Deep Learning for Image Stitching

Ing. Petr Šilling, FIT BUT

Supervisor: Ing. Michal Španěl, Ph.D.   Consultant: Ing. Oldřich Kodym, Ph.D.

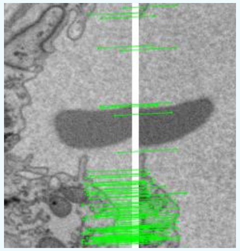
BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Motivation and Goals

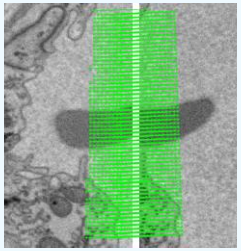
Image stitching is essential for reconstructing volumes of biological samples from overlapping tiles of electron microscopy (EM) images. Current EM stitching methods generally rely on traditional methods, such as SIFT. Consequently, they may struggle with repetitive patterns, poor texture, and high-resolution images – all of which are extremely common in EM.

Fortunately, recent developments indicate that convolutional neural networks (CNNs) can improve stitching accuracy by learning discriminative features directly from training images. Considering the potential of CNNs, the thesis proposes DEMIS, a novel EM image stitching tool based on LoFTR, an attention-based feature matching network.

SIFT: 85 matches

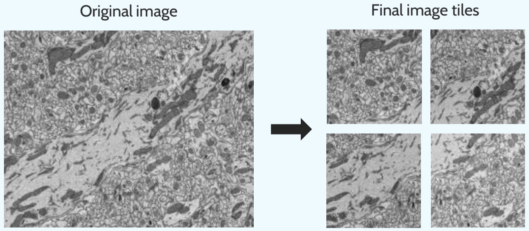


LoFTR: 624 matches



Synthetic Dataset

To train and evaluate DEMIS, we prepare a synthetic dataset by manually selecting 424 distinct high-quality EM images publicly available on EMPIAR or Cell Image Library. 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.



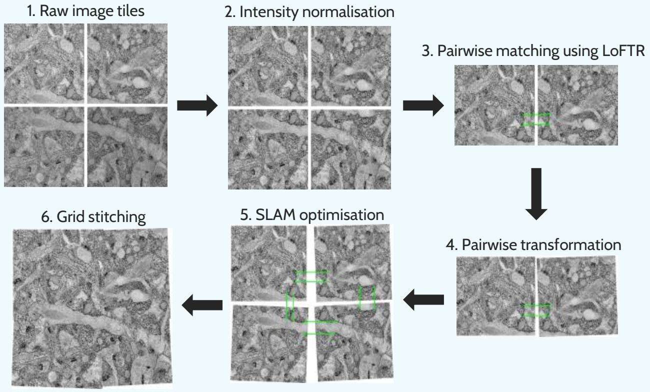
Proposed Solution

The proposed DEMIS tool is inspired by the workflow used by traditional image stitching solutions. It expects a grid of overlapping images as its input. The brightness and contrast of the input images are first normalised to aid feature detection and mask future image tile boundaries.

Then, for each pair of adjacent images, features are detected and matched by LoFTR. LoFTR is a Transformer-based CNN designed to simultaneously detect and match features between pairs of overlapping images. LoFTR was pre-trained on conventional photography, where it achieved better and more robust results than traditional stitching methods. For DEMIS, we propose to fine-tune LoFTR on EM images using the proposed synthetic dataset.

Subsequently, a SLAM graph is constructed based on the expected grid structure and initial pairwise transformations estimated from the detected feature matches. Vertices in the graph represent images in the grid and edges the transformations between them. The initial transformations are then optimised globally.

Finally, the grid is stitched into a single image by gradually applying the optimised transformations to individual image tiles in the grid.


1. Raw image tiles
2. Intensity normalisation
3. Pairwise matching using LoFTR
4. Pairwise transformation
5. SLAM optimisation
6. Grid stitching

Results and Conclusions

The DEMIS tool was experimentally evaluated on (1) the evaluation images of the proposed synthetic dataset and (2) on two challenging real-life EM datasets, which were provided by TESCAN 3DIM, s.r.o. The experiments compared a traditional stitching solution using SIFT, the DEMIS tool using only pre-trained LoFTR, and the DEMIS tool using LoFTR fine-tuned on the proposed synthetic dataset.

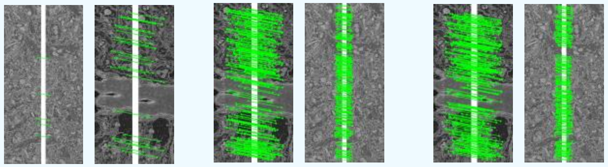
The results of the experiments on the proposed synthetic dataset, displayed in the table on the right, show that the DEMIS tool generates more accurate feature matches than a comparable method based on SIFT. Moreover, the results show that, on average, the DEMIS tool finds 12% more feature matches than the SIFT baseline.

However, image quality metrics, such as PSNR and SSIM, corroborate the improved performance only when evaluating the stitching of individual image pairs, not the whole grids. Brightness and contrast differences between tiles could explain this behaviour.

The experiments on real-life datasets of images with small overlap regions and high resolution demonstrate significantly higher stitching robustness than SIFT. Sample results from this experiment are shown in the bottom right corner.

Overall, the results suggest that deep learning methods could be beneficial for EM imaging, for example, by allowing the use of smaller tile overlaps and, consequently, increasing imaging speeds.

Metric (average on 1306 images)	SIFT baseline	Pre-trained	Fine-tuned
Feature matching time	80.44 ms	83.52 ms	83.18 ms
Feature matches found	694	770	778
Outlier feature matches	5	3	2
Inlier reprojection error	2.86 px	2.90 px	2.82 px
Corner error AUC at 5 px	42.70 %	40.99%	46.02 %
PSNR of image pairs	22.07	21.88	22.33
PSNR of complete grids	15.22	15.23	15.04
SSIM of image pairs	0.68	0.67	0.69
SSIM of complete grids	0.22	0.22	0.21



SIFT baseline
465 + 25 matches

Pre-trained DEMIS
3872 + 1321 matches

Fine-tuned DEMIS
4044 + 1329 matches