UNPAIRED IMAGE AND SEGMENTATION ENHANCEMENT OF X-RAY HOLOGRAPHIC NANOTOMOGRAPHY FOR CONNECTOMICS

Jeffrey L. Rhoades* Tri Nguyen* Mukul Narwani* Brian Reicher†
Aaron T. Kuan* Alexandra Pacureanu‡ Wei-Chung Lee* Jan Funke$^{\oplus}$

* Harvard Medical School, Cambridge, MA
† Northeastern University, Boston, MA
‡ European Synchrotron Radiation Facility, Grenoble, France
$^{\oplus}$ HHMI Janelia, Ashburn, VA

ABSTRACT

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words, and should be identical to the abstract text submitted electronically along with the paper cover sheet. All manuscripts must be in English, printed in black ink.

Index Terms**—** One, two, three, four, five

1. INTRODUCTION

While instance and semantic segmentations play a crucial role in many biological imaging studies and clinical applications, generating high-quality (HQ) segmentations remains time-consuming and resource constrained. The time intensive nature of ground truth (GT) generation for training segmentation networks has previously been compounded by a need to create GT for every new dataset, due to variations in imaging and sample preparation conditions between acquisitions. Additionally, segmentation had previously been restricted to the imaging regime in which the GT was generated, constraining the ability to segment datasets which may be particularly laborious for human annotators.

Recent work aims to allow segmentation strategies to be trained on one dataset, but be applied on another, similar dataset, without new GT generation [1]. However, this has generally focused on data in very proximal regimes - that is, data imaged at similar resolutions in similar manners, using similar preparation methods. Other work from denoising literature has begun to present impressive results inferring HQ images from low-quality (LQ) data [2, 3]. More recent work examines the potential for domain adaptation of GT datasets, for instance, via disentangling ”content” and ”style” features of images[**?**]. With the following work, we recognize the potential to not only infer HQ images from LQ data, but to generate HQ segmentations from LQ data, without novel GT annotations. In our imaging regime, X-ray Holographic Nanotomography (XNH), this has the added benefit of allowing for a dramatic increase in imaging throughput.

Variation in image quality can result from many factors, such as the effective pixel size, imaging dwell time, source coherence, sample stability, or the quantum efficiency of the camera sensor. In our case, by increasing the effective pixel size of our imaging 3-fold, from 30nm to 90nm, we are able to increase imaging throughput more than 27-fold, due to the increased field of view (FOV). Additionally, instead of acquiring holograms at multiple sample-to-sensor distances, as is typical for XNH to better constrain the inverse reconstruction problem, we can limit our acquisition to a single distance. Combined, this results in an 108-fold increase in throughput for the data we present. However, many important structures, such as unmyelinated neurites, become significantly more difficult to distinguish at lower magnifications, making manual annotation for GT generation onerously cumbersome. Thus, we sought to explore the possibility of inferring important, HQ features in the latent space of LQ data. ??INCLUDE FIGURE 1 FOR DATA EXAMPLES?

2. RELATED WORK

Original cycle-GAN: [4]

Distinction between paired and unpaired methods.

2.1. Paired Methods

2.1.0.1. Image Restoration

Content-Aware Reconstruction (CARE): [2]

Pix2Pix

SUPER-RESOLUTION RECONSTRUCTION OF TRANSMISSION ELECTRON MICROSCOPY IMAGES USING DEEP LEARNING: [5]

Correspondence: rhoades@g.harvard.edu

A neural lens for super-resolution biological imaging (seems like a low quality paper): [6]

2.1.0.2. *Segmentation*

2.2. Unpaired Methods

CycleGAN

2.2.0.1. *Image Restoration*

Deep Learning for Isotropic Super-Resolution from Non-isotropic 3D Electron Microscopy: [3]

Transfer Learning Analysis of Image Processing Workflows for Electron Microscopy Datasets: [7]

PSSR: [8]

2.2.0.2. *Segmentation*

Segmentation Enhanced Cycle Consistent Network: [1]

- very similar to our method
- translation learning limited to available GT, might be problematic if only little GT available
- not applied to X-ray of different resolution, but to EM of similar resolution

SECGAN uses a cyclegan to translate raw EM data from a domain with no segmentation ground truth to one with plenty of ground truth and a trained segmentation network. It achieves this by adding a loss signal from a discriminator designed to discriminate between segmentations produced from either real data or data generated from the cycleGAN constraining the network to produce images that look identical to the segmentation network. The downside of this approach is that it encourages the cycleGAN to embed features for the segmentation network to use (refer to satellite-map issue?) in making predictions (Can provide a sentence on our SplitGAN architecture here). [check this] They found that the biases introduced by the network averaged out over long FOVs and would only be a cause for concern with narrow FOVs. And discrepancies in resolution between source and target volumes degraded the performance of SECGAN which we aim to alleviate with our approach.

3. METHODS

In training pairs of generators, in this case deep convolutional neural networks, to be cycle-consistent in transferring images between high- and low-quality regimes, two potentially useful generators are created: 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 high-quality (HQ) images with ground truth (GT) to segment low-quality (LQ) images.

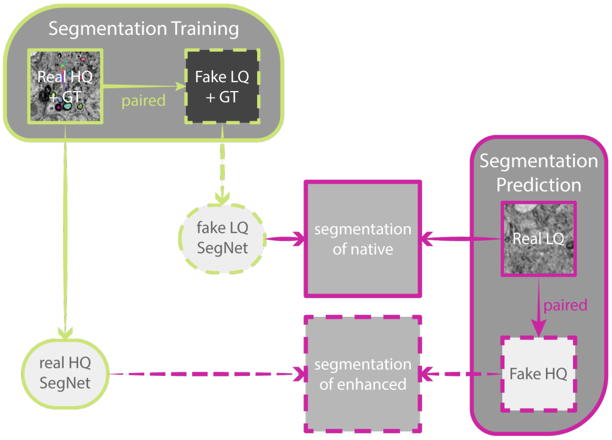

Fig. 1. Schematic of proposed segmentation approaches.

3.1. Segmentation of Enhanced Data

Perhaps the more intuitive approach utilizes the low2high quality generator to convert real LQ data into *fake* HQ images. Segmentation approaches trained on existing HQ data with GT can then be easily leveraged to segment the *fake* HQ renderings. In theory, this domain adaptation approach should allow for segmentation of novel data without retraining segmentation networks or generating additional GT.

3.2. 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, and then train segmentation approaches 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 approach to determine whether the segmentation task might be more easily learned than the regression required for predicting a HQ image from a LQ input.

3.3. Split-Loss CycleGAN Modification

In addition to the above experiments, we chose to examine the potential harm or benefit of a modification to the well-known CycleGAN's loss function[4]. In the original formulation, the L1 loss for both the high2low2high and low2high2low data paths, the portion of the loss most directly responsible for the cycle-consistency, is used in calculating the gradients for both the high2low and low2high generators. Put another way, the generators are both incentivized to embed information about the original, or *real* images into the intermediate, or *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 quite profound and honestly a bit impressive (see Figure 2b), but we wondered if it would aid us

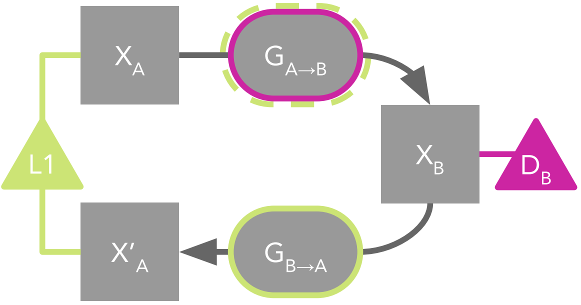

Fig. 2. CycleGAN loss visualization of the original version ("linked") and our variant ("split").

in our pursuit of enhancing images, or at least enhancing their segmentation. Thus, 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*. 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. As seen in Figure 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.

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 training networks for segmentation of native data, this might allow networks to segment *fake*-LQ images based on embedded features that are not present in actual LQ images (i.e. the test set). Furthermore, such embedded features could allow a low2high generator a shortcut to minimizing the L1-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 often 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. An example would be a dense field of myelinated axons getting 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.

3.4. Datasets

Three separate image volumes from a single sample of mouse cerebellum lobule V, stained with 2% Osmium, were acquired at the European Synchrotron Facility's (ESRF) id16a beamline. One volume was produced by reconstruction[9] from holograms imaged from 1300 angles, at a single sample-to-sensor distance resulting in 90nm/voxel images. This 2048x2048x2048 90nm ”overview” scan contained both of the following, higher-resolution volumes, which we will assign letters A and B for ease of reference. The 3216x3216x2048 30nm volumes, A and B, were each reconstructed from holograms imaged from 1900 angles, at 4 different sample-to-sensor distances. (Use of multiple distances in holography better constrains the inverse problem, thereby improving reconstruction quality.)

Volumes were subdivided and aligned, and the 90nm volume linearly interpolated to match the 30nm volumes, resulting in 3 separate 1024x1024x1024 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 ground truth (GT) segmentations were produced by an iterative process of 1) manually painting voxels using WebKnossos, 2) training affinity prediction networks to produce denser segmentations, 3) correcting these segmentations with manual tracing and additional voxel painting, then 4) repeating from step (2). GT for validation and testing cutouts, taken from volume B, were produced by manual skeleton tracing using WebKnossos. These skeletons were then rasterized into image volumes for Variation of Information (VoI) comparisons to segmentations produced by the tested networks.

3.5. Implementation Details

Cycle-Consistent Generative Adversarial Network models were implemented in PyTorch, using Gunpowder to dynamically resample and augment training images, save and monitor progress. Valid 2D UNet[**?**] generators with 3 downsampling steps were fed batches of 512x512-pixel slices, which were randomly rotated, flipped, and warped to increase diversity of the training set. PatchGAN Discriminators[**?**] 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 rate were chosen empirically to prevent discriminators from learning to quickly and thus contributing progressively little to training the generators. All networks were trained for 100k steps on single NVIDIA A100 GPUs, and each type (i.e. Split vs. Linked) was trained 3 times from 3 separate random seeds. The geometric mean of the unweighted losses for each network were used to select the best model checkpoint for evaluation. Daisy[**?**] was subsequently used to efficiently produce predictions on the full 1024x1024x1024 volumes. *Fake* LQ volumes were produced for training and validation datasets, and and *fake* HQ volumes were predicted for the test set. Results are presented for the networks that produced the best results in segmentation tests.

Multi-task Local Shape Descriptor and long-range pixel-affinity prediction networks[?] consisted of valid 3D UNets trained with PyTorch and Gunpowder, as described above, but with 196x196x196-voxel volumes and an Adam optimizer with initial learning rate of $\alpha = 5 \times 10^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$. All networks were trained for 100k steps on single NVIDIA A100 GPUs. Segmentations were produced by Mutex Watershed[?] agglomeration of the long-range affinity predictions, following rendering of network predictions with Daisy, similar to above. 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 (i.e. 90k, 95k, and 100k training steps), using the sum of VoI merge and split scores.

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

4. RESULTS

4.1. 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, and the naive approach of training on real HQ and predicting on real LQ, respectively. As expected, the naive approach performed worst overall, as measured by the sum of Variation of Information (VoI) split and merge error scores, and so we take it to accurately represent the ”floor” of the tested methods. It is worth noting, that the naive approach did perform best regarding split errors, but this is unsurprising considering that a score of 0 could be achieved for this measure by simply classifying every voxel as separate instances. That is, while a high score in this measure can indeed correspond to oversegmentation, a low score can similarly correspond to an undersegmentation.

Our intended ”ceiling” performed remarkably poorly, second worst for merge errors (naive was worst) and third worst overall. Upon further reflection, however, we realize this is well explained by the fact that the test and training volumes are from independent acquisitions, 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.

4.2. Paired

Here, our ”paired” approach corresponds to a use case in which researchers are able to gather both a large LQ volume (our 90nm/voxel 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 LQ data. This use case is restricted, however, by the necessity of paired acquisitions for every large volume. As expected, this approach performed particularly well, achieving the best merge score, and second best overall VoI score.

Type	Training	Prediction	Merge	Split	Sum
Naive	real HQ	real LQ	5.598	**0.061**	5.659
Ideal	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. Variation of Information scores on the test volume for the described approaches. (Best results are in bold.)

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. Normalized Root Mean Squared Error (NRMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) scores are given comparing the different image volumes used for training and prediction. (Best results are in bold.)

4.3. Unpaired #1: Training Segmentation on Fake LQ

As expected, *fake* LQ 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). The Linked CycleGAN performed best on all of these measures, of the three, and similarly produced the third best overall VoI score. Note that similarity scores were computed for the segmentation training cutout, as these fake LQ volumes were used for training segmentation networks.

4.4. Unpaired #2: Predicting Segmentation on Fake HQ

Interestingly, for the test cutout, the real LQ data was more similar to the real HQ data. As noted above, the training and test cutouts come from independent 30nm/voxel acquisitions. The CycleGANs were trained on HQ data from the volume used for the training cutout, and so it is not necessarily surprising that they did not end up matching the HQ data on the test cutout. In light of this, the fact that the fake HQ produced some of the best segmentation results is not surprising. That is, the fake HQ was produced by networks trained to match the image statistics of the same data used for training the HQ segmentation network. 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 VoI.

4.5. Linked vs. Split CycleGANs

In addition to our experiments with XNH data, we applied the Split version of the CycleGAN to one of the datasets used in the original paper[4], translating between satellite images and map images. 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 them being present in the map images. As expected, this effect was not observed in the case of the Split CycleGAN (see Figure 3).

5. CONCLUSIONS

5.1. Conclusions

As seen in Table 1, the linked (i.e. original) CycleGAN formulation 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

5.1.1. Sub-subheadings

Sub-subheadings, as in this paragraph, are discouraged. However, if you must use them, they should appear in lower case (initial word capitalized) and start at the left margin on a separate line, with paragraph text beginning on the following line. They should be in italics.

6. PRINTING YOUR PAPER

If the last page of your paper is only partially filled, arrange the columns so that they are evenly balanced if possible, rather than having one long column.

In LaTeX, to start a new column (but not a new page) and help balance the last-page column lengths, you can use the command “\pagebreak” as demonstrated on this page (see the LaTeX source below).

7. PAGE NUMBERING

Please do **not** paginate your paper. Page numbers, session numbers, and conference identification will be inserted when the paper is included in the proceedings.

8. ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. If you use color, make sure that the color figures are clear when printed on a black-only printer.

Since there are many ways, often incompatible, of including images (e.g., with experimental results) in a LaTeX document, below is an example of how to do this [10].

9. FOOTNOTES

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

10. COPYRIGHT FORMS

You must include your fully completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings. The copyright form is available as a Word file, a PDF file, and an HTML file. You can also use the form sent with your author kit.

11. REFERENCING

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [11].

12. COMPLIANCE WITH ETHICAL STANDARDS

IEEE-ISBI supports the standard requirements on the use of animal and human subjects for scientific and biomedical research. For all IEEE ISBI papers reporting data from studies involving human and/or animal subjects, formal review and approval, or formal review and waiver, by an appropriate institutional review board or ethics committee is required and should be stated in the papers. For those investigators whose Institutions do not have formal ethics review committees, the principles outlined in the Helsinki Declaration of 1975, as revised in 2000, should be followed.

Reporting on compliance with ethical standards is required (irrespective of whether ethical approval was needed for the study) in the paper. Authors are responsible for correctness of the statements provided in the manuscript. Examples of appropriate statements include:

- "This is a numerical simulation study for which no ethical approval was required."
- "This research study was conducted retrospectively using human subject data made available in open access by (Source information). Ethical approval was not required as confirmed by the license attached with the open access data."
- "This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of University B (Date.../No. ...)."

13. ACKNOWLEDGMENTS

IEEE-ISBI supports the disclosure of financial support for the project as well as any financial and personal relationships of the author that could create even the appearance of bias in the published work. The authors must disclose any agency or individual that provided financial support for the work as well as any personal or financial or employment relationship between any author and the sources of financial support for the work.

Other types of acknowledgements can also be listed in this section.

Reporting on real or potential conflicts of interests, or the absence thereof, is required in the paper. Authors are responsible for correctness of the statements provided in the manuscript. Examples of appropriate statements include:

- "No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose."
- "This work was supported by [. . .] (Grant numbers) and [. . .]. Author X has served on advisory boards for Company Y."
- "Author X is partially funded by Y. Author Z is a Founder and Director for Company C."

14. REFERENCES

[1] Michał Januszewski and Viren Jain, "Segmentation-enhanced cyclegan," *bioRxiv*, p. 548081, 2019.

[2] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nature methods*, vol. 15, no. 12, pp. 1090–1097, 2018.

[3] Larissa Heinrich, John A Bogovic, and Stephan Saalfeld, "Deep learning for isotropic super-resolution from non-isotropic 3d electron microscopy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 135–143.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[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)*. IEEE, 2019, pp. 548–551.

[6] James A Grant-Jacob, Benita S Mackay, James AG Baker, Yunhui Xie, Daniel J Heath, Matthew Loxham, Robert W Eason, and Ben Mills, "A neural lens for super-resolution biological imaging," *Journal of Physics Communications*, vol. 3, no. 6, pp. 065004, 2019.

[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*. IEEE, 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, et al., "Deep learning-based point-scanning super-resolution imaging," *Nature methods*, vol. 18, no. 4, pp. 406–416, 2021.

[9] 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, et al., "Dense neuronal reconstruction through x-ray holographic nano-tomography," *Nature neuroscience*, vol. 23, no. 12, pp. 1637–1643, 2020.

[10] A.B. Smith, C.D. Jones, and E.F. Roberts, "Article title," *Journal*, vol. 62, pp. 291–294, January 1920.

[11] C.D. Jones, A.B. Smith, and E.F. Roberts, "Paper title," in *Proceedings Title*. IEEE, 2003, vol. II, pp. 803–806.

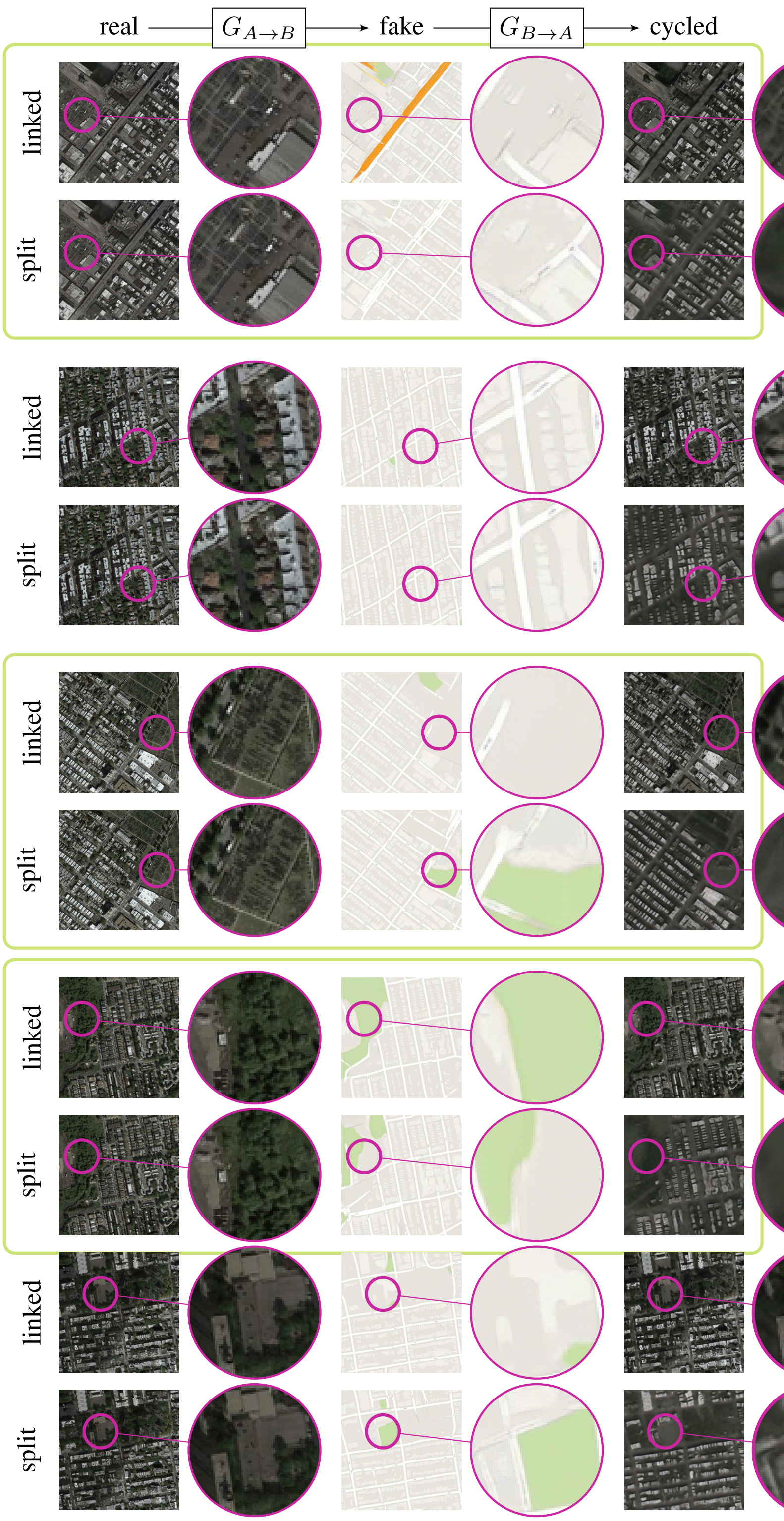

Fig. 3. Comparison of cycled images using the original CycleGAN ("linked") and the variation we propose ("split") on the example of satellite and map images. The **linked** version of the CycleGAN (*i.e.*, the original CycleGAN) encourages the generators to collaborate, which leads to the embedding of detail in the fake map images generated by the $G_{A\to B}$ generator. This detail is then used by the $G_{B\to A}$ generator, to recover the original satellite image, even filling in details that should not be recoverable from a map image. The **split** variant (see Section 3.3 and Figure 2) does not propagate gradient