Leveraging Denoising Diffusion Probabilistic Models for Segmentation in Volume Electron Microscopy

Larissa Heinrich, Stephan Saalfeld

janelia.ai

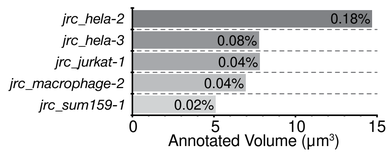
Denoising Diffusion Probabilistic Models (DDPMs) have proven to be powerful generative models for image generation while being simple to train compared to other alternatives such as GANs. We are exploring the application of DDPMs in bioimage segmentation. In particular, we perform our investigation on the example of segmentation of subcellular structures in volume electron microscopy.

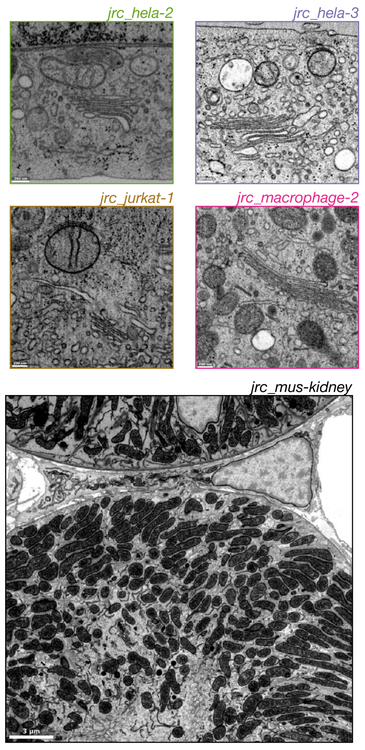
Obtaining specialized training data remains a significant and time-consuming barrier in finetuning networks to new datasets and improving segmentation performance for underrepresented classes. Our work investigates the potential of using DDPMs to supplement training data for segmentation - ranging from additional data augmentation to generating examples with classifier-free guidance under various conditioning signals. Our approach is designed to leverage the abundant unlabeled data during the training process, thereby facilitating domain adaptation and class balancing.

While still in its early stages, our research aims to assess the utility of integrating generative models, specifically DDPMs, in segmentation of subcellular structures in volume electron microscopy. This exploration offers a promising avenue for reducing the human annotation effort in bioimage segmentation.

Challenges in Segmentation of Volume Electron Microscopy

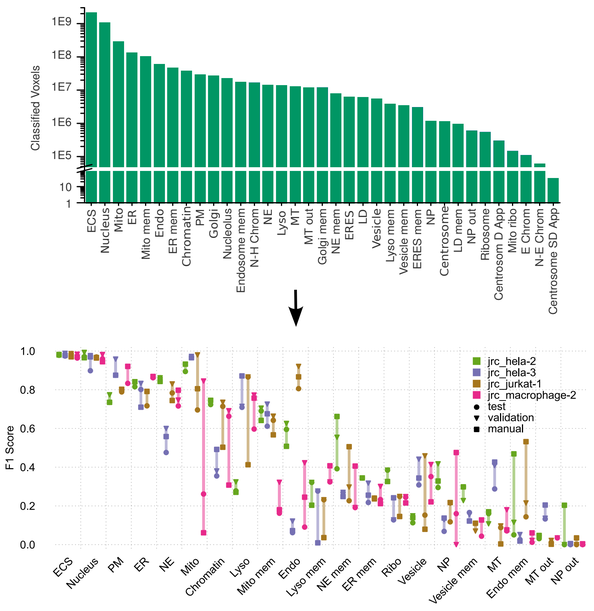
Annotating ground truth in volume EM data is challenging and time-consuming while the data sets are large and diverse. It is therefore impractical to accurately capture the data's variability with ground truth annotation.
Approaches that can leverage the vast amount of unannotated data have the potential to significantly reduce the necessary amount of training data.





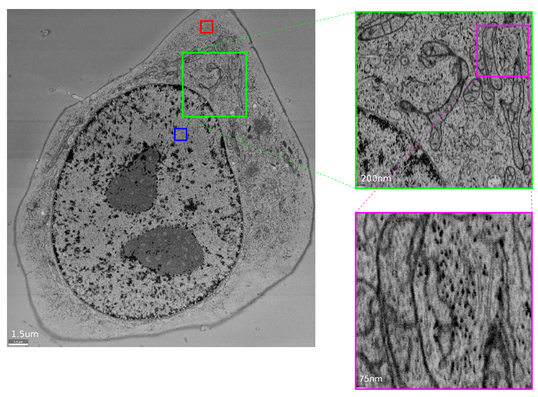
So far, organelle segmentation networks do not work out of the box for data from a somewhat different domain. Many datasets require some additional groundtruth to facilitate finetuning. While, anecdotally, it can be hard to predict whether and how much additional data needs to be annotated for satisfactory performance, the variability between datasets is evident by visual comparison.

So far, we have addressed this problem by aiming to collect a large and diverse set of ground truth. This strategy could be supplemented by domain adaptation methods that can learn something about the peculiarities of new datasets without ground truth. By decreasing the annotation burden for each individual dataset we could increase the diversity of datasets with some ground truth.

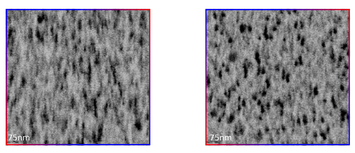


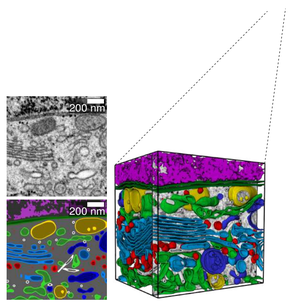
There are a multitude of discernable subcellular structures whose segmentation in volume EM data would be beneficial for biological research. However, these structures occur at vastly different rates, leading to highly unbalanced training data when annotated jointly. The imbalance of such training data affects performance for the less common classes in panoptic segmentation.

The subcellular structures of interest can have very different scales. Further, even for small structures it is often pertinent to know the larger-scale context to unambiguously identify them. Networks that operate across these scales and provide a large receptive field are needed to address this challenge.



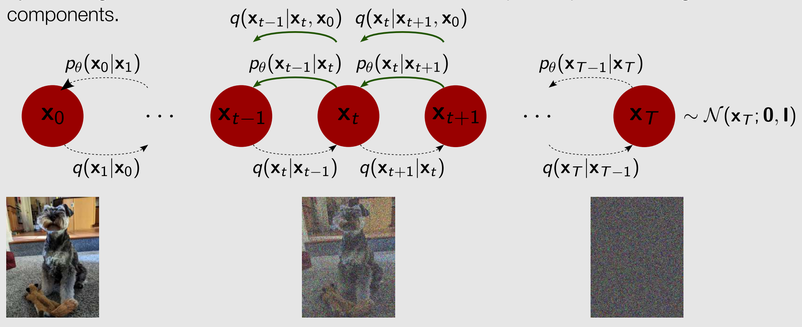
Can you guess which patch was cropped from the nucleus and which one from the cytosol?





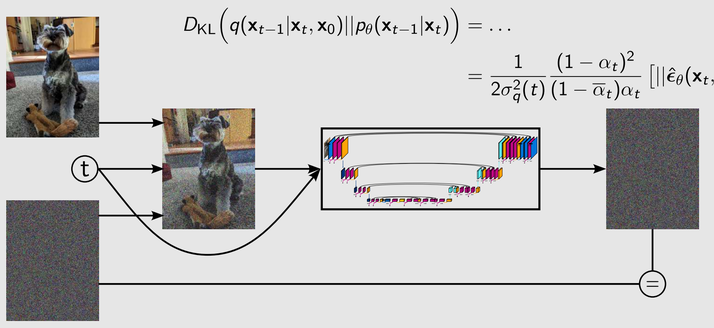
Background: Denoising Diffusion Probabilistic Models

DDPMs are a type of generative network that work by iteratively refining a an initial sample of standard Gaussian noise, progressively reducing noise levels. This corresponds to reversing a forward noise diffusion process. Importantly, the reverse transition probablities can be trained by teaching a network that is conditioned on the timestep to separate the signal and noise components.

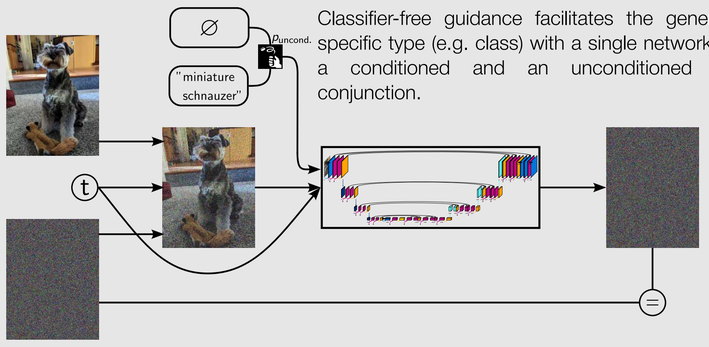


$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I})$$

$$D_{\mathrm{KL}}\Big(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)\Big) = \dots$$
$$= \frac{1}{2\sigma_q^2(t)}\frac{(1-\alpha_t)^2}{(1-\overline{\alpha}_t)\alpha_t}\left[||\hat{\epsilon}_\theta(\mathbf{x}_t, t) - \epsilon_0||_2^2\right]$$



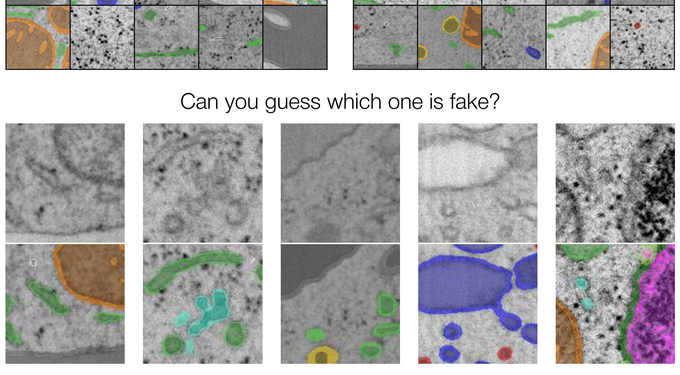
Background: Classifier-free Guidance



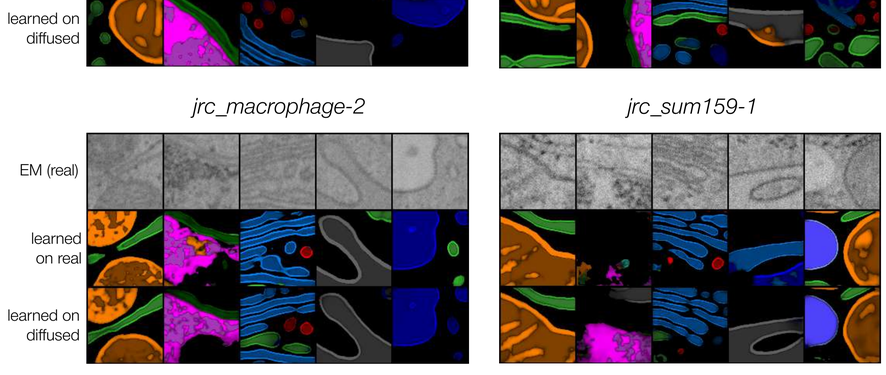
Classifier-free guidance facilitates the generation of a specific type (e.g. class) with a single network by training a conditioned and an unconditioned DDPM in conjunction.

$$\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t|y) = \nabla_{\mathbf{x}_t}\log\frac{p(\mathbf{x}_t)p(y|\mathbf{x}_t)}{p(y)}$$
$$= \nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t) + \underbrace{\gamma}_{\uparrow}\nabla_{\mathbf{x}_t}\log p(y|\mathbf{x}_t)$$
$$\text{class}$$
$$\Longrightarrow \nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t) + \gamma\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t|y) - \gamma\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t)$$
$$= (1-\gamma)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t) + \gamma\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t|y)$$

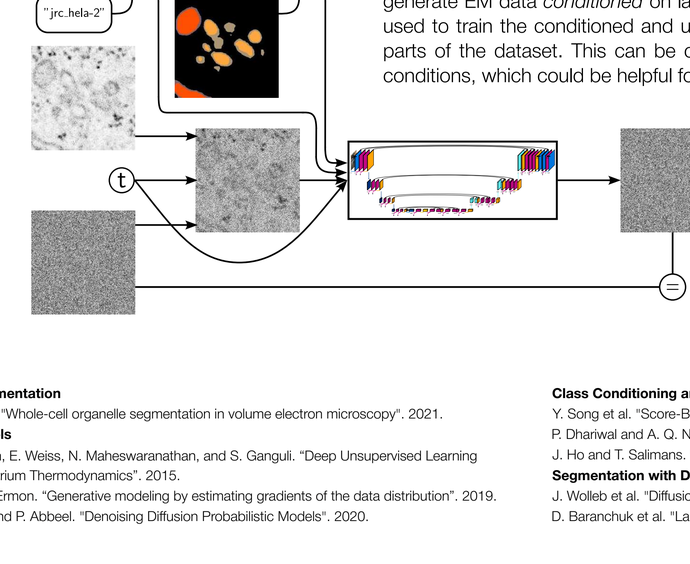
Generating Volume EM Data with Diffusion Networks



Generating Organelle Label Data with Diffusion Networks



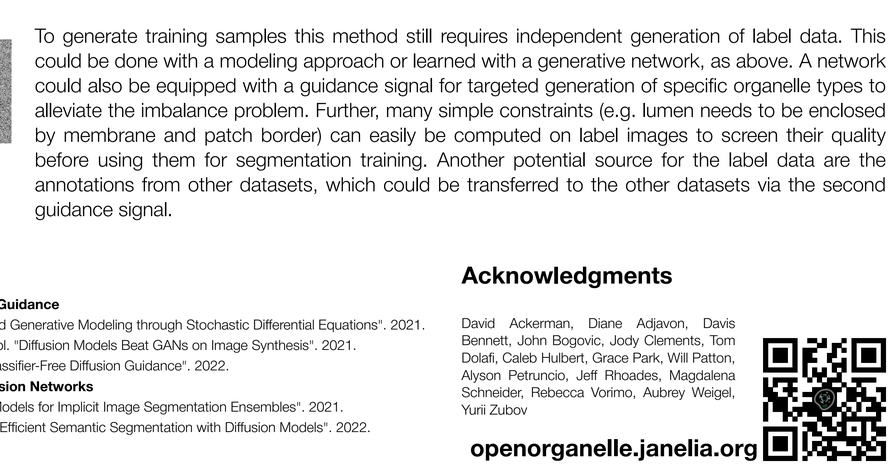
Generating EM + Label Pairs with Diffusion Networks



As a first test for the utility of diffusion networks for the generation of training examples, we use pairs of patches generated by a diffusion network such as this one to train a 2D segmentation network with the same architecture as the DPPM's.

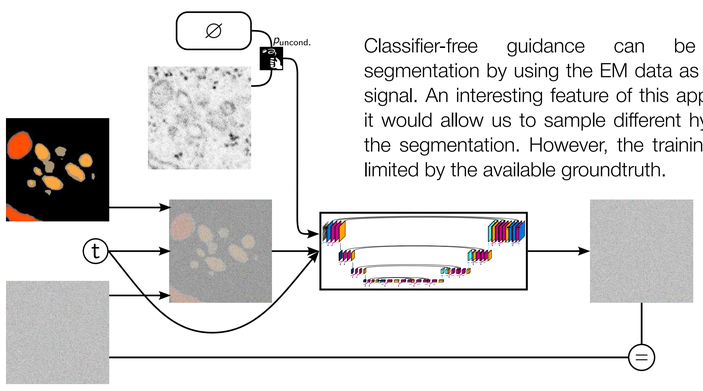
Training a 2D Segmentation Network with Generated EM + Label Pairs

We compare the results of a segmentation network trained on pairs generated by a diffusion network to an equivalent network trained with the same data as was used for training the DDPM. Note that the lack of 3D context and small patch size makes this a very challenging segmentation problem.



The good quality of the segmentations trained from diffused training data suggests that the diffusion networks serve as a useful form of data augmentation. The diffusion network and the segmentation network from real data only use a very minimal set of augmentations. While these initial experiments are promising, a quantification and further exploration are needed to draw robust conclusions.

Outlook: Classifier-free Guidance for Segmentation and Training Data Generation



Classifier-free guidance can be used for segmentation by using the EM data as the guidance signal. An interesting feature of this approach is that it would allow us to sample different hypotheses for the segmentation. However, the training data is still limited by the available groundtruth.

A more promising approach to generating training data for a segmentation network than simply generating EM/label pairs as above is to generate EM data *conditioned* on label data. This idea has several potential advantages: Since in classifier-free guidance the same network is used to train the conditioned and unconditioned classifier, training examples for the unconditioned case can be sampled from the unlabeled parts of the dataset. This can be combined with other guidance signals, such as the dataset id or various sample attributes and imaging conditions, which could be helpful for domain adaptation.

To generate training samples this method still requires independent generation of label data. This could be done with a modeling approach or learned with a generative network, as above. A network could also be equipped with a guidance signal for targeted generation of specific organelle types to alleviate the imbalance problem. Further, many simple constraints (e.g. lumen needs to be enclosed by membrane and patch border) can easily be computed on label images to screen their quality before using them for segmentation training. Another potential source for the label data are the annotations from other datasets, which could be transferred to the other datasets via the second guidance signal.

References

Data

jrc_hela-2: doi.org/10.25378/janelia.13114211; doi.org/10.25378/janelia.13108343
jrc_hela-3: doi.org/10.25378/janelia.13114244; doi.org/10.25378/janelia.13117586
jrc_jurkat-1: doi.org/10.25378/janelia.13114259; doi.org/10.25378/janelia.13117697
jrc_macrophage-2: doi.org/10.25378/janelia.13114343; doi.org/10.25378/janelia.13117745
jrc_sum159-1: doi.org/10.25378/janelia.13114352; doi.org/10.25378/janelia.13118156
jrc_mus-kidney: https://doi.org/10.25378/janelia.16913035
L. Heinrich, J. A. Bogovic and S. Saalfeld. "Deep learning for isotropic super-resolution from non-isotropic 3D electron microscopy"

Organelle Segmentation

L. Heinrich et al. "Whole-cell organelle segmentation in volume electron microscopy". 2021.

Diffusion Models

J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. "Deep Unsupervised Learning using Nonequilibrium Thermodynamics". 2015.
Y. Song and S. Ermon. "Generative modeling by estimating gradients of the data distribution". 2019.
J. Ho, A. Jain, and P. Abbeel. "Denoising Diffusion Probabilistic Models". 2020.

Class Conditioning and Guidance

Y. Song et al. "Score-Based Generative Modeling through Stochastic Differential Equations". 2021.
P. Dhariwal and A. Q. Nichol. "Diffusion Models Beat GANs on Image Synthesis". 2021.
J. Ho and T. Salimans. "Classifier-Free Diffusion Guidance". 2022.

Segmentation with Diffusion Networks

J. Wolleb et al. "Diffusion Models for Implicit Image Segmentation Ensembles". 2021.
D. Baranchuk et al. "Label-Efficient Semantic Segmentation with Diffusion Models". 2022.

Acknowledgments

David Ackerman, Diane Adjavon, Davis Bennett, John Bogovic, Jody Clements, Tom Dolafi, Caleb Hulbert, Grace Park, Will Patton, Alyson Petruncio, Jeff Rhoades, Magdalena Schneider, Rebecca Vorimo, Aubrey Weigel, Yurii Zubov

openorganelle.janelia.org