UNPAIRED IMAGE ENHANCEMENT FOR NEURITE SEGMENTATION IN X-RAY TOMOGRAPHY

Jeff L. Rhoades[⊗,⊕] Arlo Sheridan[⊖] Mukul Narwani[⊗] Brian Reicher[†]
Mark Larson[†] Shuhan Xie[†] Tri Nguyen[⊗] Aaron Kuan[⊗] Alexandra Pacureanu[‡]
Wei-Chung Allen Lee[⊗,⋆] Jan Funke[⊕,⋆]

[⊗] Harvard Medical School, Boston, MA
[⊕] HHMI Janelia, Ashburn, VA
[⊖] Salk Institute, La Jolla, CA
[†] Northeastern University, Boston, MA
[‡] ESRF, The European Synchrotron, Grenoble, France
[⋆] authors contributed equally

ABSTRACT

As the field of connectomics strives to tackle questions regarding increasingly large neuronal circuits, technologies improving imaging throughput will be vital. X-Ray Holographic Nanotomography (XNH) may play a key role, by allowing for fast, non-destructive, multi-resolution imaging. XNH is well suited for rapidly imaging large tissue volumes, with throughput easily increased at the cost of resolution and image quality. We therefore set out to systematically examine the potential for cycle-consistent generative adversarial networks (CycleGANs) to facilitate high-quality segmentation from low-quality data. Additionally, we introduce the Split CycleGAN, a modification of the original formulation designed to prevent collaboration between generators that could result in hidden features in generated images. We find that our new formulation, as well as the original CycleGAN, both improve segmentation results over the naive case, allowing for an approximately 64-fold imaging speed up.

Index Terms— connectomics, X-ray holographic nanotomography, domain adaptation, image enhancement, generative adversarial networks

1. INTRODUCTION

As we seek to build connectomic datasets of increasingly large circuits, improving imaging throughput is one of the biggest hurdles. In our imaging regime, X-ray Holographic Nanotomography (XNH) of mouse cerebellum, by increasing the effective pixel size 3-fold and acquiring holograms at only a single sample-to-sensor distance, we are able to increase imaging throughput more than 108-fold. However, many important structures, such as unmyelinated neurites, become significantly more difficult to distinguish under these conditions, making manual annotation barely feasible. While image segmentations play a crucial role in many biological studies and clinical applications, generating high-quality (HQ) segmentations remains time-consuming and resource constrained – particularly when ground truth (GT) generation is difficult, as in our high-throughput XNH dataset. Thus, we sought to infer important, HQ features in the latent space of low-quality (LQ) data.

Recent work using deep learning aims to allow segmentation strategies to be trained on one dataset, then applied to another similar dataset, without new GT generation[1]. Denoising, image enhancement and domain adaptation strategies can generally be divided into paired and unpaired approaches. Paired approaches, such as Content Aware Reconstruction (CARE)[2,3] or Pix2Pix[4] have shown impressive results[5,3], but require paired HQ and LQ data to train, limiting their usability. Unpaired techniques, on the other hand, often build upon Cycle-consistent Generative Adversarial Networks (CycleGANs)[6] or similar strategies to generate fake training pairs[1,7,8]. Recently this has included systems designed to disentangle "content" and "style" features of input images[9], or strict engineering of the process generating fake LQ from real HQ for training[8]. With the following work, we carefully examine the potential to not only infer HQ images from LQ data, but to generate HQ segmentations from LQ data, without novel GT annotations.

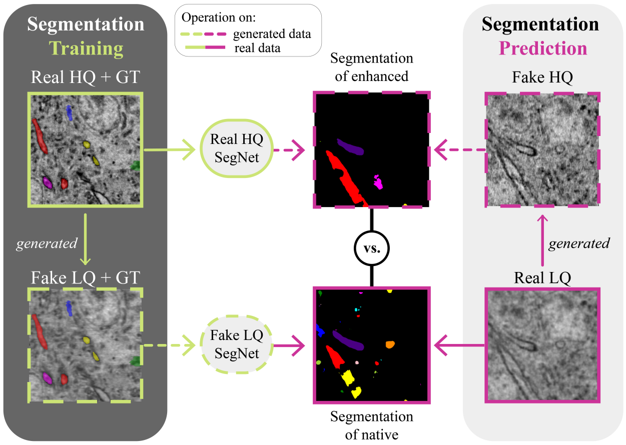
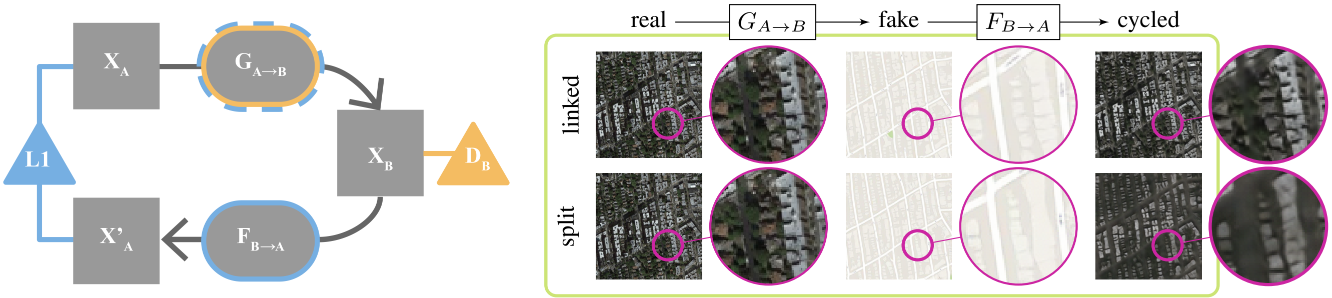


Fig. 1: **Segmentation approaches.** A multi-task local shape descriptor and pixel affinity prediction network (SegNet) is trained on real high-quality (HQ) data with volumetric groundtruth (GT), then used to segment fake HQ produced by a low2high generator from a CycleGAN. These segmentations are compared with those created by a SegNet trained on fake low-quality (LQ) data with GT, generated from the HQ with GT by a high2low generator. This fake LQ SegNet is then applied to real LQ data.

We utilized CycleGANs to train pairs of generators to translate images between HQ and LQ regimes, without the need for paired data. Additionally, we performed an ablation experiment on the classic CycleGAN architecture, which has demonstrated the capacity to embed hidden features in intermediate images to facilitate cycle-consistency[6] (see Fig. 2b). Thus, we sought to disrupt this effect to examine its impact on facilitating domain adaptation for the purpose of segmenting our dataset.

Correspondence: `wei-chung_lee@hms.harvard.edu`
`funkej@janelia.hhmi.org`



(a) (b)

Fig. 2: **Split vs. Linked CycleGANs.** (a) Colors of generator outlines indicate losses driving parameter optimization. Dotted blue outline indicates L1 loss is not available to the $G_{A\to B}$ generator in our Split CycleGAN. D_B indicates discriminator loss. Structure is symmetrical for $B \to A \to B$. (b) Comparison of Split and Linked CycleGANs on the satellite-map dataset (see Zhu et al. 2017). Linked CycleGAN (*i.e.* original) encourages generators to collaborate, leading to embedding of detail in fake map images generated by $G_{A\to B}$. This detail is used by $G_{B\to A}$ to recover the original satellite image, including details that should not be recoverable from a map image. The Split variant (see Section 2.3) does not propagate gradient contributions from the cycle loss on A to the $G_{A\to B}$ generator. As a consequence, generated fake maps are more faithful to the true distribution of map images, at the expense of less accurate cycled images.

2. METHODS

Training pairs of generators to translate images between high- and low-quality regimes creates two potentially useful generators: one that translates LQ images to HQ versions (low2high), and another that degrades HQ data to resemble a LQ likeness (high2low). As such, we chose to explore the potential utility of each for the purpose of leveraging HQ images with GT to segment LQ images (see Fig. 1).

2.1. Segmentation of enhanced data

One approach utilizes the low2high generator to convert real LQ data into *fake* HQ images. Segmentation approaches trained on existing HQ data with GT can then be leveraged to segment the fake HQ renderings. This domain adaptation approach should allow for segmentation of novel data without retraining segmentation networks or generating additional GT, similar to previous work[1].

2.2. Segmentation of native data

An alternative approach is to use the high2low generator to create fake LQ images with paired GT, from existing HQ data with GT. Segmentation approaches can then be trained to parse the fake LQ images in hopes that their performance will transfer to real LQ images. The drawback is that this approach necessitates both training of a high2low generator, as well as retraining a segmentation network. Nevertheless, we examine this method to determine whether segmenting LQ data might be more easily learned than the regression required for predicting a HQ image from a LQ input.

2.3. Split-loss CycleGAN modification

We also chose to examine the effects of a modification to the well-known CycleGAN's loss function[6]. In the original formulation, the L1 cycle losses for both $X_A \to X_B \to X'_A$ and $Y_B \to Y_A \to Y'_B$ translations (between image domains A and B) are used in calculating the gradients for both the $G_{A\to B}$ and $F_{B\to A}$ generators:

$$\mathcal{L}_{cycle}(G,F) = ||G(F(Y_B)) - Y_B||_1 + ||F(G(X_A)) - X_A||_1$$

Thus, generators are incentivized to embed information about the original, *real* images into the intermediate, *fake* images, in order to help the second generator in the sequence best reconstruct the original image in its *cycled* output. The effect of this embedding is impressive (see Fig. 2b), but we wondered if it would aid us in our pursuit of enhancing images, or at least enhancing their segmentation. We experimented with severing the L1 loss gradients such that they are only available to the second generator in the sequence - the one responsible for the transition from *fake* to *cycled* (see Fig. 2a). We call this the Split Loss, yielding the Split CycleGAN. We will refer to the original formulation of the CycleGAN as the Linked CycleGAN from hereon for clarity. The cycle losses for the two generators in the Split formulation are computed and their gradients propagated separately, as follows:

$$\mathcal{L}_{cycle}(G) = ||G(F(Y_B)) - Y_B||_1$$

$$\mathcal{L}_{cycle}(F) = ||F(G(X_A)) - X_A||_1$$

While impressive cycling may be facilitated by embedding features undetectable to discriminators, human or otherwise, it runs the risk of embedding features from HQ images into an approximated LQ regime. For the proposed strategy of optimizing networks for segmentation of native data, this might train networks to segment fake LQ images based on embedded features that are not present in real LQ images. Furthermore, such embedded features could allow a low2high generator a shortcut to minimizing the cycle loss that does not generalize to actual LQ inputs. For instance, we and others[1] have observed generators that solve the dual-optimization problem (discriminator and cycle-consistency) by producing fake images that are roughly inverted in intensities, such that they match the general image statistics of the target regime, thus fooling the discriminator, but do not resemble the actual biological structures of the original image. For example, a dense field of myelinated axons might be converted to appear like the cytoplasm and organelles of a cell body. The second generator in the sequence, nevertheless, can simply learn to invert the previous image back to the original regime, thus also minimizing the cycle-consistency loss. By severing the information flow as described, we hope to minimize such corrupting collusion between generators.

2.4. Datasets

Three separate image volumes from a single sample of mouse cerebellum lobule V, prepared as previously described[11], were acquired at the ESRF's ID16A beamline. One volume was produced by reconstruction[11,12] from holograms imaged at 1300 angles, at a single sample-to-sensor distance resulting in 90nm/voxel images. This 2048^3 (roughly 184μm per side) LQ volume contained both of the following, higher-resolution volumes, which we will assign letters A and B for ease of reference. The 3216^2x2048 30nm/voxel volumes (roughly 96μm by 96μm by 61μm), A and B, were each reconstructed from holograms imaged at 1900 angles, at 4 different sample-to-sensor distances. (Use of multiple distances in holography better constrains the inverse problem, thereby improving reconstruction quality[12].) This amounts to an approximately 64-fold increase in imaging throughput for the 90nm/voxel volume over the 30nm/voxel scans.

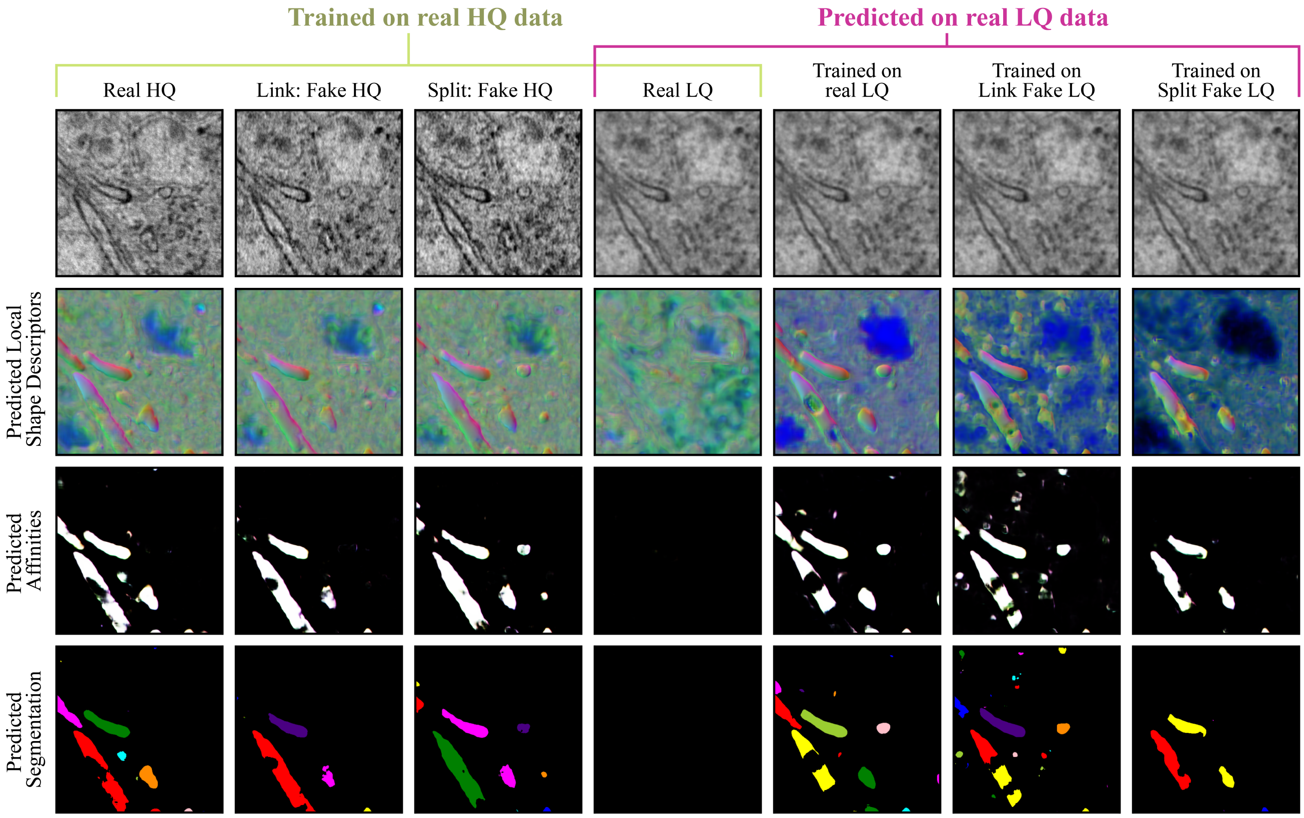


Fig. 3: **Segmentation Examples.** Segmentation is produced by Mutex watershed agglomeration[10] of long-range pixel affinities. Left 4 columns are produced by a segmentation network trained on real high-quality (HQ) data, and applied to the indicated datasets. Right 4 columns are predictions on real low-quality (LQ) data, produced by networks trained on the indicated datasets.

Volumes were subdivided and the 90nm volume aligned and linearly interpolated with Elastix[13,14] to match the 30nm volumes, resulting in 3 separate 1024^3 cutouts, which were used for training, validating, and testing segmentation networks. A subsection of one 30nm volume, A, was taken for training segmentation networks. Sparse volumetric GT of myelinated and unmyelinated neurites was produced by an iterative process of 1) manually painting voxels using webKnossos[15], 2) training networks to produce denser segmentations, 3) correcting these segmentations with manual tracing and additional voxel painting, then 4) repeating from step (2). This was done for 5% of the central 600^3 voxels of the training volume. GT for validation and testing cutouts, taken from HQ volume B, were produced by manual skeleton tracing using webKnossos. Skeletons were rasterized into image volumes for Variation of Information (VoI) comparisons to segmentations.

2.5. Implementation details

CycleGANs were implemented in PyTorch, using Gunpowder[16] to dynamically resample and augment training images, save and monitor progress. 2D UNet[17] generators with 3 downsampling steps and valid padding were fed batches of three 512^2-pixel slices, which were randomly rotated, flipped, and warped to increase diversity of the training set. Training images were drawn randomly from the LQ volume and HQ volume A. PatchGAN Discriminators[4] with 4 layers provided discrimination loss, which was combined with a smooth L1 cycle-loss, with the cycle loss weighted 3x more than the discrimination loss. Adam optimizers were used with initial learning rates of $\alpha = 4 \times 10^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\alpha = 4 \times 10^{-6}$, $\beta_1 = 0.95$, $\beta_2 = 0.999$ for generators and discriminators, respectively. The difference in initial learning rates was chosen empirically to prevent discriminators from learning too quickly and thus contributing progressively little to training the generators. All networks were trained for 100k steps and each type (*i.e.* Split vs. Linked) was trained once for each of 3 random seeds (3, 13, and 42). The geometric mean of the unweighted losses for each network were used to select the best model checkpoint for evaluation. Daisy[18] was subsequently used to efficiently produce predictions on the full 1024^3 volumes. Fake LQ volumes were produced for training and validation datasets, and fake HQ volumes were predicted for the test set. Results are presented for the networks that performed best in segmentation tests.

Segmentation was performed using a multi-task Local Shape Descriptor (LSD) network with a long-range affinity neighborhood, a variation of the original LSD approach[19]. These networks were trained for 100k steps similar to above, but with 196^3-voxel inputs and an Adam optimizer with initial learning rate of $\alpha = 5 \times 10^{-5}$. Following rendering of network predictions with Daisy, long-range affinities were used for attractive and repulsive weights in Mutex watershed[10] to produce final segmentations. Networks were validated on the same type of data they were trained on (i.e. real HQ, real LQ or fake LQ) for the last three checkpoints of training, using Variation of Information (VoI) to select the best performing checkpoints. These variations of the original implementation empirically produced more usable segmentations than classical short-range affinity predictions and seeded watershed with hierarchical agglomeration.

Training, validation, and prediction has been implemented in a manner designed to promote easy re-use and extension for other enhancement and segmentation experiments. Code, data, and documentation are available at github.com/htem/raygun.

Type	Training	Prediction	Merge	Split	Sum
Naive	real HQ	real LQ	5.598	**0.061**	5.659
	real HQ	real HQ	4.565	0.537	5.102
Paired	real LQ	real LQ	**3.950**	0.686	4.636
Linked	fake LQ	real LQ	3.955	0.693	4.648
	real HQ	fake HQ	4.066	0.567	**4.633**
Split	fake LQ	real LQ	4.504	0.611	5.115
	real HQ	fake HQ	4.089	0.652	4.741

Table 1: **Quantitative Segmentation Results.** Variation of Information scores on the test volume for the described approaches. Results for best performing networks are shown, with best overall results for each measure in bold. HQ = High-quality (30nm) and LQ = Low-quality (90nm). "Sum" corresponds to the sum of split and merge scores. Naive high-on-high, trained on a cutout from HQ volume A, and predicted on a cutout from HQ volume B, performed poorly, reflecting the differences in the image statistics between the 2 acquisitions. Naive high-on-low produced the best split score by severely undersegmenting. Despite the Split CycleGAN's ablation, the fake HQ it produced still outperformed both naive methods.

3. RESULTS

3.1. Linked vs. Split CycleGANs

We applied our Split CycleGAN to the satellite2maps dataset used in the original CycleGAN paper[6]. The original authors noted the phenomenon we also observed, that details such as cars and trees could be recovered from the fake map images by the Linked CycleGAN generators, despite no obvious details depicting their presence in the map images. Our goal was to verify that the "splitting" of the loss prevented this. As seen in Fig. 2b, disrupting the information flow from the L1 loss to the first generator in the sequence resulted in a loss of cycling performance on the satellite2maps dataset.

3.2. Baselines

We chose two baselines meant to represent the "ceiling" and "floor" of potential results, corresponding to what we expected to be the ideal scenario, training segmentation on real HQ and predicting on real HQ (Naive high-on-high), and the naive approach of training on real HQ and predicting on real LQ (Naive high-on-low), respectively. As expected, the naive high-on-low approach performed worst overall, as measured by the sum of Variation of Information (VoI)[20] split and merge error scores (5.659), and so we take it to accurately represent the "floor" of the tested methods (see Table 1). Though this approach did perform best regarding split errors (0.061), this is unsurprising considering that classifying every voxel as background would receive a perfect score of 0. That is, while a high split error score can indeed correspond to oversegmentation, a low score can similarly correspond to undersegmentation.

Our intended "ceiling" performed poorly, second worst for merge errors (4.565) (naive high-on-low was worst with 5.598) and third worst overall (5.102). However, this is likely due to the fact that the training and test volumes are from independent acquisitions (HQ volumes A and B respectively), and so vary in image statistics, despite being taken from the same sample, on the same device, at the same resolution, within hours of one another. This is a testament to how difficult it is for trained segmentation networks to generalize.

3.3. Paired

Here, our "paired" approach corresponds to a use case in which researchers are able to gather both a large LQ volume and a HQ subvolume for annotation and training. This allows GT to be constructed using HQ data, then used to train segmentation networks on real LQ data. This use case is restricted, however, by the necessity of paired acquisitions for every large volume. As expected, this approach performed well, achieving the best merge score (3.95), and second best overall VoI score (4.636).

3.4. Unpaired #1: Training segmentation on fake LQ

As expected, fake LQ training volumes produced by both Split and Linked CycleGANs were more similar to the real LQ data than was the real HQ, as measured by lower normalized root mean squared errors (NRMSE), higher peak signal to noise ratios (PSNR), and higher structural similarity indices (SSIM) (see Table 2). The Linked CycleGAN performed best on all of these measures and similarly produced the third best overall VoI score (4.648).

Volume	Target	Test	NRMSE	PSNR	SSIM
Train	real LQ	real HQ	0.201	19.170	0.355
		Linked: fake LQ	**0.133**	**22.726**	**0.602**
		Split: fake LQ	0.158	21.234	0.557
Test	real HQ	real LQ	**0.213**	**18.664**	**0.379**
		Linked: fake HQ	0.264	16.777	0.303
		Split: fake HQ	0.265	16.766	0.279

Table 2: **Comparing Generated Images.** Normalized Root Mean Squared Error (NRMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) scores comparing the different image volumes used for training and prediction. Results for best performing networks are shown, with best overall results in bold. HQ = High-quality (30nm) and LQ = Low-quality (90nm). On the training volume (HQ volume A), both CycleGANs produced fake LQ more similar to the real LQ than was the real HQ fed to the generators. On the test volume (HQ volume B), the real LQ was more similar to the real HQ by all metrics, compared to the fake HQ data. This likely reflects that the low2high generators were trained to produce fake HQ matching the image statistics from HQ volume A, which differ from that of HQ volume B (a separate acquisition).

3.5. Unpaired #2: Predicting segmentation on fake HQ

For the test cutout, the real LQ data was more similar to the real HQ data than were any of the CycleGAN HQ predictions (see Table 2). However, as noted above, the training and test cutouts come from independent HQ acquisitions, and so it is not surprising that the CycleGAN HQ predictions did not end up matching the HQ data on the test cutout. The fake HQ was produced by networks trained to match the image statistics of the same data used for training the HQ segmentation network, so it is also not surprising that the fake HQ produced some of the best segmentation results. Indeed, we see that the VoI scores for segmentation of the fake HQ is better than that of the real HQ data, with the fake HQ from the Linked CycleGAN producing the best overall of the tested methods (4.633), followed closely by that of the Split CycleGAN (4.741).

4. CONCLUSIONS

The Linked CycleGAN (*i.e.* original) performed well facilitating segmentation in both enhanced and native regimes. This indicates that, at least in the tested case, any feature-embedding generators do during training did not seem to detract from test performance. However, it is worth noting that the fake HQ produced by our Split CycleGAN followed close behind the paired approach and the results from the Linked HQ predictions. It is possible the split approach might outcompete the linked in other use cases.

Use of GANs for improving imaging yields for the purposes of gathering biological data is a blossoming field, but we must take great care to avoid inadvertently labeling network hallucinations as scientific findings. Ablations, such as that producing the Split CycleGAN, which intentionally limit the capacity of networks to embed hidden features, may serve a vital role in preventing misuse of deep learning technologies in biological research. We have demonstrated that, despite such an ablation, networks can still allow for a 64-fold improvement in imaging speed, which will facilitate acquisition of otherwise unmanageably large datasets.

5. COMPLIANCE WITH ETHICAL STANDARDS

Harvard Medical School Institutional Animal Care and Use Committee (IACUC) approved the procedures in accordance with guidelines of the National Institutes of Health (NIH) (IS00000753).

6. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Graduate Research Fellowship (2140743) and Janelia's Visiting Scientist Program for JLR, by NIH (MH128949) to WCAL and AP, and by the European Research Council (852455) to AP. We acknowledge ESRF for granting beamtime for the proposal LS2892. AS was supported by the Howard Hughes Medical Institute, the Waitt Foundation, Core Grant application NCI CCSG (CA014195), NIH (R21 DC018237), NSF NeuroNex Award (2014862).

7. REFERENCES

[1] Michał Januszewski and Viren Jain, "Segmentation-Enhanced CycleGAN," Feb. 2019.

[2] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods*, vol. 15, no. 12, pp. 1090–1097, Dec. 2018.

[3] Larissa Heinrich, John A Bogovic, and Stephan Saalfeld, "Deep learning for isotropic Super-Resolution from Non-Isotropic 3D electron microscopy," June 2017.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-Image translation with conditional adversarial networks," Nov. 2016.

[5] Amit Suveer, Anindya Gupta, Gustaf Kylberg, and Ida-Maria Sintorn, "Super-Resolution reconstruction of transmission electron microscopy images using deep learning," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 548–551.

[6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired Image-to-Image translation using Cycle-Consistent adversarial networks," Mar. 2017.

[7] Erik C Johnson, Luis M Rodriguez, Raphael Norman-Tenazas, Daniel Xenes, and William R Gray-Roncal, "Transfer learning analysis of image processing workflows for electron microscopy datasets," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, Nov. 2019, pp. 1197–1201.

[8] Linjing Fang, Fred Monroe, Sammy Weiser Novak, Lyndsey Kirk, Cara R Schiavon, Seungyoon B Yu, Tong Zhang, Melissa Wu, Kyle Kastner, Alaa Abdel Latif, Zijun Lin, Andrew Shaw, Yoshiyuki Kubota, John Mendenhall, Zhao Zhang, Gulcin Pekkurnaz, Kristen Harris, Jeremy Howard, and Uri Manor, "Deep learning-based point-scanning super-resolution imaging," *Nat. Methods*, vol. 18, no. 4, pp. 406–416, Apr. 2021.

[9] Linkai Peng, Li Lin, Pujin Cheng, Ziqi Huang, and Xiaoying Tang, "Unsupervised domain adaptation for Cross-Modality retinal vessel segmentation via disentangling representation style transfer and collaborative consistency learning," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Mar. 2022, pp. 1–5.

[10] Steffen Wolf, Constantin Pape, Alberto Bailoni, Nasim Rahaman, Anna Kreshuk, Ullrich Köthe, and Fred A Hamprecht, "The mutex watershed: Efficient, parameter-free image partitioning," in *Computer Vision – ECCV 2018*, Lecture notes in computer science, pp. 571–587. Springer International Publishing, Cham, 2018.

[11] Aaron T Kuan, Jasper S Phelps, Logan A Thomas, Tri M Nguyen, Julie Han, Chiao-Lin Chen, Anthony W Azevedo, John C Tuthill, Jan Funke, Peter Cloetens, Alexandra Pacureanu, and Wei-Chung Allen Lee, "Dense neuronal reconstruction through x-ray holographic nano-tomography," *Nat. Neurosci.*, vol. 23, no. 12, pp. 1637–1643, Dec. 2020.

[12] P Cloetens, W Ludwig, J Baruchel, D Van Dyck, J Van Landuyt, J P Guigay, and M Schlenker, "Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays," *Appl. Phys. Lett.*, vol. 75, no. 19, pp. 2912–2914, Nov. 1999.

[13] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien P W Pluim, "elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[14] Denis P Shamonin, Esther E Bron, Boudewijn P F Lelieveldt, Marion Smits, Stefan Klein, Marius Staring, and Alzheimer's Disease Neuroimaging Initiative, "Fast parallel image registration on CPU and GPU for diagnostic classification of alzheimer's disease," *Front. Neuroinform.*, vol. 7, pp. 50, 2013.

[15] Kevin M Boergens, Manuel Berning, Tom Bocklisch, Dominic Bräunlein, Florian Drawitsch, Johannes Frohnhofen, Tom Herold, Philipp Otto, Norman Rzepka, Thomas Werkmeister, Daniel Werner, Georg Wiese, Heiko Wissler, and Moritz Helmstaedter, "webknossos: efficient online 3D data annotation for connectomics," *Nat. Methods*, vol. 14, no. 7, pp. 691–694, July 2017.

[16] Jan Funke, "gunpowder: A library to facilitate machine learning on multi-dimensional images," .

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, pp. 234–241, Springer International Publishing.

[18] Tri Nguyen and Caroline Malin-Mayor and William Patton and Jan Funke, "daisy: Block-wise task scheduling for large nd volumes," 2022.

[19] Arlo Sheridan, Tri M Nguyen, Diptodip Deb, Wei-Chung Allen Lee, Stephan Saalfeld, Srinivas C Turaga, Uri Manor, and Jan Funke, "Local shape descriptors for neuron segmentation," *Nat. Methods*, vol. 20, no. 2, pp. 295–303, Feb. 2023.

[20] Marina Meilă, "Comparing clusterings—an information based distance," *J. Multivar. Anal.*, vol. 98, no. 5, pp. 873–895, May 2007.