BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023
http://excel.fit.vutbr.cz

Deep Learning for Electron Microscopy Image Stitching

Petr Šilling*

Abstract
Image stitching is an essential technique for reconstructing volumes of biological samples from overlapping tiles of electron microscopy (EM) images. Current volume EM stitching methods generally rely on handcrafted features, such as those produced by SIFT. However, recent developments indicate that convolutional neural networks (CNNs) can improve stitching accuracy by learning discriminative features directly from training images. In this paper, we apply deep learning stitching techniques to volume EM images in an attempt to improve the performance of conventional methods. Experiments on a synthetically generated dataset of volume EM images show overall accuracy similar to SIFT while achieving greater robustness on images with low-quality texture or small overlap regions. The results suggest that deep learning approaches could be beneficial for EM imaging, e.g., by enabling the use of smaller tile overlaps.

*xsilli01@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Image stitching is the process of combining overlapping images into a single image with a wide field of view (FOV). Stitching is crucial for sample reconstruction in volume electron microscopy (EM), where large grids of overlapping images are produced to capture samples that cannot fit under the FOV of a single electron microscope.

Current volume EM stitching methods [1, 2, 3, 4] rely on traditional approaches, such as SIFT [5]. Consequently, they may struggle with repetitive patterns, poor texture, and high-resolution images – all of which are common in volume EM. Hence, accurate stitching often requires large image overlaps, which slows down the speed of imaging and increases data sizes.

Motivated by the above issues and recent developments in feature detection and matching using convolutional neural networks (CNNs) [6, 7, 8], we propose DEMIS[1], a novel EM image stitching tool that utilises LoFTR [8], a deep learning feature matching network. Furthermore, we propose a novel synthetic dataset generated from 424 manually selected and publicly available EM images with high resolution.

We use the synthetic dataset to evaluate DEMIS on common image quality metrics, achieving results similar to SIFT-based approaches, while demonstrating increased robustness on images with low-quality texture, high resolution, or smaller overlaps. The results suggest that while areas for improvement remain, CNNs could be an important asset for reducing overlaps between volume EM images without compromising stitching accuracy.

[1] https://github.com/PSilling/demis

2. Motivation

In volume EM, it is common for large samples to be imaged in parts, forming a grid of overlapping images. The grid must then be stitched to form the final image. Some current methods stitch the grid using different image correlation techniques [2, 3]. However, these methods have the disadvantage of being able to represent only translations accurately [3]. Therefore, in this paper, we focus on methods that employ SIFT [5] features for image stitching [1, 4].

The general SIFT stitching pipeline works as follows [9]. First, SIFT features of the stitched images are detected and described. Second, matches are established between features of overlapping images (e.g., by searching for nearest neighbours). Based on these matches, homographies that relate the overlapping images are estimated using RANSAC [10]. Finally, the homography estimates are refined using global optimisation and the images are stitched together accordingly.

Unfortunately, as can be seen in Figure 1 [11], SIFT may have issues detecting features, e.g., in images with low-quality texture. Hence, in this paper, we propose to replace SIFT with deep learning methods.

3. Synthetic Dataset

To be able to evaluate the proposed solution, we prepare a synthetic dataset by manually selecting 424 distinct high-quality and high-resolution EM images publicly available on EMPIAR [2] or CIL [3]. As shown in Figure 2 [12], each selected image is divided into a grid of overlapping tiles of size 1024×1024 pixels. Additionally, random brightness and contrast changes, random rotation and translation, and Gaussian noise are applied to each tile. Of the resulting 8339 images, 1306 were selected for evaluation purposes.

4. Proposed Solution

The proposed solution, the DEMIS tool, is based on the standard feature-based stitching pipeline described in Section 2. However, it replaces feature detection and matching with LoFTR [8] (explained in Section 5). In particular, DEMIS processes grids of overlapping EM images in the following way.

First, the brightness and contrast of the raw input images are normalised to aid feature detection and mask future image tile boundaries. Second, for each pair of adjacent images, features are detected and matched by LoFTR. The matches are then used to estimate the initial homographies between neighbouring images. For homography estimation, we utilise the default OpenCV[4] implementation based on RANSAC [10].

Subsequently, a SLAM graph is constructed using the `graphslam`[5] library. In the graph, each image tile from the grid is represented by a single vertex placed at the expected tile position (determined by expected overlaps). The adjacent vertices (i.e., image tiles) are then connected by edges corresponding to the homography-induced relative position change. Only rotation and translation parameters can be represented this way and, hence, we estimate only those. Doing so is generally not an issue for EM images.

Finally, after the SLAM graph is optimised, the corrected homographies are extracted from the graph and the grid is stitched accordingly. The whole process is illustrated by Figure 3 [13].

5. Local Feature Transformer (LoFTR)

LoFTR [8] is a Transformer-based [14] network designed to simultaneously detect and match features between pairs of input images. As can be seen in Figure 4 [8], it has four sequential components. First, a CNN extracts coarse- and fine-level feature maps. Second, positionally encoded coarse-level feature maps are processed by the main Transformer module. Next, a differentiable matching layer matches the transformed features, generating a match confidence matrix. Finally, coarse-level matches are selected using a specified threshold and further refined with the help of fine-level feature maps.

6. Results on Evaluation Images

To evaluate DEMIS, we compare its performance on various stitching quality metrics against a baseline that uses SIFT instead of LoFTR. All experiments are run on the evaluation images from Section 3. The results are displayed in Table 1.

The results show that LoFTR finds, on average, 18% more matches than SIFT. The increased match count suggests greater overall robustness of the approach. This is demonstrated on the high-resolution image pairs from Figure 5 and Figure 6 (both provided by TESCAN 3DIM), where SIFT detects significantly fewer matches than LoFTR. Moreover, the matches are more accurate based on mean reprojection errors.

Furthermore, the results show that LoFTR achieves similar performance in terms of homography estimation accuracy and final perceived stitched image quality (measured by PSNR, SSIM [15], FSIM [16], and BRISQUE [17]). This suggests that the increase in robustness does not come at the cost of a performance decrease in other areas.

7. Conclusions

This paper presented DEMIS, a novel EM image stitching tool based on LoFTR feature matching, and a novel synthetic dataset of EM images. Experiments showed greater robustness while achieving accuracy similar to traditional stitching approaches. Future work could focus on creating a more challenging evaluation dataset and on proper domain adaptation of LoFTR to EM imagery.

Acknowledgements

I would like to thank my supervisor Ing. Michal Španěl, Ph.D., and my consultant, Ing. Oldřich Kodym, Ph.D., for their help and professional guidance.

[2] https://www.ebi.ac.uk/empiar/
[3] http://cellimagelibrary.org/
[4] https://opencv.org/
[5] https://github.com/JeffLIrion/python-graphslam/

References

[1] Albert Cardona, Stephan Saalfeld, Johannes Schindelin, Ignacio Arganda-Carreras, Stephan Preibisch, Mark Longair, Pavel Tomancak, Volker Hartenstein, and Rodney J. Douglas. Trakem2 software for neural circuit reconstruction. *PLOS ONE*, 7(6):e38011, June 2012.

[2] Joe Chalfoun, Michael Majurski, Tim Blattner, Kiran Bhadriraju, Walid Keyrouz, Peter Bajcsy, and Mary Brady. Mist: Accurate and scalable microscopy image stitching tool with stage modeling and error minimization. *Scientific Reports*, 7:4988, July 2017.

[3] Jeremy L. Muhlich, Yu-An Chen, Clarence Yapp, Douglas Russell, Sandro Santagata, and Peter K. Sorger. Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ashlar. *Bioinformatics*, 38:4613–4621, October 2022.

[4] Gayathri Mahalingam, Russel Torres, Daniel Kapner, Eric T. Trautman, Tim Fliss, Shamishtaa Seshamani, Eric Perlman, Rob Young, Samuel Kinn, JoAnn Buchanan, Marc M. Takeno, Wenjing Yin, Daniel J. Bumbarger, Ryder P. Gwinn, Julie Nyhus, Ed Lein, Steven J. Smith, R. Clay Reid, Khaled A. Khairy, Stephan Saalfeld, Forrest Collman, and Nuno Macarico da Costa. A scalable and modular automated pipeline for stitching of large electron microscopy datasets. *eLife*, 11:e76534, July 2022.

[5] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[6] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Conference on Computer Vision and Pattern Recognition Workshops*, CVPRW, pages 337–349, Salt Lake City, UT, USA, June 2018. IEEE.

[7] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Conference on Computer Vision and Pattern Recognition*, CVPR, pages 4937–4946, Seattle, WA, USA, June 2020. IEEE.

[8] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *Conference on Computer Vision and Pattern Recognition*, CVPR, pages 8918–8927, Nashville, TN, USA, June 2021. IEEE.

[9] Richard Szeliski. *Computer Vision*. Texts in Computer Science. Springer, Cham, Switzerland, 2 edition, January 2022.

[10] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[11] Rohan M. Lewis, Harikesan Baskaran, Jools Green, Stanimir Tashev, Eleni Paleologou, Emma M. Lofthouse, Jane K. Cleal, Anton Page, David S. Chatelet, Patricia Goggin, and Bram G. Sengers. Empiar. online, March 2022. Dataset. EMPIAR-10967.

[12] Eric Bushong and Tom Deerinck. The cell image library. online. Dataset. CCDB:8192.

[13] Alain C. Burette, Richard Weinberg, John Crum, Mark Ellisman, and Maryann Martone. The cell image library. online, January 2000. Dataset. CCDB:8641.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, and Rob Fergus, editors, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS, pages 6000–6010, Red Hook, NY, USA, December 2017. Curran Associates Inc.

[15] Wang Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 14(4):600–612, April 2004.

[16] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, January 2011.

[17] Anish Mittal, Anush K. Moorthy, and Alan C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, August 2012.