In Review, , pp. 1–8

doi:

Transfer learning improves performance in volumetric electron microscopy organelle segmentation across tissues

Ronald Xie,[1,2,3,] Ben Mulcahy,[4] Ali Darbandi,[4] Sagar Marwah,[4] Fez Ali,[1] Yuna Lee,[1] Gunes Parlakgul,[5] Gokhan Hotamisligil,[6,7] Bo Wang,[2,8,11,12,] Sonya MacParland,[8,9] Mei Zhen[3,4] and Gary D. Bader[1,3,4,12,13]

[1]The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada, [2]Peter Munk Cardiac Centre and Joint Department of Medical Imaging, University Health Network, Toronto, Canada, [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, [4]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, [5]University of California, Berkeley, Berkeley, CA, USA, [6]Sabri Ülker Center of Metabolic Research and Department of Molecular Metabolism, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [7]Broad Institute of MIT and Harvard, Cambridge, MA, USA, [8]Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada, [9]Ajmera Transplant Centre, Toronto General Research Institute, University Health Network, Toronto, Ontario, Canada, [10]Department of Physiology, University of Toronto, Toronto, Ontario, Canada, Vector Institute, Toronto, Canada, [12]Department of Computer Science, University of Toronto, Toronto, Canada and [13]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Volumetric electron microscopy (VEM) enables nanoscale resolution three-dimensional (3D) imaging of biological samples. Identification and labeling of organelles, cells and other structures in the image volume is required for image interpretation, but manual labeling is extremely time-consuming. This can be automated using deep learning segmentation algorithms, but these traditionally require substantial manual annotation for training and typically these labelled datasets are limited or unavailable for a new sample. We show that transfer learning can help address this challenge. By pre-training on VEM data from multiple mammalian tissues and organelle types and then fine-tuning on a target dataset, we segment multiple organelles at high performance, yet require a relatively small amount of new training data. We benchmark our method on three published VEM datasets and a new rat liver dataset we imaged over a 56x56x11μm volume measuring 7000x7000x219px using serial block face scanning electron microscopy (SBF-SEM) with corresponding manually labelled mitochondria and endoplasmic reticulum structures. Our rat liver dataset's raw image volume, manual ground truth annotation, and model predictions are freely shared at `github.com/Xrioen/cross-tissue-transfer-learning-in-VEM`. We hope that implementation of transfer learning in deep learning models will facilitate VEM data analysis in diverse biological applications.

Key words: volumetric electron microscopy, transfer learning, organelle segmentation

Introduction

Volumetric electron microscopy (VEM) generates nanoscale resolution images of cellular structures in 3D at tissue scale, using serial 2D sample sectioning. VEM has been widely used in the field of connectomics to study neuronal circuits in the brain[25, 1, 28], with increasing use in other tissues, such as pancreas, liver, and bladder[29, 24, 21].

Accurate segmentation of cellular structures in the captured image volumes is often the first step to downstream analysis[16]. Traditionally, segmentation is performed manually, which is extremely laborious and requires expert knowledge. Furthermore, VEM data can be generated using several methods including array tomography[22], FIB-SEM[12] and SBF-SEM[7], which produces images with different resolution and other characteristics. The data are also prone to diverse artifacts including misalignment of adjacent 2D images, brightness differences and out of focus regions, which complicates segmentation[4]. These issues prompted the community to develop various methods to automate the segmentation process[2, 14, 17, 9, 11]. The U-net architecture has shown excellent image segmentation performance applied to a wide variety of biological and medical image types[27, 23, 13]. U-nets have also been adapted to automate organelle segmentation in 3D VEM and have achieved excellent performance [17, 21, 18, 11, 5]. However, these models still require a large and diverse set of manual labels to train. Furthermore, these labels cannot be reused across image data from different tissues using traditional supervised machine learning[19], making it difficult to establish VEM with new

cellular structure and tissue types. Transfer learning reuses neural network parameters trained on previously labelled data and then fine-tunes the learned features on a target task. As many features such as shapes, textures and edges that can be learned in one object segmentation task are useful for a diverse range of segmentation tasks, pretraining data does not need to be the same cellular structure or tissue. Transfer learning has been successfully used to reduce manual label needs in a range of image analysis application areas[6], but extensive comparison between the transferability across different tissues or classes has not yet been studied in VEM.

We evaluate transfer learning performance by training a 3D U-net following the architecture of Lee et al. (3D ResU-Net) on multiple cellular structure label types (mitochondria, endoplasmic reticulum, lipid droplets and neurites) from existing urinary bladder[21], mouse cortex[1], and mouse liver[24] VEM datasets as well as a new liver VEM dataset we generated. We show that transfer learning can substantially reduce the need for manual labeling of VEM data by reusing labels across datasets and segmentation tasks. We make available our model, as well as our rat VEM data and manually labeled mitochondria and endoplasmic reticulum (ER) masks covering a 3500x3500x36px volume along with our model predictions for the entire volume (56x56x11μm, 8nm/pixel resolution, 219 serial 7000x7000 pixel images at 50nm per section). In addition, we manually corrected these model predictions and release the resulting segmented structures as a community resource.

Results

Pretrained 3D ResU-net outperforms a baseline acros varying amounts of target task training data

To implement transfer learning for VEM data, we pretrain a 3D ResU-net model from randomly initialized weights using images and segmentation labels from public data (Figure 1). We then transfer the model by fine-tuning on a target task for a VEM image dataset of interest. Finally, the fine-tuned model is used to make predictions for the abundant unlabelled images in the target task. To measure the magnitude of the performance gain from transfer learning, we compare model performance with and without transfer learning and evaluate under three different pretraining and target task scenarios (Figure 2): 1) the image volumes are from different tissues but the objects being segmented are similar (both mitochondria segmentation), 2) the image volumes are from the same tissue but the objects being segmented are different (endoplasmic reticulum vs. mitochondria segmentation), 3) the image volumes are from different tissues and the objects being segmented are dissimilar (neurite segmentation vs. mitochondria segmentation).

The amount of target training data used is varied at five different levels (100%, 75%, 50%, 25%, 10%) to simulate training in data scarce scenarios, and to examine the extent transfer learning can help reduce manual labeling requirements for new datasets. Each experimental condition is repeated three times with either different randomly initialized weights (for the baseline) or different pretrained models from the same pretraining dataset (for transfer learning). We evaluate performance using intersection-over-union (IoU) scores comparing predicted with known annotation pixels when the final models are applied to the target task test dataset. In all three cases, transfer learning consistently performs better than baseline on the target task, particularly when the amount of training data of the target task is low (Figure 2).

As an additional test, we visualized the difference in the resulting segmentation of our 3D ResU-Net pretrained on the SNEMI3D-N (3D Segmentation of neurites in EM Images) dataset vs. randomly initialized baseline when only using 25% (432 patches) of the total available ground truth data for the Urocell mitochondria segmentation task (Figure 3). The resulting IoU was 0.811 vs. 0.904 for the randomly initialized and transferred models, respectively. Mispredictions with less than 50% overlap with ground truths are highlighted in red in Figure 3. There are substantial false positives in the prediction of the randomly initialized model compared to minor mispredicted segments in the transferred model.

Transfer learning confers substantial increase in performance across the majority of tested pretraining and fine-tuning task pairs

We next systematically evaluated how different pretraining and downstream task combinations affect the final performance of the resulting 3D ResU-Net (Table 1). Our results show that pretraining from VEM segmentation tasks in general boosts performance of the target task, with the exception of one combination (Urocell-M transferred to MouseLiver-ER), regardless of how different the pretraining and target tasks are (Table 1). The highest average performance gain occurs with the mitochondria segmentation pretraining task from the mouse liver dataset (MouseLiver-M) with all target tasks (28% over random initialization).

Transfer learning enables accurate mitochondria and ER segmentation of the rat liver dataset with limited task specific ground truths

Using the findings of Table 1, we chose the best performing pretraining task for the RatLiver-M and RatLiver-ER segmentation tasks, which are the SNEMI3D-N and MouseLiver-ER tasks, respectively, and fine-tuned on 120 patches of the corresponding task specific ground truth labels that we manually annotated. As an additional test of our results on this data, we compared the fine-tuned and pretrained only models to a model using both (transfer learning) for the RatLiver-M and RatLiver-ER tasks (Figure 4), and confirmed by manual visual inspection that transfer learning did better than each approach separately. We applied this to our whole rat liver image (56x56x11μm volume, 8nm/pixel resolution, 219 serial 7000x7000 pixel images at 50nm per section) to automatically annotate mitochondria and ER segments (Figure 5).

Discussion

In this study we investigated the benefit of transfer learning on organelle segmentation across 3D high resolution volumetric electron microscopy data taken from mouse cortex, mouse liver, mouse urinary bladder and rat liver tissues. Accurate and scalable segmentation of these organelles is critical for downstream analysis for biological insights.

We show that transfer learning benefits segmentation in three different cases: when the pretraining task and the target task are trained from images from different domains but the segmentation task is similar; when the pretraining task and the target task are trained from images from the same domain but

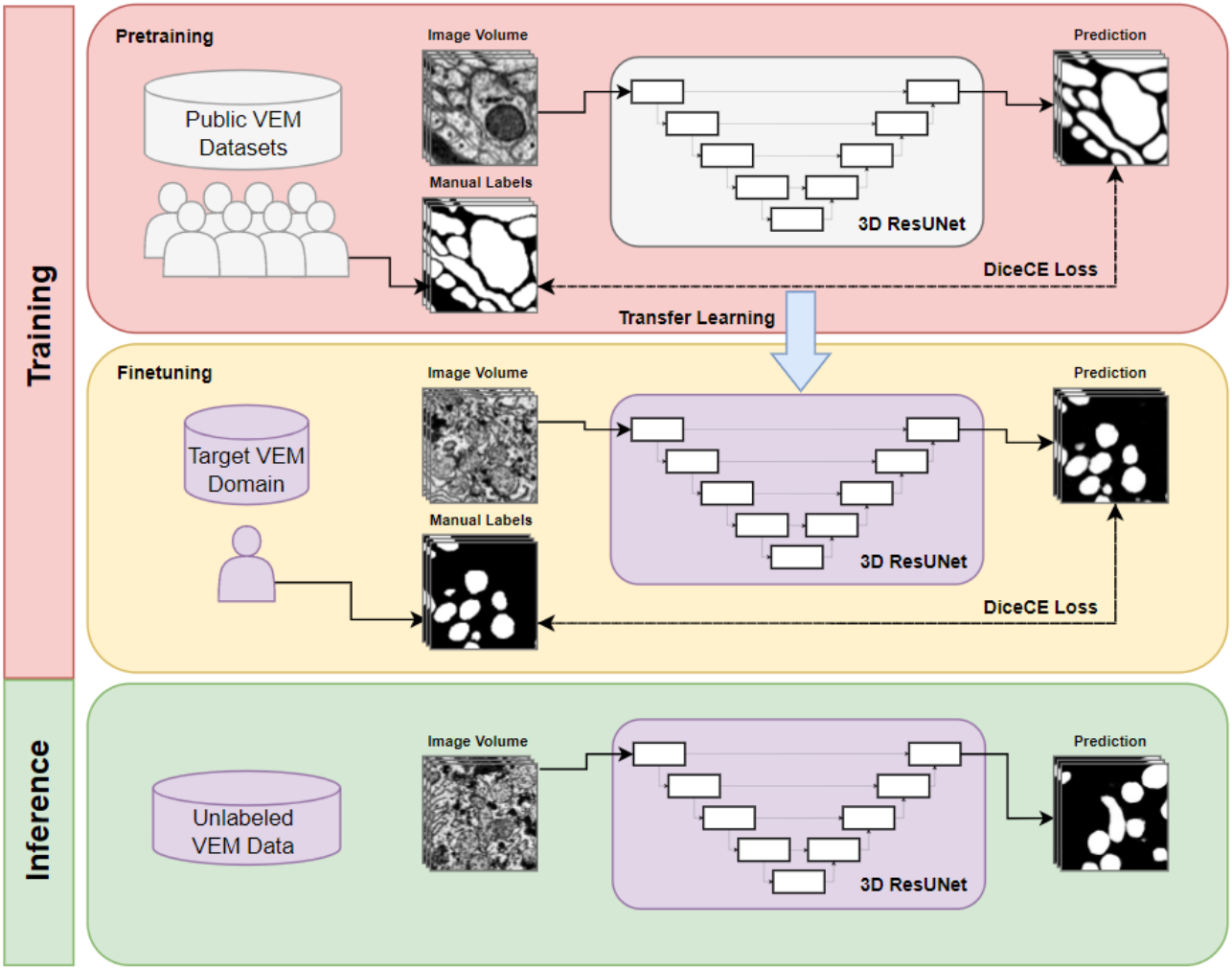

Fig. 1. Transfer learning across domains. A randomly initialized deep learning model is first pretrained on a task with abundant labels from a different EM domain and subsequently fine-tuned on a target task where labels may be scarce. The transferred knowledge ideally alleviates the requirement for abundant labels for the target task and achieves better performance.

Table 1. Transfer learning generally improves performance for different pretraining and fine-tuning task combinations. The 3D ResU-net is pretrained on a given dataset and task (rows). The pretrained model is then transferred and fine-tuned using 100 randomly selected patches from each target task (columns) for fine-tuning. The reported values in the table are the intersection over union (IoU) of predicted versus ground truth segmentations. The baseline values are randomly initialized deep learning models trained on the target task without transfer learning. Gray boxes represent the performance of the model when given 100% of the target task training data and serve as an upper bound comparison. Bolded values are combinations that produced the best performance for each target task-test pair.

	SNEMI3D-N	Urocell-M	MouseLiver-M	MouseLiver-ER	MouseLiver-LD	RatLiver-M	RatLiver-ER
NoPretrain	60.68	65.71	60.84	66.35	52.75	76.81	73.14
SNEMI3D-N	80.81	88.24	92.54	72.23	80.43	**92.37**	76.04
Urocell-M	71.18	90.44	89.84	66.17	**82.83**	88.03	74.93
MouseLiver-M	**77.07**	**89.81**	94.82	**74.36**	81.80	92.08	76.43
MouseLiver-ER	76.60	87.15	**93.75**	78.02	67.62	92.04	**77.12**
MouseLiver-LD	69.73	85.00	89.52	67.46	84.72	87.76	75.01

the segmentation task is different; and when the pretraining task and the target task are trained from images from different domains and the segmentation task is also different (Figure 2). Furthermore, we showed that this performance increase becomes increasingly larger when the amount of training data available in the target task is reduced from 100% to 10%. These results show that there is generalizable knowledge across different VEM datasets. Similarly, there is generalizable knowledge across different segmentation tasks in these VEM datasets.

To manually annotate a 3500x3500x18px volume of images at 8x8x30nm resolution for both mitochondria and ER would take roughly 100 hours (based on our manual image annotation throughput), with time varying with the task difficulty, the

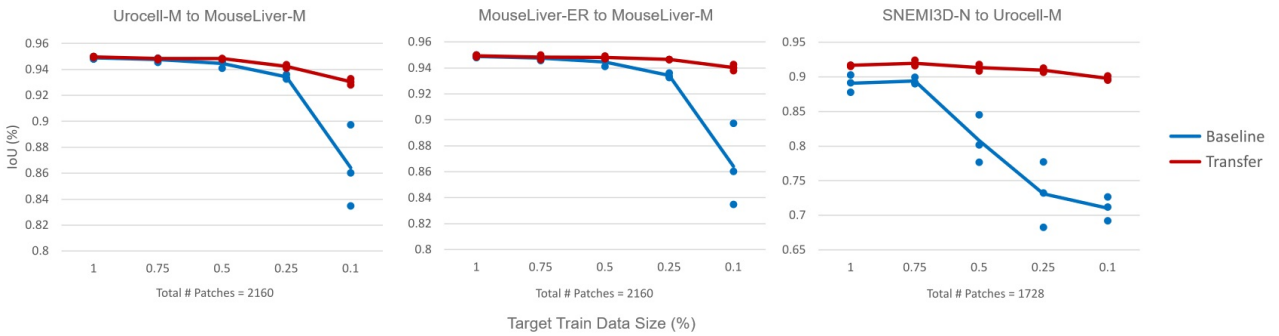

Fig. 2. Transfer learning outperforms a random initialized baseline without transfer learning across three representative pretraining and target task scenarios. For each plot, "Baseline" values refer to the intersection-over-union (IoU) metric of the resulting segmentations from a model trained from scratch using randomly initialized weights compared to ground truth labels in a pixel-wise fashion, without using transfer learning. "Transfer" values refer to the IoU of the resulting segmentations from a model fine-tuned using the respective pretrained model compared to ground truth labels. For each pretraining and target task pair experiment, we vary the amount of training data available for the target task at five different levels (100%, 75%, 50%, 25%, 10%). The transferred networks surpass the performance of baseline particularly in training data scarce scenarios.

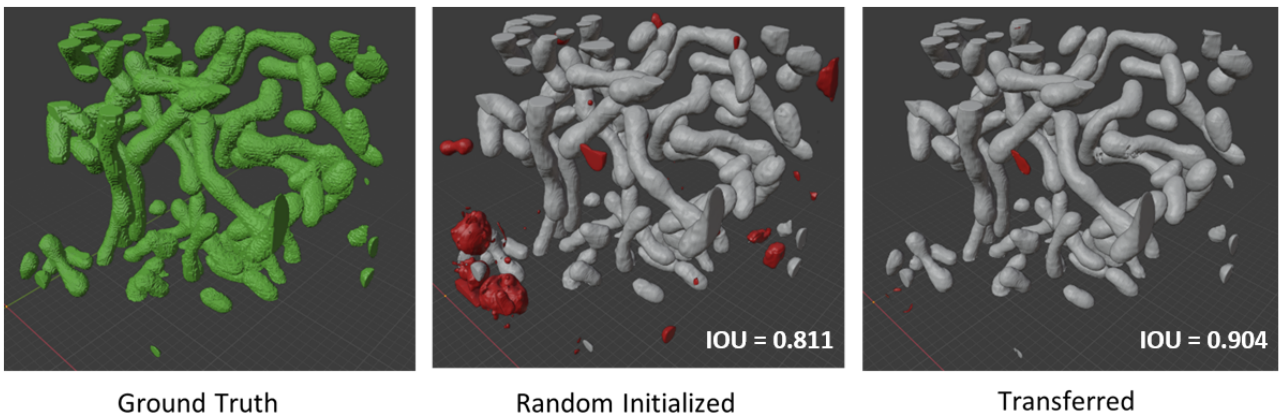

Fig. 3. Transfer learning improves mitochondria segmentation in 3D. Binarized output of mitochondria segmentation model predictions. Large mispredictions (¡ 50% overlap) are colored in red. Both the randomly initialized model and the transferred model were trained with only 25% of training data. The transferred model makes noticeably fewer mispredictions.

Table 2. Data sources and their attributes. Datasets are normalized together by applying contrast limited adaptive histogram equalization (CLAHE[26]) and Gaussian blur before rescaling to the appropriate resolution and patched into the input size of the 3D ResU-net (160x160x18). The labels N, M, ER, and LD stand for neurites, mitochondria, endoplasmic reticulum, and lipid droplet, respectively.

	SNEMI3D	Urocell	Mouse Liver	Rat Liver
Method	SEM	FIB-SEM	FIB-SEM	SBF-SEM
Organism	Mouse	Mouse	Mouse	Rat
Tissue	Cortex	Urothelium	Liver	Liver
Image Resolution (nm)	6x6x30	16x16x15	8x8x8	8x8x50
Rescaled Resolution (nm)	6x6x30	10.2x10.2x15	32x32x40	28x28x50
Train Size (# Patches)	3388	1728	2160	120
Test Size (# Patches)	484	432	2160	120
Labels	N	M	M, ER, LD	M, ER

density of the organelle in the particular field of view, the ability of the individual annotator and image quality. Our results show promise to alleviate the time consuming manual annotation bottleneck. Instead of manually annotating the whole image, as little as 10% of the image can be manually annotated and then this can be used to fine-tune an existing pre-trained model to predict the rest of the segmentation. This is especially important for applications of VEM in new tissue types, as there may be a lack of publicly available domain specific datasets to train neural networks using standard approaches that don't use transfer learning.

We study this phenomenon by systematically evaluating the performance of models across different pretraining and fine-tuning task pairs, particularly including a newly acquired rat liver dataset with minimal manual labels as a practical

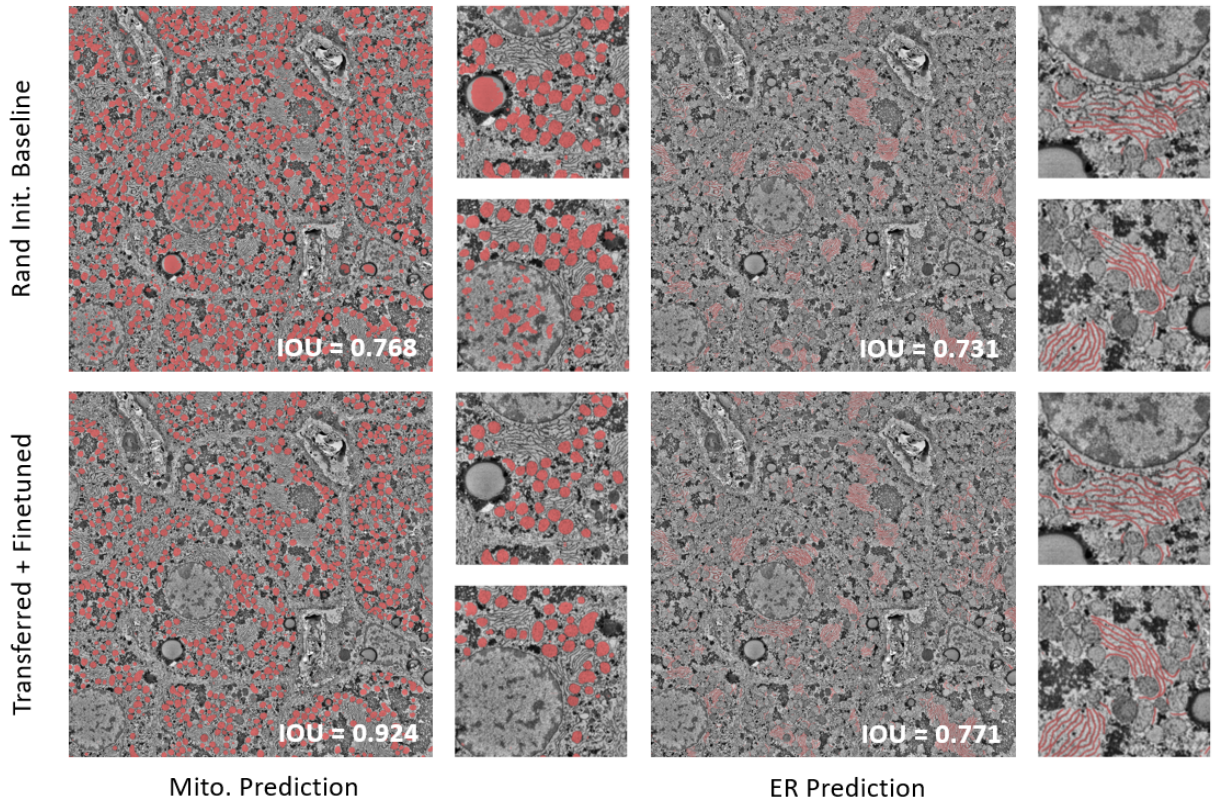

Fig. 4. Transfer learning outperforms randomly initialized baseline for both mitochondria and ER prediction tasks in the rat liver dataset. Rand Init. Baseline: Rat liver dataset inference on previously unseen slice using a 3D ResU-Net with randomly initialized weights, trained on 18 slices of manual ground truths of mitochondria (Left) and ER (Right). Transferred + Finetuned: Rat liver dataset inference on unseen slice using a model first pretrained on the SNEMI3D-N and MouseLiver-ER datasets, and then fine-tuned on the same 18 slices of manual segmentation masks for mitochondria (Left) and ER (Right), respectively. Red pixels indicate model predictions. Smaller image patches are zoom-ins of a region in the larger image to more clearly show the organelle structures. Zoom-ins are matched in the top and bottom figure sections. IoU values are shown for each image, with higher values being better.

scenario and proof of concept (Table 1). The target task is fine-tuned using 100 randomly selected patches of 160x160x18, to simulate a data scarce scenario. This represents approximately 18 hours of annotation time extrapolated from our own experience with annotating the ER and mitochondria of the rat liver dataset. There are a few observations of interest. First, all pretraining and target task combinations outperform the randomly initialized baseline with the exception of one case where there is a small decrease in performance. This demonstrates the value of naively pretraining segmentation and annotation models on VEM datasets in general. This makes sense as concepts such as shape and edge detection are expected to be generally useful for segmentation of biological structures. Second, the MouseLiver-M pretraining task is consistently the best performing after transferring across all target tasks compared to other pretraining tasks. This may be because mitochondria segmentation provides a rich training signal in that it requires both good modeling of shape to distinguish from other non-rotund objects and structures, as well as good modeling of texture and edges to distinguish between rotund objects as well as between ER and mitochondria cristae. Third, for the rat liver dataset, the SNEMI3D-N pretraining task somewhat outperformed MouseLiver-M. This is rather unexpected as the SNEMI3D-N data (mouse brain, neurites) is both cross domain and cross task compared to the rat liver dataset. This may be because the SNEMI3D-N neurite segmentation task visually resembles the shape and size of mitochondria and contains potentially more variability for better generalization. This result highlights the potential of using relatively widely available connectomics VEM datasets to improve training of less widely available non-connectomics related segmentation tasks. In contrast, the MouseLiver-ER pretraining task confers the most benefit on the RatLiver-ER segmentation task, likely due to task similarity between the pair. This shows that while pretraining on VEM datasets in general benefits the performance substantially, it can still be beneficial for the pretraining task to be more similar to the target task.

Overall, our experiments highlight the promise of using transfer learning to improve automatic segmentation performance across diverse organelles and tissue types. Transfer learning is particularly effective in data scarce scenarios, demonstrating its utility in emerging application areas of VEM. We also release our model and newly collected rat liver data and its accompanying manual labels as a resource for the community to further facilitate the adaptation and encourage the wider use of volumetric electron microscopy in new biological domains.

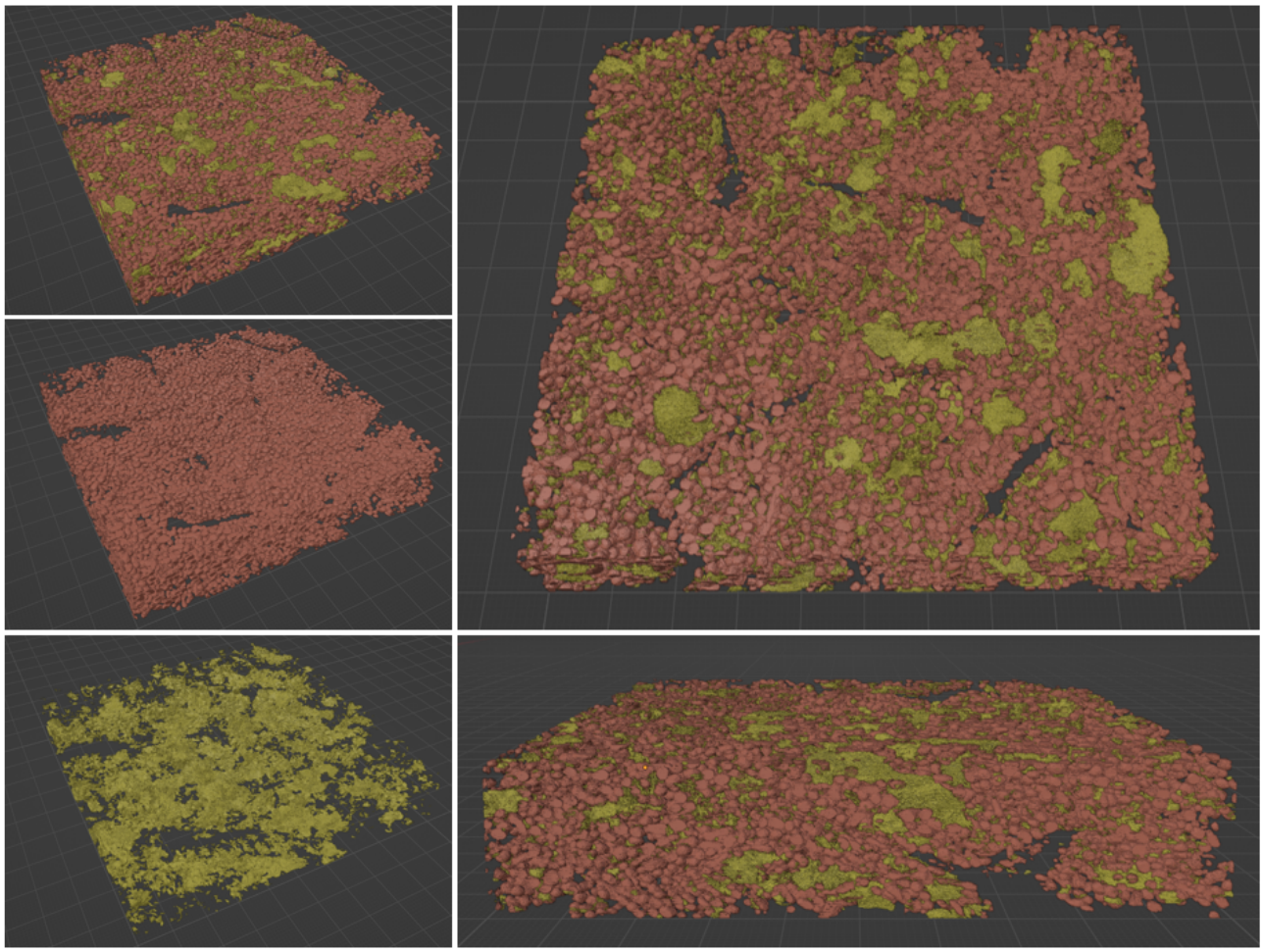

Fig. 5. Automatic segmentation of mitochondria and ER of the entire rat liver volume using transfer learning and 3D ResU-Net. 3D ResU-Nets were pretrained using the best pretraining and target pair in Table 1, respectively, for mitochondria and ER segmentation of the rat liver dataset. The resulting model is used to infer annotation on all 219 serial sections of the dataset. The figures show the resulting 3D mesh of the ER and mitochondria predictions. Red pixels label mitochondria and yellow pixels label ER regions. Left: Perspective view of the segmentation prediction for both mitochondria and ER (Top), mitochondria only (middle), and ER only (bottom). Right: Top down (Top) and side view (bottom) of the combined mitochondria and ER predictions on the entire rat liver volume.

Methods

Rat liver sample preparation

A 4 month old male rat weighing 180g was sacrificed and perfused with 2% glutaraldehyde and 2% paraformaldehyde in 0.15M sodium cacodylate buffer (pH7.4) through the left ventricle at a hydraulic pressure of 90 cm for 5 minutes. The liver was removed, cut into 2-5mm chunks, and stored in the same fixative solution at 4°C until needed. Secondary fixation and staining were performed using a modification of the fBROPA method, without formalin[10]. Briefly, samples were incubated with 2% OsO 4 + 1.5% potassium ferrocyanide in 0.15M sodium cacodylate buffer for 90 minutes at room temperature, switched to 1% OsO 4 in 0.15M sodium cacodylate buffer for 90 minutes, then washed with 0.1M sodium cacodylate buffer. Samples were incubated with 320mM pyrogallol in ddH 2 O for 30 minutes, washed with 0.15M sodium cacodylate buffer, incubated with 40mM OsO 4 in distilled water for 90 minutes, washed and stored in ddH 2 O overnight. The following day they were stained in Walton's lead aspartate for 60 min at 60°C, washed in ddH 2 O, dehydrated with an ethanol series followed by propylene oxide, then infiltrated with a graded series of Epon resin and cured at 60°C for 24h. Samples were trimmed and mounted on a stub that was transferred to a Gatan 3View stage inside a Zeiss Gemini SEM. 219 sections were imaged at a thickness of 50nm, resolution of 8nm/pixel, and tile dimensions of 7000x7000 pixels (56x56µm).

Datasets and preprocessing

Table 2 describes the datasets used to benchmark transfer learning across domains.

The SNEMI3D neurite dataset was downloaded from their challenge website[1]. There are 100 images in total, with the first 72 used for training and the last 18 used for testing. Slices 73-82 were discarded to limit information leak between training and testing sets. The dataset is anisotropic with 6nm XY resolution and 30nm Z resolution.

The Urocell mitochondria dataset was downloaded from its GitHub repository[21]. The dataset also contains Golgi and lysosome labels, but they were not used for our experiments

because they are relatively sparse. The released data is already binned by a factor of three to produce a near isotropic dataset. However for our purposes, we upsampled the dataset by a factor of two in the X and Y dimensions to reproduce the anisotropic nature to match the other datasets used in our study, as we assumed this would facilitate transfer learning between datasets.

The mouse liver dataset was downloaded from EMPIAR[8]. The dataset contains 5638 serial images, each with an image size of 12000x8000 pixels. We used ground truth labels for mitochondria, ER and lipid droplets every fifth slice in the Z dimension to create anisotropicity in the dataset to match the SNEMI3D and the rat liver datasets. We also use only 180 such slices starting at slice 1000 and 2000 respectively for the training and testing set to simulate data scarce scenario for which transfer learning is practically useful.

All datasets were processed using contrast limited adaptive histogram equalization (CLAHE[26]) before the pixel values were scaled and centered. The datasets were scaled accordingly to keep the size of the objects inside one patch consistent between tasks. The rescaled volumes were then tiled into 160x160x18px patches before inputting into our neural network models.

Training and evaluation

The neural network architecture used was adapted from the best performing model of the SNEMI3D competition[1] (Figure S1). The input dimension of this network is 160x160x18. It consists of a downsampling and upsampling branch consisting of 5 layers with 28, 36, 48, 64 and 80 channels respectively and a 2D max pooling layer between each layer. Each layer consists of a 2D convolution followed by two 3D convolutions, with a residual connection between the first and third convolutional module. As the dataset is anisotropic, the maxpooling was only performed in the XY direction, with the dimensions of Z remaining the same throughout. Furthermore, a single 5x5x1 convolution and corresponding transpose convolution is applied prior to the aforementioned convolutions in the first layer and before the output layer respectively, which helps alleviate the anisotropic nature of the input volumes.

The model was trained using the DiceCE (Dice and cross entropy) loss and the ADAM optimizer[15] with an initial learning rate of 0.002. In preliminary work, we found that a unified DiceCE loss which combines cross entropy loss and Dice loss is more robust across different domains and training tasks compared to either Dice or cross entropy loss alone (results not shown). Dice loss gave poorer performance than cross entropy when labels are balanced, whereas cross entropy loss alone was less consistent in more imbalanced prediction tasks (results not shown). Input patches are randomly subjected to flips, rotations, brightness, and contrast augmentations during training to increase the robustness of the model. Training was performed on a single V100 Nvidia graphics card and repeated for 100 epochs or until convergence.

For evaluation, we measure the intersection over union (IoU). Similar improvements are observed in terms of accuracy, positive predictive value (PPV), and true positive rate (TPR).

For visualization of the model's segmentation predictions, we first binarize the pixel-wise predicted logits with a threshold of 0.5, followed by conversion into mesh via the marching cubes algorithm[20]. Finally, the resulting mesh is visualized using Blender[3].

References

1. Snemi3d segmentation challenge. `http://brainiac2.mit.edu/SNEMI3D/home`. Accessed: 2021-07-01.
2. Thorsten Beier, Constantin Pape, Nasim Rahaman, Timo Prange, Stuart Berg, Davi D Bock, Albert Cardona, Graham W Knott, Stephen M Plaza, Louis K Scheffer, et al. Multicut brings automated neurite segmentation closer to human performance. *Nature methods*, 14(2):101–102, 2017.
3. Blender Foundation. Blender. https://www.blender.org/, 2022.
4. S Borrett and L Hughes. Reporting methods for processing and analysis of data from serial block face scanning electron microscopy. *Journal of microscopy*, 263(1):3–9, 2016.
5. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
6. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
7. Winfried Denk and Heinz Horstmann. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS biology*, 2(11):e329, 2004.
8. Electron Microscopy Public Image Archive. Empiar-10791, Year of access.
9. Jan Funke, Fabian Tschopp, William Grisaitis, Arlo Sheridan, Chandan Singh, Stephan Saalfeld, and Srinivas C Turaga. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1669–1680, 2018.
10. Christel Genoud, Benjamin Titze, Alexandra Graff-Meyer, and Rainer W Friedrich. Fast homogeneous en bloc staining of large tissue samples for volume electron microscopy. *Frontiers in Neuroanatomy*, 12:76, 2018.
11. Larissa Heinrich, Davis Bennett, David Ackerman, Woohyun Park, John Bogovic, Nils Eckstein, Alyson Petruncio, Jody Clements, Song Pang, C Shan Xu, et al. Whole-cell organelle segmentation in volume electron microscopy. *Nature*, 599(7883):141–146, 2021.
12. Jurgen AW Heymann, Mike Hayles, Ingo Gestmann, Lucille A Giannuzzi, Ben Lich, and Sriram Subramaniam. Site-specific 3d imaging of cells and tissues with a dual beam microscope. *Journal of structural biology*, 155(1):63–73, 2006.
13. Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
14. Michał Januszewski, Jörgen Kornfeld, Peter H Li, Art Pope, Tim Blakely, Larry Lindsey, Jeremy Maitin-Shepard, Mike Tyka, Winfried Denk, and Viren Jain. High-precision automated reconstruction of neurons with flood-filling networks. *Nature methods*, 15(8):605–610, 2018.
15. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
16. Jörgen Kornfeld and Winfried Denk. Progress and remaining challenges in high-throughput volume electron

microscopy. *Current opinion in neurobiology*, 50:261–267, 2018.

17. Kisuk Lee, Jonathan Zung, Peter Li, Viren Jain, and H Sebastian Seung. Superhuman accuracy on the snemi3d connectomics challenge. *arXiv preprint arXiv:1706.00120*, 2017.
18. Mingxing Li, Chang Chen, Xiaoyu Liu, Wei Huang, Yueyi Zhang, and Zhiwei Xiong. Advanced deep networks for 3d mitochondria instance segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.
19. Jing Liu, Linlin Li, Yang Yang, Bei Hong, Xi Chen, Qiwei Xie, and Hua Han. Automatic reconstruction of mitochondria and endoplasmic reticulum in electron microscopy volumes by deep learning. *Frontiers in neuroscience*, 14:599, 2020.
20. William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM Siggraph Computer Graphics*, 21(4):163–169, 1987.
21. Manca Žerovnik Mekuč, Ciril Bohak, Samo Hudoklin, Byeong Hak Kim, Min Young Kim, Matija Marolt, et al. Automatic segmentation of mitochondria and endolysosomes in volumetric electron microscopy data. *Computers in biology and medicine*, 119:103693, 2020.
22. Kristina D Micheva and Stephen J Smith. Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron*, 55(1):25–36, 2007.
23. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
24. Güneş Parlakgül, Ana Paula Arruda, Song Pang, Erika Cagampan, Nina Min, Ekin Güney, Grace Yankun Lee, Karen Inouye, Harald F Hess, C Shan Xu, et al. Regulation of liver subcellular architecture controls metabolic homeostasis. *Nature*, 603(7902):736–742, 2022.
25. Christopher J Peddie, Christel Genoud, Anna Kreshuk, Kimberly Meechan, Kristina D Micheva, Kedar Narayan, Constantin Pape, Robert G Parton, Nicole L Schieber, Yannick Schwab, et al. Volume electron microscopy. *Nature Reviews Methods Primers*, 2(1):1–23, 2022.
26. Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38:35–44, 2004.
27. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
28. Daniel Witvliet, Ben Mulcahy, James K Mitchell, Yaron Meirovitch, Daniel R Berger, Yuelong Wu, Yufang Liu, Wan Xian Koh, Rajeev Parvathala, Douglas Holmyard, et al. Connectomes across development reveal principles of brain maturation. *Nature*, 596(7871):257–261, 2021.
29. C Shan Xu, Song Pang, Gleb Shtengel, Andreas Müller, Alex T Ritter, Huxley K Hoffman, Shin-ya Takemura, Zhiyuan Lu, H Amalia Pasolli, Nirmala Iyer, et al. An open-access volume electron microscopy atlas of whole cells and tissues. *Nature*, 599(7883):147–151, 2021.

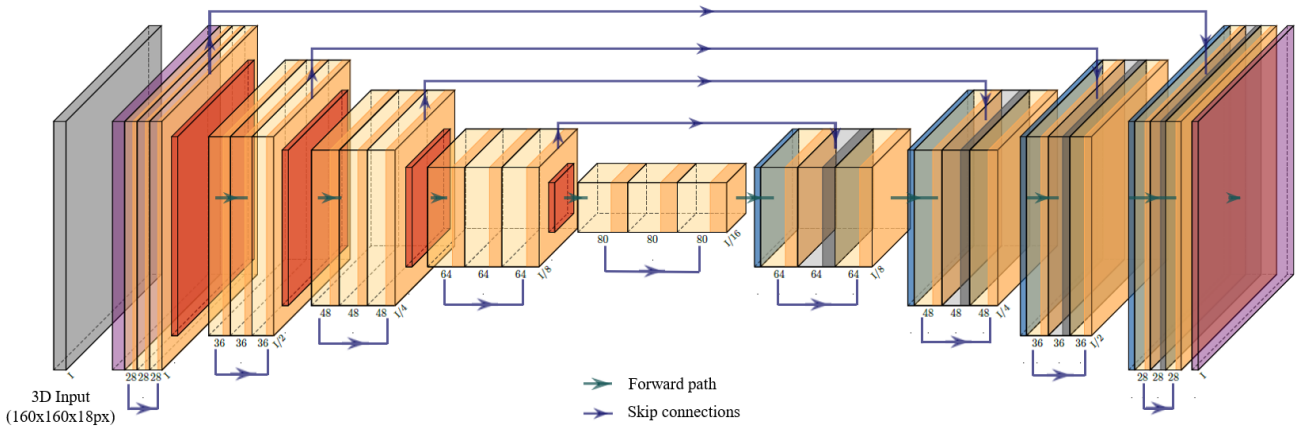

Fig. S1. 3D residual U-net architecture used in this work: The input dimension of our neural network is 160x160x18. The input first passes through a 5x5x1 convolution before going through a downsampling and upsampling branch of 5 layers with 28, 36, 48, 64 and 80 channels respectively with skip connections connecting the downsampling and upsampling branch in each layer and a 2D max pooling layer between each layer. Each layer consists of a 2D convolution followed by two 3D convolutions, with a residual connection between the first and third convolutional module. Finally, the features are passed through a final 5x5x1 transpose convolution layer followed by softmax to generate segmentation predictions.