DENOISING DIFFUSION PROBABILISTIC MODELS FOR IMAGE INPAINTING OF CELL DISTRIBUTIONS IN THE HUMAN BRAIN

Jan-Oliver Kropp[1,2], Christian Schiffer[1,2], Katrin Amunts[1,3], Timo Dickscheid[1,2,4]

1 Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Germany
2 Helmholtz AI, Research Centre Jülich, Germany
3 Cécile & Oskar Vogt Institute of Brain Research, University Hospital Düsseldorf, Germany
4 Institute of Computer Science, Heinrich-Heine-University Düsseldorf, Germany

Visit us!

j.kropp@fz-juelich.de

Introduction

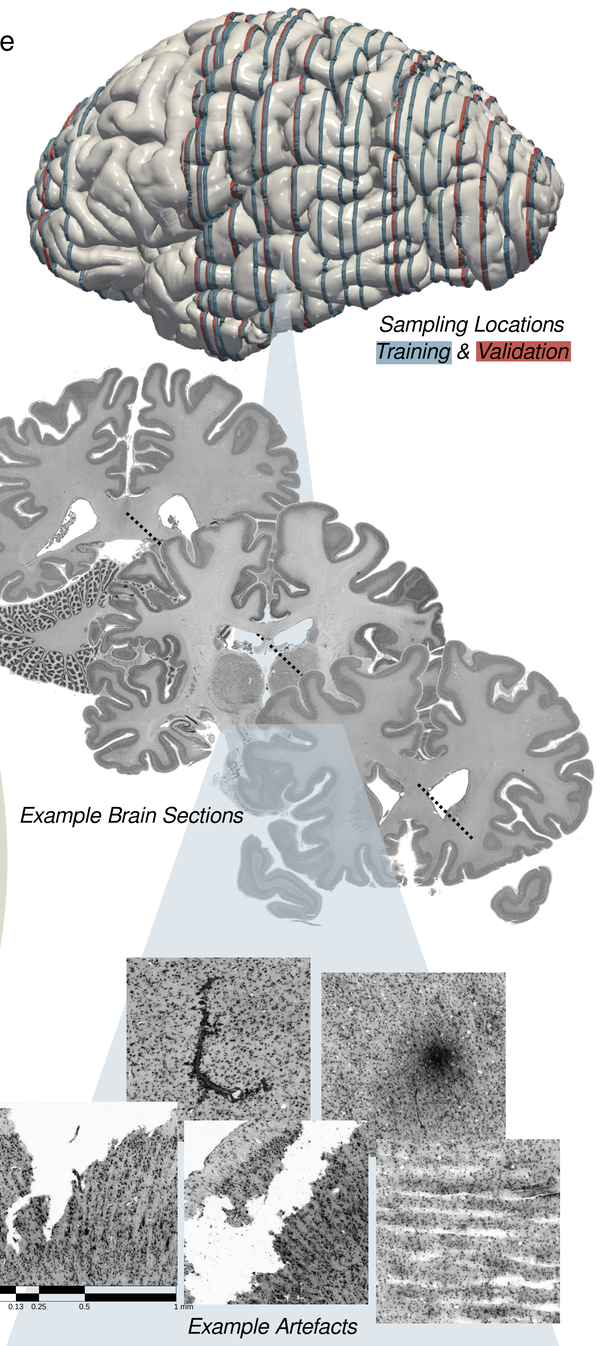
Recent advances in imaging and machine learning enable the imaging of whole human brains at **cellular-level**, where novel methods automate the **cytoarchitectonic mapping** [1] and **detection of cell bodies** [2].
However, the presence of processing induced **artefacts** in the image data caused by missing sections, tears in the tissue or staining errors, need to be addressed to facilitate the analysis.

We propose a method to **fill in missing information** in the high-resolution microscopic scans of histological brain sections by using **denoising diffusion probabilistic models**.

Data and Artefacts

The training and validation data were derived from the **BigBrain** dataset [3] at 1micron resolution. All the data was **screened for artefacts** to ensure **that only intact** data was used to train and validate the model.

64 sections of the BigBrain dataset were used for **training** and **10** sections were used for **validation**.

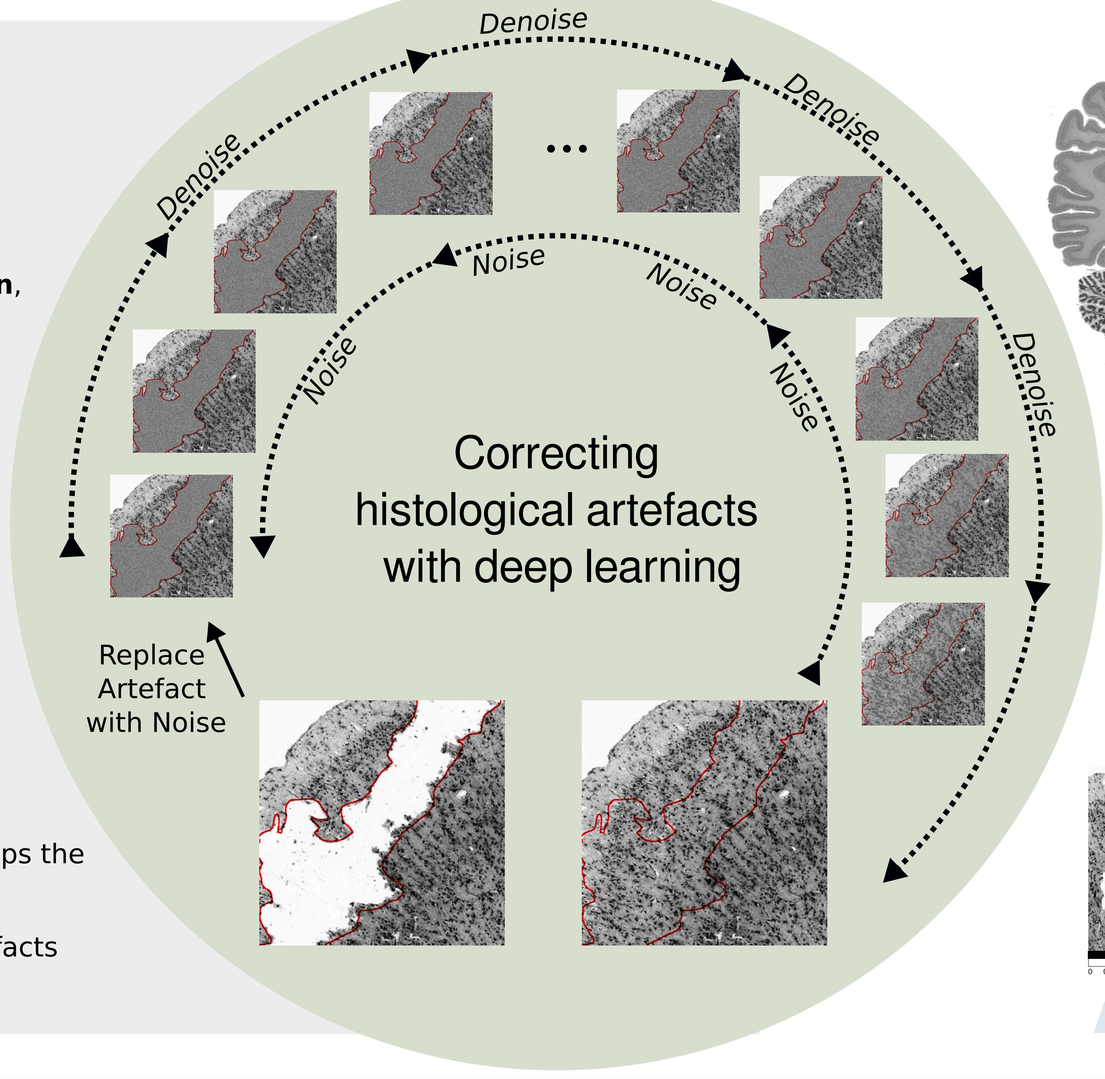


Denoising Diffusion Probabilistic Models

- **DDPM**s [4] map between noise and data distributions by gradually removing noise
- The model architecture closely follows [4] a deep **residual** U-Net with **Self-Attention**, **SiLU** activation functions**, 256** diffusion steps and **a cosine noise schedule**
- Our DDPM was trained on 1.25 million histological image patches of size 1 mm^2 @ 1 micron resolution
- The model was trained for 12 hours on 64 GPUs on JURECA-DC [5].

RePaint

- **RePaint** [6] extends our model to fix artefacts and fill in missing data
- The DDPM gets conditioned on the known part in every diffusion step
- To increase the correlation the methods loops the diffusion steps multiple times
- This method needs annotations of the artefacts (here done by hand)



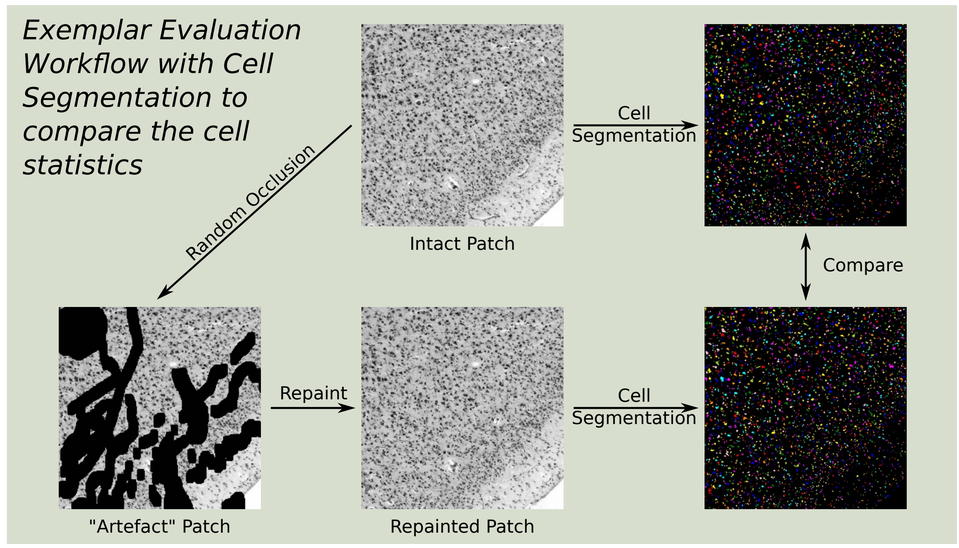
How do we know that the generated data matches the cell distribution?

Evaluating the Cell Distribution with Downstream Models

- Estimating the cell distribution is **intractable**
- Compare characteristics
  - **Density**
  - **Size**
  - **Position**
  - **Arrangement**

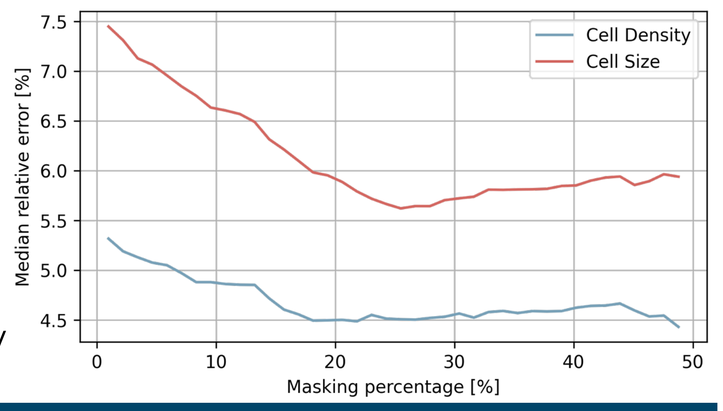
  } **Cytoarchitecture**

- Repaint **known areas** with our diffusion model in **intact patches**
- We evaluated the model on **10,000 patches**
- The repaint areas were randomised ranging from **5 to 50%** of a given patch
- **Established models** are used to extract charactersitics



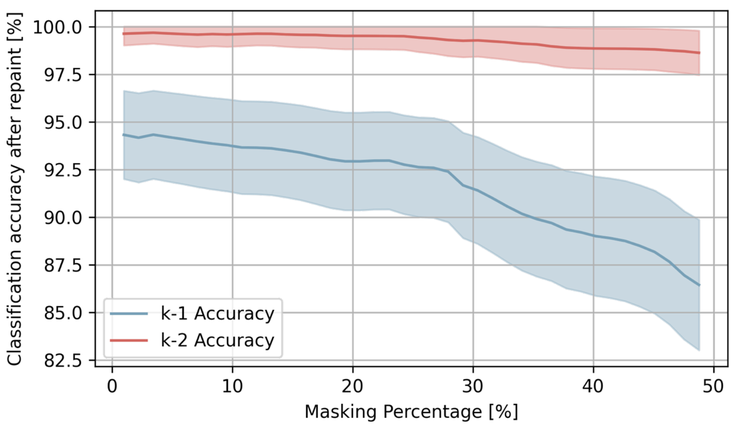
Cell Statistics - Density and Size

- Segment the cells with the CPN [2] and measure **cell density and size**
- Compare the average error in **cell density** and **size** before and after the repaint
- Our model closely follows the distribution with median relative **errors below 8%**
- Relative errors are slightly larger for small regions, as slight inaccuracies are relatively larger



Classification Flips - Cytoarchitecture

- Classify the images using [1] to compare the **cytoarchitecture**
- **Classification flips** occur, when the predicted class changes for the repaint
- Very few classification flips occur with small inpainted regions, increasing for larger areas to ~15%
- When accounting for the second likeliest guess, the accuracy increases to above 98%



Conclusion & Future Work

- We propose an evaluation method for generated **light-microscopy images of human brain sections**
- Our DDPM is able to accurately fill in missing data, within an error margin of **less than 8% for cell statistics**
- We aim to incoporate information from **neighbouring sections** using **crossattention** to improve the classification consistency
- Our models enables an **automatic correction** of processing artefacts at **full 1 mircon resolution**

This project received funding from the European Union's Horizon 2020 Research and Innovation Programme, grant agreement 101147319 (EBRAINS 2.0 Project), and the Helmholtz Association port-folio theme "Supercomputing and Modeling for the Human Brain", and the Helmholtz Association's Initiative and Networking Fund through the Helmholtz International BigBrain Analytics and Learning Laboratory (HIBALL) under the Helmholtz International Lab grant agreement InterLabs-0015, and by HELMHOLTZ IMAGING, a platform of the Helmholtz Information & Data Science Incubator [X-BRAIN, grant number: ZT-I-PF-4-061]. Computing time was granted through JARA on the supercomputer JURECA-DC at Jülich Supercomputing Centre.

[1] Schiffer, C et al. (2021). Contrastive Representation Learning for Whole Brain Cytoarchitectonic Mapping in Histological Human Brain Sections, ISBI
[2] Upschulte, E., Harmeling, S., Amunts, K., & Dickscheid, T. (2022). Contour proposal networks for biomedical instance segmentation. Medical image analysis , _77_, 102371.
[3] Katrin Amunts et al. BigBrain: An Ultrahigh-Resolution 3D Human Brain Model. Science340, 1472-1475(2013)
[4] Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34, 8780-8794.
[5] Jülich Supercomputing Centre. (2021). JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre.
[6] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF CVPR