Segmentation of 3D volume images for connectomics

CMS Research project

Oleh Shkalikov (5102818)

Supervisors: Jannik Irmai, David Stein
Chairholder: Prof. Dr. Bjoern Andres

TU Dresden, MLCV chair

14 March, 2024

The original dataset

The original image volumes are acquired by a focused ion beam scanning electron microscope (FIB-SEM) of the CA1 hippocampus region of the brain. The raw data has a shape $2048 \times 1536 \times 1065$ and initially were used by [Lucchi et al., 2013, Lucchi et al., 2011] for mitochondria segmentation task.

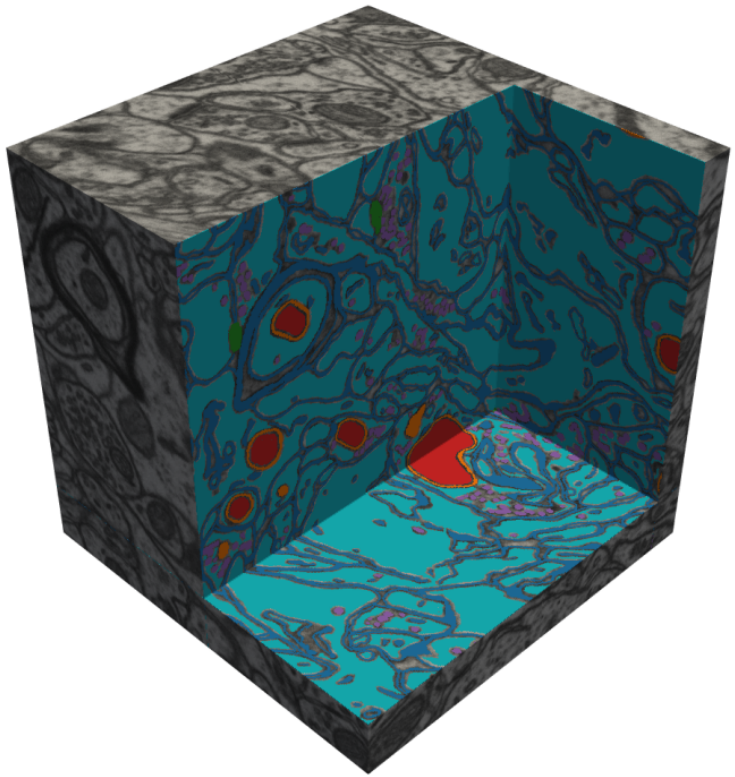

Figure 1: Connectomics volume image with provided labels[a]

[a] Credits to Jannik Irmai

Labeling

The actual labeling has been performed on 9 slices of subvolumes into the following classes:

1. cell cytoplasm
2. cell membrane
3. mitochondrion
4. mitochondrion membrane
5. synapse
6. vesicle
7. undefined (in the case where annotator was uncertain about the correct label)

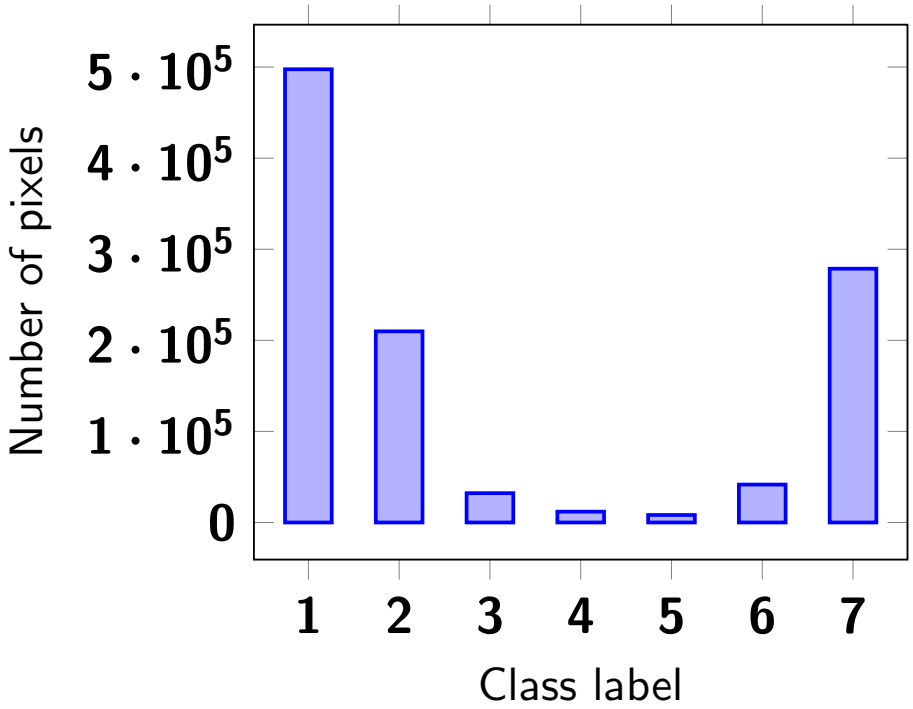

Figure 2: The distribution of labels for the train split

Example of a labeled slice

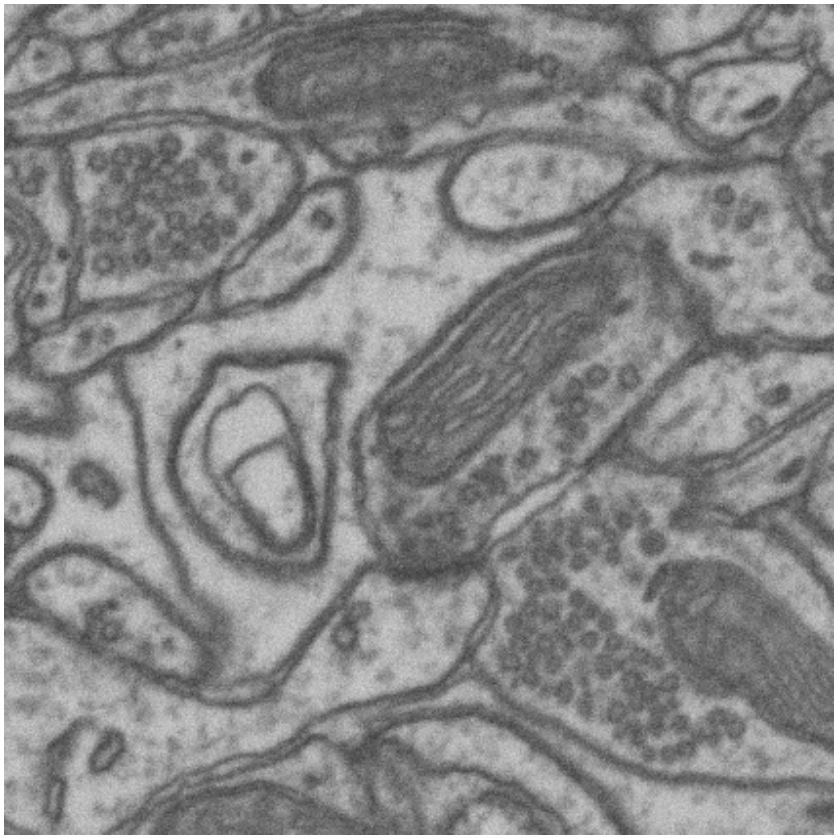

Figure 3: Original XY test slice

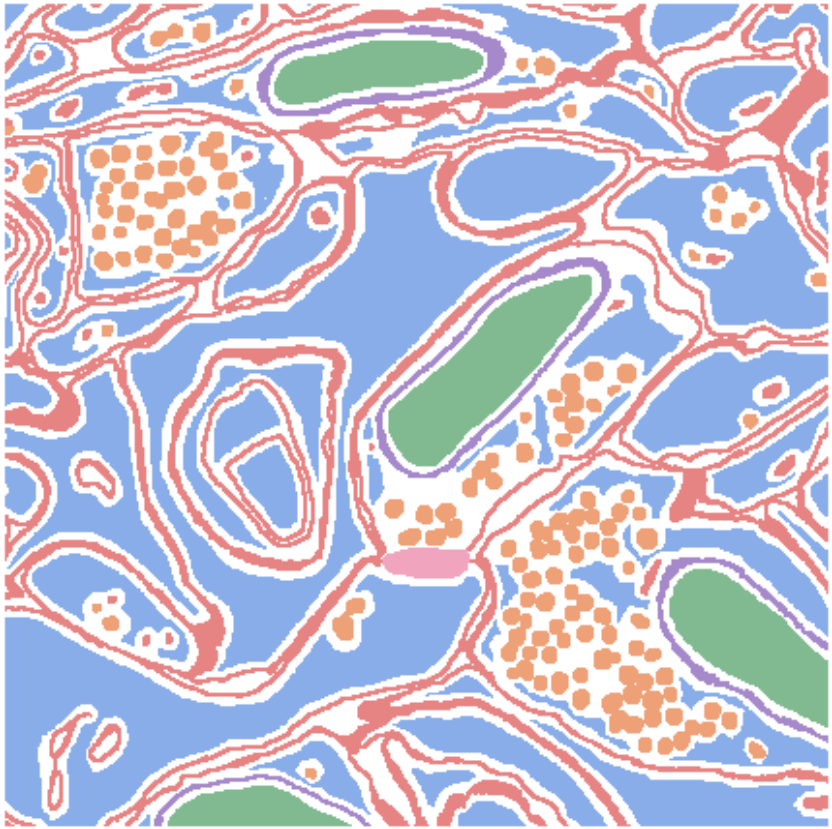

Figure 4: Labels for the test XY slice

Distributions of intensities

Are data normally distributed?

Statistical test of normality fails, so there is no evidence that data are normally distributed

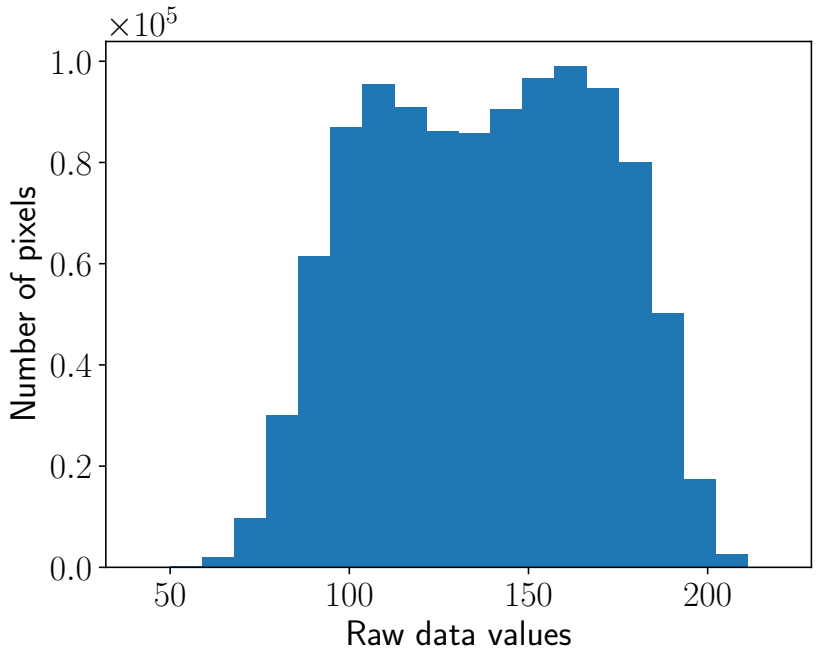

Figure 5: Connectomics volume image with provided labels

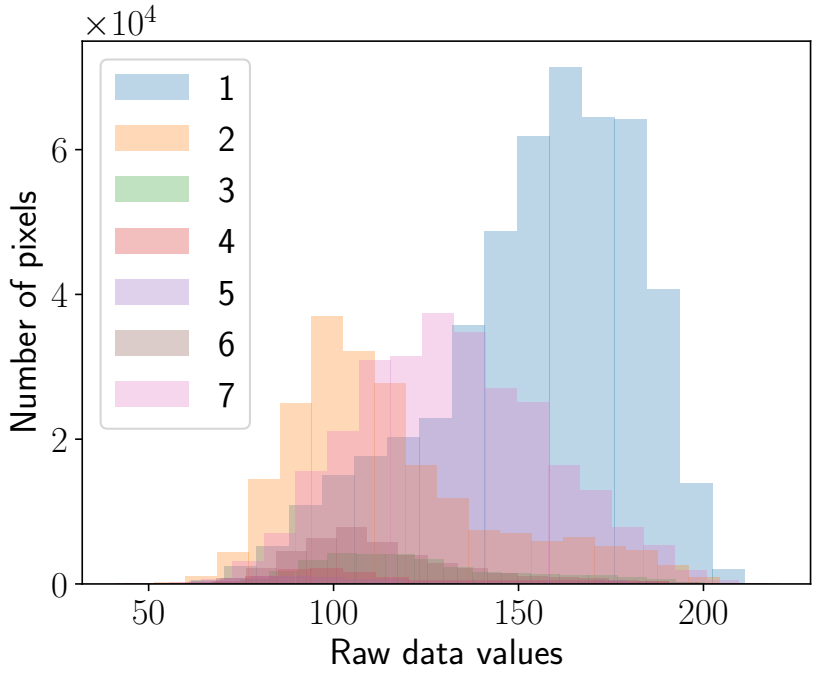

Figure 6: Connectomics volume image with provided labels

Difference in intensities

Means are different

Means of all class except cell membrane and mitochondrion membrane are different with high confidence levels.

Label	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	-	0	0	0	0	0	0
2	0	-	0	0.175	0	0	0
3	0	0	-	0	0	0	0
4	0	0.175	0	-	0	0	0.0003
5	0	0	0	0	-	0	0.0029
6	0	0	0	0	0	-	0
7	0	0	0	0.0003	0.0029	0	-

Table 1: P-values of t-test for different labels combination

Limitations of the dataset

- Labeling by non expert $\Rightarrow$ *there may be some mistakes in labeling*
- Limited size *(even for artificial data generation via GANs)*
- Most interesting and hardest regions are unlabeled
- Labels only for slices $\Rightarrow$ *can not use SOTA segmentation models*
- One label per voxel $\Rightarrow$ *can not perform multi label classification*
- High class imbalance

Voxelwise classification models

First of all we evaluate classical ML models where input subvolumes represent only 1 voxel which model should classify

We use the following ML models:

- Decision tree
- Random forest
- Ada boosting on trees
- Gradient boosting on trees

In addition to intensities we add features computed by filtering and input slice such as:

- Sobel's x and y derivatives
- Prewitt's x and y derivatives
- Laplace
- Gradient magnitudes from sobel and prewitt gradients

Convolutional neural networks

CNN are SOTA models for task like this because they have 2 important advantages over classical models:

1. Downsampling layers reduce dimensionality of features
2. They learn features and we don't need to invent them by hand

All our 6 types of CNNs architectures differ only in backbones.

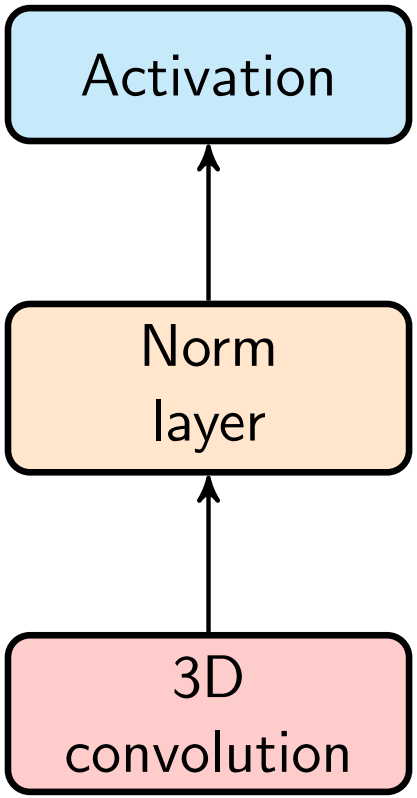

Figure 7: Convolutional block

Handcrafted CNN backbones

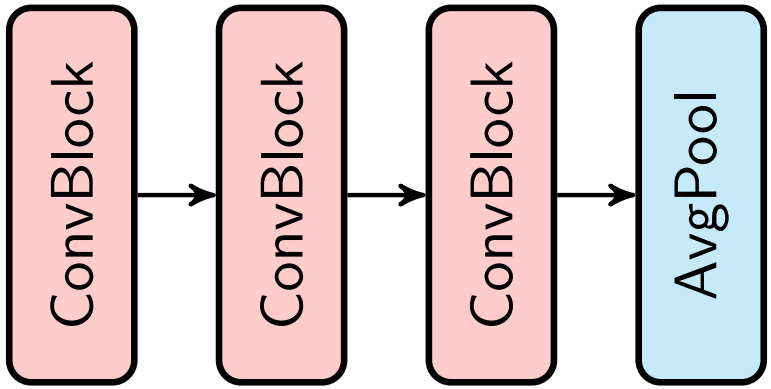

Figure 8: *3layer* architecture. Kernel sizes are $3 \times 3 \times 3$, strides are 1, activations are ReLU and an input shape is $8 \times 8 \times 8$

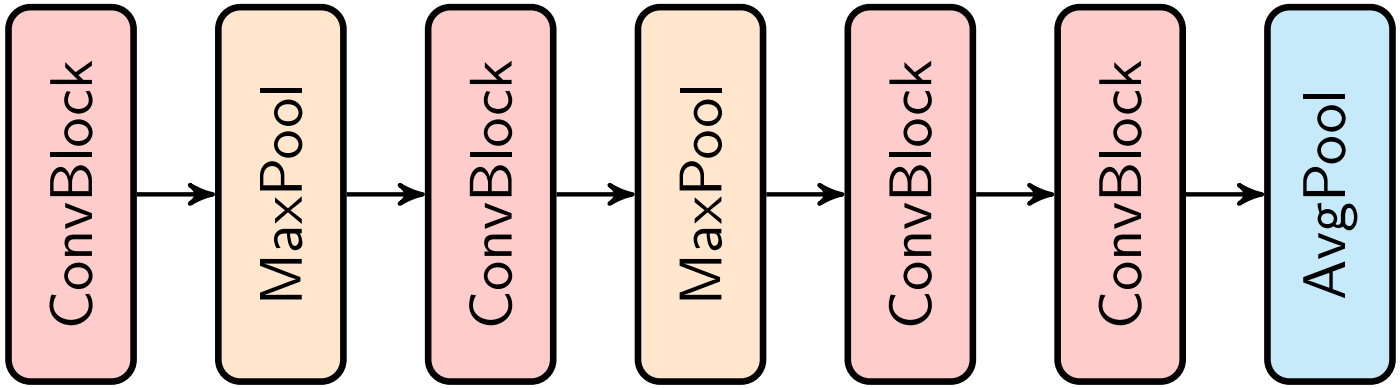

Figure 9: *4conv* architecture. Kernel sizes are $3 \times 3 \times 3$, strides are 1, activations are GELU and an input shape is $32 \times 32 \times 32$

Handcrafted CNN backbones

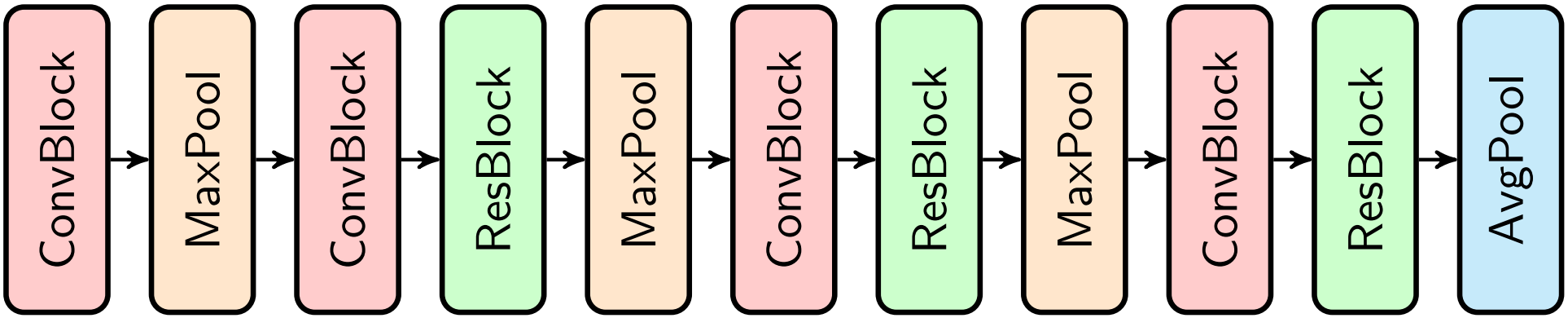

Figure 10: *4res* architecture. Kernel sizes are $3 \times 3 \times 3$, paddings are 1, activations are GELU and an input shapes are $32 \times 32 \times 32$ or $16 \times 16 \times 16$

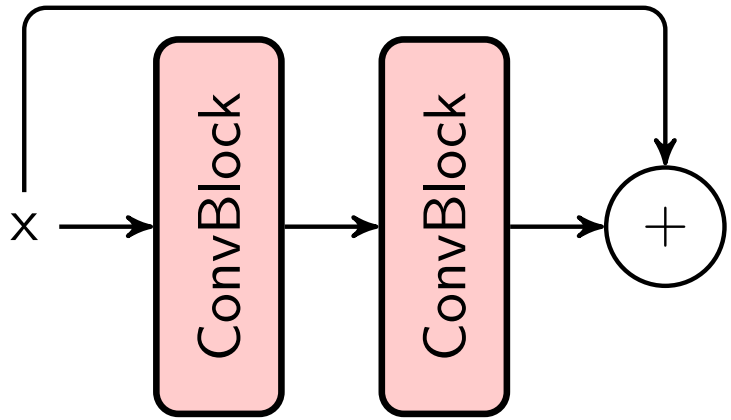

Figure 11: ResNet block of the *4res* architecture

Handcrafted CNN backbones

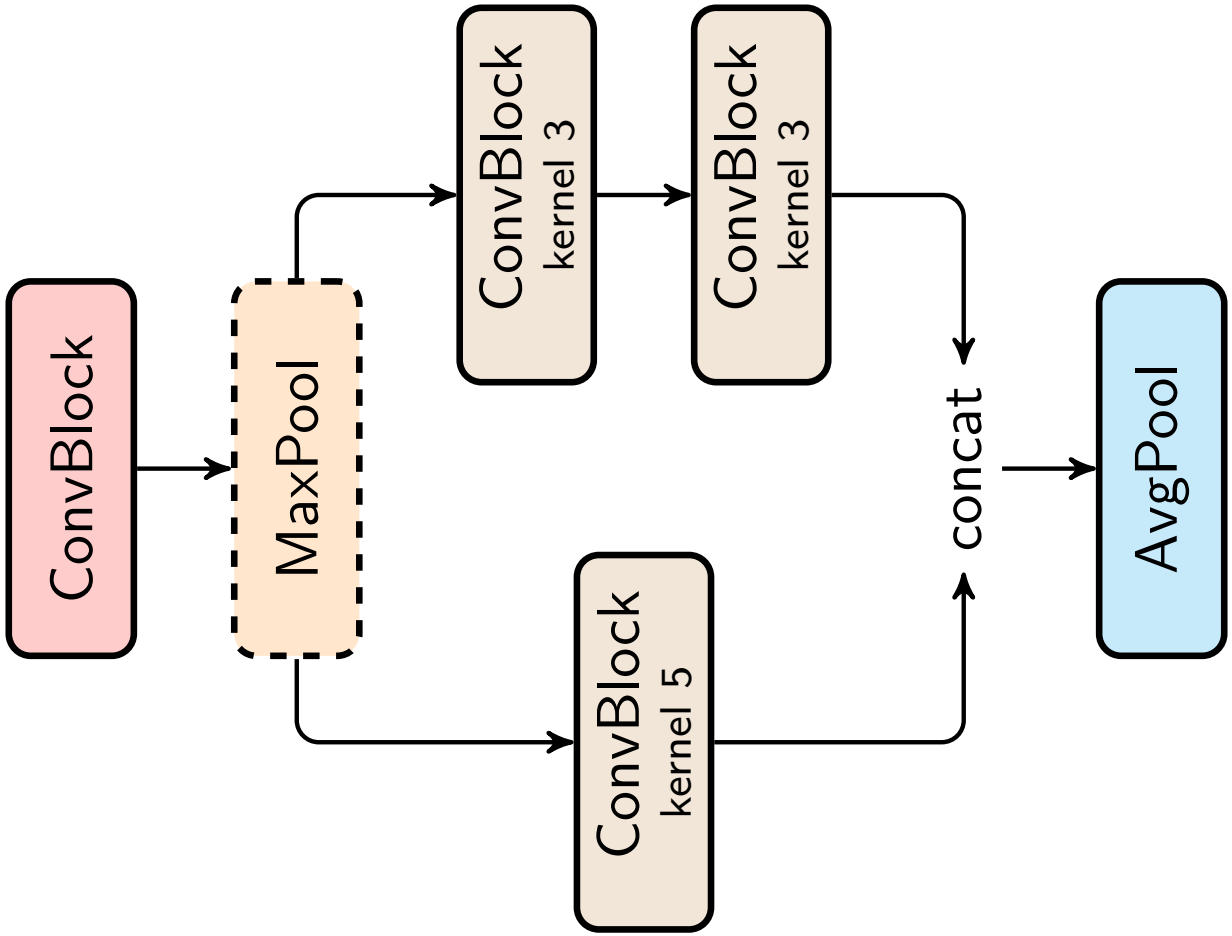

Figure 12: *dil* architectures. Conv blocks in the middle have dilation 3, activations are ReLU and an input shapes are $32 \times 32 \times 32$ or $16 \times 16 \times 16$ (in this case max pooling is omitted)

ConvNext

We took a SOTA 2D CNN ConvNext [Liu et al., 2022] and change 2D convolutions with 3D.

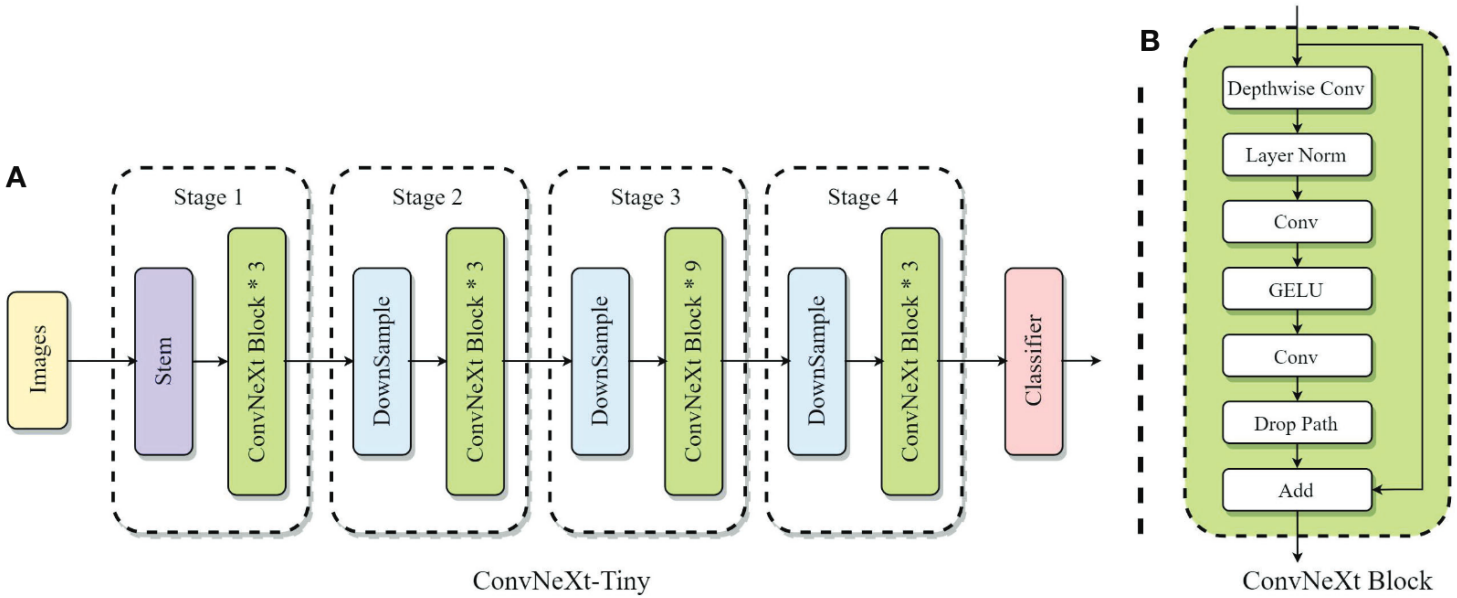

Figure 13: ConvNext-tiny architecture[1]

[1]Image source: https://www.researchgate.net/figure/A-ConvNeXt-Tiny-overall-network-structure-B-ConvNeXt-Block-structure_fig2_367224906

Pretraining techniques

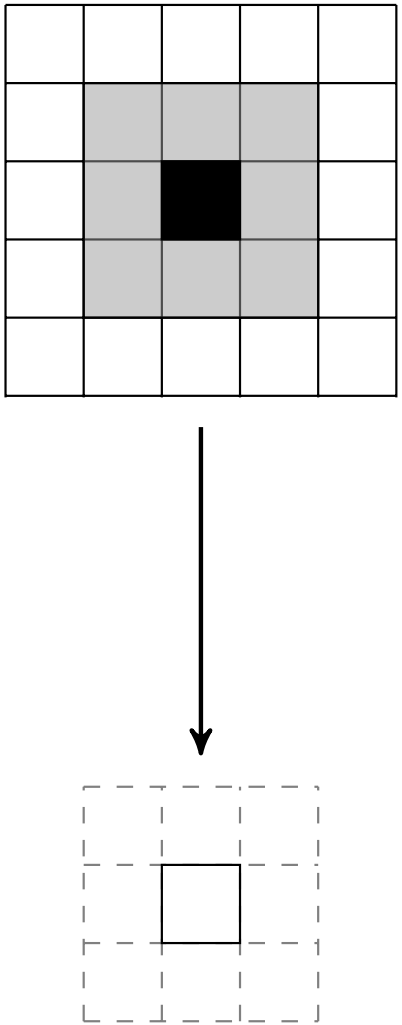

Figure 14: Center region regression

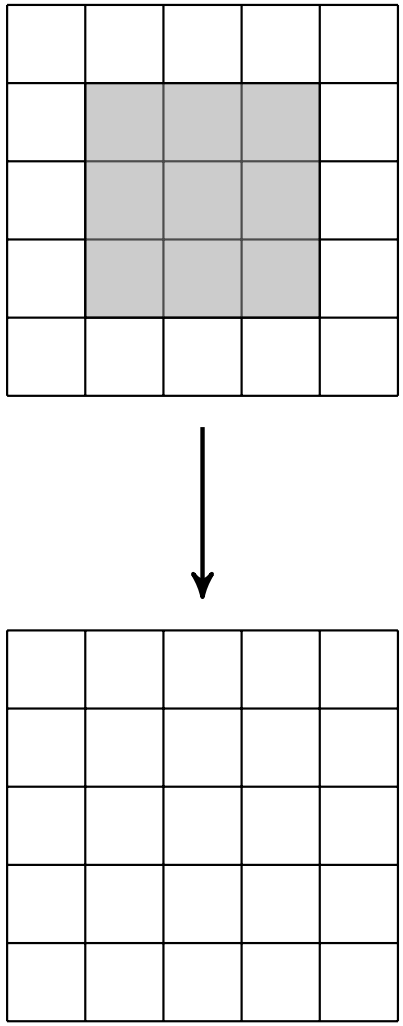

Figure 15: Volume reconstruction via variational auto encoder

Tiled predictions

Motivation

Predictions for adjacent center voxels share a lot of input data. Can we use properties of convolution to reduce number of computation?

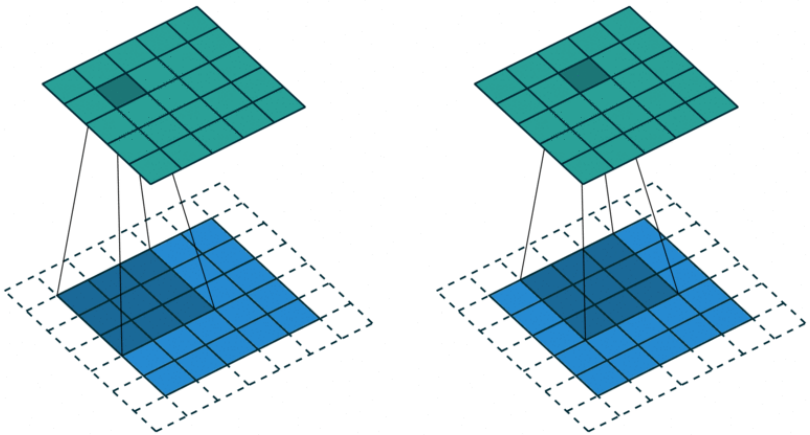

Figure 16: Convolution visualization[a]

Let s and o be an standard input and output shape, m – additional extent, d – downsampling parameter

$$\begin{cases} \frac{s+m}{d} = o + m \\ \frac{s}{d} = o \end{cases} \implies d = 1$$

No downsampling

It is possible iff we don't use any layers with stride > 1

[a] Images source: https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d

Postprocessing with use of CRF

The energy function for CRF is the following:

$$E(\mathbf{x}) = \sum_{i \in V} \theta_i(\mathbf{x_i}) + \sum_{ij \in E} \theta_{ij}(\mathbf{x_i}, \mathbf{x_j})$$

where $G = (V, E)$ – **complete** graph where every voxel $v \in V$ is a node, $\mathbf{x}_i$ – one hot encoding of labeling of the voxel i, $\theta_i = -\ln p_i$ – unary term depended on the predicted probabilities p_i of labels.

$$\begin{aligned} \theta_{ij}(\mathbf{x_i}, \mathbf{x_j}) = & \mu_1(\mathbf{x_i}, \mathbf{x_j}) \exp\left(-\frac{\|c_i - c_j\|^2}{2\theta_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\theta_\beta^2}\right) + \\ & \mu_2(\mathbf{x_i}, \mathbf{x_j}) \exp\left(-\frac{\|c_i - c_j\|^2}{2\theta_\gamma^2}\right) \end{aligned} \tag{1}$$

where $\mu_k(\mathbf{x_i}, \mathbf{x_j})$ – labels incompatibility term, c_i – position of the voxel i, $I_i \in [0, 1]$ – normalized intensity of voxel i, $\theta_\alpha, \theta_\beta, \theta_\gamma$ – parameters of the importance of corresponding difference in intensity and position.

Voxelwise models evaluation

Model	Features	F1	AP	$\varkappa$
Decision tree	1	0.278	0.345	0.656
	8	0.317	0.261	0.584
Random forest	1	0.278	0.345	0.655
	8	**0.319**	0.315	0.618
AdaBoost	1	0.277	0.219	0.654
	8	0.281	0.212	0.671
GradientBoost	1	0.278	0.346	0.656
	8	0.282	**0.355**	**0.676**

Table 2: Metrics for voxelwise models

CNN models evaluation

Model	**Input shape**	**Loss**	**F1**	**AP**	$\varkappa$
4res	32	CE	**0.913**	**0.962**	**0.939**
4conv	32	CE	0.906	0.952	0.938
4res_ln	32	CE	0.905	0.959	0.937
dil	32	CE	0.885	0.95	0.923
4res	16	CE	0.883	0.933	0.926
4res_ln	16	CE	0.851	0.925	0.918
dil	32	Focal ($\gamma = 2$)	0.828	0.886	0.894

Table 3: Metrics for top 7 CNN models

Efficiency of pretraining

Model	Pretr. type	Input shape	Loss	F1
ConvNext	CR1	32	CE	**0.741**
	CR0		CE	0.682
	-		CE	0.701
ConvNext	CR0	32	Focal ($\gamma = 2$)	**0.687**
	-		Focal ($\gamma = 2$)	0.531
	-		Focal ($\gamma = 4$)	0.492
ConvNext	CR1	16	CE	0.603
	-		CE	**0.728**

Table 4: Comparison of ConvNext models with and without pretraining. CRN means center voxel regression with padding equals to N

Quality of postprocessing: metrics

Model	**Input shape**	**Loss**	**F1**		$\mathcal{K}$	
4res	32	CE	0.901	-0.012	**0.937**	-0.002
4conv	32	CE	**0.902**	-0.004	0.935	-0.003
4res_ln	32	CE	0.897	-0.008	0.930	-0.007
dil	32	CE	0.828	-0.057	0.868	-0.064
4res	16	CE	0.878	-0.005	0.922	-0.004
4res_ln	16	CE	0.846	-0.005	0.916	-0.002
dil	32	Focal ($\gamma = 2$)	0.828	0	0.893	-0.001

Table 5: Metrics for postprpocesed outputs of top 7 CNN models. The difference with metrics for raw prediction are denoted right of the metric values

Quality of postprocessing: biological structures

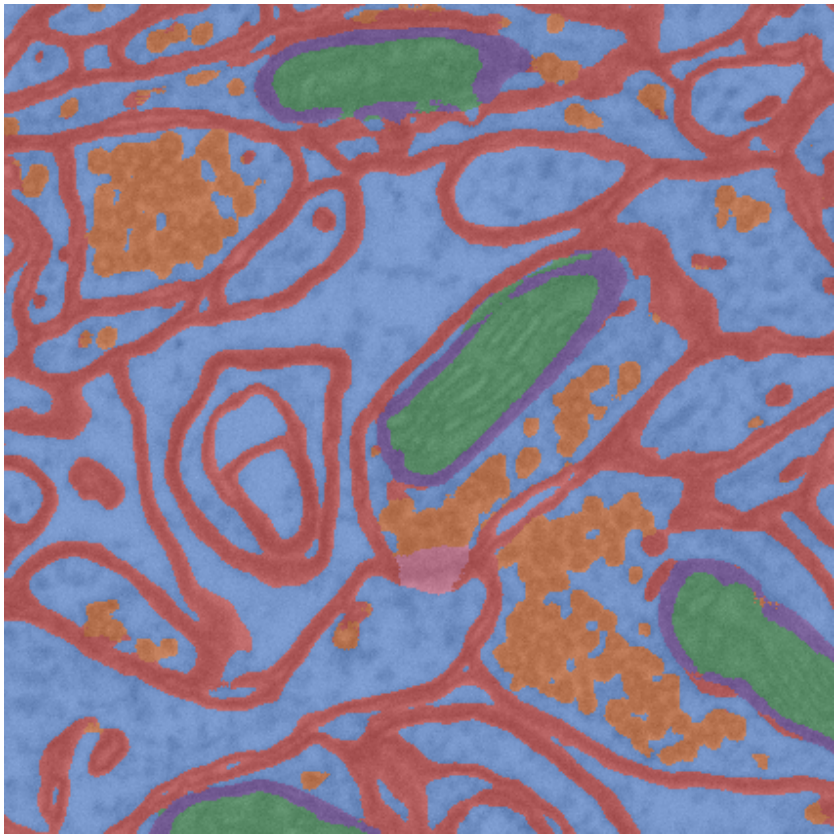

Figure 17: raw predictions

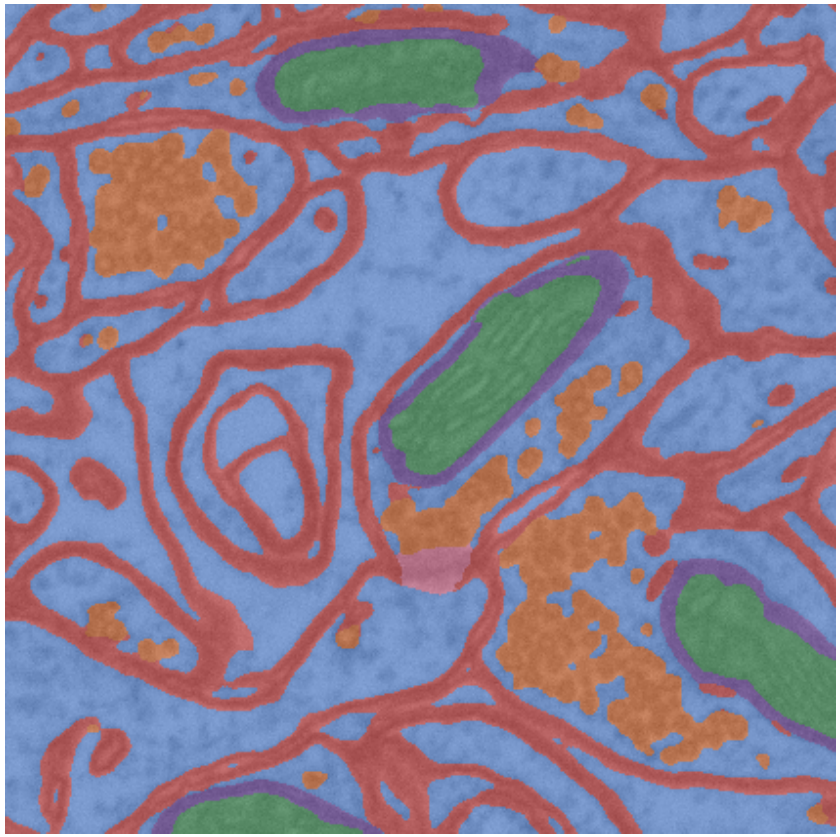

Figure 18: CRF postprpocesed

Context free grammars

CFGs describe structural properties of the text.
For example the grammar for balanced parentheses in the Chomsky normal form is the following:

- $S \to SS|LR$
- $L \to aS|a$
- $R \to Sb|b$
- $a \to ($
- $b \to)$

S					
$\varnothing$	$\varnothing$				
S	$\varnothing$	$\varnothing$			
L	R	$\varnothing$	$\varnothing$		
$\varnothing$	S	$\varnothing$	$\varnothing$	S	
L, a	L, a	R, b	R, b	L, a	R, b
(	(	)	)	(	)

Figure 19: Example of the CYK algorithm for CFG for balanced parentheses

N-dimensional CFGs

[Schlesinger and Hlavác, 2013] proposed an extension of CFGs and CYK for N dimensional case. We can try to use it for connectomics:

- To check the structural validity of the predictions
- For refining / training: let's make a stochastic grammar and in a case of fail backtrace to voxels that broke the structure and correct them

Problem of intractablility

The main problem in this regard is to come up with grammar for connectomics, because the number of rules grows very fast with increasing an input size

Conclusions

- Using of CNNs is reasonable
- Relatively small residual networks work better than bigger models
- The optimal input shape for used models is $32 \times 32 \times 32$
- The postprocessing with use of CRF improves the structural property of predictions
- The computed metrics may not represent a real quality of models because of unlabeled regions in the dataset
- Structural properties of connectomics can help us improve the quality of segmentation

References I

[Liu et al., 2022] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986.

[Lucchi et al., 2013] Lucchi, A., Li, Y., and Fua, P. (2013). Learning for structured prediction using approximate subgradient descent with working sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1987–1994.

[Lucchi et al., 2011] Lucchi, A., Smith, K., Achanta, R., Knott, G., and Fua, P. (2011). Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *IEEE transactions on medical imaging*, 31(2):474–486.

References II

[Schlesinger and Hlavác, 2013] Schlesinger, M. I. and Hlavác, V. (2013). *Ten lectures on statistical and structural pattern recognition*, volume 24. Springer Science & Business Media.