequently, the emphasis should lie primarily on learning this discrimination rather than matching specific colors with the neuronal structures. Thus, the particular colors should not be considered indicative of the type of neuronal cells or their functional attributes, but merely as a practical tool aiding in the overall visualization and interpretation.

Challenges.   Some important insights for future studies can be drawn examining ground-truth masks at the pixel level, revealing significant characteristics that impact the training process.

The two classes, namely cells (1) and background (0), exhibit an extreme **class imbalance**, with background pixels being overwhelmingly predominant, typically exceeding cell pixels by over a factor of 100 (cf. Table 1, signal %). These observations highlight the necessity for specialized training strategies to address this pronounced class imbalance and effectively learn the pixel classification.

Additional challenges are associated with the macroscopic content of the images. The Fluorescent Neuronal Cells data showcase a diverse collection of 11704 subnuclear neuronal structures, varying in shape, size, and extension (cf Table 1, *area, Feret diameter* and *equivalent diameter* columns). The distribution of these structures across the collections is uneven, with some images containing numerous cells while others are devoid of them. Consequently, the model needs to be flexible enough to handle both scenarios.

Furthermore, despite considerable efforts to stabilize the acquisition procedure, several technical challenges persist. Firstly, there is a **high variability in terms of color, saturation, and contrast** from one image to another. For instance, there are instances where the tissues absorb some of the markers (see Fig. 5b,e,g), causing irrelevant compounds to emit light which is then captured by the microscope. Consequently, the background's hue may shift towards values similar to those of faint neuronal cells (see Fig. 5b–f). In such circumstances, relying solely on pixel intensity is insufficient to distinguish between signal and background, necessitating the consideration of additional characteristics such as saturation and contrast. However, even the analysis of these characteristics is not straightforward, as fluorescent emissions are naturally unstable, leading to fluctuations in the saturation levels exhibited by cell pixels (cf. Figure 5a–c or Fig. 5f,g).

Moreover, the substructures of interest have a fluid nature. Also, the shot can capture different two-dimensional sections depending on how the cells are oriented within the tissues. As a consequence, the **size and the shape of the stained cells can vary significantly** (cf. objects dimension in Fig. 2d,f), further complicating the discrimination between cells and the background.

Another challenge arises from the occasional presence of accumulations of fluorophore in narrow areas, resulting in emissions that closely resemble those of cells. These **artifacts** can manifest as small areas, such as point artifacts and filaments, or larger structures, like lateral stripes (see Fig. 5b,g). Again, their presence hampers the detection task, making the recognition and the understanding of cells structure and size mandatory for the model.

A further source of complexity is represented by **overcrowding** (Fig. 5a,d,f,g). When several cells are close-by, maybe partially overlapping, precisely localizing cell boundaries can be challenging, thus requiring adjustments to prevent the model from merging nearby cells into single agglomerations.

Last but not least, in some occasions the recognition of cells may be ambiguous even for human operators (cf. *marked* and *non-marked* instances in Fig. 5a,d,f,g). Of course, this poses an issue of intrinsic **subjectivity** in the annotation process, which in turn affects both the training and assessment phases.

By and large, all of these factors make the recognition and counting tasks harder and complicate the learning process. Likewise, borderline annotations hinder model assessment as their subjectivity introduces irreducible noise in the evaluation.
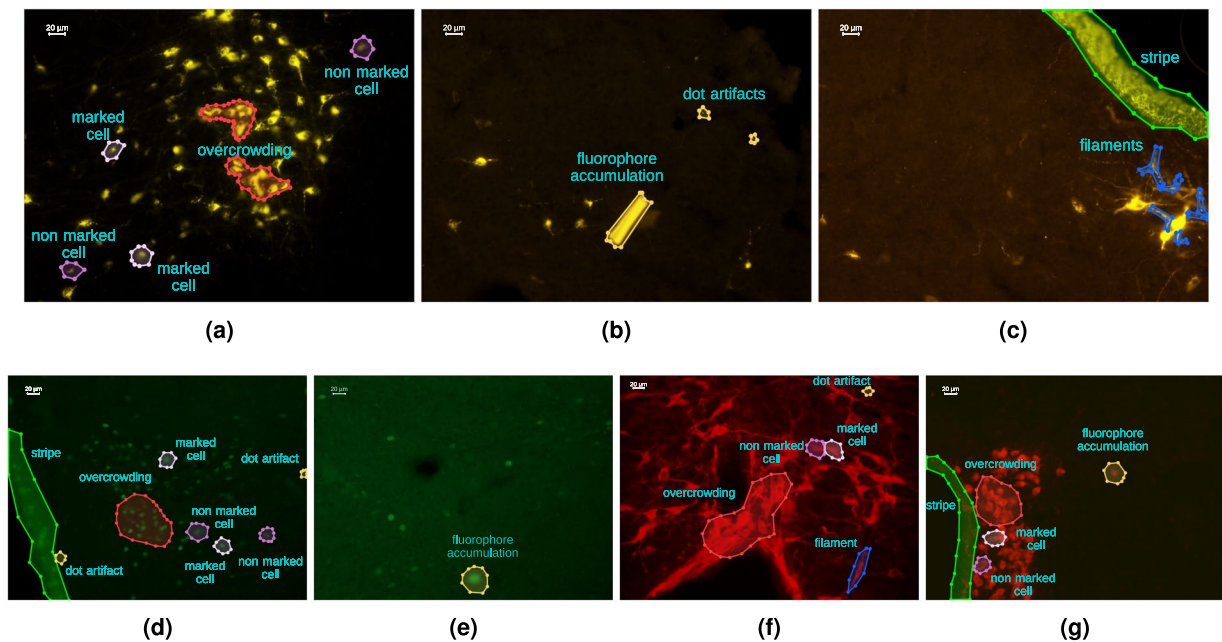
Fig. 5 Challenges. FNC data present several difficulties to take into account during their analysis. Common challenges are represented by *overcrowding* (**a,d,f,g**), *ambiguity* (**a,d,f**), and *artifacts* (**b–g**).

Limitations. While the Fluorescent Neuronal Cells data encompass diverse images in many aspects, they exhibit reduced heterogeneity in some respects.

First, all the images were collected by a single research laboratory using fixed experimental conditions and acquisition setups. Despite being representative of standard procedures[35,36], the lack of diversity in this regard may lead to suboptimal generalization when applied to data collected under different settings.

Furthermore, the images were captured using epifluorescence microscopy, resulting in noisier images and no variability in terms of acquisition technologies employed. However, we believe the reduced image quality may actually represent an interesting assessment scenario due to its more challenging nature. Indeed, it is reasonable to assume that pre-training on FNC data[13] should generalize well to modern equipment like confocal microscopy, where better image definition, sharper object boundaries and improved signal-to-noise ratio should ease the recognition task.

Another limitation lies in the lack of diversity in the cell types depicted and the animal species involved. Our dataset only focuses on subcellular components of rodent neurons. This might potentially impact the generalization of the models to different use cases and restrict their application to other cell types or animal species.

Code availability

The code associated with this work is available on GitHub (https://github.com/clissa/fluocells-scientific-data). The repository contains utils to:

- perform **data operations** (`dataOps/`: i) converting raw *TIFF* images into *PNG* with metadata, ii) recreating expected data folders structure, iii) convert VIA annotation to binary masks, iv) encode binary masks into various annotation formats and types, v) preprocess yellow masks from previous FNC version[24])
- implement deep learning **modelling** strategies (`fluocells/models/`: contains network blocks to implement c-ResUnet architecture[6]; `compute_metrics.py`, `evaluate.py` and `training.py`: contain utils to implement model training and evaluation)
- explore, analyze and evaluate models interactively (`notebooks/`: contains jupyter notebooks with examples of how to deal with standard stages of data analysis, namely i) exploratory data analysis, ii) implementation of model architecture and training pipeline, and iii) experiments

References

1. Dentico, D. *et al.* C-fos expression in preoptic nuclei as a marker of sleep rebound in the rat. *European Journal of Neuroscience* **30**, 651–661, https://doi.org/10.1111/j.1460-9568.2009.06848.x (2009).
2. Luppi, M. *et al.* C-fos expression in the limbic thalamus following thermoregulatory and wake–sleep changes in the rat. *Experimental Brain Research* **237**, 1397–1407, https://doi.org/10.1007/s00221-019-05521-2 (2019).
3. Chiocchetti, R. *et al.* Phosphorylated tau protein in the myenteric plexus of the ileum and colon of normothermic rats and during synthetic torpor. *Cell and Tissue Research* **384**, 287–299, https://doi.org/10.1007/s00441-020-03328-0 (2021).
4. Hitrec, T. *et al.* Neural control of fasting-induced torpor in mice. *Scientific Reports* **9**, https://doi.org/10.1038/s41598-019-51841-2 (2019).

5. Clissa, L., Macaluso, A. & Zoccoli, A. Optimizing deep learning models for cell recognition in fluorescence microscopy: The impact of loss functions on performance and generalization. In Foresti, G. L., Fusiello, A. & Hancock, E. (eds.) *Image Analysis and Processing - ICIAP 2023 Workshops*, 179–190, https://doi.org/10.1007/978-3-031-51023-6_16 (Springer Nature Switzerland, Cham, 2024).

6. Morelli, R. *et al.* Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. *Scientific Reports* **11**, 22920, https://doi.org/10.1038/s41598-021-01929-5 (2021).

7. Cao, Y., Liu, S., Peng, Y. & Li, J. Denseunet: densely connected unet for electron microscopy image segmentation. *IET Image Processing* **14**, 2682–2689, https://doi.org/10.1049/iet-ipr.2019.1527 (2020).

8. Riccio, D., Brancati, N., Frucci, M. & Gragnaniello, D. A new unsupervised approach for segmenting and counting cells in high-throughput microscopy image sets. *IEEE Journal of Biomedical and Health Informatics* **PP**, 1–1, https://doi.org/10.1109/JBHI.2018.2817485 (2018).

9. Kumar, N. *et al.* A multi-organ nucleus segmentation challenge. *IEEE Transactions on Medical Imaging* **39**, 1380–1391, https://doi.org/10.1109/TMI.2019.2947628 (2020).

10. Ouyang, C. *et al.* Causality-inspired single-source domain generalization for medical image segmentation. *IEEE Transactions on Medical Imaging* **42**, 1095–1106, https://doi.org/10.1109/TMI.2022.3224067 (2023).

11. Poon, C., Teikari, P., Rachmadi, M. F., Skibbe, H. & Hynynen, K. A dataset of rodent cerebrovasculature from *in vivo* multiphoton fluorescence microscopy imaging. *Scientific Data* **10**, 141, https://doi.org/10.1038/s41597-023-02048-8 (2023).

12. Xie, J., Kiefel, M., Sun, M.-T. & Geiger, A. Semantic instance annotation of street scenes by 3d to 2d label transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.401 (2016).

13. Clissa, L. *et al.* Fluorescent neuronal cells v2. *AMS Acta* https://doi.org/10.6092/unibo.amsacta/7347 (2023).

14. Haq, M. M. & Huang, J. Adversarial domain adaptation for cell segmentation. In *Medical Imaging with Deep Learning*, 277–287, http://proceedings.mlr.press/v121/haq20a.html (PMLR, 2020).

15. Brieu, N. *et al.* Domain adaptation-based augmentation for weakly supervised nuclei detection. Preprint at https://arxiv.org/abs/1907.04681 (2019).

16. Raza, S. E. A. *et al.* Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis* **52**, 160–173, https://doi.org/10.1016/j.media.2018.12.003 (2019).

17. Taschner-Mandl, S. *et al.* An annotated fluorescence image dataset for training nuclear segmentation methods. https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BSST265 (2020).

18. Waithe, D. *et al. Fluorescence Microscopy Data for Cellular Detection using Object Detection Networks.* https://doi.org/10.5281/zenodo.2548493 (2019).

19. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* **18**, 100–106, https://doi.org/10.1038/s41592-020-01018-x (2021).

20. Kovács, K. Measurement of immediate-early gene activation-c-fos and beyond. *Journal of neuroendocrinology* **20**, 665–672, https://doi.org/10.1111/j.1365-2826.2008.01734.x (2008).

21. Lencer, W. I. & Tsai, B. The intracellular voyage of cholera toxin: going retro. *Trends in biochemical sciences* **28**, 639–645, https://doi.org/10.1016/j.tibs.2003.10.002 (2003).

22. Dutta, A., Gupta, A. & Zissermann, A. VGG image annotator (VIA). http://www.robots.ox.ac.uk/~vgg/software/via/. Version: 2.0.12, Accessed: 2023 (2016).

23. Dutta, A. & Zisserman, A. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, https://doi.org/10.1145/3343031.3350535 (ACM, New York, NY, USA, 2019).

24. Clissa, L. *et al.* Fluorescent neuronal cells. *AMS Acta* https://doi.org/10.6092/unibo.amsacta/6706 (2021).

25. Lin, T.-Y. *et al.* Microsoft COCO: common objects in context. *In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham.* https://doi.org/10.1007/978-3-319-10602-1_48 (2015).

26. Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. The pascal visual object classes (VOC) challenge. *International journal of computer vision* **88**, 303–338, https://doi.org/10.1007/s11263-009-0275-4 (2010).

27. Pabst, W. & Gregorova, E. Characterization of particles and particle systems. *ICT Prague* **122**, 122 (2007).

28. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, 240–248, https://doi.org/10.1007/2F978-3-319-67558-9_28 (Springer, 2017).

29. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015.* https://arxiv.org/abs/1412.6980 (2015).

30. Smith, L. N. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. Preprint at https://arxiv.org/abs/1803.09820 (2018).

31. Howard, J. & Gugger, S. Fastai: a layered API for deep learning. *Information* **11**, 108, https://doi.org/10.3390/info11020108 (2020).

32. Smith, L. N. Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, 464–472, https://doi.org/10.1109/WACV.2017.58 (IEEE, 2017).

33. Smith, L. N. & Topin, N. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, vol. 11006, 369–386, https://doi.org/10.1117/12.2520589 (SPIE, 2019).

34. Kirillov, A., He, K., Girshick, R., Rother, C. & Dollár, P. Panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9404–9413, https://doi.org/10.1109/CVPR.2019.00963 (2019).

35. Shi, Y. *et al.* 5-HT7 receptors expressed in the mouse parafacial region are not required for respiratory chemosensitivity. *The Journal of Physiology* **600**, 2789–2811, https://doi.org/10.1113/jp282279 (2022).

36. Cano, G., Hernan, S. L. & Sved, A. F. Centrally projecting edinger-westphal nucleus in the control of sympathetic outflow and energy homeostasis. *Brain Sciences* **11**, 1005, https://doi.org/10.3390/2Fbrainsci11081005 (2021).

Acknowledgements

Author contributions

L.C. conceived the idea of curating and sharing the annotated data. M.L., R.A., M.C. and T.H. designed the biological study and M.L., A.O., E.P. and L.T. collected the original microscopic fluorescent images. L.C. defined and orchestrated the annotation pipeline and instruments. M.L., A.O., E.P. and L.T. performed data annotation and review. L.C., A.M. and R.M. implemented the code for data conversion, pre-processing, exploration,

visualization and modelling. L.C. and A.M. performed technical validation. L.C. and A.M. drafted the manuscript. L.C., A.M., R.M., M.L., A.O., E.P. and L.T. proofread and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.C.

Reprints and permissions information is available at www.nature.com/reprints.

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