2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

CAESAR: SEGMENT-WISE ALIGNMENT METHOD FOR SOLVING DISCONTINUOUS DEFORMATIONS

Sergiy Popovych[⋆] *J. Alexander Bae*[⋆] *H. Sebastian Seung*[⋆]

[⋆] Princeton University, Princeton NJ 08544, USA

ABSTRACT

Images obtained from serial section electron microscopy can contain defects that create discontinuous tissue deformation. Fixing such defects during image registration is especially challenging, as classical block matching registration techniques assume smooth motion within each block, and ConvNet based registration techniques must rely on smoothness assumption during training.

We propose Caesar, a divide-and-conquer technique that breaks registered images into segments, such that most of discontinuity is confined to segment boundaries. Then, we align the segments independently and stitch the results back together. We provide extensive experimental evaluation on brain tissue serial section microscopy data that shows that segment-wise alignment reduces the average misalignment area around defects by 6-10x.

1. INTRODUCTION

In serial section electron microscopy (ssEM), a brain volume is cut into a series of ultrathin sections, each of which is imaged via electron microscopy. Then the 2D images are aligned to create a 3D image stack, from which the connections between neurons can be reconstructed.[1] A caveat is that image alignment can be surprisingly challenging for several reasons. First, ssEM images have become very large and unwieldy. Second, the deformations of the images are nonaffine, due to physical deformations of the sections as well as distortions introduced by imaging. Third, highly accurate alignment is required for automated segmentation of neurons. A thin axon can be less than 100 nm in diameter, and a misalignment greater than 100 nm can cause the axon to become discontinuous within the image stack, which typically leads to an error in the automated segmentation.

The difficulty of aligning serial section images motivated the development of block face electron microscopy (bfEM)[3, 4, 5]. In this approach, ultrathin sections are sequentially removed from the block. Each time a fresh block face is exposed, and is imaged by scanning EM. Because the block deforms so little, bfEM yields a series of 2D images that require relatively little processing to align into a 3D stack. While bfEM has many strong points, ssEM has remained popular [6], aided by rapid improvements in the state of the art in image alignment. There are a number of software packages for ssEM image alignment, including TrakEM2 [7], AlignTK [8], NCR Tools [9], FijiBento [10], and EM_aligner [11]. The packages are typically based on some method of identifying correspondences between pairs of image points, as well as some method of combining these correspondences to produce a global alignment. The packages are capable of aligning smoothly deformed images, but fail with discontinuous deformations.

Even when cut by an expert, ultrathin sections contain occasional cracks and folds (Fig. 1). In the worst case, a series may contain numerous cracks and folds in virtually every section, in which case the series is typically discarded as useless. On either side of a crack, the section deforms away from the crack. On either side of a fold, the section deforms toward the fold. These deformations lead to misalignment. If one applies the above-mentioned software packages, the misalignment is eliminated far away from the crack or fold. However, a misaligned region surrounding the crack or fold still remains, and this region can be quite wide (Fig. 2).[2] If such misalignments could be corrected, ssEM datasets with many cracks and folds would become more useful for science.

The problem is that the deformation is discontinuous at a crack or fold, and the existing alignment methods assume that the deformation is smooth. Recently proposed ConvNet-based alignment methods[12] can potentially correct wider range of deformation types. ConvNet-based alignment methods, as inspired by recent work in optical flow prediction[13, 14], rely on brightness constancy and motion smoothness assumptions during training. The smoothness assumption penalizes the network for predicting non-smooth displacement fields, which is necessary for successful self-supervised training, but is not appropriate for cracks and folds.

In this work, we propose segmenting the image at the cracks and folds, aligning the segments independently, and then stitching the results back together. We call the proposed method Caesar. Because motion of each side of the defect is

[1] ssEM is also applied to other biological tissues, but we are particularly interested in the application to reconstructing neuronal connections in brain tissue. Serial sectioning is also used with light microscopy [1, 2] as well as EM, and our work may be relevant in this context also.

[2] Some image data is also "devoured" at a fold. Our goal is not to recover such missing data, but to correct the misalignment surrounding a fold.

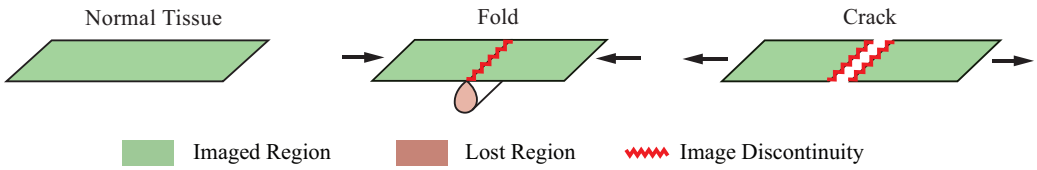


Fig. 1: Crack and Fold Mechanics

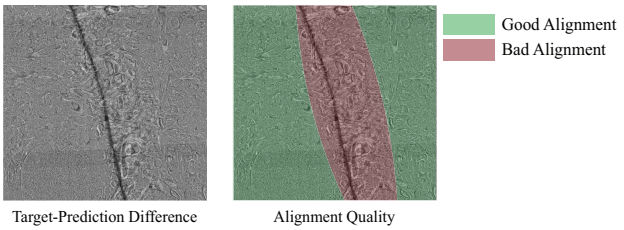
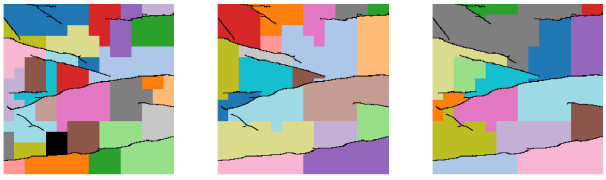


Fig. 2: Illustration of a fold causing a large area of misalignment. Images show pixel-wise difference between the source and the target image, with brighter pixels representing large difference, darker values representing and black values representing the fold itself. Uniformly gray areas far away from fold are well aligned, and high contrast areal around the fold are aligned poorly.



(a) Pad 1% Image Width **(b)** Pad 3% Image Width **(c)** Pad 7% Image Width

Fig. 3: Segmentation Examples for different padding thresholds. Larger padding values lead to bigger segments. Best seen in color. Fold masks are gathered from EM images, 1024 x 1024 in resolution of 1024 x 1024 nm^2.

continuous, aligning one side at a time does not require the aligner to predict discontinuous motion, which simplifies the alignment task. We show that a ConvNet can be trained to successfully predict discontinuity locations in images. Based on discontinuity locations predicted by the ConvNet we break the image into segments such that most of the discontinuity is confined to segment boundaries. The chosen alignment method can then be applied to each segment independently. This is similar to the approach commonly taken in semantic segmentation, where an image is first divided into segments delineating object boundaries, and then each segment is classified independently in order to identify the object within the segment. We provide empirical evaluation of segment-wise alignment method on a brain EM serial section data set with a large number folds. Although our empirical evaluation mostly focuses on folds, not cracks, the proposed technique can be applied to cracks without loss of generality. We show that Caesar can reduce the average width of misalignment areas by 6-10x.

2. METHODS

2.1. Detection

Folds have distinctive shape and color in the EM image as they are expressed as dark lines. Therefore, it is possible to train a ConvNet to detect the folds automatically. Fold detector (FoldNet) architecture is a simplified version of U-Net [15]. The architecture consists of 9 levels, with 3 convolutions each. The fold detector takes EM image as an input and outputs likelihood of fold for each pixel.

Ground truth data consists of 2400 EM images of 1024 x 1024 pixels each downsampled to 16 nm pixel resolution. 1600 of the samples were used for training, 400 of the samples were used for validation, and 400 of the samples were used for the test set. Binary cross entropy was used as the loss function in the training. Flip and rotation augmentation were introduced to represent diverse inputs.

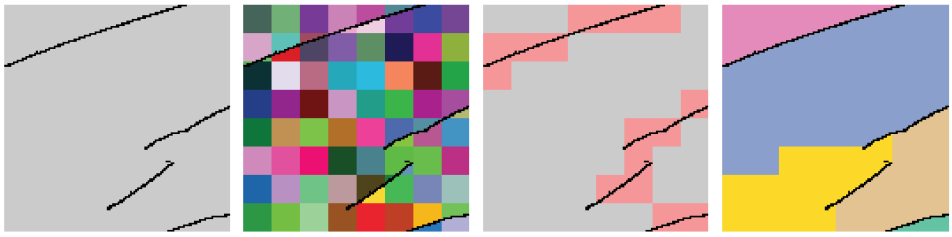
2.2. Segmentation

After identifying the defect locations with a ConvNet, we divide the image into segments. The goal is to produce a segmentation such that no segment will cross across a defect boundary. The tissue within each segment is going to be aligned independently, with padding. Padding is necessary to produce consistent alignment across segment boundaries.

2.3. Acknowledgements

After identifying the defect locations with a ConvNet, we divide the image into segments. The goal is to produce a segmentation such that no segment will cross across a defect boundary. The tissue within each segment is going to be aligned independently, with padding. Padding is necessary to produce consistent alignment across segment boundaries.

Because existing alignment methods expect a rectangular image as input, aligning each segment is computationally equivalent to aligning the bounding box of the segment. If two segments have overlapping bounding boxes, the amount of overlap is going to correspond to the amount of repeated computation. Good defect segmentation would minimize the computation overhead by generating segments with minimal total amount of bounding box overlap. Total overhead can be computed as the sum of areas of segment bounding boxes di-



(a) Fold Mask (b) Oversegmentation (c) Conflicting Atoms (d) Agglomeration

Fig. 4: Segmentation algorithm walk-through

vided by the total area of the image, which we will refer to as *overlap factor*.

Segment padding plays an important role in defect segmentation. Without padding, an optimal, trivial solution would be to treat each pixel as an independent segment. The area of the bounding box for each pixel would be equivalent to the area of the pixel itself, leading to overlap factor of 1. The padding requirement encourages constructing larger segments and makes trivial solutions prohibitive.

We propose segmenting the image by breaking the image into a large number of small segments (over-segmentation), and then repeatedly merge those segments into larger ones (agglomeration). We generate the over-segmentation by chunking the image with a uniform grid and and identifying connected components within each chunk (Fig. 4 (b)). We will refer to these initial segments as atoms. In order for the agglomeration stage not to merge atoms across defect boundaries, we mark pairs of atoms which are directly separated by a defect as mutually conflicting (Fig. 4 (c)). During agglomeration, two segments can only be merged if none of their atoms are mutually conflicting. Agglomeration stage greedily merges the smallest segment with a neighbor that leads to biggest reduction of the overlap factor. Note that the final agglomeration result is dependent on the padding amount, as larger padding encourages larger segments (Fig. 3).

This segmentation algorithm complexity is $O(W * H)$, where W and H are width and height of the image.

3. EXPERIMENTS

3.1. Fold Detection

Fig. 5 shows example outputs of FoldNet (b). As it is described in the example outputs, baseline fold detector is decent for the folds that are distinctive but it's vulnerable to dimmer folds unlike FoldNet. To evaluate the performance, we dilated the masks first in order to remove the noise in the ground truth. Then we calculated pixel-wise precision-recall, Jaccard index, and Dice index to evaluate the performance. As shown in Fig. 5(c), our FoldNet outperforms the baseline detector with optimum F1 score of 0.861 (1). The difference is more dominant in Jaccard index and Dice index. Table 1 shows, average Jaccard index and Dice index of FoldNet are 0.825 and 0.873 respectively while baseline detector's are 0.329 and 0.386. This is because baseline detector confuses many dark pixels as fold while FoldNet is more robust to this so FoldNet is able to detect folds more exclusively.

Table 1: Fold detection evaluation.

Method	F1	Jac ($\pm$ std)	Dice ($\pm$ std)
Baseline	0.71	0.32 ± 0.36	0.36 ± 0.39
FoldNet*	**0.86**	$\mathbf{0.82 \pm 0.23}$	$\mathbf{0.87 \pm 0.22}$

FoldNet* is the network with optimum threshold in Fig. 5.

3.2. Alignment

We evaluate the effectiveness of segment-wise alignment on coarse alignment of brain tissue EM images. The goal of coarse alignment is to correct the largest displacements in the stack, which is why it is usually done on a heavily downsampled images. In this case, our coarse alignment methods were applied to images downsampled to 1024 x 1024 nm^2 pixel resolution.

We evaluate the method paired with both ConvNet based and block matching based methods. The ConvNet was trained on cutouts from 2000 images 2048 x 2048 px images. Training images contained a large amount of folds, averaging over 24 folds per image. For block matching, tile size and step size parameters were set to the values that produced best MSE between aligned images in the evaluation stack. The evaluation was performed on 600 images, 2048 x 2048 px, averaging 7 folds per image.

We use 2 evaluation metrics – the average width of misalignment area around folds and percentage of folds that could not be resolved perfectly. We find the width of misalignment area by moving focusing on 10 px stripe of tissue around the fold and moving the stripe away until the average pixel error within the stripe is less than or equal to the average pixel error in the whole image. If the stripe does not need to be moved backward, we judge the fold to be resolved perfectly.

When paired with ConvNet ailment, Caesar decreases the number of unresolved defects by 4.5x, while reducing the av-

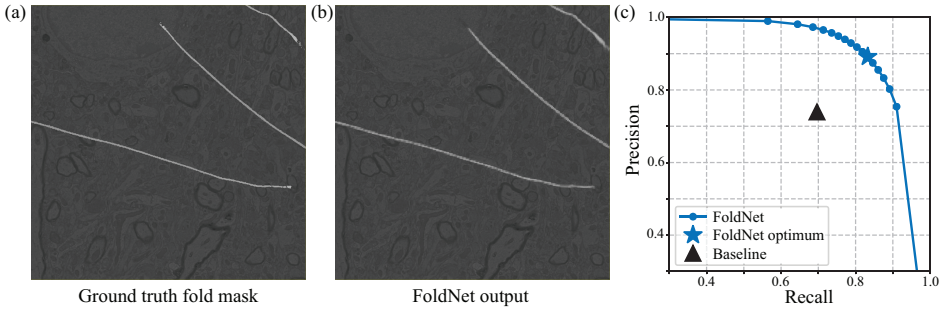


Fig. 5: (a) Ground truth target fold mask. (b) FoldNet output fold mask. (c) Precision-recall curve of FoldNet and Baseline.

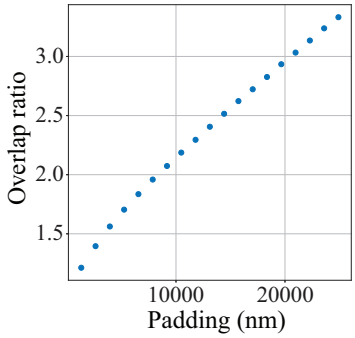


Fig. 6: Dependence between segment padding and overlap ratio.

Table 2: Alignment Quality Comparison

Method	Unresolved Defects	Misalign Width (nm)
ConvAlign	23.9%	$6.1 * 10^4$
ConvAlign + Caesar	4.6%	$1.0 * 10^4$
Improvement	4.5x	6.1x
BlockMatch	72.9%	$2.3 * 10^5$
BlockMatch + Caesar	20.3%	$2.1 * 10^4$
Improvement	3.6x	10.9x

erage misalignment width of the remaining defects by 6.1x, and (Table 2). When paired with block matching, Caesar decreases decreases the number of unresolved defects by 3.6x, the average misalignment width of remaining misalignments by 10.9x. Overall, we find that ConvNet alignment performs better around folds than block matching, both with and without Caesar. This is consistent with our expectations, since ConvNet is more suited for solving discontinuous defects. Although no discontinuous defects need to be solved within a segment with Caesar, block matching still often fails to accurately align thin segments of stretched tissue. The best obtained average misalignment width of remaining defects is $10^4 nm$, which corresponds to 10 pixels at 1024 x 1024 pixel resolution.

3.3. Segmentation

Figure 6 presents the average overlap ratio achieved by the segmentation procedure as a function of the padding amount. The data is obtained from running segmentation on 800 EM 1024 x 1024 px images. Overlap ratio represents the computational overhead introduced by segment-wise alignment method. The padding ranged from 5 px to 25 px at 1024 x 1024 nm resolution. Figure 6 shows that overlap ratio grows linearly with padding, ranging from 1.2x to 3.3x.

4. CONCLUSION

In this work, we present a section-wise method of aligning ssEM images in presence of discontinuous defects. We detect defect locations with a ConvNet, segment the image so that no segment contains tissue on both sides of a single fold, and then align each of the segments independently. Because the segments do not contain discontinuous defects, aligning them is easier both with block matching and with ConvNet methods. We empirically show that segment-wise alignment significantly decreases the misalignment areas around defects.

5. ACKNOWLEDGEMENTS

This work has been supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DoI/IBC) contract number D16PC0005. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/IBC, or the U.S. Government. We also want to thank Allen Institute for Brain Science for producing the data used in our experiments.

6. REFERENCES

[1] Kristina D Micheva and Stephen J Smith, "Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits," *Neuron*, vol. 55, no. 1, pp. 25–36, 2007.

[2] Vadim Pinskiy, Jamie Jones, Alexander S Tolpygo, Neil Franciotti, Kevin Weber, and Partha P Mitra, "High-throughput method of whole-brain sectioning, using the tape-transfer technique," *PloS one*, vol. 10, no. 7, pp. e0102363, 2015.

[3] Winfried Denk and Heinz Horstmann, "Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure," *PLoS biology*, vol. 2, no. 11, pp. e329, 2004.

[4] Graham Knott, Herschel Marchman, David Wall, and Ben Lich, "Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling," *Journal of Neuroscience*, vol. 28, no. 12, pp. 2959–2964, 2008.

[5] Kenneth Jeffrey Hayworth, David Peale, Michal Januszewski, Graham Knott, Zhiyuan Lu, C Shan Xu, and Harald F Hess, "Gcib-sem: A path to 10 nm isotropic imaging of cubic millimeter volumes," *bioRxiv*, p. 563239, 2019.

[6] Jörgen Kornfeld and Winfried Denk, "Progress and remaining challenges in high-throughput volume electron microscopy," *Current opinion in neurobiology*, vol. 50, pp. 261–267, 2018.

[7] Stephan Saalfeld, Richard Fetter, Albert Cardona, and Pavel Tomancak, "Elastic volume reconstruction from series of ultra-thin microscopy sections," *Nature methods*, vol. 9, no. 7, pp. 717, 2012.

[8] Arthur W Wetzel, Jennifer Bakal, Markus Dittrich, David GC Hildebrand, Josh L Morgan, and Jeff W Lichtman, "Registering large volume serial-section electron microscopy image sets for neural circuit reconstruction using fft signal whitening," in *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2016, pp. 1–10.

[9] James R Anderson, Bryan W Jones, Carl B Watt, Margaret V Shaw, Jia-Hui Yang, David DeMill, James S Lauritzen, Yanhua Lin, Kevin D Rapp, David Mastronarde, et al., "Exploring the retinal connectome," *Molecular vision*, vol. 17, pp. 355, 2011.

[10] Maximilian Joesch, David Mankus, Masahito Yamagata, Ali Shahbazi, Richard Schalek, Adi Suissa-Peleg, Markus Meister, Jeff W Lichtman, Walter J Scheirer, and Joshua R Sanes, "Reconstruction of genetically identified neurons imaged by serial-section electron microscopy," *Elife*, vol. 5, pp. e15015, 2016.

[11] Khaled Khairy, Gennady Denisov, and Stephan Saalfeld, "Joint deformable registration of large em image volumes: A matrix solver approach," *arXiv preprint arXiv:1804.10019*, 2018.

[12] Adrian V. Dalca, Guha Balakrishnan, John V. Guttag, and Mert R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *MICCAI*, 2018.

[13] Jason J. Yu, Adam W. Harley, and Konstantinos G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *Computer Vision - ECCV 2016 Workshops, Part 3*, 2016.

[14] Wei-Sheng Lai, Jia-Bin Huang, and Ming-Hsuan Yang, "Semi-supervised learning for optical flow with generative adversarial networks," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., pp. 354–364. Curran Associates, Inc., 2017.

[15] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2015, vol. 9351 of *LNCS*, pp. 234–241, Springer, (available on arXiv:1505.04597 [cs.CV]).