

A Weakly Supervised Pre-processing Pipeline for Multi-Modal Image Pair Generation for Image Registration

Daksh Daksh

Max Planck Institute for Polymer Research

Anke Kaltbeitzel

Max Planck Institute for Polymer Research

Gunnar Glaßer

Max Planck Institute for Polymer Research

Katharina Landfester

Max Planck Institute for Polymer Research

Ingo Lieberwirth

`lieberw@mpip-mainz.mpg.de`

Max Planck Institute for Polymer Research

Research Article

Keywords:

Posted Date: August 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10279879/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Weakly Supervised Pre-processing Pipeline for Multi-Modal Image Pair Generation for Image Registration

Daksh Daksh ¹, Anke Kaltbeitzel ¹, Gunnar Glaßer ¹, Katharina Landfester ¹, Ingo Lieberwirth ^{1*}

¹Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany.

*Corresponding author (Email: lieberw@mpip-mainz.mpg.de, Tel.: +49 6131 379 580)
contacts:

Daksh Daksh: daksh@mpip-mainz.mpg.de

Anke Kaltbeitzel: kaltbeit@mpip-mainz.mpg.de

Gunnar Glaßer: glasser@mpip-mainz.mpg.de

Katharina Landfester: landfest@mpip-mainz.mpg.de

Abstract

Recent advances in AI-based image analysis have substantially improved the efficiency of extracting meaningful information from large-scale image datasets. However, the effectiveness of these methods strongly depends on the quality of input data, necessitating robust and standardized image preprocessing steps to ensure accurate and reproducible results. In this study, we present a tailored preprocessing workflow designed to address the specific challenges associated with the registration of multimodal microscopy data. The core objective of this workflow is to prepare for the precise spatial alignment of images acquired from disparate imaging modalities, specifically electron microscopy (EM) and light microscopy (LM). Due to fundamental differences in contrast mechanisms, resolution, and field of view, direct correlation between such images remains a non-trivial task. Our approach aims to generate image pairs representing the same region of interest, thereby facilitating accurate registration. To this end, we implement a normalized cross-correlation-based template matching algorithm to localize corresponding features across modalities. In addition, to reconcile disparities in pixel dimensions, we exploit the presence of common structural landmarks in both image types, which serve as internal standard for calibrating the pixel size. Pixel size calibration constitutes

a critical step in the preprocessing pipeline, as it directly impacts the performance of template matching and downstream registration procedures. Notably, this step currently requires manual supervision, as there is a lack of computational tools capable of autonomously adjusting pixel dimensions across heterogeneous imaging datasets. Upon completion of these preprocessing steps, the resulting image pairs are suitable for subsequent image analysis tasks, including registration, segmentation, and quantitative assessment. Moreover, the proposed weakly supervised preprocessing framework offers potential for integration into automated microscopy platforms, thereby enabling unsupervised, high-throughput processing of multimodal image data.

Introduction

Image pre-processing¹ is a fundamental and crucial step in the field of computer vision² and image analysis.^{3,4} It involves a series of techniques used to prepare raw image data for analysis and model training. It is the process of preparing raw image data to feed into algorithms or machine learning models⁵ for further analysis, learning, or decision-making. Since images in their raw form can contain various challenges such as noise, inconsistencies, and other relevant information, the image preprocessing aims to remove such hindrances and enhance the relevant features of an image, which ultimately helps in improving the performance of computing algorithms.⁶ The image preprocessing can include operations like resizing, normalization, noise reduction, image augmentation, and different transformations.^{7, 8, 1} That means by transforming the input data into a more suitable form, image preprocessing ensures that models can better understand and learn from the visual information, leading to more accurate predictions and insights. In this context, effective pre-processing is essential for achieving robust and reliable results in tasks like object detection, image classification, image segmentation and recognition for example in different biological and medical image datasets.^{6,}

4, 9

Here, we are going to address the image preprocessing for the registration of correlated light and electron microscopy images (CLEM).¹⁰ In microscopy, images which come from diverse microscopes have different resolutions, pixel sizes, modalities, unique colour schemes, and aspect ratios.^{7, 11} While conventional light and confocal laser scanning microscopes have a physical resolution limit of approximately 200 nm, an electron microscope can easily achieve sub-nm resolution.¹² On the other hand, the LM is capable to localize specific molecules using fluorescence labelling, whereas the EM displays the entire morphological structure, but without

molecular selectivity. Therefore, since both microscopes deliver complementary local information on the sample, it is of interest to combine both images using image registration approaches to achieve added value from the strengths of both methods.

Modern microscopes can acquire images with quite large pixel dimensions; images with 4k x 4k pixels and larger are becoming standard nowadays. In order to give the acquired images a meaningful dimension, the microscope is calibrated at each magnification yielding a defined pixel size, which denotes the real space dimension of one pixel in a unit of e.g. nm/pixel. This can then be used to measure real space distances of any object in the image. For LM images, the conducive lower limit of the pixel size is in the order 100 nm/pixel, as this approximately reflects half the physical resolution limit. Lower pixel sizes will not yield any information, but instead e.g. increase the acquisition time and consumes unnecessary memory. For EM imaging, on the other side, pixel sizes can be as low as some 0.1 nm/pixel for extremely high magnifications.^{7, 13, 10}

Another important definition for the registration is the field of view (FoV) of an image. It refers to the depicted area visible in an image. This means that two multimodal images with the same FoV should display the same image area, regardless of the pixel size.

The aim of the image preprocessing approach in the case of multimodal input images is to precisely adjust the pixel size, the image dimensions and the FoV of both input images to generate an image pair, suitable for subsequent registration algorithms. The usual starting situation with CLEM data is that a source LM image with a large FoV is used and smaller areas, so-called templates, are then captured using EM. The aim of the preprocessing is to crop out the exact FoV from the LM image source, matching the EM template image and resize the images to identical image dimensions. The resulting image pair can then serve as optimal input for a subsequent machine learning registration process.

In order to identify and subsequently crop a template from a given source image, a template matching approach is ideal. Template matching is a computer vision technique that compares portions of an image (templates) with the source image to find the region that best matches the template^{14, 15, 16}. While it is a basic method compared to more advanced image recognition techniques, template matching can be useful in certain situations where simple alignment between the template EM and source LM images is required.

Therefore, we implemented a template matching approach into our weakly supervised image pre-processing pipeline. It can quickly locate the initial position of the template image within

the source image. Moreover, the presented preprocessing is a general approach and can easily be transferred to any multimodal image system.¹

Results and Discussions

To address the fundamental challenges associated with raw image handling and to improve the overall imaging workflow, we propose a novel image preprocessing pipeline (Figure 1). The pipeline is hardware-independent, easy to integrate into existing imaging systems, and designed to be broadly applicable.

In the first step, the different pixel size calibrations, which originate directly from the physical calibration of both microscopes as metadata, must be matched. However, the initial microscope calibrations can deviate considerably from each other. I) In order to obtain a precisely matching calibration of both images, we need a common feature in both images from which we can determine the pixel size ratio (PSR) between both images. II) The LM and EM images together with the PSR are the input for the following, unsupervised pipeline. III) In order to generate a suitable template from the EM input image, we reduce the EM image dimensions by binning down according to the calculated PSR value, while the LM image serves as the source image. IV) Subsequently, we apply the area based template matching to localize the template position within source image. V) Finally, the identified area is cropped from the LM source image and resized using the determined PSR value. Together with the raw EM image, this results in an image pair that shares the same field of view (FoV) and has a calibrated, matching pixel size—while preserving the original resolution of the EM image.

The core of this approach is based on template matching, which has the unique ability to search for and locate a template image within a larger source image.^{14, 16} We tested two template matching approaches to solve the multi-modal problem for generating an image pair, either using normalized cross-correlation method or normalized correlation coefficient method.^{14, 15} Here, we ultimately chose the latter method due to its superior performance compared to the cross-correlation approach, particularly in cases where the template and image regions exhibit differing lighting contrasts, as it effectively normalizes intensity variations. This process determines the spatial location of a specified pattern / template via a pixel-wise

correlation between the source input image and a predefined template encapsulating the target structure. The template is translated by u discrete units along the x-axis and v units along the y-axis of the image, with similarity metrics computed over the template region at each spatial offset. The standard normalized cross correlation $R(u,v)$ in template matching is mathematically represented as:

$$R(u, v) = \frac{\sum \sum ((I(x+u, y+v) - I') (T(x, y) - T'))}{\sqrt{\sum (I(x+u, y+v) - I')^2} \cdot \sqrt{\sum (T(x, y) - T')^2}}$$

, where I' and T' is the mean region in the source and template image respectively.

As the multi-modal images are influenced by intensity variations, the mean subtraction needs to be avoided. This is the key strategy here that we are not going to compensate for intensity change, as it has to be determined through the low dimensional image. Therefore, the standard normalized cross correlation coefficient is modified as:

$$R'(u, v) = \frac{\sum \sum (I(x+u, y+v) T(x, y))}{\sqrt{\sum (I(x+u, y+v))^2} \cdot \sqrt{\sum (T(x, y))^2}}$$

With the help of the normalized correlation coefficient template matching, we can match the template image within the source image. The obtained matching result provides the precise location of the region that has to be cropped out from the LM image.

This pre-processing pipeline can be implemented for generating any multi-modal image pair of identical features e.g. in medical, biological or satellite image analysis.^{1, 13} Moreover, it is truly independent of any image metadata parameters.

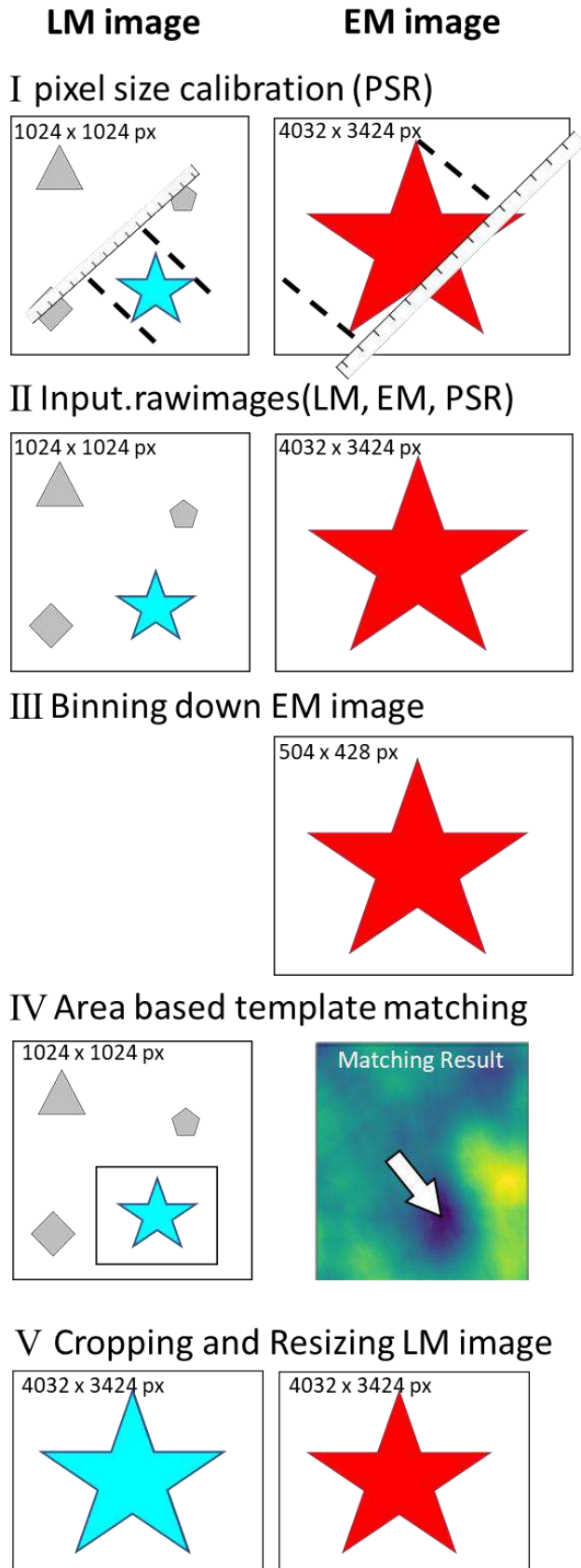


Fig. 1: Graphical representation of the workflow of our weakly supervised image pre-processing algorithm. I) Before applying the automated pre-processing algorithm, the exact pixel size calibration ratio (PSR) between the two input images has to be determined manually by using internal, common landmarks. II) the input feed comprises of the LM and EM raw images of

different image dimensions and the PSR in addition. Usually, the image dimension of the EM image is significantly larger than that of the LM image. III) The first step of the pre-processing algorithm is to bin down the EM raw image to a pixel dimension using the calculated PSR. IV) With this reduced EM image the area based template matching is applied using normalized cross-correlation. V) After successful matching, the respective area is cropped from the raw LM image and is finally resized to the raw EM image dimension using the calculated PSR.

In Figure 2, the initial situation for the two input images is displayed. In columns a and b the initial input EM and LM images are shown. As can be seen, the FoV in the EM image is significantly smaller than the FoV of the LM image. However, the task of our presented workflow is to precisely locate and crop the EM image FoV from the LM image (the red box in Fig. 2). As for a successful application for subsequent machine learning registration approaches, the generated image pairs need to have an identical FoV and equal pixel dimensions. This requires a thorough determination of the pixel size ratio (PSR) between both image modalities. The PSR between the two microscopes is defined as the quotient of LM and EM pixel size. During our analyses, we found that the pixel size calibrations, as coming from the two different microscope metadata, do not match precisely, as clearly visible when comparing panels 1a and 1c. Hence, the cell appears to be much larger in the LM image compared to the EM image and the cropping window does not match (white arrows).

Moreover, the two EM-images presented in Figure 2 were acquired at different magnifications having a microscope pixel calibration of 15.7 nm/px and 12.7 nm/px for the upper and the lower EM image, respectively. In combination with the LM resolution of 83 nm/px, this yields a PSR of 5.28 and 6.55, respectively. Column c displays the result of resizing the cropped LM images using the microscope metadata obtained PSR. It clearly shows a discrepancy of the FoV with respect to the EM image. However, because we are aiming to extract the same FoV from both images, we can use an internal, landmark based calibration standard to match the pixel size of both images. The process of calibrating the PSR is by simply comparing the distance of identifiable landmarks in both images, as indicated by the red arrows in Fig. 2. Here, the ratio of the distances in pixels is used to calculate the PSR, which finally leads to an equivalent FoV, with the landmark distances having the same pixel length, as displayed in column d, with the manually calculated PSR values of 5.1253 and 6.095, which deviate from the microscope calibrations by 3 and 7 %, respectively.

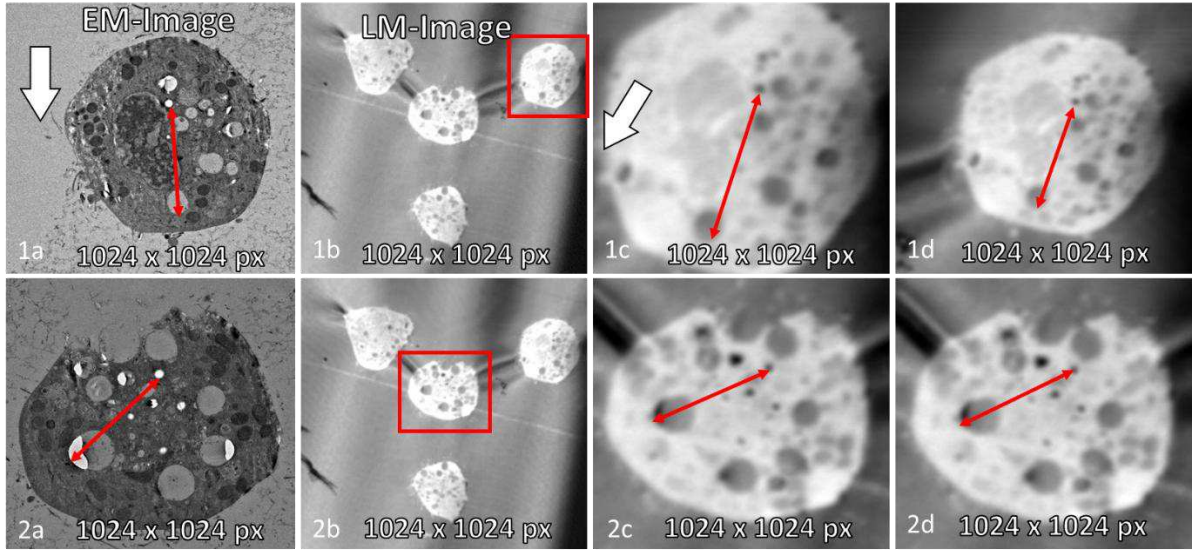


Fig. 2: Pixel size calibration process: Columns a) and b) display the raw EM and LM input images. It should be noted, that the EM images have been acquired at different magnifications. As can be seen the FoV of the LM image is considerably larger than the EM image; the area corresponding to the EM image a) is marked by the red box in the LM image b). Both input images have the same image dimension of 1024 x 1024 px. In order to maintain the image dimensions, we need to rescale the marked area in the LM image to achieve the same image dimension of the EM input. c) Rescaling the LM image using the microscope pixel calibration values yields a distinct deviation in the FoV for the upper panel images, as can be seen by the white arrows. d) Manually calibrated pixel size ratio by using common landmarks in both on the images. This finally yield a match in the FoV of both images. Although 2c) visually seems to display the similar FoV, the measured length of the landmarks (red lines in 2a and 2c) deviate by 7 % and needs to be recalibrated as shown in 2d, where the mismatch of the landmarks is compensated.

Subsequently, a template for the area based template matching has to be prepared from the lower FoV input image, in this case from the EM image. This involves downscaling the EM image to a certain factor. Figure 3 displays the influence of this downscaling factor, which has to be carefully determined. Depending on the downscaling factor, the image dimension of the generated template (Fig. 3 b) can be adjusted. Here, we used different downscaling factors, ranging from 1/7 to 1/10, yielding template dimensions between 732 x 508 px and 512 x 354 pixels, respectively. With these templates we applied the area based template matching using normalized cross correlation coefficient method. The respective results for the crop-out in the LM image are displayed in Fig. 3 c. Obviously, the 1/10 downscaling approach yields a cut-off in the FoV and so does the 1/9 downscaling. On the other side, the 1/7 downsizing leads to a mislocalization (blue box in Fig. 3 c). The optimal result is achieved using a downsizing factor of 1/8 (Figure 3 d). This corresponds exactly to the PSR value of 8.05 between the two input images. Therefore, our observations show, that exact resizing of the template with regards to the source-image is an essential prerequisite for successful template matching. Hence, we implement the PSR as the crucial parameter for any resizing procedures. At this point, it is worth mentioning again, that the determination of the PSR still is a manual process in our

workflow. The development of an unsupervised, automatic PSR determination will be the last step towards a fully autonomous image pair generation.

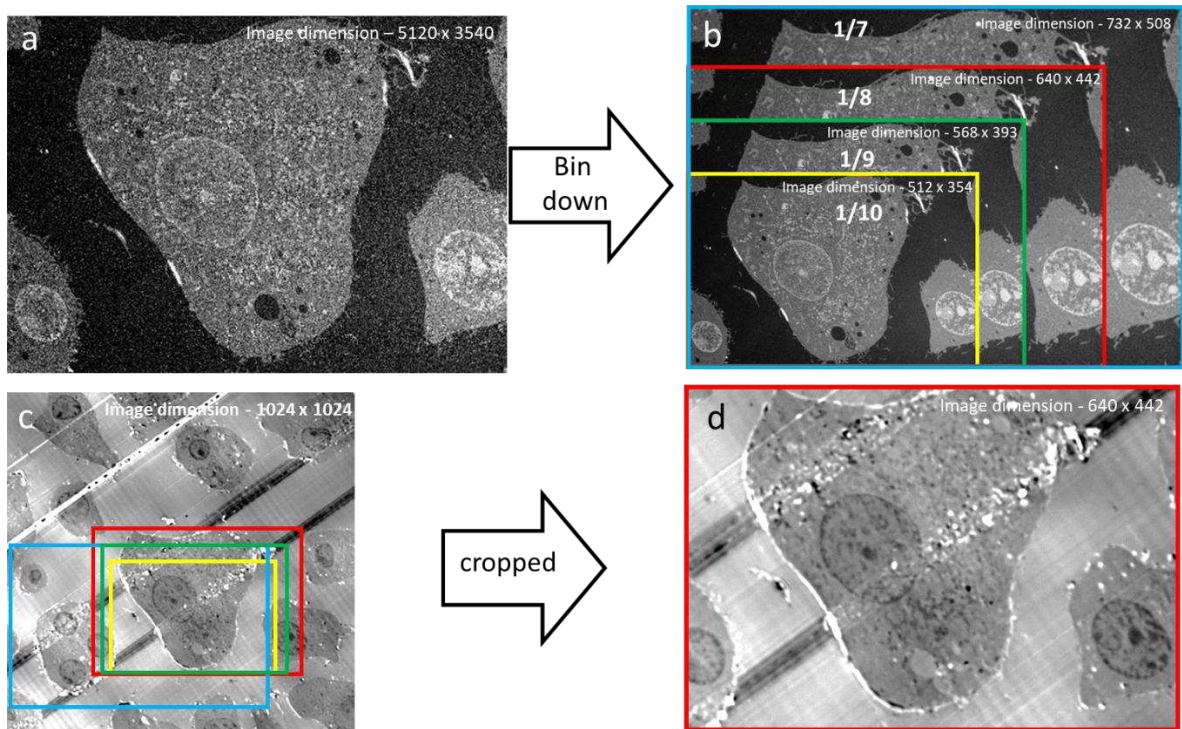


Fig. 3: Binning, matching and cropping operations: The raw EM input image a) has to be binned down before the area based matching. However, the binning factor plays a crucial role for successful matching. b) Testing different bin-down factors (1/7 to 1/10) for the EM-image. c) The matching results obtained for the different binned down EM images are indicated by the coloured boxes. As can be seen, the best matching result is obtained for the 1/8 times binning. This equals exactly the PSR value. For higher bin-down factors the matching fails, as indicated by the yellow and green box. d) displays the cropping result showing a perfect matching between LM and EM image feature.

Usually, template matching approaches are applied on single modal datasets, i.e. that the template image is extracted from the source image itself.^{14, 16} Therefore, there are naturally no differences between the template and the source in terms of pixel size and information content. However, if the template and source originate from different microscopes, then these two have different characteristics, especially in terms of pixel size. As we have demonstrated in Fig. 3, the precise resizing of the template with regards to the source image pixel size is of utmost importance. Moreover, due to the multimodal nature of source and template, the two input images are looking quite different and yet our matching approach is successful. The underlying key factor for this is passing the template with a gaussian filter, which enhances its characteristic features and reduces image noise. These enhanced features finally enable successful localization of the template position within the source image, which finally results in the generation of a multi-modal image pair that can be further used in different applications.

Figure 4 shows two more examples of cropping an EM template image from an LM image source. After resizing the LM image, the image pair has identical image dimensions, e.g. the image dimension of the raw EM input image and common landmarks in both images have the same distance (yellow markers in Fig. 4). However, close inspection of this image pair reveals a very subtle translational shift between both images, as indicated by the white arrows. Moreover, as can be concluded from the example in the lower pane of Fig. 4, the area based template matching can even successfully handle artefacts, which appear only in one of the input images. The bright areas in Fig. 4 d are caused by the wrinkling of the thin section specimen during sample preparation and thus induce a high signal in the reflected channel of the confocal Laser Scanning Microscope (cLSM). Accordingly, the template matching approach is very robust, even if one of the input images contains additional features, such as the wrinkles in this case.

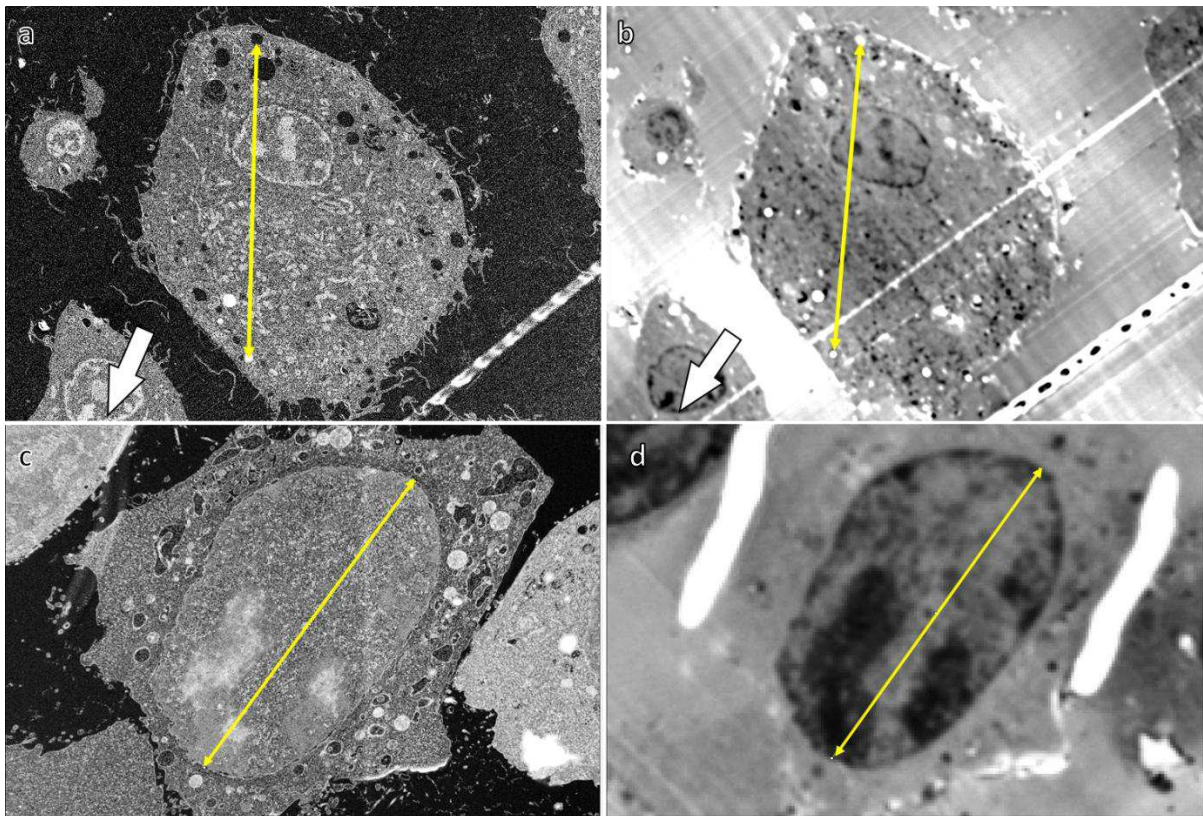


Fig. 4: Examples of processed image pairs, readily precisely cropped and pixel size calibrated. Images a and c display the input EM image, with pixel dimensions of 5120 x 3540 px, b and d show the precisely cropped area from the raw LM image with equal dimension of 5120 x 3540 px. Due to the prior determination of the exact pixel size ratio between both input images, the LM and EM images have exactly the same pixel calibration. This can be easily checked by measuring the distance between common features, as shown by the yellow line. These two distances in the two pairs of images have the same length in pixels (landmark distance). Furthermore, it is worth mentioning that the area-based template matching still works, even if the image

pairs differ significantly due to artefacts, as can be seen from the two bright objects in d. These are imaging artefacts caused by wrinkling of the thin sections in the LM, which are not visualized by the EM.

The upscaled and cropped LM image (including all channels, reflected and fluorescence) with correspondence to the FoV of the original EM image contribute an image pair, that can be further used for image registration applications. As displayed in Fig. 5, the simple overlay of the image pair already yields a reasonable registration. To illustrate the correspondence between the two images, we have superimposed the EM and the reflected channel of the LM image. This representation is the best way to recognize shifts between the two images and the slight misalignments of the simple overlay are clearly visible, as indicated by the white arrows in Figure 5. Using the fluorescence channel, however, it is almost impossible to judge the quality and precision of the overlay, as can be seen in Figs. S1 and S2.

In order to demonstrate the use of this pre-processing in the framework of an automated registration routine, we have applied our custom made machine learning based registration software (<https://keeper.mpdl.mpg.de/f/6ad35c6b234f45838a08/>) to the generated image pairs. As shown in the right panel of Fig. 5, the slight mismatch between the input images can subsequently be eliminated by machine learning based approaches. However, we found that even for machine learning approaches a low mismatch error below 1 % is required for successful image registration. Therefore, the precise rescaling of the images in a pre-processing step is of utmost importance. The total computation time of our pipeline from starting to end is less than a minute. Hence, it is fast, easy to implement and reliable on a common desktop PC. We have developed our supervised image pre-processing pipeline in the Jupyter Notebook using Python on the Windows 11 Pro desktop. The configuration consists of 11th Gen Intel(R) core™ CPU i9-11900k @ 3.50 GHz 3.50 GHz with 128 GB RAM having Nvidia GeForce RTX 3080 Graphics. As a note, we want to highlight that all the operations of our pipeline execute only on the CPU core, hence no GPU compatibility is required.

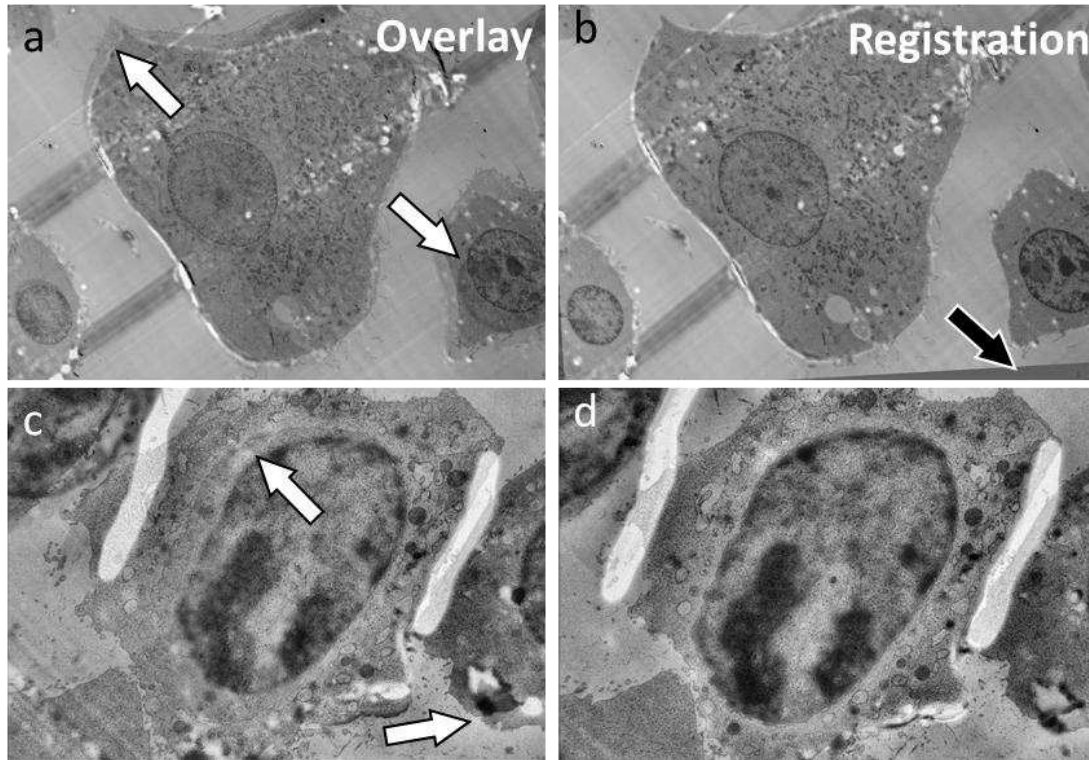


Fig. 5: Registration results of cropped image pairs from Fig 3 a and d and Fig. 4 c and d. Here, the images display an overlay of the EM and the reflective channel of the LM image. This type of representation makes it possible to identify shifts between the two images. In the left column (a and c) the cropped image pairs are simply overlaid, without any further image registration. The white arrows indicate features, where the offset between both images is clearly visible. In the right column (b and d), we applied our machine learning registration software to align both input images, yielding a precise matching. Even a slight rotation between the two input images can be compensated for with such an approach (black arrow).

During the development of this weakly supervised preprocessing pipeline, we have focused on two aspects in particular: Time efficiency and improving accuracy with easy implementation on different operating systems (OS) e.g. Windows and Linux, therefore enabling straightforward operation.

Conclusions

We are presenting a novel and easily implemented weakly supervised image pre-processing algorithm that efficiently generates image-pairs of different modalities and imaging systems. It not only solves the challenges and hurdles of having consistent resolution but also serves as starting point for analysis of image processing. It is worth stating that the precise determination of the PSR between both image modalities is of utmost importance for all subsequent image analysis procedures and that the pixel calibration metadata coming from both microscopes may not be sufficiently accurate. By enabling the seamless integration of images of different resolutions, our pipeline enhances multimodal analysis, facilitates high-throughput processing, improves data accuracy, and supports the development of multiscale models. Additionally, our

pre-processing system might indicate the general structure for workflows that are more efficient, reduce human error, and enable innovative research that requires precise and accurate image fusion.

Data Availability

All image data and the python code are available from the authors upon reasonable request.

Code Availability

All Scripts and image datasets related to this manuscript is available from the corresponding author upon reasonable request.

Acknowledgements

The authors would like to express their gratitude to Katrin Kirchoff, Katja Klein, Renate van Brandwijk-Peterhans, Christoph Sieber, Richard DaCosta Marques, Shen Han, Yeliz Yangazoglu, Alina Heck and Konrad Maxeiner for providing and preparing the samples for microscopy. For their help with setting up the IT infrastructure, we would like to thank Stefan Schuhmacher.

DD acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG; project number 213555243 grand no. SFB1066). Financial support from the Max-Planck Society (MPG) is acknowledged as well.

References

1. Salvi, M., Acharya, U.R., Molinari, F. & Meiburger, K.M. The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine* **128**, 104129 (2021).
2. Szeliski, R. Computer vision: algorithms and applications. (Springer Nature, 2022).
3. Petrou, M.M. & Petrou, C. Image processing: the fundamentals. (John Wiley & Sons, 2010).
4. Van der Walt, S. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
5. Mahesh, B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*.*[Internet]* **9**, 381-386 (2020).
6. Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience* **2018**, 7068349 (2018).
7. Merchant, F. & Castleman, K. Microscope image processing. (Academic press, 2022).
8. Bow, S.T. Pattern recognition and image preprocessing. (CRC press, 2002).
9. Zhou, Z.-H. Machine learning. (Springer nature, 2021).
10. Goshtasby, A.A. Image registration: Principles, tools and methods. (Springer Science & Business Media, 2012).
11. Goodhew, P.J. & Humphreys, J. Electron microscopy and analysis. (CRC press, 2000).
12. Murphy, D.B. & Davidson, M.W. Fundamentals of light microscopy and electronic imaging. (John Wiley & Sons, 2012).
13. Chen, J. et al. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615 (2022).
14. Brunelli, R. Template matching techniques in computer vision: theory and practice. (John Wiley & Sons, 2009).
15. Briechle, K. & Hanebeck, U.D. in Optical pattern recognition XII, Vol. 4387 95-102 (SPIE, 2001).
16. Hashemi, N.S., Aghdam, R.B., Ghiasi, A.S.B. & Fatemi, P. Template matching advances and applications in image analysis. *arXiv preprint arXiv:1610.07231* (2016).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIILV2.docx](#)