Journal Pre-proof

Two-stage error detection to improve electron microscopy image mosaicking

Hongyu Ge, Jiahao Shi, Shuohong Wang, Donglai Wei,
Jiancheng Yang, Ao Cheng, Richard Schalek, Jun Guo, Jeff Lichtman,
Lirong Wang, Ruobing Zhang

PII: S0010-4825(24)00540-7
DOI: https://doi.org/10.1016/j.compbiomed.2024.108456
Reference: CBM 108456

To appear in: *Computers in Biology and Medicine*

Received date : 9 October 2023
Revised date : 26 March 2024
Accepted date : 7 April 2024

Please cite this article as: H. Ge, J. Shi, S. Wang et al., Two-stage error detection to improve electron microscopy image mosaicking, *Computers in Biology and Medicine* (2024), doi: https://doi.org/10.1016/j.compbiomed.2024.108456.

Two-stage Error Detection to Improve Electron Microscopy Image Mosaicking

Hongyu Ge[a,c,1], Jiahao Shi[b,c,1], Shuohong Wang[d], Donglai Wei[e], Jiancheng Yang[f], Ao Cheng[a,c], Richard Schalek[d], Jun Guo[g], Jeff Lichtman[d], Lirong Wang[a] and Ruobing Zhang[c,g,*]

[a]School of Electronic and Information Engineering, Soochow University, Suzhou 215009, China

[b]School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230026, China

[c]Jiangsu Key Laboratory of Medical Optics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China

[d]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138 USA

[e]Department of Computer Science, Boston College, Boston, MA 02467, USA

[f]Shanghai Jiao Tong University, Shanghai 200240, China, and also with EPFL, Lausanne 1015, Switzerland

[g]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

ARTICLE INFO

Keywords:
electron microscopy, image stitching, keypoint features, stitching assessment

ABSTRACT

Large-scale electron microscopy (EM) has enabled the reconstruction of brain connectomes at the synaptic level by serially scanning over massive areas of sample sections. The acquired big EM data sets raise the great challenge of image mosaicking at high accuracy. Currently, it simply follows the conventional algorithms designed for natural images, which are usually composed of only a few tiles, using a single type of keypoint feature that would sacrifice speed for stronger performance. Even so, in the process of stitching hundreds of thousands of tiles for large EM data, errors are still inevitable and diverse. Moreover, there has not yet been an appropriate metric to quantitatively evaluate the stitching of biomedical EM images. Here we propose a two-stage error detection method to improve the EM image mosaicking. It firstly uses point-based error detection in combination with a hybrid feature framework to expedite the stitching computation while maintaining high accuracy. Following is the second detection of unresolved errors with a newly designed metric of EM stitched image quality assessment (EMSIQA). The novel detection-based mosaicking pipeline is tested on large EM data sets and proven to be more effective and as accurate when compared with existing methods.

1. Introduction

The reconstruction of neural circuits through the imaging of serial ultra-thin sections of brain tissues at nanometer-range resolution with 2D large-scale electron microscopy (EM), employing serial sectioning techniques such as serial section scanning electron microscopy (ssSEM), has emerged as a critical and effective method for connectomic studies [1, 2, 3, 4]. The mosaicking of a substantial number of imaging tiles within the region of interest (ROI) into a cohesive 2D EM image is indispensable due to the inherent limitations of the size of the field of view (mfov).

The mosaicking task of EM images for connectomic studies encounters the challenge of balancing high speed and high precision. The inherently high resolution of EM imaging results in substantial amounts of data, imposing stringent requirements on stitching speed. Moreover, this substantial amounts of data exacerbates the already demanding accuracy requirements imposed by downstream alignment and segmentation tasks [5]. In contrast to generic nature images, the parallax and distortion inherent in EM images are often mitigated by opting for a smaller tile size. However, this strategic decision amplifies the number of tiles and consequently escalates the computational burden for stitching. To mitigate this computational load, it is customary to reduce the overlap area, yet this approach engenders heightened challenges in the mosaicking process.

In brief, natural image stitching puts the emphasis on minimizing local geometric misalignment, improving transition smoothness, and hiding the seam between parallax images [6, 7, 8, 9, 10, 11, 12, 13]. In contrast, the mosaicking of large-scale EM images can be satisfied with nearly rigid transformation but instead greatly suffer from long computation time and inevitable errors through the enormous amount of data.

The process of mosaicking electron microscopy (EM) images involves several steps, including feature extraction, matching, outlier rejection, and global optimization to derive the necessary transformations [14] (Fig. 1). A key consideration in this process is the trade-off between accuracy and speed in feature extraction. While faster methods like ORB [15] prioritize speed over accuracy, more accurate techniques like SIFT [16] require longer computation times [17]. However, the lack of standardized evaluations for EM image mosaicking makes it challenging to quantitatively compare feature performance. As a result, researchers often

*Corresponding author

hyge4869@stu.suda.edu.cn (H. Ge); beholder@mail.ustc.edu.cn (J. Shi); wangsh@fas.harvard.edu (S. Wang); donglai.wei@bc.edu (D. Wei); jekyll4168@sjtu.edu.cn (J. Yang); aoc.7@outlook.com (A. Cheng); rschalek@mcb.harvard.edu (R. Schalek); guoj@iai.ustc.edu.cn (J. Guo); jeff@mcb.harvard.edu (J. Lichtman); wanglirong@suda.edu.cn (L. Wang); zhangrb@sibet.ac.cn (R. Zhang)

ORCID(s):

[1]These authors contributed to the work equallly and should be regarded as co-first authors

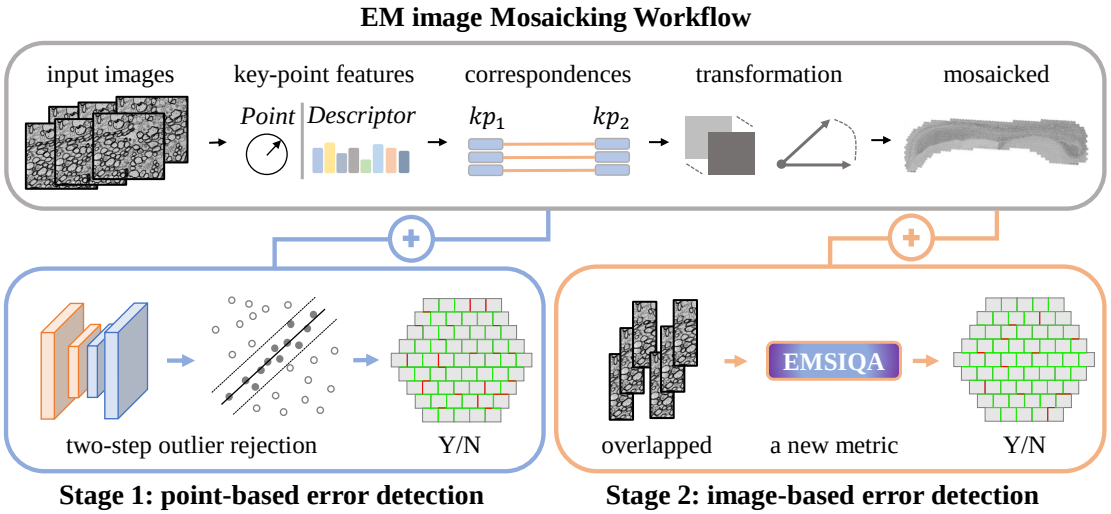


Figure 1: Overview of the image mosaicking pipeline with the proposed two-stage error detection. Upper, the conventional EM image mosaicking workflow. Lower, the two-stage error detection that is added. In the Y/N insets, the red (Y) and green (N) line segments indicate whether a stitching error exists on the border of two tiles.

rely on qualitative assessments, which can be laborious and subjective. Despite the preference for accuracy, errors in mosaicking remain challenging to detect, especially given the large volume of data and time required for analysis. Efforts to enhance keypoint features and transformation models have been ongoing, but no method has yet achieved an optimal balance between speed and accuracy. In response to this challenge, we propose a novel approach that focuses on error detection and iterative feature refinement.

We designed a two-stage error detection pipeline. In the first stage, correspondences derived from a hybrid feature framework undergo scrutiny from a point-based error detection method prior to image rendering. Subsequently, the detected errors are utilized to iteratively prioritize a feature with heightened accuracy for handling the flawed tiles. Then, in the second stage, the mosaicked images undergo evaluation using a novel EM stitched image quality assessment (EMSIQA) metric to identify any remaining errors. In essence, the approach involves leveraging fast features to maximize computational speed, while simultaneously employing error detection methods and exploring accuracy-focused features to ensure precision. We tested the detection-based biological EM image mosaicking pipeline on large data sets of mouse brain from multibeam SEM and mouse glioblastoma from single-beam SEM, and demonstrated high accuracy and significantly shortened processing time.

2. Related work

For image mosaicking, current pipelines first match key points for each pair of overlapped images, estimate the transformation for each image tile with a global optimization approach [14, 18], render the mosaicked image, and assess the stitched image quality. Below we review the computational costly keypoint matching step and the final image quality assessment step.

2.1. Keypoint Matching for Image Stitching

Image keypoints. One major time-consuming step in the image stitching pipeline is keypoint detection. Since the scale-invariant feature transform (SIFT) [16] was proposed by Lowe et al in 1999 and widely applied in many computer vision tasks like stitching, registration, and template matching, many handcrafted features have been proposed to improve in either accuracy or speed. Speeded-Up Robust Features (SURF) [19] was developed as a faster replacement of SIFT by replacing the Difference of Gaussian (DoG) with Hessian matrix and squeezing the dimensions of descriptors to speed up the matching. Oriented FAST and Rotated BRIEF (ORB) [15] further accelerated the extraction, reaching up to a 100-fold speed increase of SIFT in theory, but its robustness is not as good as SIFT and SURF. AKAZE [20], proposed as the accelerated version of KAZE [21], adds FED (Fast Explicit Diffusion) to the pyramid framework and the utilization of non-linear scale space makes it more stable than SIFT or SURF. BRISK [22] was proposed to achieve a high-quality performance albeit at a dramatically lower computational cost. In recent years, learning-based features emerged to take advantage of GPU parallel computation. Learned Invariant Feature Transform (LIFT) [12] used convolutional neural networks (CNNs) to implement detector, orientation estimator, and descriptor. However, a CNN-like network is only weakly invariant to the rotation, which limits its application in many tasks.

Matching outlier rejection. Sparse feature extraction algorithms pick out the points that are distinctive and robust

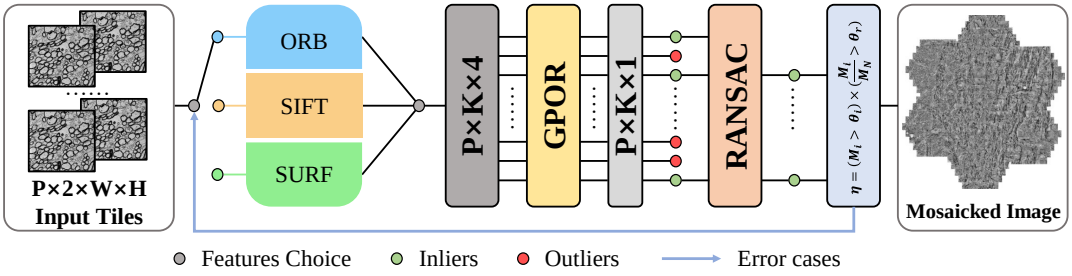


Figure 2: The first-stage point-based error detection and the hybrid feature framework. Given **P** overlapped tile pairs with shape $[W \times H]$, a feature extraction algorithm with the highest speed, which is ORB in our experiment, is first used to generate the tentative correspondences. Assisted by modified learning-based global-perception outlier rejection (GPOR) [23] and RANSAC [24], potential errors in correspondences are detected. Then, slower but more accurate extraction and matching algorithms, such as SIFT and SURF, are applied to erroneous tile pairs to improve the stitching quality.

to transformation and then give each key point a high-dimensional descriptor. By calculating the distance of descriptors in an image pair, each key point in one image will be linked to the closest point in the other, and we refer to this point pair as a correspondence. Therefore, the challenge is to find the correct geometric transformation out of massive erroneous correspondences (outliers). RANSAC [24] is an old but effective algorithm to reject outliers [25]. By iteratively selecting random points, the fitted model is applied to check how many points are potentially inliers until a model that can include the most correspondences is achieved. MLESAC [26], a generalization of RANSAC, maximizes the likelihood rather than just the number of inliers. PROSAC [27] optimizes the speed from the perspective of sampling. Also, deep learning is introduced to make up for the ignorance of global geometric information. Choy *et al.* [28] further explored the outlier rejection in high-dimensional space powered by the Minkowski engine [29]. Yi *et al.* [23] drew lessons from the processing of disordered points in PointNet [30], and proposed a context normalization module to extract the inliers with global perception, which we call as Global-Perception Outlier Rejection (GPOR).

2.2. Stitched Image Quality Assessment (SIQA)

Different from other computer vision tasks like classification or segmentation, it is almost impossible to manually label the ground truth of the stitching of two naturally acquired images. Thus, researchers tend to compare the structure of interest in the overlapped region. One simple way to assess the stitching is to adopt classical image quality metrics, such as peak signal-to-noise ratio (PSNR) [31], structural similarity (SSIM) [32], and normalized cross-correlation (NCC) [33]. However, these methods are not designed for the evaluation of image stitching and ignore the different importance of various types of errors between the stitched images. Qureshi *et al.* [34] quantified the geometric and photometric qualities separately of a stitched image and named the geometric part HFI-SSIM. Yang *et al.* [35] fused a perceptual geometric error metric and a local structure-guided metric into one. Tian *et al.* [36] took consideration of six different stitching distortion types and trained an assessment model by SVR [37]. Furthermore, Ullah [38] took advantage of mask R-CNN [39] to build a three-fold deep learning-based no-reference stitched image quality assessment called DLNR-SIQA.

3. Methods

3.1. Framework Overview

Our error detection framework has two stages (Fig. 1). In the first stage, we adapt and integrate the previously proposed GPOR into a hybrid feature selection framework aimed at striking a harmonious equilibrium between speed and accuracy. In the second stage, we introduce and implement a novel metric that more comprehensively incorporates the image characteristics specific to biomedical EM data. This metric enables the identification of any persisting errors and facilitates an accurate assessment of the mosaicking quality.

3.2. Stage 1: Key point Matches Error Detection

Among image features, SIFT is known to have high-quality matches with costly computation while ORB is faster to compute with a significant drop in match quality. It is a straightforward idea to first try ORB and later try SIFT if the ORB match quality is not sufficient. However, it is challenging to design a reliable metric for keypoint matches to know when to switch to a different image feature.

Given a chosen feature and an image pair to stitch, we can have two statistics: M_n, the number of all matches between the image pair and M_i, the number of inlier matches chosen by RANSAC. A commonly used binary heuristic variable, η, to determine if the matches are good or not can be defined by

$$\eta = (M_i > \theta_i) \cap (\frac{M_i}{M_n} > \theta_r), \tag{1}$$

where θ_i demands big enough number of inlier matches and θ_r demands high enough ratio of inlier matches. When

the matches are not good which means the chosen feature failed, the value of η will be False. Intuitively, when θ_i is small, there are not enough matches to robustly estimate the transformation matrix; when θ_r is small, the image may have ambiguous structures leading to non-consensus matches.

However, for EM images, the initial keypoint matches are noisy, which makes the θ_r unstable for the selection.

We thus designed a combined approach to detect potential stitching errors before global optimization and rendering, by filtering the output inliers from the GPOR with an additional RANSAC and calculating the acceptance ratio.

In this work, we adopt a learning-based outlier rejection algorithm proposed by Yi et al [23]. This algorithm involves considering image pairs $(I, I^{'})$ and their corresponding essential matrices E to extract the set of correspondences X associated with E. The challenge of outlier rejection can be addressed by designing a deep network that encodes a map f parameterized by Φ, which

$$W = f_{\Phi}(X), E = g(X, W). \tag{2}$$

The $W = [\Omega_1, ..., \Omega_N]$ is the output of the network f_{Φ}, where $\Omega_i \in [0, 1]$ represents the score assigned to correspondence x_i, and $\Omega_i = 1$ indicates x_i as an inlier. The function g filters correspondences X based on W and computes the essential matrix E from the filtered X.

In order to individually consider each correspondence within the broader global context, allowing for the encoding of camera motion, the feature map is normalized based on its distribution following each perceptron. The network utilized in this study is a 12-layer ResNet, with each layer comprising two consecutive blocks comprising a Perceptron featuring 128 neurons sharing weights for every correspondence, a Context Normalization layer, a Batch Normalization layer, and a Rectified Linear Unit (ReLU).

The training of this network employs a hybrid loss function comprising a classification loss to reject outliers and a regression loss to predict the essential matrix. Since there is no requirement to estimate the transformation matrix for each image pair, we solely utilize the classification loss function.

$$\mathcal{L}(\Phi) = \sum_{k=1}^{P} \mathcal{L}(\Phi, \mathbf{x}_k) \tag{3}$$

where Φ are the network parameters and $\mathbf{x}_k$ is the set of putative correspondences for image pair k. Given a set of N putative correspondences $\mathbf{x}_k$ and their respective labels $\mathbf{y}_k = [y_k^1, \dots, y_k^N]$ where $y_k^i \in 0, 1$, and $y_k^i = 1$ denotes that the i-th correspondence is an inlier, our outlier classification error is

$$\mathcal{L}(\Phi, \mathbf{x}_k) = \frac{1}{N} \gamma_k^i H(y_k^i, S(o_k^i)), \tag{4}$$

where o_k^i is the linear output of the last layer for the i-th correspondence in training pair k, S is the logistic function used in conjunction with the binary cross entropy H, and γ_k^i is the per-label weight to balance positive and negative examples.

As shown in Fig. 2, this network accepts the input correspondences with shape [Batch, 4, K] and outputs the likelihood ranging in (0, 1) for every correspondence shaped as [Batch, 1, K] to estimate the probability to be an inlier. With such keypoint match error removal, we empirically find the common image feature selection method (Eqn. 1) becomes more effective due to a more stable inlier match ratio M_i/M_n.

3.3. Stage 2: Stitched Image Error Detection

In the multi-step processing of biomedical EM images, image mosaicking is an upstream step to assist later three-dimensional registration and segmentation. The primary goal is to make every biological structure well-stitched at the pixel level. In comparison, the visualization factors like the photometric quality have less effect on the downstream analysis. Furthermore, since the structures in 2D images are used to reconstruct the 3D volume, any trick to blandish the eyes such as multi-band blending [40] should not be applied to avoid hidden errors. Thus, the principles of evaluating the stitching result should (1) pay the most attention to cellular structures, (2) ignore the photometric quality and (3) be prior to fusion or blending. Given the stitched left and right image pair I_A and I_B, we design a new SIQA score that is customized for EM images with the downstream segmentation task in mind, termed EMSIQA, for which we take the factors below into consideration.

(a) Deformation magnitude. Traditional SIQA methods are sensitive to the change of image appearances, *e.g.,* out-of-focus blur and brightness, between the pair of images, even if there is no geometric change (Fig. 3a-b). To focus on the geometric matching quality for the stitched image pair, the proposed EMSIQA method computes the average deformation field magnitude, defined as

$$E_{flow}(I_A, I_B, \Omega) = \frac{1}{N} \sum_{i \in \Omega} \sqrt{u_i^2 + v_i^2} \tag{5}$$

where N is the number of pixels, u and v denote the horizontal and vertical values of the optical flow between the pair of input images, and Ω represents the region of valid pixels.

(b) Border structure. As illustrated in Fig. 3 c-d, due to the imaging noise, there is non-zero deformation on cell texture, which can overwhelm the deformation field magnitude on the cell and organelle borders that are critical for the downstream segmentation task. Thus, we designed the EMSIQA to focus on the important border features. As the labeling of precise boundaries of cellular compartments leads to the challenging segmentation task, we herein use a fast and simple method that is very effective in scenarios with low-precision requirements. OTSU threshold segmentation [42] maximizes the contrast between foreground and background to find the most appropriate segmentation threshold. We added a median filter to decrease the noise and were able to obtain a binary mask that coarsely outlined the cellular structures. Thus, we choose the region of deformation field Ω_B for image I_B as

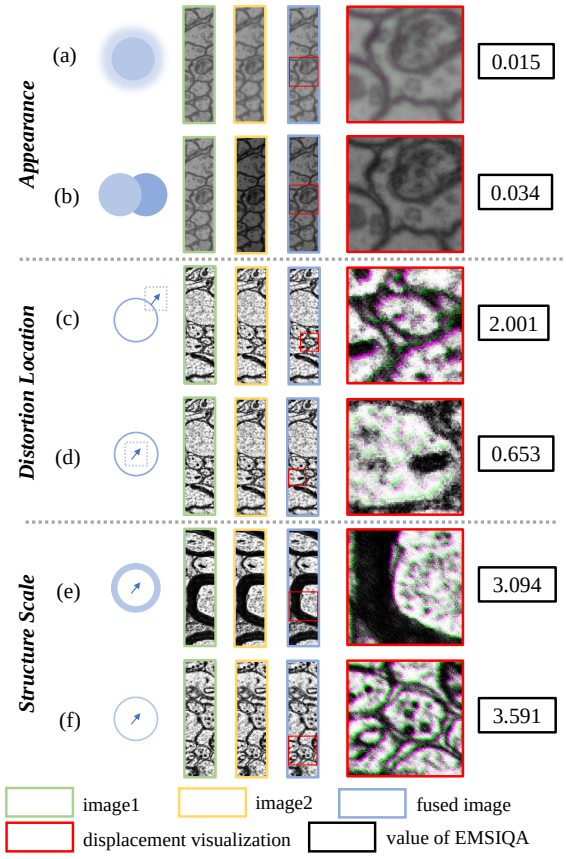


Figure 3: Typical scenarios in EM image mosaicking. (a) One image has an out-of-focus blur simulated by a Gaussian blur. (b) The pair is different only in brightness. (c) Local distortion on the boundary membranes of cellular compartments, no global translation. (d) Local distortion in the info-less cytosolic area inside the cellular compartments, no global translation. (e) Thick membranes, 2-pixel vertical and horizontal global translation. (f) Thin membranes, 2-pixel vertical and horizontal global translation.

$$\Omega_B = \text{Median-Filter}(\text{OTSU}(I_B)) \tag{6}$$

(c) Border matching. Although geometric error can quantitatively describe the displacement in pixels, it cannot represent the mismatch of biological structures relative to their scales which are extensively diverse among different cellular compartments. In other words, the same pixel displacement in big and small cellular structures can cause different effects on the registration and segmentation that follows (Fig. 3 e-f). Inspired by segmentation algorithms, we adopted Dice index [43] to quantify the matching of the border structures of the pair of images:

$$\text{DICE}(\Omega_A, \Omega_B) = \frac{2 * (\Omega_A \cap \Omega_B)}{\Omega_A + \Omega_B} \tag{7}$$

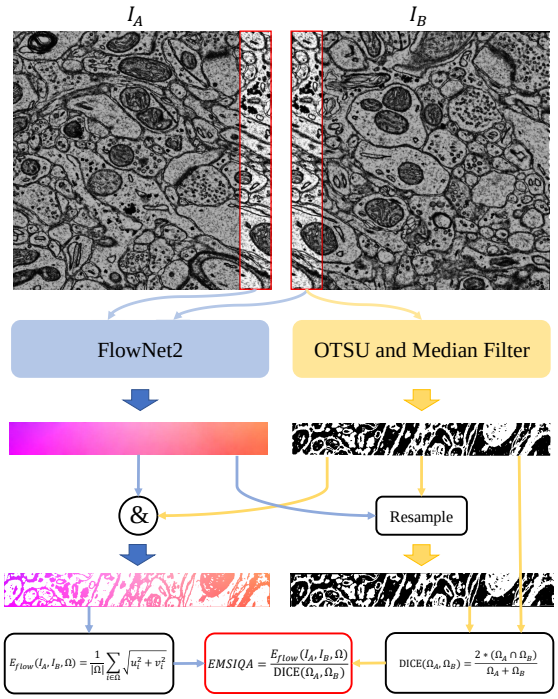


Figure 4: EMSIQA computation. EMSIQA is a novel metric to evaluate electron microscopy stitching. Guided by optical flows (via FlowNet2 [41]) and boundaries (via OTSU segmentation [42]), it measures the geometric error normalized to the biological structure.

where the greater the similarity between the image pairs I_A and I_B, the higher the resulting value. Similarly, when there is a consistent pixel displacement, indicating equal divergence between I_A and I_B, any displacement observed in smaller cellular structures *e.g.,* thin membranes will exert a stronger influence, resulting in a diminished unaffected region and consequently a reduced Dice index. In other words, the larger cellular structures have a substantial ($\Omega_A \cap \Omega_B$) under similar displacement conditions, which results in a higher Dice index:

$$\frac{2 * (\Omega_A \cap \Omega_B)}{\Omega_A + \Omega_B} = \frac{2}{1 + \frac{Dis}{\Omega_A \cap \Omega_B}} \tag{8}$$

$$Dis = \Omega_A \cap \bar{\Omega_B} + \bar{\Omega_A} \cap \Omega_B \tag{9}$$

As shown in Fig. 4, for each image pair to be stitched together, we crop out the overlapping area from the two images, respectively, calculate the average geometric error in pixels of all cellular structures, and then divide it by a penalty item that represents the structure matching of the two overlapping areas. We call this metric EMSIQA (EM stitched image quality assessment) and formulate it as:

$$\text{EMSIQA}(I_A, I_B) = \frac{E_{flow}(I_A, I_B, \Omega_B)}{\text{DICE}(\Omega_{A'}, \Omega_B)} \tag{10}$$

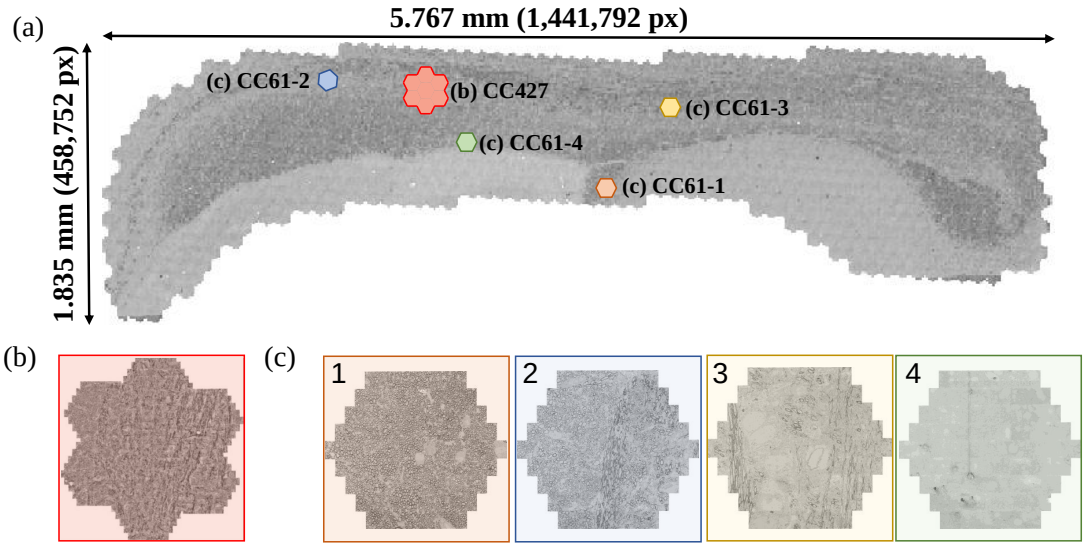


Figure 5: CC50K dataset gallery. (a) CC50k dataset, a complete 2D cross-section of adult mouse corpus callosum consisting of $458,752 \times 1,441,792$ pixels. (b) The 7-mFoV sub-dataset CC427. (c) Four 1-mFoV subsets named CC61-1 to 4, each containing 61 tiles.

where $I_{A'}(x, y) = I_A(x + u, y + v)$ is the warping of I_A by the optical flow between the pair of images.

3.4. Implementation Details

We tested the pre-trained FlowNet2[1] [41] on image pairs with known displacement and found it sufficiently precise and robust. Thus, when computing the optical flow for EMSIQA evaluation, we directly adopted the pre-trained model of FlowNet2 [41]. To organize the large-scale EM data, we adopted the data structure used in TrakEM2 [44] and the workflow of rh-aligner[2] [45] with modifications. We implemented the GPOR referencing Yi et al [3] [23] using PyTorch. To train the model, we set Adam as the optimizer with a learning rate equal to 0.00005 and set the batch size to 32. We only preserved the classification loss since the weighted 8-point algorithm does not match the workflow of multiple-image stitching. Other arguments were kept unchanged to Yi *et al.* [23]. When extracting features, we set the number of ORB features to be close to the average value of those extracted by SIFT or SURF. Commonly, when the image is low-textured, the number of key points extracted by SIFT or SURF will drastically decrease while ORB will keep constant or close to the number we pre-set. In the error detection step, when setting the thresholds of the acceptance ratio and the number of inliers, we took into consideration the image size and the type of features. In our experiment, we set the number of ORB features for one tile to be 50000 and regard a pair as a stitching error when the acceptance ratio is lower than 0.9 or the inliers number is below 40 or 20 for ORB and SIFT, respectively by experience. For the execution order of the features, we set ORB as the first choice to perform the simplest yet fastest key point extraction. SIFT, as the second option, will take over where ORB fails and more accurate correspondences are required. In some very low-texture regions, SURF will serve as the last choice to extract more feature points than SIFT.

In our experiment, we tested all algorithms on a workstation equipped with Intel Core i9-9920X and one Nvidia RTX2080Ti (11GB memory). Due to the different scheduling strategies when using OpenCV [46], we used schedtool[4] on the Linux platform and ran the tests on a single processor to ensure the fairness. While processing the complete large-scale CC50k dataset, we used a multiprocessing module and PyTorch multiprocessing module to accelerate the traditional keypoint extraction methods and deep learning-based outlier rejection methods, respectively.

4. Results

4.1. Datasets

The presented real datasets were approved by the Experimental Animal Ethics Committee of Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences. The dataset CC50K was collected from mouse corpus callosum on September 25th, 2018(NO.2018-A30). The ST793 dataset was collected from the mouse striatum on September 22nd, 2022. The GBM9 dataset was obtained from mouse glioblastoma on November 18th, 2020. We used a 61-beam scanning electron microscope (Zeiss MultiSEM 505) for acquiring CC50K and ST793 images, capable

[1] https://github.com/NVIDIA/flownet2-pytorch
[2] https://github.com/Rhoana/rh_aligner
[3] https://github.com/vcg-uvic/learned-correspondence-release
[4] https://github.com/freequaos/schedtool

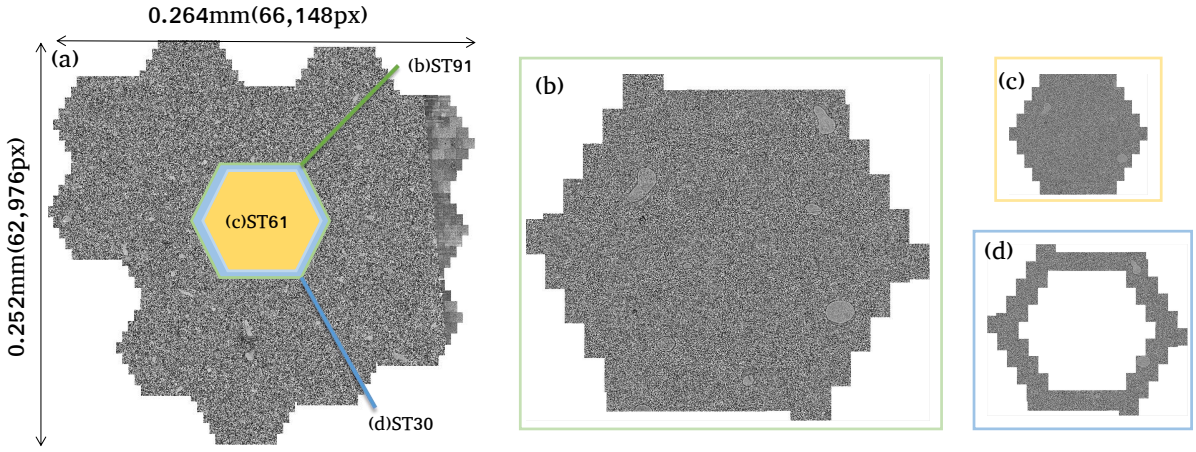
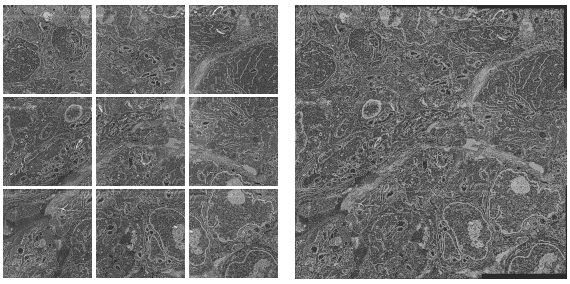


Figure 6: ST793 dataset gallery. (a) The ST793 dataset comprises a segment of the adult mouse striatum, featuring 17 mFovs with dimensions of $62,976 \times 66,148$ pixels. (b) The sub-dataset ST91 comprises a complete mFov with 61 tiles and 30 tiles from surrounding mFovs, exhibiting overlap with the entire mFov. (c) The complete mFov consists of 61 tiles. (d) The 24 boundary tiles from the complete mFov and 30 tiles from surrounding mFovs.



(a) imgs before stitching (b) stitching result(proposed)

Figure 7: GBM9 dataset. Part of a section of adult mouse glioblastoma cell with 9 tiles, $3,000 \times 3,000$ pixels per tile. (a) the images before stitching. (b) the stitching result of proposed.

of simultaneously capturing multiple tiles. The images of GBM9 were acquired using a single-beam scanning electron microscope (Zeiss GeminiSEM 300).

Besides, we generated two sets of synthetic data for the training and evaluation of GPOR. The dataset for evaluating GPOR will be detailed and introduced in sec 4.2. Without the metadata like camera poses in natural images, it is difficult to make a real dataset for training when processing the EM images because we do not have the ground truth of $\mathbf{L}_{i,j}$ according to the epipolar distance. We take advantage of the large area of EM images to configure a method to generate synthetic datasets that can simulate the real training data with ground truth. First, we choose a set of large 2D EM images and randomly select a pixel, used as the left-top corner of training image 1a. Then, an affine matrix is generated to transform the training image 1a to the corresponding area of training image 1b'. This area is usually not a rectangle so we need to solve another matrix to transform the whole large EM image in order to obtain a rectangle training image 1b of the same dimension with training image 1a. Please refer to the appendix A for more details about generating the synthetic datasets.

CC-train. To get the best performance on the real data, we cropped image pairs from the below **CC50k** and made a synthetic dataset for training. The overlap rate is set to be between 0.03 and 0.1 and we added an extra mask on every image because the later features matching step only works on an approximately overlapping rectangle. This training set contains 9226 pairs of images and each pair contains extracted 1000 correspondences.

CC50k. In the dataset acquired by the 61-beam SEM, one multifield of view (mFoV) consists of 61 tiles shaped in [2724, 3128], each scanned by an individual electron beam. We chose a complete cross-section of mouse corpus callosum containing 826 mFoVs and 50,386 tiles to test our framework (Fig. 5 (a)). The physical resolution is 4 nm/px so the about 10 mm^2 area contains over six hundred billion pixels. The overlap rate between tiles was set to be 3% when acquiring the images. This dataset is the superset of CC427 and CC61.

CC427. To promote testing efficiency, we cropped out 7 adjacent mFoVs with 427 tiles from CC50k (Fig. 5 (b)). This 4-billion-pixel subset is used to test the performance of different features on a large-scale EM dataset.

CC61. To evaluate the generalization performance of the proposed method, we also cropped out 4 subsets, each containing one mFoV with 61 tiles (Fig. 5 (c)). These mFoVs come from different areas of the CC50k, with different cellular structures or image contrasts.

ST793. In the dataset acquired by the 61-beam SEM, a multifield of view (mFoV) comprises 61 tiles, each shaped in [3376, 3876], with individual electron beams scanning each tile(Fig. 6 (a)). We selected a mouse striatum section with 13 mFovs and 793 tiles. The physical resolution is 4 nm/px,

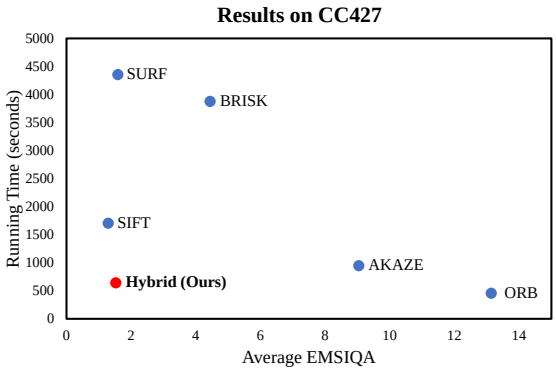


Figure 8: Stitching performance (measured in EMSIQA) and running time tested on the CC427 dataset. The proposed hybrid feature framework achieved an optimized balance between performance and speed. Note that we constrained the computing resource to one processor to ensure fairness. SURF ran slower than SIFT in the OpenCV implementation, which is contrary to the expectation.

Table 1
EMSIQA results and running time on **CC427** dataset, a real-world dataset containing 427 tiles.

Method	All↓	Top10%↓	Top20%↓	Top50%↓	Time↓
SIFT [16]	1.289	0.325	0.437	0.747	28'25"
SURF [19]	1.588	0.327	0.444	0.783	72'34"
ORB [15]	13.372	0.565	0.922	2.541	7'36"
AKAZE [20]	9.040	0.478	0.807	1.988	12'48"
BRISK [22]	4.441	0.386	0.546	1.083	64'36"
ORB+GP.	1.972	0.339	0.464	0.830	10'14"
Hyb. w/o. GP.	17.421	0.756	1.305	3.673	9'22"
Proposed	1.523	0.289	0.388	0.715	10'41"

with an 8% overlap between mFovs and 1 µm between tiles during image acquisition. This dataset serves as the superset of ST91.

ST91 In SEM image stitching, tile pairs within the same mFov and between different mFovs yield distinct results. Typically, stitching errors occur between tiles inter-mFovs. So We choose a complete mFov at the midpoint of the section and tiles from surrounding mFovs that overlap with this entire mFov (Fig. 6 (b)). In this sub-dataset, there are 91 tiles.

GBM9 This 3×3 tiles mouse glioblastoma dataset is consist of 9 tiles (Fig. 7).

4.2. Point-based Error Detection: EM- Feature

Assisted by error detection, we can first use the faster feature to obtain the preliminary inliers, and then optimize the potential wrong pairs using a slower feature with stronger performance. As shown in Table 1, we recorded the mean EMSIQA of CC427 dataset to test the overall performance, and Top 10%, Top 20%, and Top 50% mean EMSIQA to evaluate how well the top stitched pairs perform. As depicted in Table 4, we documented the mean EMSIQA for the ST91 dataset, evaluating the stitching results within mFov and between mFovs (Intra and Inter mean EMSIQA). Similarly, Table 5 displays the mean EMSIQA for the GBM9 dataset.

Table 2
A Comparison of RANSAC [24] and GPOR on **EMPAIR3000**, a synthetic dataset including 1k pairs in the overlap rate range of 20%-30%, 30%-40%, and 40%-50%, respectively

	20%-30%		30%-40%		40%-50%	
	RAN.	GPOR	RAN.	GPOR	RAN.	GPOR
Inliers	13.5	**42.0**	61.0	**86.1**	139.6	**144.6**
Accuracy	0.966	**0.998**	0.974	**0.999**	0.996	**0.998**
Precision	0.477	**0.958**	0.886	**0.984**	**0.995**	0.990
Recall	0.198	**0.998**	0.649	**0.997**	0.963	**0.999**

GPOR vs. RANSAC. In order to evaluate the performance of GPOR on image pairs with different overlap rates, we made a synthetic dataset containing 3k pairs of four different overlap rate ranges, 1k for each range (Table 2). We constrained the displacement of four corner points within 50 pixels to simulate the nearly rigid transformation. For each image, the dimension is 1024×1024, and we set 1k feature points for SIFT. We found 40.6, 85.1, and 143.3 inliers on average in the four subsets, respectively. As shown in Table 2, on EMPair3000 which includes different overlap rates, GPOR outperforms RANSAC in most cases. Because of the imbalance of inliers and outliers (i.e., outliers are much more than inliers), the disparity of accuracy is insignificant compared with precision and recall. Especially in the groups with lower than 30% overlapping area, GPOR has a great advantage. Considering all recall values are larger than 0.99, we believe that in the case of nearly rigid transformation and small overlapping areas, which apply to most multibeam EM images, GPOR significantly outperforms RANSAC.

Proposed vs. Single Feature. In previous workflow, people tend to select SIFT to guarantee higher stitching precision at the cost of speed. Our experiment result from CC427 shows that sometimes accuracy and speed are mutually compatible goals, shown in Fig. 8. By detecting the errors from ORB with GPOR and RANSAC, and replacing it with SIFT, the Top10%, Top20%, and Top50% averaging EMSIQA all exceed the result of pure SIFT or SURF and the mean EMSIQA of all tile pairs is very close to them (Table 1). Meanwhile, under the same computation resource, the running time including extracting and matching features, is cut down to nearly one-third of SIFT.

Proposed vs. without Hybrid Features. GPOR can only work when there are enough good correspondences, which is not guaranteed when handling sparse key points. The comparison of *Proposed* and *ORB+GPOR* in Table 1 indicates that pure ORB assisted by GPOR cannot achieve the performance of ORB hybrid with SIFT.

Proposed vs. Error Detection without GPOR. We also explored the case of simply using either the filter rate or the number of accepted correspondences via RANSAC as the error detection criterion, instead of applying GPOR. The results are listed in the row of 'Hyb. w/o GPOR' in Table 1, indicating they do not work well.

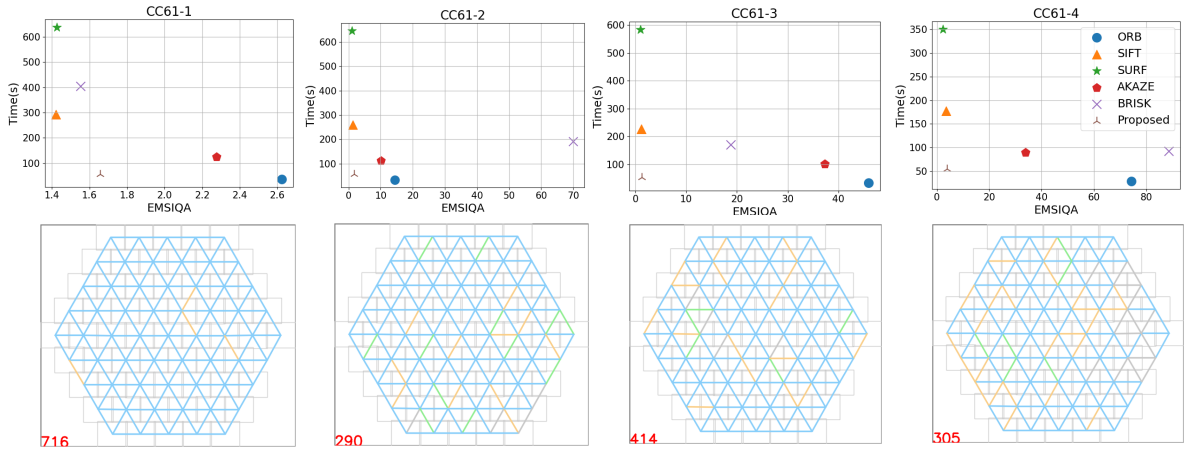


Figure 9: Visualization of the feature characteristics on the CC61 dataset. Upper: speed-accuracy trade-off. Lower: spatial visualization of the hybrid feature adopted by the framework on different tile pairs. The color of the short line connecting the center of two tiles represents the final chosen features. Blue, yellow, and green denote ORB, SIFT, and SURF, respectively.

Generalization performance. To assess the generalization performance, our method performs well on CC61, ST91, and GBM9, as demonstrated in Table 1, Table 4, and Table 5. CC61 has 4 sets of 61-tile EM images with varied cellular structures and contrasts. ST91 includes an mFov with 61 tiles and 30 tiles from surrounding mFovs. GBM9 is a glioblastoma dataset with 9 tiles acquired by a single-beam scanning electron microscope. As described in section 4.1, CC61-1 is relatively easy because of the abundant and uniform axon bundles, thus providing sufficient keypoint features to extract. The cellular structures have different scales in CC61-2. CC61-3 covers many cytons with low texture. CC61-4 has a low contrast compared to the other regions. We generate plots for time, mean EMSIQA, and the distribution of hybrid features in Fig. 9. These illustrate that our method demonstrates excellent performance across these datasets, with notable improvement on more challenging data.

4.3. Image-based Error Detection: EMSIQA

We designed three typical scenarios, including six image pairs as shown in Fig. 3 to compare EMSIQA with other assessments. The *Appearance* represents the blurred images by focus inaccuracy and the images of different signal intensities. The *Distortion Location* shows the image deformation that occurred on the boundary membranes of biological structures or in the information-less areas inside the cellular compartments. And the *Structure Scale* discerns the thick and thin cellular membrane structures. For (a) and (b), the values are very close to zero, which indicates that the proposed metric is nearly invariant to the blur and brightness change. In (c) and (d), we added distortion on boundary membranes and cytosolic areas, respectively. It can clearly discriminate the influence of the distortion when it is distributed on different structures. (e) is an image with a thick cellular structure while (f) shows a thin one. Although they have the same pixel-wise displacement, the mismatching of cellular structure is more serious when it is thin and thus the value is larger, proving that EMSIQA is sensitive to resolution-independent structure matching. As shown in Table 3, PSNR is sensitive to the brightness change. SSIM increases when the structure becomes thinner. NCC does not clearly discriminate the effect of distortion on boundary membranes or cytosolic areas. HFI-SSIM outputs obviously unreasonable values when evaluating the thick and thin structures. In conclusion, compared with classical IQAs and HFI-SSIM designed for stitching, the proposed EMSIQA gives more reasonable results under common scenarios in EM image mosaicking.

4.4. Application on Ultra-large 2D Image

We applied the proposed framework on **CC50k** to test the performance, speed, and robustness of our framework dealing with ultra-large multi-tile EM images. Fig. 5 (a) exhibits the overview of the stitched complete cross-section of corpus callosum. As shown in Fig. 10 (a), the Top10%, Top20%, and Top50% EMSIQA can achieve 0.337, 0.459, and 0.829, respectively. Over 80% of the overlapping areas have a value below 3. Fig. 10 (b) shows three stitching areas in detail for every EMSIQA range. Fig. 10 (c) illustrates the EMSIQA distribution among the whole section. In most regions, the values are kept at a relevant low level like the blue box, while in some areas lack texture, and the performance is still not perfect, like the red box.

Table 3
Comparison of EMSIQA and other assessments (PSNR [31], NCC [33], SSIM [32] and HFI-SSIM [34]) in the three scenarios (6 image pairs) shown in Fig. 3. $E_G(struc.)$ means the geometric error of the cellular structures and $E_G(whole)$ denotes the geometric error of the whole image. The result shows that the proposed EMSIQA can better evaluate the stitching of EM images. ✓ indicates where the metric can serve as a successful assessment, while × denotes not.

	PSNR↑	NCC↑	SSIM↑	HFI-SSIM↑	EMSIQA (Proposed)			
					Overall↓	$E_G(struc.)$↓	$E_G(whole)$↓	M_S↑
Appearance	×	✓	×	×	✓			
Fig. 3 (a) blur	30.290	0.959	0.812	0.001	0.015	0.015	0.015	0.993
Fig. 3 (b) brightness change	11.873	1.000	0.751	0.999	0.034	0.033	0.035	0.991
Distortion Location	×	×	×	✓	✓			
Fig. 3 (c) boundary membranes	15.793	0.856	0.687	0.377	2.001	1.784	0.954	0.687
Fig. 3 (d) cytosolic areas	16.826	0.890	0.661	0.700	0.653	0.624	1.086	0.955
Structure Scale	✓	✓	×	×	✓			
Fig. 3 (e) thick structure	14.347	0.865	0.252	0.004	3.094	2.792	2.789	0.902
Fig. 3 (f) thin structure	12.662	0.711	0.270	0.004	3.591	2.806	2.793	0.781

5. Conclusion and Limitation

Contribution. To address the challenge in large-scale EM image mosaicking, we proposed a two-stage error detection method to assess the mosaicking in and after the processing. The first stage combines the learning-based GPOR and the classical RANSAC to examine the key point matches and detect the potential stitching errors before the time-consuming global optimization and image rendering. We proposed a hybrid feature framework, where the first stage is point-based error detection, to comprehensively optimize stitching speed and accuracy. The second stage takes advantage of a newly designed measurement of EM stitched image quality assessment (EMSIQA) to detect unsolved errors and to comprehensively evaluate the stitching result. Experiment results showed that our framework can significantly reduce the computation time compared with existing single-feature workflows, and meanwhile attain excellent stitching quality. The application of our framework to ultra-large multi-tile EM images of the adult mouse's striatum, glioblastoma, and corpus callosum showcased outstanding performance and robustness in mosaicking extensive and diverse EM images.

Limitations. The proposed hybrid feature framework takes advantage of different key features and achieves an optimized balance between performance and speed. However, naturally, it cannot surpass the upper limit of performance of the chosen features. Given the extensibility of our framework, more advanced features proposed in the future can be added to the hybrid features to achieve further improvement.

In the experiment **GPOR vs. RANSAC**(Sec 4.2) using synthetic dataset, the GPOR method adopted in this work can reach a high speed using GPUs since the feature numbers of all images are set to be equal at the data preparation stage. However, the computation speed did not reach our expectation on the three real EM datasets because the number of correspondences varies a lot in different tile pairs, which means we can not concatenate them together into a batch and compute them in parallel.

Besides, our framework is specially optimized for the images acquired by the multibeam SEM with small tile size, of which the non-linear distortion is negligible in most cases for stitching. Thus we here did not discuss the distortion nor apply our proposed scheme on deformed EM images where elastic transformation is needed.

Ethical Use of Laboratory Animals

All animal experiments in this paper comply with the Guidelines for Ethical Conduct in the Care and Use of Animals in Research (Guo-Ke-Fa-Cai-Zi [2016] No. 398) and the Code of Conduct in Laboratory Animal Management of Jiangsu Province. All procedures were performed in compliance with relevant laws and institutional guidelines of the Suzhou Institute of Biomedical Engineering and Technology, and the appropriate institutional committee approved them.

Conflict of Interest Statement

The Authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) did not use generative AI or AI-assisted technologies in the writing process.

Funding

This work was supported by the National Natural Science Foundation of China (32271430), and the CAS Project for Young Scientists in Basic Research (Grant No. YSBR-067).

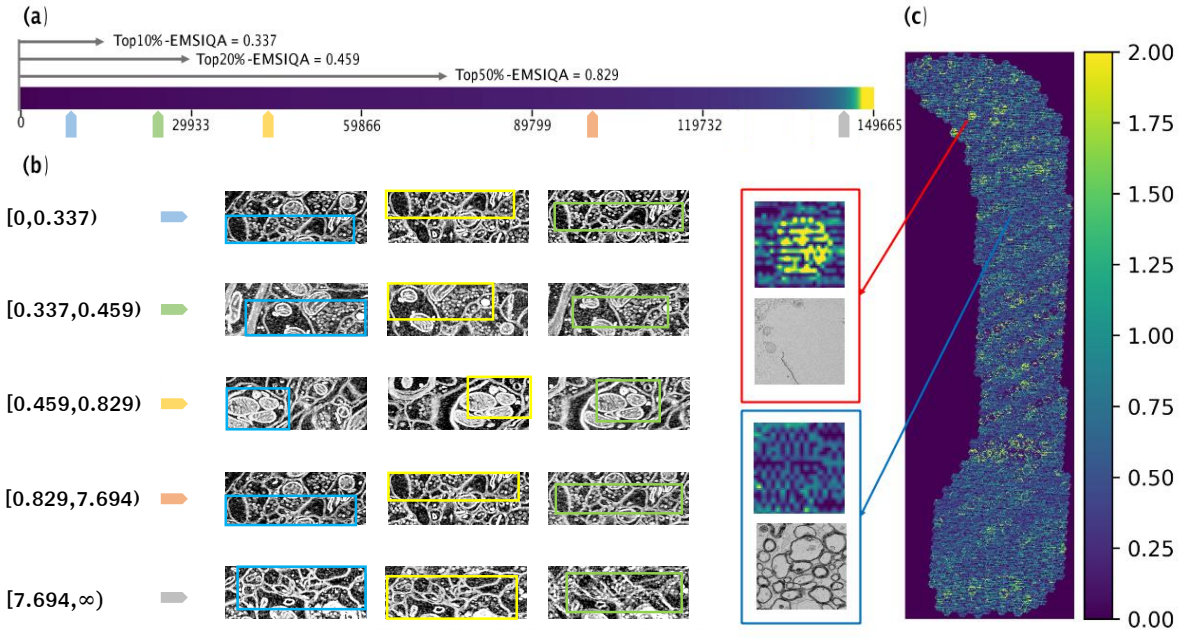


Figure 10: Visualized EMSIQA score distribution and stitching result of CC50k. (a) The EMSIQA distribution is in ascending order among 149666 overlapping areas and three cutoff point values. (b) The stitching results in different EMSIQA score ranges. The final image in each row is the stitching result of the first two images. (c) The EMSIQA distribution across the whole CC50k image, with two typical regions called out, which are mosaicked bad and well, respectively. A log function on the values.

Acknowledgment

We thank Prof. Lingxiao Zhao at the Suzhou Institute of Biomedical Engineering and Technology for his insightful discussion on SEM image stitching methods. Additionally, we thank Prof. Minxuan Sun, also from the Suzhou Institute of Biomedical Engineering and Technology, for providing the glioblastoma sample.

Appendix

A. Details of making the synthetic dataset

Cropping from a large image Supposing that the size of the large image and cropped image are [H, W] and [h,w], respectively. We first randomly select a point $[disp\_x_{src}, disp\_y_{src}]$ as the left-top of image 1, then a random affine matrix M is generated to represent the transformation between image 1 and image 2. We can formulate the coordinates of one point in these two images' coordinate systems as:

$$\mathbf{M} \cdot \mathbf{Pt}_{src}^{coor_1} = \mathbf{Pt}_{src}^{coor_2} \tag{A.1}$$

Next, we need to calculate $\mathbf{Pt}_{dst}$ in the large image's coordinate system. We can choose three corner points in image 2: [0, 0], [0, h], [w, 0] as the anchors. Thus, their coordinates in image 1 can be written as:

$$\mathbf{Pt}_{dst}^{coor_1} = \mathbf{M}^{-1} \cdot \mathbf{Pt}_{dst}^{coor_2} \tag{A.2}$$

According to the location of image 1 in the large image, we can get the global coordinates:

$$\mathbf{Pt}_{dst}^{coor_g} = \mathbf{M}^{-1} \cdot \mathbf{Pt}_{dst}^{coor_1} + [disp\_x_{src}, disp\_y_{src}, 0]^T \tag{A.3}$$

However, the parallelogram solved by the three points in the global coordinate system cannot be directly cropped because in most cases it is not a rectangle so we have to transform the large image using the matrix:

$$\mathbf{M}_{large2small} = getAffine(\mathbf{Pt}_{dst}^{coor_g}, \mathbf{Pt}_{dst}^{coor_1}) \tag{A.4}$$

Then, cropping the transformed large image using $M_{large2small}$ into [h, w] can generate image 2.

Normalization In real EM data, the dimensions of tiles are not constant so the learning-based outlier rejection model has to be trained under a normalized coordinate system. We constrain the coordinates in the range of [-1, 1] following:

$$\begin{aligned} x_1^{norm} &= \frac{2x_1}{w_1} - 1, \quad y_1^{norm} = \frac{2y_1}{h_1} - 1, \\ x_2^{norm} &= \frac{2x_2}{w_2} - 1, \quad y_2^{norm} = \frac{2y_2}{h_2} - 1 \end{aligned} \tag{A.5}$$

Supposing that the affine matrix M is:

$$M = \begin{pmatrix} f11 & f12 & f13 \\ f21 & f22 & f23 \end{pmatrix} \tag{A.6}$$

Table 4
EMSIQA results and running time on **ST91** dataset, a real-world dataset containing 91 tiles.

Method	All↓	Intra%↓	Inter%↓	Time↓
SIFT [16]	3.28	3.18	3.58	39'43"
SURF [19]	3.44	3.46	3.39	31'34"
ORB [15]	42.58	45.95	32.85	8'05"
AKAZE [20]	24.01	30.02	6.67	13'43"
BRISK [22]	22.24	25.99	11.39	17'48"
Proposed	4.31	4.27	4.41	15'43"

Table 5
EMSIQA results and running time on **GBM9** dataset, a real-world dataset containing 9 tiles.

Method	All↓	Best%↓	Worst%↓	Time↓
SIFT [16]	3.20	0.49	3.66	16"
SURF [19]	3.11	0.51	3.66	15"
ORB [15]	5.62	0.65	11.65	6"
AKAZE [20]	3.40	0.59	4.15	11"
BRISK [22]	3.27	0.47	3.84	14"
Proposed	3.21	0.57	3.72	10"

then the corresponding elements in normalized matrix M^{norm} can be formulated as:

$$\begin{aligned}
[0,0] &= f11 \times \frac{w1}{w2} \\
[0,1] &= f12 \times \frac{h1}{w2} \\
[0,2] &= f11 \times \frac{w1}{w2} + f12 \times \frac{h1}{w2} + f13 \times \frac{2}{w2} - 1 \\
[1,0] &= f21 \times \frac{w1}{h2} \\
[1,1] &= f22 \times \frac{h1}{h2} \\
[1,2] &= f21 \times \frac{w1}{h2} + f22 \times \frac{h1}{h2} + f23 \times \frac{2}{h2} - 1
\end{aligned} \tag{A.7}$$

B. EMSIQA results and running time for the ST91 and GBM9 datasets

Here we present the results of datasets **ST91** and **GBM9** in Table 4 and Table 5.

References

[1] A. Aswath, A. Alsahaf, B. N. Giepmans, and G. Azzopardi, "Segmentation in large-scale cellular electron microscopy with deep learning: A literature survey," *Medical image analysis*, p. 102920, 2023.

[2] K. L. Briggman and D. D. Bock, "Volume electron microscopy for neuronal circuit reconstruction," *Current Opinion in Neurobiology*, vol. 22, no. 1, pp. 154–161, 2012. Neurotechnology.

[3] N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. e. A. Conchello, S. Knowles-Barley, D. Lee, A. Vāzquez Reina, V. Kaynig, T. R. Jones, *et al.*, "Saturated reconstruction of a volume of neocortex," *Cell*, vol. 162, no. 3, pp. 648–661, 2015.

[4] L. Swanson and J. Lichtman, "From cajal to connectome and beyond," *Annual review of neuroscience*, vol. 39, pp. 197–216, 07 2016.

[5] J. L. Muhlich, Y.-A. Chen, C. Yapp, D. Russell, S. Santagata, and P. K. Sorger, "Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ashlar," *Bioinformatics*, vol. 38, no. 19, pp. 4613–4621, 2022.

[6] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3254–3261, 2014.

[7] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving dlt," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2339–2346, 2013.

[8] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *CVPR 2011*, pp. 49–56, 2011.

[9] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, "Smoothly varying affine stitching," in *CVPR 2011*, pp. 345–352, 2011.

[10] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep image homography estimation," 2016.

[11] L. Nie, C. Lin, K. Liao, M. Liu, and Y. Zhao, "A view-free image stitching network based on global homography," *Journal of Visual Communication and Image Representation*, vol. 73, p. 102950, 2020.

[12] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 467–483, Springer International Publishing, 2016.

[13] C. Zhao, Z. Cao, C. Li, X. Li, and J. Yang, "Nm-net: Mining reliable neighbors for robust feature correspondences," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 215–224, 2019.

[14] S. Saalfeld, A. Cardona, V. Hartenstein, and P. Tomančák, "As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets," *Bioinformatics*, vol. 26, pp. i57–i63, 06 2010.

[15] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International Conference on Computer Vision*, pp. 2564–2571, 2011.

[16] D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 vol.2, 1999.

[17] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "Lf-net: Learning local features from images," *CoRR*, vol. abs/1805.09662, 2018.

[18] T. Tasdizen, P. Koshevoy, B. C. Grimm, J. R. Anderson, B. W. Jones, C. B. Watt, R. T. Whitaker, and R. E. Marc, "Automatic mosaicking and volume assembly for high-throughput serial-section transmission electron microscopy," *Journal of Neuroscience Methods*, vol. 193, no. 1, pp. 132–144, 2010.

[19] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[20] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 34, no. 7, pp. 1281–1298, 2011.

[21] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *European conference on computer vision*, pp. 214–227, Springer, 2012.

[22] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *2011 International conference on computer vision*, pp. 2548–2555, Ieee, 2011.

[23] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2666–2674, 2018.

[24] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," vol. 24, p. 381–395, June 1981.

[25] Y. Liu, J. Tian, R. Hu, B. Yang, S. Liu, L. Yin, and W. Zheng, "Improved feature point pair purification algorithm based on sift during endoscope image stitching," *Frontiers in Neurorobotics*, vol. 16,

p. 840594, 2022.

[26] P. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[27] O. Chum and J. Matas, "Matching with prosac - progressive sample consensus," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 220–226 vol. 1, 2005.

[28] C. Choy, J. Lee, R. Ranftl, J. Park, and V. Koltun, "High-dimensional convolutional networks for geometric pattern recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11227–11236, 2020.

[29] C. Choy, J. Gwak, and S. Savarese, "4d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3075–3084, 2019.

[30] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85, 2017.

[31] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in *Optical Pattern Recognition XII*, vol. 4387, pp. 95–102, International Society for Optics and Photonics, 2001.

[34] H. Qureshi, M. Khan, R. Hafiz, Y. Cho, and J. Cha, "Quantitative quality assessment of stitched panoramic images," *IET image processing*, vol. 6, no. 9, pp. 1348–1358, 2012.

[35] L. Yang, Z. Tan, Z. Huang, and G. Cheung, "A content-aware metric for stitched panoramic image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 2487–2494, 2017.

[36] C. Tian, X. Chai, and F. Shao, "Stitched image quality assessment based on local measurement errors and global statistical properties," *Journal of Visual Communication and Image Representation*, vol. 81, p. 103324, 2021.

[37] M. Awad and R. Khanna, "Support vector regression," in *Efficient learning machines*, pp. 67–80, Springer, 2015.

[38] H. Ullah, M. Irfan, K. Han, and J. W. Lee, "Dlnr-siqa: Deep learning-based no-reference stitched image quality assessment," *Sensors*, vol. 20, no. 22, p. 6457, 2020.

[39] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[40] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics (TOG)*, vol. 2, no. 4, pp. 217–236, 1983.

[41] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2462–2470, 2017.

[42] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[43] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[44] A. Cardona, S. Saalfeld, J. Schindelin, I. Arganda-Carreras, S. Preibisch, M. Longair, P. Tomancak, V. Hartenstein, and R. J. Douglas, "Trakem2 software for neural circuit reconstruction," *PloS one*, vol. 7, no. 6, p. e38011, 2012.

[45] D. Haehn, J. Hoffer, B. Matejek, A. Suissa-Peleg, A. K. Al-Awami, L. Kamentsky, F. Gonda, E. Meng, W. Zhang, R. Schalek, *et al.*, "Scalable interactive visualization for connectomics," in *Informatics*, vol. 4, p. 29, Multidisciplinary Digital Publishing Institute, 2017.

[46] G. Bradski, "The opencv library.," *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

●Balancing electron microscopy mosaicking speed and accuracy is achievable.

●Integrating varied features aids in precise, efficient EM image stitching.

●EM stitched image quality assessment ensures precise stitching evaluation.

The Authors declare no conflict of interest.