

# Deep Learning Approach for Layer-Specific Segmentation of the Olfactory Bulb in X-ray Phase-Contrast Tomography

V.A. Karyakina<sup>1</sup>, D.V. Polevoy<sup>1,2,3</sup>, I. Bukreeva<sup>4</sup>, O. Junemann<sup>4</sup>, S.V. Saveliev<sup>5</sup>, M.V. Chukalina<sup>1,3</sup>

<sup>1</sup> Smart Engines Service LLC, 117312, Russia, Moscow, 60-Letiya Oktyabrya pr. 9;

<sup>2</sup> Institute for Information Transmission Problems, RAS, 127051, Russia, Moscow, Bolshoy Karetny per. 19, build.1;

<sup>3</sup> Federal Research Center "Computer Science and Control", RAS, 119333, Russia, Moscow, Vavilova Str. 44, build.2;

<sup>4</sup> Institute of Nanotechnology CNR, Rome unit, 00185, Italy, Rome, Piazzale Aldo Moro 5;

<sup>5</sup> Avtsyn Research Institute of Human Morphology of Federal State Budgetary Scientific Institution "Petrovsky National Research Centre of Surgery", 117418, Russia, Moscow, Tsyurupy Str. 3

## Abstract

This paper addresses neural network segmentation of a human olfactory bulb sample on X-ray phase-contrast tomographic reconstruction. The olfactory bulb plays a key role in the primary processing of olfactory information. It consists of several nested cell layers, the morphometric analysis of which has important diagnostic value. However, manual segmentation of the reconstructed volume is labor-intensive and requires high qualifications, which makes the development of automated segmentation methods crucial. X-ray phase-contrast tomography provides a high-resolution reconstruction of the olfactory bulb morphological structure. The resulting reconstructions are characterized by excessive morphological details and reconstruction artifacts. These features, combined with limited data volume, visual similarity of neighboring slices, and sparse ground truth, hinder the application of standard neural network-based segmentation approaches. This paper examines the characteristics of the data under consideration and suggests a training pipeline for a convolutional neural network, including inter-slice smoothing at the data preprocessing stage, alternative strategies for splitting the data into subsets, a set of augmentations, and training on sparse sampling. The proposed adaptations achieved a Dice score (micro) value of 0.93 on the test subset. An ablation study demonstrated that each of the above-mentioned modifications independently improves segmentation quality. The presented training pipeline can be applied to the segmentation of morphological structures on tomographic images in biomedical tasks with a limited dataset and non-standard ground truth.

**Keywords:** olfactory bulb, deep learning, convolutional neural network, semantic segmentation, data curation, X-ray phase-contrast tomography

**Citation:** Karyakina VA, Polevoy DV, Bukreeva I, Junemann O, Saveliev SV, Chukalina MV. Deep Learning Approach for Layer-Specific Segmentation of the Olfactory Bulb in X-ray Phase-Contrast Tomography. *Computer Optics* 2025; 49(6): 1102-1111. DOI: 10.18287/COJ1763.

## Introduction

The olfactory bulb (OB) is an anatomical structure in the forebrain that comprises a system of concentric cell layers responsible for processing olfactory information. Morphometric analysis of the OB layers plays an important role in the study of neurodegenerative disorders [1] and the effects of COVID-19 [2].

Modern methods of semantic segmentation of biomedical images predominantly rely on neural networks [3, 4]. The most popular architectures are U-Net [5] and its numerous modifications [6, 4]. Among these developments, nn-UNet and its derived variants [7, 8] are leading the way in biomedical 3D segmentation.

Only a few studies have examined the segmentation of the OB, mainly focusing on its overall shape and volume rather than the internal layer structure [9, 10, 11]. Recently, the study [12] proposed an interactive method for volumetric segmentation of layers in the rodent OB, while the authors of [13] employed a convolutional neural network (CNN) to segment individual OB layers in the human brain. Compared to previous research, this study analyzes the human OB layer data at a micro level, which makes the segmentation process more challenging.

The internal morphological structure of the OB can be reconstructed using X-ray phase-contrast tomography (XPCT) [14, 15], which provides resolution at the cellular level. The prior work [13] describes a pipeline for semantic segmentation of a highly detailed 3D reconstruction of a single OB sample using a U-Net-based neural network. However, the severely limited amount of data for the single OB sample considered in this paper, in conjunction with high-frequency noise, low contrast, indistinct inter-layer boundaries, and the peculiarities of the ground truth (GT), requires adapting standard model training approaches to the specifics of the task.

This study analyzes the model training pipeline described in [13] and proposes improvements, considering the characteristics of the data and aiming to enhance the quality of the OB semantic segmentation (see fig. 1). In addition, we organized the XPCT Human Olfactory Bulb Dataset (XPCT-HOBD), containing data originally obtained in [13]. The XPCT-HOBD, along with trained model and scripts for data preprocessing and inference, is publicly available (see "Data and code availability").

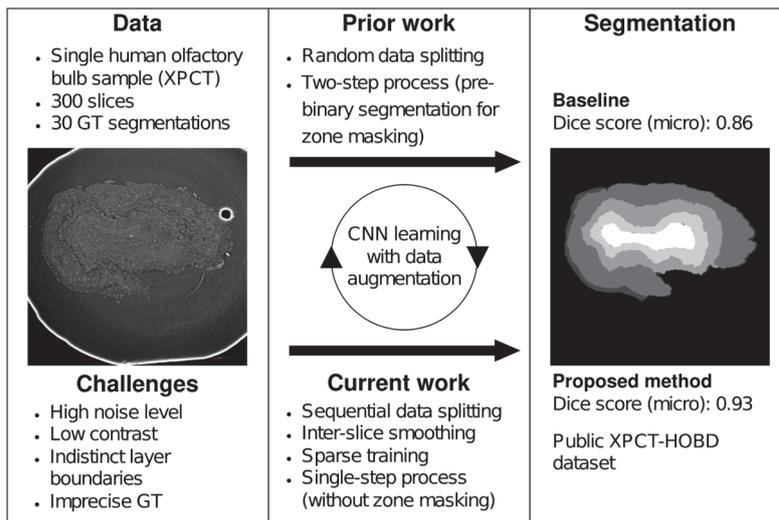


Fig. 1. Task and research outline

This article is organized as follows. The "Materials and Methods" section describes the data used, baseline approach, proposed methodology, and the ablation study conducted. The "Discussion" section is devoted to the analysis of the results obtained. The "Conclusion" section summarizes the paper.

**Materials and Methods**

Data description

For the XPCT procedure, the OB sample was preliminarily fixed in formalin, then dehydrated and embedded in paraffin blocks. Then the sample was imaged using XPCT, resulting in a series of  $2461 \times 2461$  2D tomographic slices [13]. We denote the original dataset, presented in paper [13], as  $\mathcal{D} = \{S_i\}_{i=1}^N$ , comprising  $N$  consecutive slices  $S_i$  of the single OB sample. Fig. 2a shows a 3D visualization of the OB volume, which we created by stacking  $N = 300$  sequential tomographic slices using the Vedo framework [16].

To decrease computational complexity and training time, expert-defined  $864 \times 864$  regions of interest (ROIs) were extracted from each slice and used for model learning. An example of a slice with overlaid ROI is demonstrated in fig. 2b.

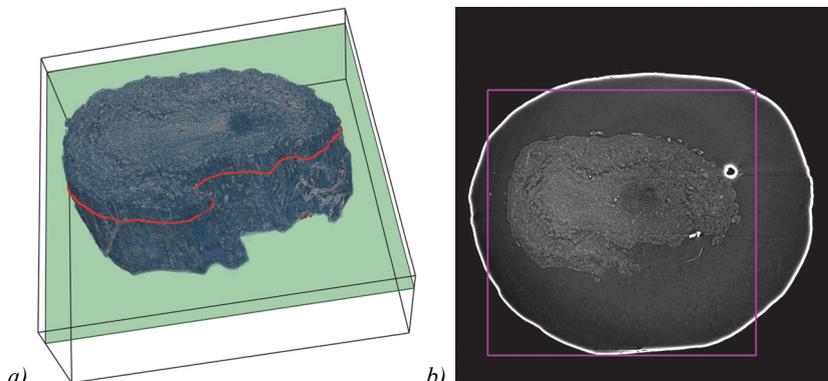


Fig. 2. Visualization: (a) 3D-volume of the OB from 300 consecutive slices; (b) OB slice with overlaid  $864 \times 864$  ROI boundaries (purple)

The OB has a multilayered morphology, including the olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), and granule cell layer (GCL). Table 1 shows the distribution of pixel classes in the OB layers.

Tab. 1. The distribution of pixel classes for olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), and granule cell layer (GCL), %

<b>Id</b>	<b>Class</b>	<b>Median</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Std</b>
0	ground	66.69	64.18	69.74	66.92	1.61
1	ONL	3.13	1.51	4.37	2.99	0.76
2	GL	14.64	11.09	15.88	14.49	0.97
3	EPL	6.88	5.73	9.31	6.98	0.84
4	MCL	3.66	2.61	6.61	4.04	1.06
5	GCL	3.16	2.63	4.26	3.23	0.45
255	ignored	0.63	0.26	1.78	0.66	0.35

Fig. 3a presents an example of the OB slice with the superimposed boundaries of these five layers. The bright white curve on the slice corresponds to the boundary between the paraffin and the surrounding air.

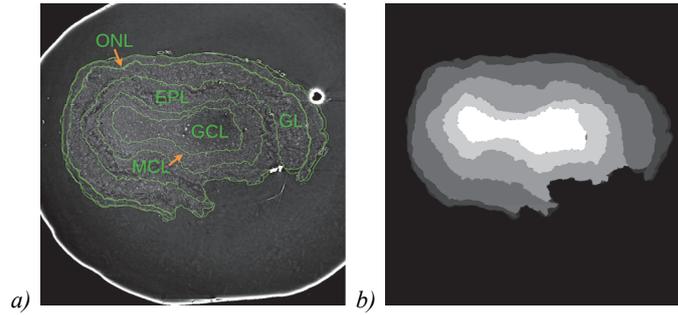


Fig. 3. Visualizations of the ROI example: (a) slice with layer boundaries; (b) the GT for this slice

GT includes five classes that correspond to the OB layers (see fig. 3b), as well as a background class labeled 0 and an "ignore" class labeled 255. For convenience, the OB layer classes are labeled from 1 to 5 towards the central OB layer.

Due to the indistinctness of the boundaries between layers on individual slices, GT for dataset  $\mathcal{D}$  was made by the expert as follows. The initial dataset  $\mathcal{D}$  splits into  $K = 30$  non-intersecting chunks:  $\mathcal{D} = \coprod_{i=1}^K \mathcal{C}_i$ . Define the chunk  $\mathcal{C}_i = \{S_j\}_{j=10i-9}^{10i}$ ,  $i \in [1,30]$ ,  $S_j \in \mathcal{D}$  as ten consecutive slices.

For each chunk  $\mathcal{C}_k$ , a synthetic image  $M_i$  is obtained as an element-by-element inter-slice maximum intensity:

$$M_i(x, y) = \max_{S_j \in \mathcal{C}_i} S_j(x, y),$$

where  $(x, y)$  denote the pixel coordinates.

GT  $G_i$  was made based on  $M_i$  and extended to all slices of the corresponding ten  $\mathcal{C}_k$ . Thus, in addition to possible deviations of layer boundaries from the true layer boundaries, in general, GT made for the synthetic image  $M_i$  may not correspond exactly to any slice in a chunk  $\mathcal{C}_i$ .

GT contains background pixels at the boundaries between layers, so such pixels have been replaced by an "ignore" class that is excluded from the calculation of the loss functions and quality score to avoid influencing the segmentation results.

The intensity values on the slices ranges from  $-1.15 \times 10^{-4}$  to  $3 \times 10^{-3}$ . However, the diagnostically significant information in the slices is concentrated within a narrow subrange of brightness. To isolate the relevant signal from the noise and artifacts, we performed a preliminary linear normalization of the intensity values  $I$ . The acceptable range was defined by expert-selected thresholds, with a lower bound  $L = 5 \times 10^{-4}$  and an upper bound  $U = 1.7 \times 10^{-3}$ .

Analysis of the histogram in fig. 4a indicates that most pixel intensities lie within the range  $[L, U]$ . The extreme values outside this range correspond to low-intensity noise at the bottom and high-intensity artifacts at the top, negative values have no physical meaning (see fig. 4b).

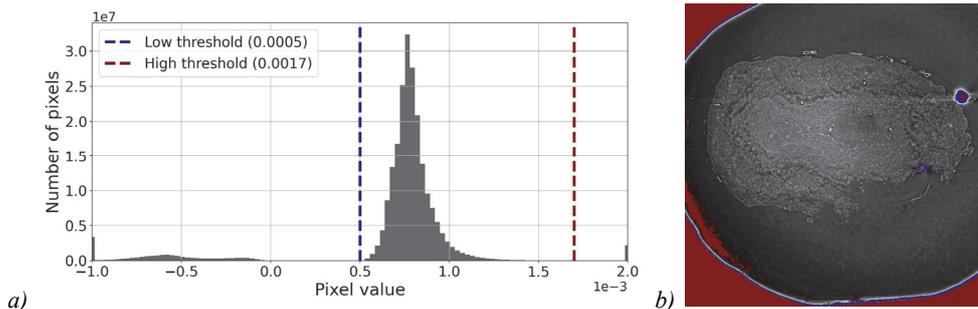


Fig. 4. (a) Histogram of the pixel distribution for  $864 \times 864$  ROIs. The blue and red dashed lines mark the thresholds  $L$  and  $U$ , respectively; (b) Visualization of the OB slice showing pixels with intensity values below  $L$  in blue and above  $U$  in red

Intensity values beyond the range  $[L, U]$  were brought to the nearest boundary:

$$I' = \begin{cases} L, & \text{if } I < L, \\ U, & \text{else if } I > U, \\ I, & \text{otherwise.} \end{cases}$$

The intensity values bounded by  $L$  and  $U$  were linearly normalized to the range  $[0,1]$  using the formula:

$$I = \frac{I' - L}{U - L}.$$

To reduce the computational load, the slices and GT were compressed by half. Then, ROIs of  $864 \times 864$  pixels were extracted from the central part of each.

### Semantic segmentation quality assessment

The Dice score [17], which is widely used in image segmentation tasks, was chosen as the segmentation quality evaluation function. One of the key advantages of the Dice score is its robustness to the class imbalance observed in the dataset under consideration (see Table 1 in the subsection "Data description").

To obtain an overall assessment of the segmentation quality, a micro-averaged version of the Dice score was used [18], in which the quality is calculated without prior averaging at the individual class level. The Dice score with micro-averaging is calculated as the harmonic mean of  $Precision_{micro}$  and  $Recall_{micro}$  with micro-averaging:

$$Precision_{micro} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i)},$$

$$Recall_{micro} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FN_i)},$$

$$Dice_{micro} = \frac{2 \cdot Precision_{micro} \cdot Recall_{micro}}{Precision_{micro} + Recall_{micro}},$$

where  $TP_i$ ,  $FP_i$  and  $FN_i$  are the numbers of true positive, false positive and false negative pixels corresponding to the  $i$ -th class.

### Baseline training pipeline

The paper [13] describes a neural network model training pipeline for automated semantic segmentation of the OB layers using the U-Net [5] architecture.

The data were divided into training, validation, and test subsets randomly in the ratio of 240/30/30. To improve the model's generalization ability, the following augmentations were applied: random mirroring, rotation by a random angle from  $-90$  to  $90$  degrees, and the addition of Gaussian noise with zero mean and standard deviation of  $10^{-4}$ .

The training employed the RMSProp (Root Mean Square Propagation) optimizer with a momentum of 0.9. The learning rate was decreased by a factor of 10 whenever the validation metric stopped improving or the loss plateaued for two consecutive epochs. The authors used L2 regularization with "weight decay" of  $10^{-8}$  and the batch size of 1.

The OB segmentation was performed in two steps. First, the OB was separated from the background with binary segmentation [13]. Next, the resulting binary segmentations (see fig. 5b) were applied to the original slices (see fig. 5a). Slices with masked backgrounds were passed to the model input for layer segmentation (see fig. 5c).

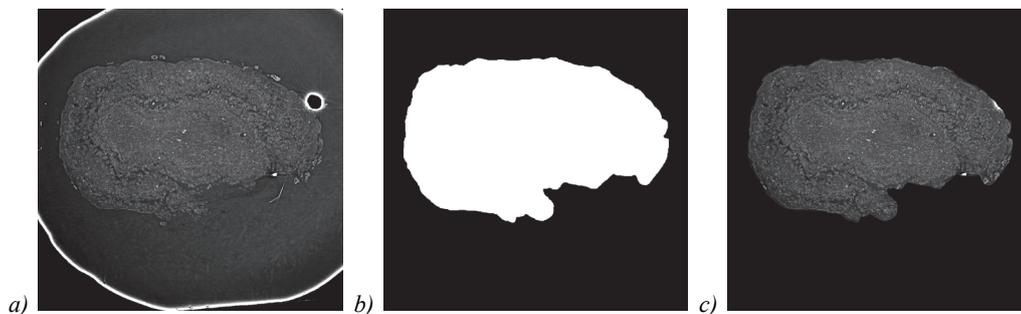


Fig. 5. OB masking: (a) ROI from the original slice; (b) OB binary mask; (c) masking result

### Data splitting strategies

The dataset  $\mathcal{D}$  is split into three disjoint subsets of slices:

$$\mathcal{D} = \mathcal{D}_l \coprod \mathcal{D}_v \coprod \mathcal{D}_t,$$

where  $\mathcal{D}_l$ ,  $\mathcal{D}_v$ , and  $\mathcal{D}_t$  denote the training, validation, and test subsets, respectively, and  $N_l = |\mathcal{D}_l|$ ,  $N_v = |\mathcal{D}_v|$ ,  $N_t = |\mathcal{D}_t|$ .

Let define the data splitting strategies as  $\mathcal{S}_r$ ,  $\mathcal{S}_b$ ,  $\mathcal{S}_e$  and  $\mathcal{S}_n$ , where  $n$  is the number of the starting validation chunk,  $n \in \mathbb{N} \cup \{0\}$ .  $\mathcal{S}_r$  refers to a random split, where slices are assigned to training, validation, and test subsets randomly.  $\mathcal{S}_b$ ,  $\mathcal{S}_e$ , and  $\mathcal{S}_n$  refer to sequential splitting strategies that are designed to minimize the number of pairs of adjacent slices from different subsets and isolate the slices of the test subset from those of the training subset. With sequential splitting, each subset consists of entire chunks:  $K_l$  chunks in the training subset,  $K_v$  in the validation subset, and  $K_t$  in the test subset, so that  $K = K_l + K_v + K_t = 30$ .

In the sequential splitting  $\mathcal{S}_b$ , the test subset occupies the first  $K_t$  chunks of the volume, and the validation subset takes the next  $K_v$  chunks. The remaining  $K_l$  chunks are used for training.

$\mathcal{S}_n$  positions the test subset in the interior of the volume and surrounds it with two equal-sized validation parts, located immediately before and after the test subset.  $K_v$  is required to be even so that both validation parts contain the same number of chunks. All remaining chunks form the training subset.

$\mathcal{S}_e$  places the training subset at the beginning, followed by the validation subset, while the last  $K_t$  chunks are assigned to the test subset.

A formal description of data splitting strategies with a visualization is presented in Table 2.

Tab. 2. Data splitting strategies

Splitting	Visualization	Subsets
$\mathcal{S}_r$		$\mathcal{D}_l = \text{Random}(\mathcal{D}, N_l)$ $\mathcal{D}_v = \text{Random}(\mathcal{D} \setminus \mathcal{D}_l, N_v)$ $\mathcal{D}_t = \mathcal{D} \setminus (\mathcal{D}_l \cup \mathcal{D}_v)$
$\mathcal{S}_b$		$\mathcal{D}_l = \prod_{i=K_t+K_v}^K C_i$ $\mathcal{D}_v = \prod_{i=K_t}^{K_t+K_v-1} C_i$ $\mathcal{D}_t = \prod_{i=0}^{K_t-1} C_i$
$\mathcal{S}_n$		$\mathcal{D}_l = (\prod_{i=0}^{10n-1} C_i) \cup (\prod_{j=10n+K_v+K_t}^{K-1} C_j)$ $\mathcal{D}_v = \mathcal{D} \setminus (\mathcal{D}_l \cup \mathcal{D}_t)$ $\mathcal{D}_t = \prod_{i=10n+\lfloor \frac{K_v}{2} \rfloor}^{10n+\lfloor \frac{K_v}{2} \rfloor + K_t - 1} C_i$
$\mathcal{S}_e$		$\mathcal{D}_l = \prod_{i=0}^{K_l-1} C_i$ $\mathcal{D}_v = \prod_{i=K_l}^{K_l+K_v-1} C_i$ $\mathcal{D}_t = \prod_{i=K_l+K_v}^{K-1} C_i$

**Comparison of random and proposed sequential splitting strategies**

To analyze the applicability of the strategy  $\mathcal{S}_r$  used in [13], we trained two models with the same architecture and hyperparameters as described in the same work. The data were divided into training, validation, and test subsets according to strategies  $\mathcal{S}_r$  and  $\mathcal{S}_b$ .

Following the methodology from [13], we applied the reference OB binary masks, provided in the original publication, to mask the background on the original slices.

Fig. 6 shows plots of the Dice score dynamics for these models during training over 200 epochs.

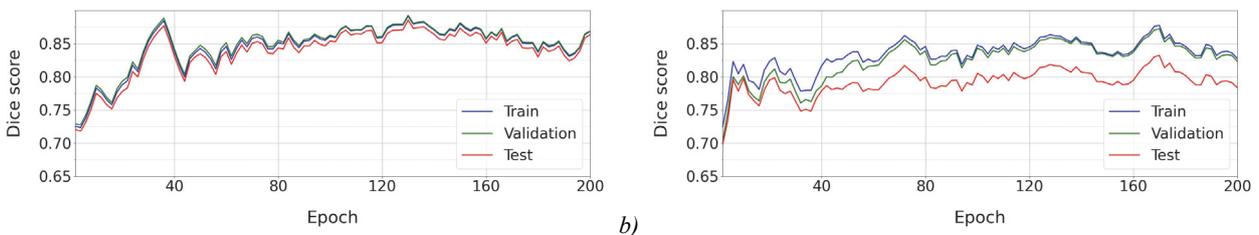


Fig. 6. Dynamics of the Dice score during model training with the pipeline from [13] using two data splitting strategies: (a)  $\mathcal{S}_r$  (random); (b)  $\mathcal{S}_b$  (sequential, with the test subset located at the beginning of the volume)

The plots show that with  $\mathcal{S}_r$  splitting (see fig. 6a), the quality scores for the training, validation, and test subsets are almost the same. In contrast, for  $\mathcal{S}_b$ , the quality for the training and test subsets has significant differences (see fig. 6b), reflecting greater independence between subsets in sequential splitting. This independence makes  $\mathcal{S}_b$  preferable to  $\mathcal{S}_r$ .

The model trained according to the pipeline, described in [13], with  $\mathcal{S}_b$  is used as the baseline in subsequent experiments.

**The proposed training pipeline**

In this paper, we propose a neural network training pipeline using  $\mathcal{S}_b$  splitting into training, validation, and test subsets in the ratio of 240/30/30. A modified version of U-Net is used as the base architecture (see fig. 7).

The classical U-Net [5] architecture has a symmetric structure, including the downstream (encoder) and upstream (decoder) parts, between which information is transferred through connections between layers of the same level (skip-connections). In the proposed modification, each encoder block utilizes residual blocks with residual connections [19]

that bypass two convolutional layers within the block. Residual connections prevent gradient vanishing and improve information transfer from earlier layers.

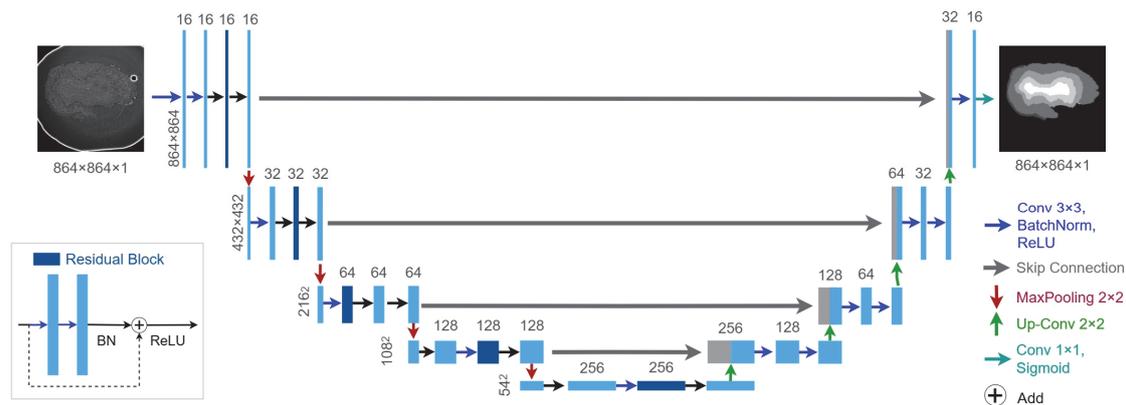


Fig. 7. Employed U-Net with residual blocks

During training, a fixed set of augmentations is applied to the "on-the-fly" data, including specular reflections, random rotations ranging from  $-180$  to  $180$  degrees, and elastic deformations – random smoothed local pixel displacements that mimic natural variations in biological structures. The network parameters are optimized using the Adam algorithm with a  $5 \times 10^{-4}$  learning rate. Since a pronounced class imbalance characterizes the data, the focal loss function [20], which enhances the contribution of rare classes and reduces the influence of dominant classes, is chosen. The proposed pipeline assumes that the model is trained for 1000 epochs with a batch size of 2.

To reduce high-frequency noise, the original tomographic slices were preprocessed with inter-slice smoothing, which averages data using a sliding window size of  $1 \times 1 \times 5$ .

Due to the visual similarity between neighboring slices, only the fourth and seventh slices from each chunk are used for training. These slices are separated by two intermediate slices, which reduces excessive correlation in the training subset. Segmentation quality assessment was performed on all slices, achieving a Dice (micro) score of 0.9325.

Class-specific Dice scores for the background and OB layers are presented in Tab. 3.

Tab. 3. Dice score for different data subsets and segmentation classes

Subset	Background	ONL	GL	EPL	MCL	GCL
train	0.993	0.936	0.975	0.954	0.943	0.972
validation	0.992	0.862	0.941	0.891	0.904	0.942
test	0.991	0.857	0.911	0.846	0.875	0.943

#### Analysis of subset order in sequential data splitting strategy

To evaluate the segmentation performance of the proposed pipeline relative to the baseline across different sequential data splitting strategies, we conducted experiments using a fixed training subset of 240 slices, a validation subset of 20 slices, and a test subset of 40 slices.

Tab. 4 compares the Dice Micro scores for the baseline and proposed pipelines across sequential splittings, with the last column ( $\Delta$ ) representing the difference between them.

The results demonstrate that, regardless of the specific sequential splitting strategy, the proposed pipeline consistently outperforms the baseline, yielding an improvement in the Dice score. Moreover, it shows greater robustness to the position of the test subset, as evidenced by the minimal variation in its performance. In contrast, the baseline model exhibits a high dependency on the test subset location, with its Dice score ranging from 0.754 to 0.942.

#### Ablation study

To evaluate the contribution of each improvement to the final segmentation quality, we conducted an ablation study comprising four experiments, each with one improvement removed from the proposed pipeline:

1. OB pre-masking using reference binary masks.
2. Training the model on initial slices, without prior inter-slice smoothing.
3. Training the model on all slices from each chunk of the training subset. To ensure comparable weight updates when training on subsets of different sizes, the total number of epochs was proportionally adjusted to the size of the training subset. Using a full training subset, for example, training was conducted for 200 epochs.
4. Training using a set of augmentations from the baseline.

All experiments used the  $\mathcal{S}_b$  data splitting strategy.

Tab. 5 reports the Dice scores values for the baseline, the full proposed pipeline, and the training pipelines with one of the improvements disabled.

Tab. 4. Comparison of Dice Micro scores between the proposed and baseline pipelines across sequential data splitting strategies. Blue, green, and red correspond to the training, validation, and test subsets, respectively

	Sequential Splitting	Proposed	Baseline	$\Delta$
$\mathcal{S}_p$		0.928	0.881	0.047
$\mathcal{S}_{00}$		0.940	0.908	0.032
$\mathcal{S}_{01}$		0.942	0.812	0.130
$\mathcal{S}_{02}$		0.937	0.927	0.010
$\mathcal{S}_{03}$		0.934	0.902	0.031
$\mathcal{S}_{04}$		0.929	0.919	0.011
$\mathcal{S}_{05}$		0.931	0.901	0.029
$\mathcal{S}_{06}$		0.945	0.921	0.024
$\mathcal{S}_{07}$		0.938	0.909	0.030
$\mathcal{S}_{08}$		0.938	0.897	0.042
$\mathcal{S}_{09}$		0.930	0.899	0.031
$\mathcal{S}_{10}$		0.935	0.754	0.181
$\mathcal{S}_{11}$		0.943	0.909	0.033
$\mathcal{S}_{12}$		0.947	0.840	0.107
$\mathcal{S}_{13}$		0.950	0.942	0.007
$\mathcal{S}_{14}$		0.950	0.941	0.009
$\mathcal{S}_{15}$		0.948	0.893	0.055
$\mathcal{S}_{16}$		0.952	0.933	0.019
$\mathcal{S}_{17}$		0.952	0.923	0.030
$\mathcal{S}_{18}$		0.953	0.886	0.067
$\mathcal{S}_{19}$		0.952	0.908	0.044
$\mathcal{S}_{20}$		0.954	0.894	0.061
$\mathcal{S}_{21}$		0.960	0.841	0.119
$\mathcal{S}_{22}$		0.960	0.882	0.078
$\mathcal{S}_{23}$		0.961	0.826	0.135
$\mathcal{S}_{24}$		0.958	0.918	0.040
$\mathcal{S}_e$		0.957	0.830	0.127

Tab. 5. Ablation study results

Improvement	Baseline	Proposed	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Pre-masking	✓	×	✓	×	×	×
Inter-slice smoothing	×	✓	✓	×	✓	✓
Sparse train	×	✓	✓	✓	×	✓
Alternative augmentations	×	✓	✓	✓	✓	×
Dice score (micro)	0.8620	<b>0.9325</b>	0.9159	0.9234	0.9164	0.9321

As shown in Tab. 5, the proposed improvements substantially enhanced the segmentation quality. Compared to the baseline, the Dice score increased from 0.8620 to 0.9325, an absolute gain of 0.07.

The largest decrease in quality occurred when the OB pre-masking step was omitted (Exp. 1), with Dice score dropping by 0.0166. This indicates that this step is optional in the proposed pipeline.

Training on the original slices (Exp. 2) results in a Dice score of 0.9234, confirming the importance of the data preprocessing step involving inter-slice smoothing. Including extreme slices from each chunk of the training subset, which weakly correspond to GT (Exp. 3), reduces the segmentation quality to 0.9164, a decrease of 0.0161. Replacing the set of augmentations in Exp. 4 with ones from the baseline leads to a negligible decrease in the Dice score.

Thus, the proposed training pipeline enhances the segmentation quality over the baseline, with each improvement contributing to the final segmentation quality.

### Discussion

In this study, we analyzed a single sample of the human olfactory bulb, which is described in detail in [13]. Due to the limited size and specificity of the dataset, many proven methods and techniques for improving the quality of neural network segmentation are either not suitable or require additional adaptation and tuning.

Promising directions for future research include selection of ROI size, refinement of the intensity normalization method, and comprehensive evaluation of subsets ordering in the sequential data splitting strategy.

A logical continuation of our work is a transition to 3D-segmentation neural networks [7], which have shown great progress in biomedical research in recent years. Typical implementations process input tensors of size  $64^3$  or larger, and the datasets contain tens or hundreds of different annotated volumes. For the present task, however, the volume of data is extremely limited, necessitating smaller and non-cubic input tensors.

The ablation study revealed that the proposed change to the set of augmentations does not significantly impact the final quality of semantic segmentation. This suggests that the typical computer vision augmentations [21] are unsuitable for this task, since the distortions are of a different nature. The results indicate that the selection of relevant augmentations for this domain remains an open question.

A manual error analysis showed that the quality of the segmentation decreases for the initial and final slices in the chunks compared to the quality indicators for the central slices. The errors are localized mainly on the inner boundaries of the layers. This observation can be explained by the fact that, in the implemented method of obtaining the GT, the boundaries between layers change dramatically between successive slices in neighboring chunks (see fig. 8).

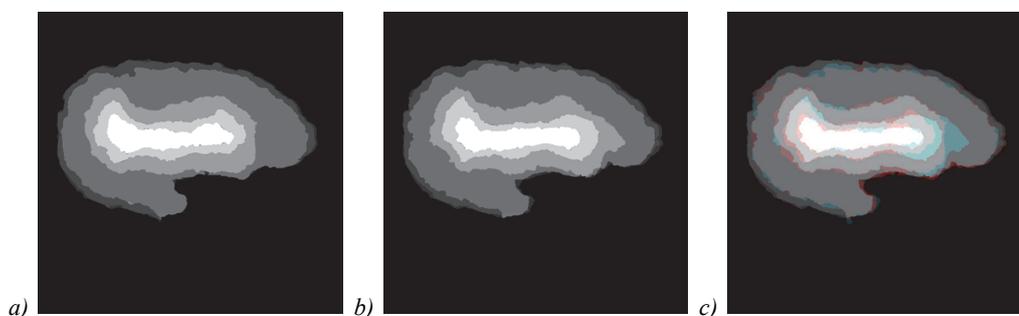


Fig. 8. Visualization of GT segmentation: (a) the GT for slice  $S_{250}$ ; (b) the GT for slice  $S_{251}$ ; (c) the difference map for slices  $S_{250}$  and  $S_{251}$ , where colored regions highlight mismatches (red – unique to slice  $S_{250}$ , blue – unique to slice  $S_{251}$ )

This effect can be mitigated by using an enhanced annotation protocol that incorporates an ignore class. Such modification enables experts to explicitly label regions as belonging to specific OB layers or as representing areas of uncertainty. This scheme more accurately corresponds to the nature of the object under study, since there are no clear boundaries between the layers.

Another promising research direction is organizing a pipeline for the iterative improvement of the GT segmentation. This can be achieved by adapting existing interactive segmentation methods to the specific data type or by involving a human in the process and providing them with visual cues on the model's confidence levels. To improve the quality of segmentation, a refined geometric model of the object can be used in the form of restrictions on the shape of spatial boundaries between layers. Violations of such restrictions may provide additional clues for expert review. One more point to consider is the development of an alternative approach for evaluation methods that do not require the use of GT.

In the long term, improving the method may involve integrating specialized layers. For instance, Hough layers [22] could efficiently extract OB layer boundaries from XPCT images while significantly reducing the number of network parameters. Bipolar morphological layers [23, 24] could be used for computational optimization.

The released dataset and methodology facilitate the subsequent development of enhanced segmentation approaches for tomographic data segmentation.

### Conclusion

This paper introduces an enhanced processing pipeline for semantic segmentation of human OB layers in XPCT reconstructions, building upon the approach from [13]. An important change in the training pipeline was revising the data splitting strategy to reduce the correlation between subsets for a more objective assessment of the segmentation quality. Unlike the pipeline mentioned above, the proposed pipeline uses a one-step approach to directly segment the OB layers without background masking. To account for the data's unique characteristics, we propose preprocessing via inter-slice brightness smoothing and sparse training, using only central slices from each chunk in the training subset. The proposed training pipeline increases the test set's Dice score (micro) value from 0.86 (baseline) to 0.93. An ablation study showed that each of these changes contributes to this increase.

### Data and code availability

The XPCT Human Olfactory Bulb Dataset (XPCT-HOBD), which includes raw slices and ground truth segmentations, has been published on Zenodo along with a pretrained neural network segmentation model and scripts for data preprocessing and inference, all freely available for download: <https://doi.org/10.5281/zenodo.15626069>.

### Ethics approval and consent to participate

The study was carried out on autopsy material obtained from the collection of the Avtsyn Research Institute of Human Morphology of the Federal State Budgetary Scientific Institution "Petrovsky National Research Centre of Surgery" (Moscow, Russian Federation). All protocols were approved by the Ethical Committee of the Research Institute of Human Morphology of the Russian Academy of Medical Sciences (now Avtsyn Research Institute of Human Morphology of Federal State Budgetary Scientific Institution "Petrovsky National Research Centre of Surgery") (No. 6A of October 19, 2009) and are in correspondence with instructions of the Declaration of Helsinki including points 7-10 for human material from 12.01.1996 with the last amendments from 19.12.2016.

### References

- [1] Ruan Y, Zheng X-Y, Zhang H-L, Zhu W, Zhu J. Olfactory dysfunctions in neurodegenerative disorders. *Journal of Neuroscience Research*. 2012; 90(9): 1693–1700. DOI: 10.1002/jnr.23054.
- [2] Stoyanov G, Petkova L, Dzhenev D, Sapundzhiev N, Todorov I. Gross and histopathology of COVID-19 with first histology report of olfactory bulb changes. *Cureus*. 2020; 12: e11912. DOI: 10.7759/cureus.11912.
- [3] Xu Y, Quan R, Xu W, Huang Y, Chen X, Liu F. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering*. 2024; 11(10): 1034. DOI: 10.3390/bioengineering11101034.
- [4] Reza A, Aghdam EK, Rauland A, Jia Y, Avval AH, Bozorgpour A, Karimjafarbigloo S, Cohen JP, Adeli E, Merhof D. Medical image segmentation review: The success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2024; 46(12): 10076–10095. DOI: 10.1109/TPAMI.2024.3435571.
- [5] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. 2015 *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2015; 234–241. DOI: 10.1007/978-3-319-24574-4\_28.
- [6] Paringer R, Mukhin A, Ilyasova N, Demin N. Neural networks application for semantic segmentation of fundus. *Computer Optics*. 2022; 46(4): 596–602. DOI: 10.18287/2412-6179-CO-1010.
- [7] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. 2021; 18(2): 203–211. DOI: 10.1038/s41592-020-01008-z.
- [8] Isensee F, Wald T, Ulrich K, Baumgartner M, Roy S, Maier-Hein K, Jaeger PF. nnU-Net revisited: A call for rigorous validation in 3D medical image segmentation. 2024 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024; 488–498. DOI: 10.1007/978-3-031-72114-4\_47.
- [9] Desser D, Assunção F, Yan X, Alves V, Fernandes FM, Hummel T. Automatic segmentation of the olfactory bulb. *Brain Sciences*. 2021; 11(9): 1141. DOI: 10.3390/brainsci11091141.
- [10] Noothout J, Postma E, Boesveldt S, de Vos B, Smeets P, Išgum I. Automatic segmentation of the olfactory bulbs in MRI. 2021 *Medical Imaging: Image Processing*. 2021. DOI: 10.1117/12.2580354.
- [11] Postma E, Noothout J, Boek W, Joshi A, Herrmann T, Hummel T, Smeets P, Išgum I, Boesveldt S. The potential for clinical application of automatic quantification of olfactory bulb volume in MRI scans using convolutional neural networks. *NeuroImage: Clinical*. 2023; 38:103411. DOI: 10.1016/j.nicl.2023.103411.
- [12] Liu X, Li A, Luo Y, Bao S, Jiang T, Li X, Yuan J, Feng Z. An interactive image segmentation method for the anatomical structures of the main olfactory bulb with micro-level resolution. *Frontiers in Neuroinformatics*. 2023; 17:1276891. DOI: 10.3389/fninf.2023.1276891.
- [13] Meshkov A, Khafizov A, Buzmakov A, Bukreeva I, Junemann O, Fratini M, Cedola A, Chukalina M, Yamaev A, Gigli G, Wilde F, Longo E, Asadchikov V, Saveliev S, Nikolaev D. Deep learning-based segmentation of post-mortem human's olfactory bulb structures in X-ray phase-contrast tomography. *Tomography*. 2022; 8(4): 1854–1868. DOI: 10.3390/tomography8040156.
- [14] Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Igor S. On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Review of Scientific Instruments*. 1996; 66(12): 5486–5492. DOI: 10.1063/1.1146073.
- [15] Arlazarov V, Nikolaev D, Arlazarov V, Chukalina M. X-ray tomography: The way from layer-by-layer radiography to computed tomography. *Computer Optics*. 2021; 45: 897–906. DOI: 10.18287/2412-6179-CO-898.
- [16] Vedo – a module for scientific analysis and visualization of 3D objects. Source: <https://pypi.org/project/vedo/2024.5.1/>.
- [17] Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945; 26(3): 297–302.
- [18] Takahashi K, Yamamoto K, Kuchiba A, Koyama T. Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Applied Intelligence*. 2022; 52(3): 4961–4972. DOI: 10.1007/s10489-021-02635-5.
- [19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016; 770–778. DOI: 10.1109/CVPR.2016.90.
- [20] Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020; 42(2): 318–327. DOI: 10.1109/TPAMI.2018.2858826.
- [21] Gayer A, Sheshkus A, Chernyshova Y. Augmentation of a training sample on the fly for training neural networks [In Russian]. *Trudy ISA RAN*. 2018; 68: 150–157. DOI: 10.14357/20790279180517.
- [22] Sheshkus A, Nikolaev D, Arlazarov VL. HoughEncoder: Neural network architecture for document image semantic segmentation. 2020 *IEEE International Conference on Image Processing (ICIP)*. 2020: 1946–1950. DOI: 10.1109/ICIP40778.2020.9191182.
- [23] Limonova E, Matveev D, Nikolaev D, Arlazarov V. Bipolar morphological neural networks: Convolution without multiplication. 2019 *International Conference on Machine Vision (ICMV)*. 2019. DOI: 10.1117/12.2559299.

[24] Limonova E. Fast and gate-efficient approximated activations for bipolar morphological neural networks. *Journal of Information Technologies and Computing Systems (JITCS)*. 2022; 2: 3–10. DOI: 10.14357/20718632220201.

---

### *Author's information*

**Varvara Alekseevna Karyakina** (b. 2003), was born in Tver, Russia. She has been working as a laboratory assistant programmer at Smart Engines Service LLC, Moscow, since 2024. In 2025, she received a Bachelor's degree in Applied Mathematics from the National University of Science and Technology "MISIS" in Moscow and started pursuing a Master's degree in Computer Science at the Moscow Institute of Physics and Technology (MIPT). Her research interests include computer vision, deep learning, and X-ray tomography. E-mail: [v.karyakina@smartengines.com](mailto:v.karyakina@smartengines.com)

**Dmitry Valerevich Polevoy** (b. 1981), received a master's degree in applied mathematics and physics and a Ph.D. degree in Computer Science from the Moscow Institute of Physics and Technology (MIPT) in 2004 and 2007, respectively. Since 2011, he has been a senior researcher at the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia, in a research group in the field of computer vision. Since 2024, he has been a senior researcher at the Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia, in a research group in the field of computer vision and X-ray tomography. Since 2011, he has been an Associate Professor with the National University of Science and Technology "MISIS". The research interests are computer vision and X-ray tomography. E-mail: [dvpsun@gmail.com](mailto:dvpsun@gmail.com)

**Inna Nikolayevna Bukreeva** (b. 1965), was born in Krasnodar region, Russia. Education: Master's degree in Physics and Engineering, specialization in Solid State Physics. Moscow Engineering Physics Institute (MEPI), 1988. Diploma of Scientific Researcher, specialization in Laser Physics, Moscow Engineering Physics Institute (MEPI). Ph.D in Physics and Mathematics, specialization in Optics, Moscow Engineering Physics Institute (MEPI), 2002. Current Position: Researcher at the Institute of Nanotechnology - CNR, Rome, Italy. Research Focus: X-ray phase contrast tomography applied to biophysics and cultural heritage. Publications: Author of over 100 scientific publications. E-mail: [inna.bukreeva@cnr.it](mailto:inna.bukreeva@cnr.it)

**Olga Andreevna Junemann** (b. 1984), was born in Moscow, Russia. Education: Master's degree in Biology, specialization in Anthropology at Moscow State University in 2009. PhD in Biology, specialization in Histology, Cytology, and Cell Biology at the Research Institute of Human Morphology in 2012. Current position: Researcher at the Avtsyn's Research Institute of Human Morphology (Moscow, Russia) and the Institute of Nanotechnology – CNR (Rome, Italy). Area of interest: brain research in norm and pathology, light and electron microscopy. E-mail: [junemann@outlook.com](mailto:junemann@outlook.com)

**Sergey Vyacheslavovich Saveliev** (b. 1959), was born in Moscow, Russia. Education: Master's degree in Biology and Chemistry at Moscow State University of Education (now the Sholokhov Moscow State University for Humanities). PhD thesis in 1987 in embryology, doctoral thesis in 1992 in embryology, professor title in 2000. Current Position: Head of the Laboratory of Nervous System Development at the Research Institute of Human Morphology, Moscow, Russia. Research Focus: Normal and pathological development of the nervous system. E-mail: [embrains@mail.ru](mailto:embrains@mail.ru)

**Marina Valerevna Chukalina** (b. 1965), was born in Cheboksary, Russia. She received a master's degree in physics and mathematics from the Moscow Physical Engineering Institute, Moscow, Russia, in 1988, and a Ph.D. degree in physics from the Institute of Microelectronics Technology and High Purity Materials, Russian Academy of Sciences, Moscow, in 1997. She has been working with the Institute of Microelectronics Technology and High Purity Materials, Russian Academy of Sciences, since 1988. Now she is a leading researcher of the Federal Research Center "Computer Science and Control", Russian Academy of Sciences. Since 2019, she has been the Head of the X-ray Computer Vision Department at Smart Engines Service LLC, Moscow. She is the author of over 150 papers. Her research interests include the development of signal and image processing tools for X-ray microscopy and tomography. E-mail: [m.chukalina@smartengines.com](mailto:m.chukalina@smartengines.com)

---

*Received June 26, 2025. The final version – November 26, 2025.*

---