Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Identification of microscopic structures in CFRP from X-ray CT based on topological data analysis

Xichan Gao, Kazuto Akagi, Daiki Kido, Mayrene Uy & Masao Kimura

To cite this article: Xichan Gao, Kazuto Akagi, Daiki Kido, Mayrene Uy & Masao Kimura (28 Oct 2025): Identification of microscopic structures in CFRP from X-ray CT based on topological data analysis, Science and Technology of Advanced Materials: Methods, DOI: 10.1080/27660400.2025.2572288

To link to this article: https://doi.org/10.1080/27660400.2025.2572288

Publisher: Taylor & Francis & The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group

Journal: *Science and Technology of Advanced Materials: Methods*

DOI: 10.1080/27660400.2025.2572288

Identification of microscopic structures in CFRP from X-ray CT based on topological data analysis

Xichan Gao[a]*, Kazuto Akagi[a]*, Daiki Kido[b,c], Mayrene Uy[b], Masao Kimura[b,c]

[a]*The Advanced Institute for Materials Research (AIMR), Tohoku University, Sendai, Japan;* [b]*Institute of Materials Structure Science (IMSS), High Energy Accelerator Research Organization (KEK), Tsukuba, Japan;* [c]*School of High Energy Accelerator Science, Dep. Materials Structure Science, The Graduate University for Advanced Studies (SOKENDAI), Tsukuba, Japan*

Xichan Gao: xichan.gao.e7@tohoku.ac.jp
The Advanced Institute for Materials Research (AIMR), Tohoku University, 2-1-1 Katahira,
Sendai, Miyagi 980-8577, Japan

Kazuto Akagi: kazuto.akagi.b5@tohoku.ac.jp
The Advanced Institute for Materials Research (AIMR), Tohoku University, 2-1-1 Katahira,
Sendai, Miyagi 980-8577, Japan

Identification of microscopic structures in CFRP from X-ray CT based on topological data analysis

ABSTRACT

X-ray computed tomography (X-ray CT) is widely used to examine the internal structure of carbon fiber-reinforced plastics (CFRPs), yet challenges remain due to the trade-off between field-of-view and spatial resolution. This study presents a framework based on topological data analysis (TDA) using persistent homology to identify carbon fiber positions and cracks in CFRPs from X-ray CT images. Degree-1 persistence diagrams (PD1), obtained via both superlevel-set and sublevel-set filtrations of grayscale X-ray CT images, enable the identification of fiber positions while avoiding artifacts in the resin matrix. The framework is also applied to the detection of particle positions in a scanning electron microscopy (SEM) image. Comparative evaluation with watershed and local thickness algorithms demonstrates the advantages of our method. Segmentation of crack regions is also achieved through a refined application of PD1 using superlevel-set filtration. This framework is applicable to a variety of grayscale images beyond X-ray CT.

KEYWORDS: CFRP; fiber positions; crack, persistent homology; TDA; X-ray CT

1. Introduction

Carbon fiber-reinforced plastics (CFRPs), composed of high-strength carbon fibers embedded in a resin matrix, are extensively used in aerospace, automotive, and robotics industries owing to their exceptional strength-to-weight and stiffness-to-weight ratios [1–5]. Despite these advantageous mechanical properties stem from the composite's hierarchical structure, the interface between fibers and the matrix is prone to trigger crack initiation and propagation, which can ultimately lead to catastrophic failure under operational stress [6–10]. Hence, understanding the microscopic structure of CFRPs**,** especially carbon fiber alignment, is crucial for enhancing durability and overall mechanical performance.

X-ray computed tomography (X-ray CT) has emerged as an indispensable, non-destructive tool for visualizing internal microstructures in three dimensions, making it well-suited for the characterization of CFRP's microstructures [11–13]. Through quantitative image analysis, CT data can reveal defect-prone regions and inform the optimization of fabrication parameters [14,15]. However, a trade-off between field-of-view and spatial resolution often limits the detection of fine structural details. Millimeter-scale field-of-views, as required in practical applications, generally restrict the spatial resolution to a submicrometer range. This constraint hampers the accurate detection of individual fiber positions and microcracks, considering that the typical diameter of carbon fiber is 5–7 μm.

To overcome these limitations, various image analysis techniques have been explored. Traditional approaches such as thresholding and edge detection—available in software platforms like ImageJ, Dragonfly, and VGSTUDIO MAX—are computationally efficient but highly dependent on image contrast, which is often insufficient for discriminating between fibers and matrix. More advanced geometric segmentation methods, such as watershed and local thickness, can improve detection but still strongly depend on outline quality and thresholding, making them vulnerable to noise and intensity inhomogeneity [16,17]. Machine learning-based techniques offer alternative solutions but require large, well-annotated datasets and can be sensitive to variations in scanning conditions.

Topological data analysis (TDA) provides a robust, label-free framework for extracting intrinsic geometric and topological features from image data [18–21]. Persistent homology is a mathematical framework that captures multi-scale topological features—such as connected components and rings—by tracking their birth and death across varying thresholds [22–25]. One of the key strengths of persistent homology lies in its robustness to image blurring and noise, enabling the reliable extraction of topological features even from barely resolved or inconsistent imaging data. Applications of persistent homology span diverse domains including biomolecular modeling and materials science [26–31]. Notably, Uy *et al.* employed persistent homology to successfully identify crack propagation for high spatial-resolved X-ray CT images of CFRP [32].

In this paper, we leverage TDA based on persistent homology to develop a robust descriptor for the quantitative characterization of microscopic structures in CFRPs. After introducing the concept of persistent homology necessary for the analysis of grayscale images, we extract particle positions from a scanning electron microscopy (SEM) image as a demonstration of the basic framework of our method. Next, we apply the method to X-ray CT data of CFRP to extract the positions of carbon fibers from a cross-sectional image with improvements for avoiding the beam-hardening artifact. The results are compared with those obtained using the watershed and local thickness algorithms. We also describe an automatic parameter optimization method for X-ray CT images of CFRP, as well as a segmentation method for crack regions.

2. Methods and materials

2.1. Specimen and X-ray CT measurement

Carbon fiber/epoxy prepregs (fiber areal weight: 190 g/m^2, resin content: 35 wt.%, fiber volume fraction: ca. 60%) were used to fabricate CFRP specimens with a layup of [90°/0°/90°/0°/90°]. The reinforcing fibers were polyacrylonitrile (PAN)-based: HYEJ25-36 fibers (Mitsubishi Rayon Co. Ltd.) with a diameter of ca. 5 μm.

The X-ray CT measurements were conducted using an Xradia Versa 520 system (Carl Zeiss Microscopy, Inc.). A cuboidal specimen with dimensions of $0.3 \times 1.0 \times 13.0\ mm^3$ was cut from CFRP. A CCD camera with a scintillator was used as the detector. Transmission images were measured in 0.225° steps for a 360° rotation using a white X-ray source generated by a tungsten target with a tube voltage of 40 keV. The resolution is determined by the configuration of the X-ray source and detector in the Xradia Versa 520 system. In this study, 20× objectives were used with binning 2. Finally, the minimum spatial resolution was as small as 0.7 μm, which was determined by the Siemens star pattern. The specimen was fixed by hanging the bottom side on the sample stage. The field-of-view measured 464 μm × 493 μm × 493 μm, and the voxel size was $0.5 \times 0.5 \times 0.5\ \mu m^3$. Each image had dimensions of 928 × 986 pixels.

2.2. Persistent homology analysis of grayscale images

We developed a methodology for extracting topological features—specifically degree 1 homology corresponding to rings—from cross sections of three-dimensional X-ray CT images of CFRP using persistent homology. These features are directly associated with the identification of fibers and cracks. The level-set approach [33–36] is employed, in which a grayscale threshold is incrementally adjusted to generate a sequence of binary images, allowing topological features to evolve across intensity levels. For an 8-bit grayscale image, decreasing the threshold from 255 to 0 produces white rings; this is termed superlevel-set filtration. Conversely, increasing the threshold from 0 to 255 reveals black rings, which is called sublevel-set filtration.

As an illustrative example, Figure 1a presents a grayscale image with four white spots, (spots 1–4), where spot 4 contains a small dark region. Each spot exhibits a 2D Gaussian distribution, with highest intensity at the center that decreases toward the edges. For sublevel-set filtration, the image initially appears almost entirely white. As the threshold increases, black regions expand and a single white region splits into four parts, each gradually becoming surrounded by a black ring. Each of these events corresponds to the birth of a black ring (Figures 1b–d). These ring structures then disappear at higher thresholds, corresponding to the death of the black rings (Figures 1e–f). The pixels where the rings are finally formed and disappear are termed birth and death positions, respectively, and are illustrated in Figure 1g. The grayscale intensities at birth and death positions are recorded as birth and death values, and a series of these birth–death pairs are plotted in a two-dimensional diagram (Figure 1h), referred to as a degree-1 persistence diagram (PD1). In contrast, for superlevel-set filtration, the intensity threshold decreases. The image initially appears almost entirely black, and a white ring appears within spot 4 (Figure 1i), shrinks (Figure 1j) and disappears (Figure 1k), first. Another white ring emerges when the four white regions become connected and the central black region is isolated (Figure 1l). As the threshold decreases further, this ring also disappears (Figure 1m). The birth and death positions corresponding to these two white rings are illustrated in Figure 1n, and summarized as PD1 in Figure 1o.

2.3. Level-set mask

A pixel mask that separates the bright and dark regions of a given grayscale image is also constructed from PD1 using superlevel-set filtration. It is schematically obtained by filling the inside of each ring, and its implementation was provided by Obayashi [37]. Figure S1 shows an example 2D SEM image of particles. This mask defines the background region and is used in this paper to refine the detection of particle and fiber positions, as well as to identify crack regions in CFRP. No additional parameters are required.

2.4. Distance transform-based analysis of binary images

In binary images, structural features are analyzed through distance transforms, which convert binary data into a distance map by measuring distances from foreground pixels (e.g., fiber, crack or particle). Topological evolution is tracked by incrementally increasing the distance threshold. During fiber and particle detection, the Euclidean distance transform is employed because its isotropic nature makes it well-suited for detecting circular features. In contrast, for crack extraction, the Manhattan distance transform—measuring distances along grid-aligned horizontal and vertical paths—is more appropriate, as it better captures the anisotropic and branched morphology of cracks.

Figure 2a shows a binary input image, where black regions are treated as target features. The transformation into a Euclidean distance map and corresponding topological changes are illustrated in Figures 2b–f (with details in Figure S2). As the threshold increases, distinct regions (termed islands) emerge, merge, or disappear. For instance, islands 1 and 2 appear at a threshold of −2.2 and merge at −2.0, resulting in a birth–death pair of (−2.2, −2.0) for island 2. Similarly, island 3 appears at −1.4 and disappears at 1.0, forming a pair (−1.4, 1.0). These events are visualized in the degree-0 persistence diagram (PD0), which captures information about the connectivity among islands. Focusing on the black regions in a binary image, when the black regions are thinned by distance transform and segmentation occurs, the corresponding values are recorded as negative death values. Conversely, when the black regions are thickened and merging occurs, the corresponding values are recorded as positive death values. Therefore, by counting the birth–death pairs with negative death values, the number of overlapping fibers can be determined. This framework is later utilized to quantify true and excessive fiber and particle detection, in addition to crack expansion.

3. Results and discussion

3.1. Application to a 2D SEM image: Identification of particle positions

We apply our method based on TDA with persistent homology to an SEM image of nanoparticles. The sample image was provided by Prof. Takanari Togashi from Yamagata University. First, the raw image is preprocessed with contrast enhancement and Gaussian blurring to suppress noise. Persistent homology is then applied using sublevel-set filtration to compute PD1, and the death points corresponding to bright regions (i.e., particles) are extracted. Most of these points are located at the particles, providing possible positions, but some incorrect signals still remain. Further refinement can be achieved by adjusting the birth and death criteria and applying a level-set mask. The detailed procedure is described below.

Preconditioning of given grayscale images: For the given image, histogram equalization—which redistributes pixel intensities so that the histogram of the image becomes approximately uniform—is applied to enhance contrast. Gaussian blurring is then applied to reduce noise. While avoiding the loss of possible candidate positions, the kernel size is chosen to be as large as possible. In this case, the optimal kernel size for Gaussian blur was determined to be 23×23. The effect of kernel size on the results is shown in Figure S3.

Detection of particle positions: PD1 is computed using sublevel-set filtration on the preconditioned grayscale image, and the death points of rings—where a bright region is surrounded by a darker region (i.e., particles)—are extracted. While most particles are successfully identified, an unexpectedly large number of non-particle points are also detected (Figure S4).

Reduction of spurious detection: As shown in Figure 3a, PD1 separates into two clusters of birth–death pairs. The color scale indicates the density of birth–death pairs in a 50×50 birth–death grid. Warmer colors represent higher density. Histograms of birth and death values (Figures 3b–c) are used to determine the criteria for birth and death, defined by the minimum between the first two main peaks. In this case, by applying the conditions of birth value (brightness) ≤ 73 and death value (brightness) ≤ 87, most of the unexpected points are removed. The results depending on the birth and death criteria are also summarized in Figure S4. Further refinement can be achieved using the level-set mask introduced in Section 2.3.

Comparison with existing methods: As a comparison with existing methods, particle position detection was also performed using the watershed algorithm and the local thickness algorithm, respectively. The three results, including that of our approach, were quantitatively evaluated using metrics such as precision, recall, and F1-score, where precision measures the ratio of correctly identified particles among all detected particles, recall measures the ratio of correctly identified particles among all true particles, and the F1-score represents the harmonic mean of precision and recall.

The watershed algorithm consists of several preconditioning steps, including Gaussian blurring, Otsu binarization, and hole filling. After these processes, the Euclidean distance transform is computed, and the local maxima of the distance are identified as seeds. To reduce spurious minima and prevent over-segmentation, h-minima suppression is applied, where shallow minima with depths less than a threshold h are removed. The seeds are expanded by watershed flooding, and region filtering is subsequently performed to remove objects that are unrealistically small or large. The final particle positions are defined as the local maxima of the distance transform within each segmented region.

For local thickness segmentation, the same preconditioning steps as in the watershed algorithm are applied. The Euclidean distance transform is then computed, and h-maxima suppression is used to eliminate shallow peaks with heights smaller than a threshold h, thereby preventing over-segmentation. For each remaining peak, the pixel with the maximum distance value is selected as particle position, and its diameter is estimated as twice the distance to the nearest boundary. Unrealistically small or large detections are filtered out using predefined diameter thresholds. For both the watershed and local thickness algorithms, a grid search is employed to optimize the parameters, and the optimized values are summarized in Table S1.

As shown in Figure 4, our method achieves the highest accuracy with a precision of 0.989, recall of 0.978 and F1-score of 0.984. In contrast, watershed and local thickness algorithms often miss particles because they rely on edge quality, which is degraded under noise and intensity inhomogeneity, leading to misclassification by h-minima or h-maxima. Our method instead tracks the evolution of topological ring structures across sublevel-sets, first detecting all possible particle positions by level-set filtration and then removing spurious ones using birth and death criteria as well as a level-set mask. Thus, the ease of controlling accuracy is also an advantage of our method.

Limit of applicability: Our method detects particle positions as comprehensively as possible using persistent homology with level-set filtration. However, some are missed when particles become fused after Gaussian blurring, or when they disappear due to poor contrast.

3.2. Application to sliced X-ray CT images: Identification of carbon fiber positions

Application to X-ray CT images of CFRP is achieved by extending the above framework. Figure 5 shows a typical image contrast-enhanced slice image perpendicular to the fiber bundles. Cross sections of carbon fibers, the resin matrix, and cracks can be observed, and the overall view is provided in Figure S5. The images of CFRP had the void artifacts which came from beam-hardening. This problem was solved by improvements to the above framework.

The procedure is summarized in Figure 6. First, the raw image is preconditioned using histogram equalization and Gaussian blurring (Figures 6a–c), in the same manner as described above. Next, the "voids" in some carbon fibers are detected and filled (Figure 6d). After a second round of Gaussian blurring (Figure 6e), possible fiber positions are extracted (Figure 6f). Finally, spurious points in resin matrix are removed (Figure 6g). This workflow is also shown as a block diagram in Figure 7.

Preconditioning of given grayscale images: For the given image, histogram equalization is applied to enhance contrast in the same manner as described in Section 3.1. While avoiding the loss of possible candidate positions, the kernel size is chosen to be as large as possible. The optimal kernel size for Gaussian blur was determined to be 5×5. The effect of kernel size is illustrated in Figure S6.

Detection of fibers with void: Due to the trade-off between field-of-view and spatial resolution, beam-hardening artifacts cause fiber centers to appear with lower intensities than their edges, producing apparent voids. Accordingly, fibers in X-ray CT images are typically classified as either solid or voided. Before detecting fiber positions, these apparent voids need to be filled in order to apply the same approach as described above. PD1 is computed using superlevel-set filtration on the preconditioned grayscale image to capture the bright rings surrounding dark regions. The possible void positions are obtained as death points. While most voids are successfully identified, numerous spurious points are also detected in resin matrix.

Reduction of spurious void detection: To remove the death points in the resin matrix, a death criterion is introduced. As shown in Figure 8, PD1 separates into two clusters of birth–death pairs, upper and lower parts. The color scale indicates the density of birth–death pairs in a 100×100 birth–death grid. In this case, the optimal threshold for the death value was determined to be 95. Birth–death pairs with death values greater than 95 are identified as voids within the fibers, while the others are removed as noise. Detection results vary with Gaussian blurring and death criterion settings, as summarized in Figure S6.

Fill voids and apply Gaussian blurring: Detected voids are filled by placing white disks at each death position in the preconditioned image without Gaussian blurring. The radius of each disk was set to 1 pixel (thus, in this case, it appears as a cross rather than a disk). Although the disk size must be manually assigned, it has little effect on the results.

Subsequently, another Gaussian blurring is applied. While avoiding the loss of possible candidate positions, the kernel size is chosen to be as large as possible. The optimal kernel size was determined to be 13×13.

Detection of fiber positions: PD1 is recomputed using sublevel-set filtration, and the death points are obtained as possible fiber positions in the same manner as in the case of particle position detection. While most fibers are correctly identified, some spurious points are also detected in the resin matrix.

Reduction of spurious fiber detection: As shown in Figure 9a, PD1 has a tail extending toward the lower birth and death region, the color scale indicates the density of birth–death pairs in a 100×100 birth–death grid. Histograms of birth and death values (Figures 9b–c) are used to determine the criteria for birth and death. In this case, by applying the conditions of birth value (brightness) ≤ 111 and death value (brightness) ≤ 156, most of the unexpected points are removed. Further refinement can be achieved using the level-set mask (Figure S7) introduced in Section 2.3. If necessary, an additional condition on the coordination number (CN) can also be applied. The condition of $\mathrm{CN} < 5$ (cutoff radius: 8.3 μm) slightly improves the removal of spurious points (Figure S8). Detection results vary with Gaussian blurring as well as the birth and death criteria, as summarized in Figures S8 and S9.

Comparison with existing methods: As a comparison with existing methods, particle position detection was also performed using the watershed algorithm and the local thickness algorithm, respectively. The three results, including that of our approach, were quantitatively evaluated using metrics such as precision, recall, and F1-score. Cropped regions are shown in Figure 10, with full-image results in Figure S10. For the watershed and local thickness methods, the same detection procedures as those used for SEM are applied to a single CT image, and the optimized parameters obtained via grid search are summarized in Table S1.

As shown in Figure 10, our method effectively detects fiber positions across a wide range of contrast, achieving a precision, a recall, and an F1-score of (0.993, 0.990, 0.992). In contrast, the watershed algorithm and the local thickness algorithm achieve (0.965, 0.957, 0.961) and (0.953, 0.962, 0.957), respectively, because both miss more true fibers and count spurious fibers in the resin matrix. It is likely due to their reliance on binary image–based distance transforms, since once a threshold is applied, there is no way to distinguish fibers (foreground) from the resin matrix (background). The watershed algorithm with a grayscale image–based distance transform was also tested, but the performance was worse, yielding (0.880, 0.863, 0.871), as shown in Figure S11. Our method, which first extracts all candidate fiber positions by combining persistent homology with level-set filtration and then removes spurious detections afterward, proved more effective.

Limit of applicability: Our method detects voids and fiber positions as comprehensively as possible using persistent homology with level-set filtration. However, when voids or fibers are fused due to adjacent bright-edge merging, caused by too severe beam-hardening artifacts, some voids or fiber positions are missed. Moreover, the removal process of spurious voids or fiber positions may unexpectedly eliminate true ones when the corresponding birth–death pairs cannot be clearly separated in PD1.

3.3. Automatic optimization of parameters

All parameters—namely the parameters in (1) histogram equalization and Gaussian blurring, (2) death criterion for detecting voids in the fibers, (3) a second round of Gaussian blurring, and (4) birth and death criteria for detecting fiber positions—can be automatically optimized from a "sufficient" number of input images in the case of X-ray CT.

The given 3D X-ray CT dataset consists of 964 sequential slices along the specimen depth. Sample images were selected every 20 slices, resulting in 49 images used for parameter optimization. The dependence of the results on the number of sample images is summarized in Figure S12.

Classification of voids in fibers and resin region: The kernel size of Gaussian blur strongly affects void detection: kernels that are too large miss voids, while kernels that are too small cause excessive and spurious detections. In this study, kernel size from 3×3 to 7×7 were tested, considering the mean fiber radius of 2.5 μm (= 5 pixels). For each kernel size, an accumulated PD1—obtained by summing the PD1 from 49 images—is computed using superlevel-set filtration. A histogram of the death values (e.g., Figure S13) is then used to determine the death criterion. The total filling number is defined as the number of voids identified in the upper PD1 segment (death value > death criterion). The procedure for counting excessive fillings is illustrated in Figure 11. Figure 11a presents an X-ray CT image, and Figure 11b displays the total filling positions in an image binarized at the death-criterion threshold, marked as disks with a radius of 2.5 μm (= 5 pixels). The corresponding PD0 is shown in Figure 11c, where the color scale indicates the density of birth–death pairs in a 20×20 birth–death grid. As described in section 2.4, a birth–death pair with a negative death value in PD0 indicates an overlap of two disks, meaning that two "voids" are located within the same fiber. This is not a reasonable result. Thus, the number of excessive detections is obtained by counting birth–death pairs with negative death values, while the number of true voids is obtained by counting those with positive death values. The optimal kernel size is determined as the one that yields the maximum true voids. In this study, a 5×5 kernel size with a death criterion of 95 is found to be optimal. Results for the various kernel sizes are provided in Figures S13 and S14.

Reduction of spurious detection: The kernel size of the Gaussian blur also affects fiber position detection: kernels that are too large miss fiber positions, while kernels that are too small cause excessive and spurious detections within a single fiber. Kernel sizes from 11×11 to 15×15 were tested, considering the mean fiber diameter of 5 µm (= 10 pixels). For each kernel size, an accumulated PD1—obtained by summing the PD1 from 49 images—was computed using sublevel-set filtration. Figure 11d shows all detected points visualized as disks with a radius of 2.5 µm (= 5 pixels), representing "fibers," although some of these detections are not actual fibers, and many disks overlap. Such excessive detections can be identified using PD0 (Figure 11e) of this binarized image by focusing only on birth–death pairs with negative death values, as described above. The optimal kernel size was determined as 13×13, which yields the maximum number of birth–death pairs with positive death values and suppresses excessive detections. Results for the various kernel sizes are provided in Figure S15.

The birth and death criteria, used to remove spurious detections in the resin matrix are determined as follows. If the accumulated PD1 separates into distinct clusters of birth–death pairs (e.g., Figure 3), the criteria are set to the minimum between the main peaks of histograms. If no clear separation is observed (e.g., Figure 9 and Figure S16), histograms of birth and death values and their second derivative curves reveal two transition zones (Figure S16): the left sides correspond to spurious detections, while the right sides represent true fiber positions. The bounds of these zones are defined by the minima and maxima of the second derivative on the left side of each histogram peak. The lower bound of transition zone in birth-value histogram defines the birth criterion, and the upper bound of transition zone in death-value histogram defines the death criterion. Spurious points can be removed using the conditions death value ≤ death and birth values ≤ birth criterion.

A dedicated program was developed to perform these processes step by step. Parameter optimization with the training set (49 images, each 928 × 986 pixels) took approximately 1 hour, while analysis of the entire X-ray CT dataset (964 images, each 928 × 986 pixels) took about 3 hours. These computation times are based on the use of a single

CPU core on a server equipped with dual Intel® Xeon® Gold 6338 processors (64 total cores) and 1.0 TiB of RAM.

3.4. Identification and detection of crack regions

The proposed technique can be widely applied to X-ray CT images obtained by *in situ* observation under strain and stress, as it can distinguish carbon fiber positions from cracks. Figure 12 shows an X-ray CT image with visible cracks. Grayscale profiles along line segments intersecting both crack and resin regions reveal intensity overlap (Figures 12a–b), indicating that intensity-based thresholding alone is insufficient for crack segmentation. To address this issue, the superlevel-set mask introduced in Section 2.3 is applied. Preprocessing begins with the raw image (Figure 12a), followed by histogram equalization (Figure 12b). After applying the superlevel-set mask (Figure 12c), the extracted regions are refined using intensity and area criteria (Figures 12d–e), and by applying the Manhattan distance transform (Figure 12f).

Preconditioning of given grayscale images: The same contrast enhancement used for fiber detection is applied here, without subsequent Gaussian blurring.

Identification of crack regions based on PD1 with superlevel-set filtration: The level-set mask of PD1 is computed at a death value of 0, and the corresponding deep dark regions are extracted. These represent cracks and other enclosed resin regions (Figure 12c).

Refinement using grayscale and area criteria: A grayscale histogram of this region (Figure S17a) reveals a high frequency of low-intensity values. The first local minimum (value = 5) is used as a shrinking criterion to refine the detected region (Figure 12d). The refined region captures most cracks but also includes noise. An area-size histogram (Figure S17b) indicates that small regions dominate, and its first derivative stabilizes above 11 pixels. Accordingly, a size threshold of 11 pixels is applied to eliminate noise (Figure 12e).

Final crack expansion: To connect disconnected segments and voids within crack regions, Manhattan step expansion is applied. This method iteratively expands cracks based on a Manhattan distance transform of the binary image (Figure S18). Only expansion steps within the PD1-defined dark region (Figure 12c) are considered. At each step, the average intensity is computed, and the expansion stops when this value reaches its maximum (Figure S17c), indicating proximity to the carbon fiber boundaries (Figure 12f).

4. Conclusion

We developed a method to extract the positions of carbon fibers from slice images of X-ray CT of CFRP based on persistent homology. The method is versatile, as we also demonstrated its ability to extract particle positions from an SEM image.

In persistent homology with level-set filtration, PD1 records birth–death pairs corresponding to rings where bright regions enclose dark regions (superlevel-set filtration) or dark regions enclose bright regions (sublevel-set filtration). By applying this to images processed with histogram equalization followed by appropriate Gaussian blurring, candidate positions for fibers or particles can be extracted with minimal omissions, based on brightness differences in the grayscale image. Subsequently, thresholds for birth and death are determined from the clustering of birth–death pairs in PD1, enabling the removal of falsely detected fibers or particles. Detection accuracy can be further refined using a level-set mask, which provides a pixel representation of non-fiber or non-particle regions (background).

The method we constructed in this way produced good results (precision, recall, and F1-score) in both particle position detection from an SEM image and fiber position detection from X-ray CT images, outperforming conventional approaches such as the watershed and local thickness algorithms. From these comparisons, it can be concluded that our method is more robust to contrast inhomogeneity than conventional methods and offers the advantage of easily adjusting accuracy according to the purpose.

When a large number of images under similar conditions, such as X-ray CT images, are available, the Gaussian blur kernel size and the birth and death criteria can be automatically optimized. Furthermore, we demonstrated that the level-set mask can also be applied to the extraction of crack regions from CFRP X-ray CT images.

This method is highly beneficial for handling large datasets. In recent years, the size of X-CT data has increased to the terabyte scale when *in situ* or dynamic observations are performed. In such cases, it is challenging to manually determine carbon fiber alignment with respect to crack propagation. our method can be applied to big data and enables automatic analysis of carbon fiber alignment.

Our method was implemented using the open-source software HomCloud v4 [38], and outlines of the scripts using the Python API are provided in the Supplementary Information.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Acknowledgements

We would like to express our gratitude to Prof. Takanari Togashi of Yamagata University for his kind provision of the SEM image of particles.

Funding

Experiments using synchrotron radiation were performed with the approval of the Photon Factory Program Advisory Committee under Proposal Nos. 2019S2-002 and 2022S2-001.

This work was partially supported by; a) JST-Mirai Program Grant Number JPMJMI22C1, and b) JSPS KAKENHI Grant Numbers JP19H00834, JP20H02028, and JP22H05109. (JST: Japan Science and Technology Agency, JSPS: Japan Society for the Promotion of Science).

Data availability statement

Data and code will be made available on request

References

[1] Zhang J, Lin G, Vaidya U, et al. Past, present and future prospective of global carbon fibre composite developments and applications. Composites Part B. 2023;250:110463. doi: 10.1016/j.compositesb.2022.110463

[2] Huda Z, Edi P. Materials selection in design of structures and engines of supersonic aircrafts: a review. Mater Des. 2013;46:552-560. doi: 10.1016/j.matdes.2012.10.001

[3] Melo EGd, Silva JCdS, Klein TB, et al. Evaluation of carbon fiber reinforced polymer – CFRP – machining by applying industrial robots. Sci Eng Compos Mater. 2021;28(1):285-298. doi: 10.1515/secm-2021-0026

[4] Lee JM, Min BJ, Park JH, et al. Design of lightweight CFRP automotive part as an alternative for steel part by thickness and lay-up optimization. Materials. 2019;12(14):2309. doi: 10.3390/ma12142309

[5] Vijayan DS, Sivasuriyan A, Devarajan P, et al. Carbon fibre-reinforced polymer (CFRP) composites in civil engineering application—a comprehensive review. Buildings. 2023;13(6):1509. doi: 10.3390/buildings13061509

[6] Friedrich K. Application of fracture mechanics to composite materials. Elsevier; 2012.

[7] Gutkin R, Pinho ST, Robinson P, et al. On the transition from shear-driven fibre compressive failure to fibre kinking in notched CFRP laminates under longitudinal compression. Compos Sci Technol. 2010;70(8):1223-1231. doi: 10.1016/j.compscitech.2010.03.010

[8] Kimura M, Watanabe T, Oshima S, et al. Nanoscale in situ observation of damage formation in carbon fiber/epoxy composites under mixed-mode loading using synchrotron radiation X-ray computed tomography. Compos Sci Technol. 2022;230:109332. doi: 10.1016/j.compscitech.2022.109332

[9] Watanabe T, Takeichi Y, Niwa Y, et al. Nanoscale in situ observations of crack initiation and propagation in carbon fiber/epoxy composites using synchrotron radiation X-ray computed tomography. Compos Sci Technol. 2020;197:108244. doi: 10.1016/j.compscitech.2020.108244

[10] Garcea SC, Sinclair I, Spearing SM. Fibre failure assessment in carbon fibre reinforced polymers under fatigue loading by synchrotron X-ray computed tomography. Compos Sci Technol. 2016;133:157-164. doi: 10.1016/j.compscitech.2016.07.030

[11] Garcea SC, Wang Y, Withers PJ. X-ray computed tomography of polymer composites. Compos Sci Technol. 2018;156:305-319. doi: 10.1016/j.compscitech.2017.10.023

[12] Dilonardo E, Nacucchi M, De Pascalis F, et al. High resolution X-ray computed tomography: a versatile non-destructive tool to characterize CFRP-based aircraft composite elements. Compos Sci Technol. 2020;192:108093. doi: 10.1016/j.compscitech.2020.108093

[13] Dilonardo E, Nacucchi M, De Pascalis F, et al. Inspection of carbon fibre reinforced polymers: 3D identification and quantification of components by X-ray CT. Appl Compos Mater. 2022;29(2):497-513. doi: 10.1007/s10443-021-09976-x

[14] Liu X, Chen F. Defects characterization in CFRP using X-ray computed tomography. Polym Polym Compos. 2016;24(2):149-154. doi: 10.1177/096739111602400210

[15] Petrò S, Reina C, Moroni G. X-ray CT-based defect evaluation of continuous CFRP additive manufacturing. J Nondestruc Eval. 2021;40(1):7. doi: 10.1007/s10921-020-00737-7

[16] Meyer F. Topographic distance and watershed lines. Signal Process. 1994;38(1):113-125. doi: 10.1016/0165-1684(94)90060-4

[17] Doube M, Kłosowski MM, Arganda-Carreras I, et al. Bonej: Free and extensible bone image analysis in ImageJ. Bone. 2010;47(6):1076-1079. doi: 10.1016/j.bone.2010.08.023

[18] Wasserman L. Topological data analysis. Annu Rev Stat Appl. 2018;5(1):501-532. doi: 10.1146/annurev-statistics-031017-100045

[19] Kunii S, Lira FA, Chiharu M, et al. Super-hierarchical and explanatory analysis of magnetization reversal process using topological data analysis. Sci Technol Adv Mater Methods. 2022;2(1):445-459. doi: 10.1080/27660400.2022.2149037

[20] Singh Y, Farrelly CM, Hathaway QA, et al. Topological data analysis in medical imaging: current state of the art. Insights Imaging. 2023;14(1):58. doi: 10.1186/s13244-023-01413-w

[21] Tone M, Shunsuke S, Sotaro K, et al. Linking structure and process in dendritic growth using persistent homology with energy analysis. Sci Technol Adv Mater Methods. 2025;5(1):2475735. doi: 10.1080/27660400.2025.2475735

[22] Otter N, Porter MA, Tillmann U, et al. A roadmap for the computation of persistent homology. EPJ Data Sci. 2017;6:1-38. doi: 10.1140/epjds/s13688-017-0109-5

[23] Shiraiwa T, Shoya K, Fabien B, et al. Exploration of outliers in strength–ductility relationship of dual-phase steels. Sci Technol Adv Mater Methods. 2022;2(1):175-197. doi: 10.1080/27660400.2022.2080483

[24] Buchet M, Hiraoka Y, Obayashi I. Persistent homology and materials informatics. Nanoinformatics. 2018:75-95. doi: 10.1007/978-981-10-7617-6

[25] Obayashi I, Hiraoka Y, Kimura M. Persistence diagrams with linear machine learning models. J Appl Comput Topol. 2018;1(3):421-449. doi: 10.1007/s41468-018-0013-5

[26] Obayashi I, Nakamura T, Hiraoka Y. Persistent homology analysis for materials research and persistent homology software: Homcloud. J Phys Soc Jpn. 2022;91(9):091013. doi: 10.7566/JPSJ.91.091013

[27] Hiraoka Y, Nakamura T, Hirata A, et al. Hierarchical structures of amorphous solids characterized by persistent homology. Proc Natl Acad Sci USA. 2016;113(26):7035-7040. doi: 10.1073/pnas.1520877113

[28] Saadatfar M, Takeuchi H, Robins V, et al. Pore configuration landscape of granular crystallization. Nature Commun. 2017;8(1):15082. doi: 10.1038/ncomms15082

[29] Xia K, Wei G-W. Persistent homology analysis of protein structure, flexibility, and folding. Int J Numer Methods Biomed Engrg. 2014;30(8):814-844. doi: 10.1002/cnm.2655

[30] Hirata A, Wada T, Obayashi I, et al. Structural changes during glass formation extracted by computational homology with machine learning. Commun Mater. 2020;1(1):98. doi: 10.1038/s43246-020-00100-3

[31] Kimura M, Obayashi I, Takeichi Y, et al. Non-empirical identification of trigger sites in heterogeneous processes using persistent homology. Sci Rep. 2018;8(1):3553. doi: 10.1038/s41598-018-21867-z

[32] Uy M, Kido D, Kimura M. A topological approach to understanding crack initiation and propagation in carbon fiber reinforced polymer composites under an opening load. Sci Rep. 2025;15(1):19441. doi: 10.1038/s41598-025-98821-3

[33] Sethian JA, Wiegmann A. Structural boundary design via level set and immersed interface methods. J Comput Phys. 2000;163(2):489-528. doi: 10.1006/jcph.2000.6581

[34] Allaire G, Jouve F, Toader A-M. Structural optimization using sensitivity analysis and a level-set method. J Comput Phys. 2004;194(1):363-393. doi: 10.1016/j.jcp.2003.09.032

[35] Yamasaki S, Nomura T, Kawamoto A, et al. A level set-based topology optimization method targeting metallic waveguide design problems. Int J Numer Methods Eng. 2011;87(9):844-868. doi: 10.1002/nme.3135

[36] Guirguis D, Hamza K, Aly M, et al. Multi-objective topology optimization of multi-component continuum structures via a kriging-interpolated level set approach. Struct Multidiscip Optim. 2015;51(3):733-748. doi: 10.1007/s00158-014-1154-3

[37] Obayashi I. Volume-optimal cycle: Tightest representative cycle of a generator in persistent homology. SIAM Journal on Applied Algebra and Geometry. 2018;2(4):508-534. doi: 10.1137/17m1159439

[38] Obayashi I, Nakamura T, Hiraoka Y. Persistent homology analysis for materials research and persistent homology software: Homcloud. J Phys Soc Jpn. 2022;91(9):091013. doi: 10.7566/JPSJ.91.091013

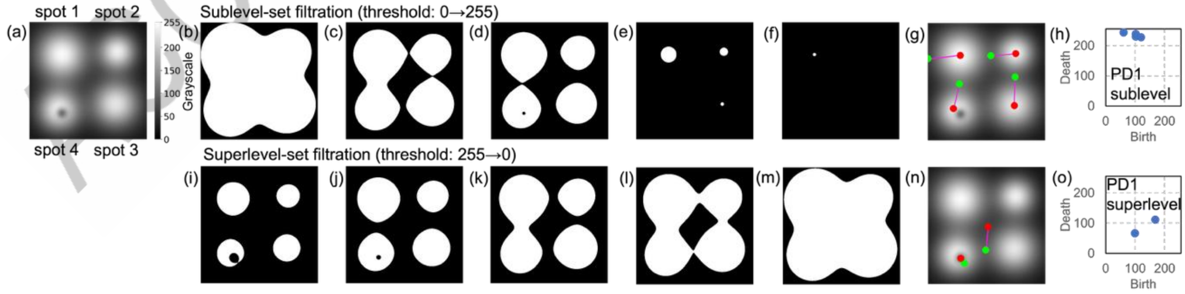
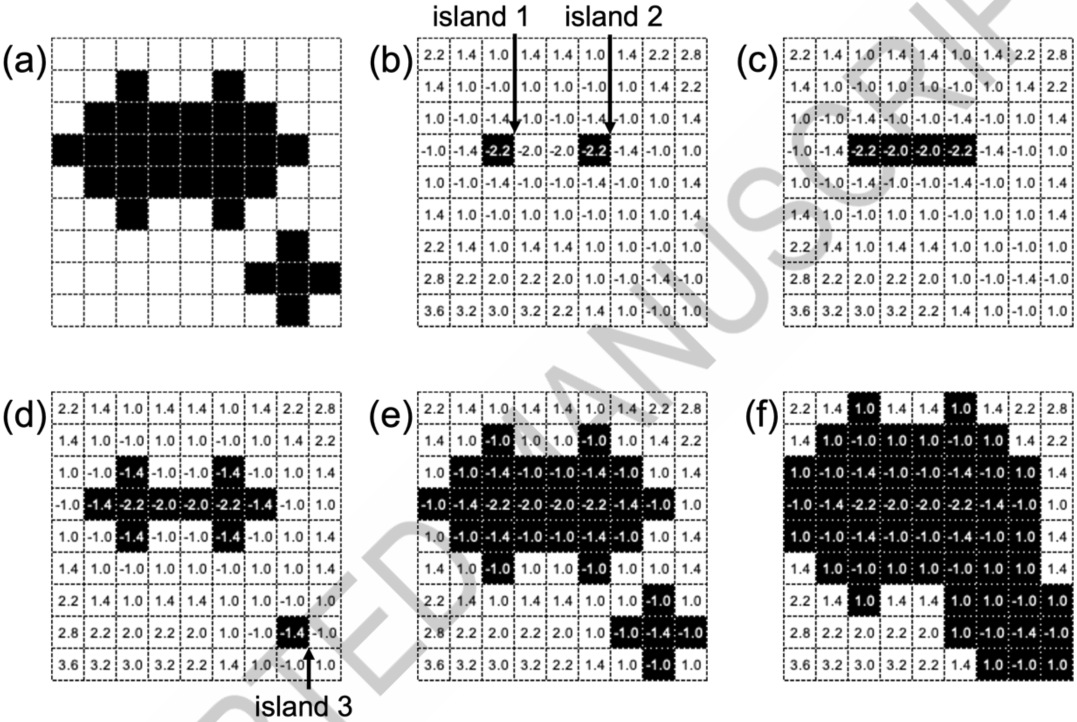


Figure 1. Characterization of rings in a grayscale image using PD1 based on sublevel- and superlevel-set filtrations. The unit of birth and death values is the intensity (brightness) in an 8-bit grayscale image. (a) Input image with four white spots (spots 1–4). (b) Birth of spot 1.

(c) Birth of spots 2 and 3. (d) Birth of spot 4. (e, f) Death of spots 4, 2 and 3. (g) Four pairs of birth (green) and death (red) positions. (h) PD1 from sublevel-set filtration. (i–k) Birth, shrinkage and death of a void in spot 4. (l, m) Birth and death of a hole enclosed by the four spots. (n) Two pairs of birth (green) and death (red) positions. (o) PD1 from superlevel-set filtration.



2.2	1.4	1.0	1.4	1.4	1.0	1.4	2.2	2.8
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.4	2.2
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
-1.0	-1.4	-2.2	-2.0	-2.0	-2.2	-1.4	-1.0	1.0
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.0	1.4
2.2	1.4	1.0	1.4	1.4	1.0	1.0	-1.0	1.0
2.8	2.2	2.0	2.2	2.0	1.0	-1.0	-1.4	-1.0
3.6	3.2	3.0	3.2	2.2	1.4	1.0	-1.0	1.0

2.2	1.4	1.0	1.4	1.4	1.0	1.4	2.2	2.8
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.4	2.2
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
-1.0	-1.4	-2.2	-2.0	-2.0	-2.2	-1.4	-1.0	1.0
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.0	1.4
2.2	1.4	1.0	1.4	1.4	1.0	1.0	-1.0	1.0
2.8	2.2	2.0	2.2	2.0	1.0	-1.0	-1.4	-1.0
3.6	3.2	3.0	3.2	2.2	1.4	1.0	-1.0	1.0

2.2	1.4	1.0	1.4	1.4	1.0	1.4	2.2	2.8
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.4	2.2
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
-1.0	-1.4	-2.2	-2.0	-2.0	-2.2	-1.4	-1.0	1.0
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.0	1.4
2.2	1.4	1.0	1.4	1.4	1.0	1.0	-1.0	1.0
2.8	2.2	2.0	2.2	2.0	1.0	-1.0	-1.4	-1.0
3.6	3.2	3.0	3.2	2.2	1.4	1.0	-1.0	1.0

2.2	1.4	1.0	1.4	1.4	1.0	1.4	2.2	2.8
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.4	2.2
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
-1.0	-1.4	-2.2	-2.0	-2.0	-2.2	-1.4	-1.0	1.0
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.0	1.4
2.2	1.4	1.0	1.4	1.4	1.0	1.0	-1.0	1.0
2.8	2.2	2.0	2.2	2.0	1.0	-1.0	-1.4	-1.0
3.6	3.2	3.0	3.2	2.2	1.4	1.0	-1.0	1.0

2.2	1.4	1.0	1.4	1.4	1.0	1.4	2.2	2.8
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.4	2.2
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
-1.0	-1.4	-2.2	-2.0	-2.0	-2.2	-1.4	-1.0	1.0
1.0	-1.0	-1.4	-1.0	-1.0	-1.4	-1.0	1.0	1.4
1.4	1.0	-1.0	1.0	1.0	-1.0	1.0	1.0	1.4
2.2	1.4	1.0	1.4	1.4	1.0	1.0	-1.0	1.0
2.8	2.2	2.0	2.2	2.0	1.0	-1.0	-1.4	-1.0
3.6	3.2	3.0	3.2	2.2	1.4	1.0	-1.0	1.0

Figure 2. Characterization of connectivity in a binary image using PD0 derived from distance transforms. The values shown in the cells represent the Euclidean distance to the center of the nearest black cell. The negative sign indicates pixels inside the black region. The birth value of PD0 is the distance recorded when black regions are thinned until each region disappears. The death value of PD0 is the distance at which the black regions split during thinning or merge during thickening. (a) Input binary image. (b) Birth of islands 1 and 2. (c) Death of island 2 with a negative death value. (d) Birth of island 3. (e) Growth of islands. (f) Death of island 3 with a positive death value.

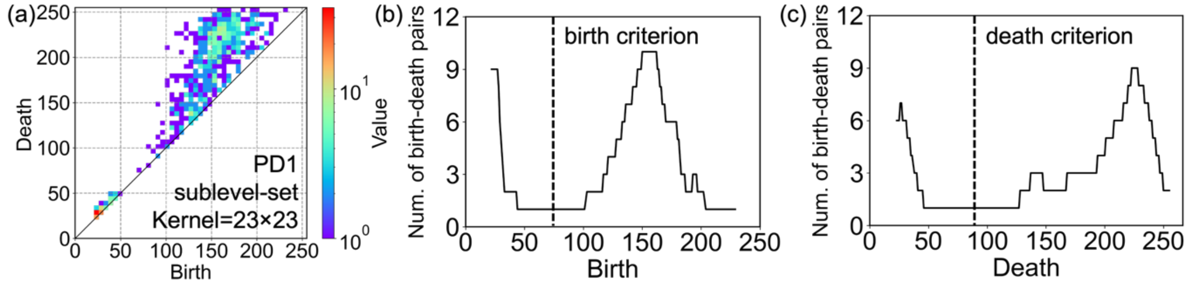


Figure 3. Determination of the birth and death criteria to eliminate spurious particle detections in an SEM image. (a) PD1 using sublevel-set filtration on a single preconditioned image (23×23 Gaussian blurring). The unit of birth and death values is the intensity (brightness). The color scale shows the density of birth–death pairs in a 50×50 birth–death grid. (b) Histogram of birth values with the corresponding birth criterion. (c) Histogram of death values with the corresponding death criterion.

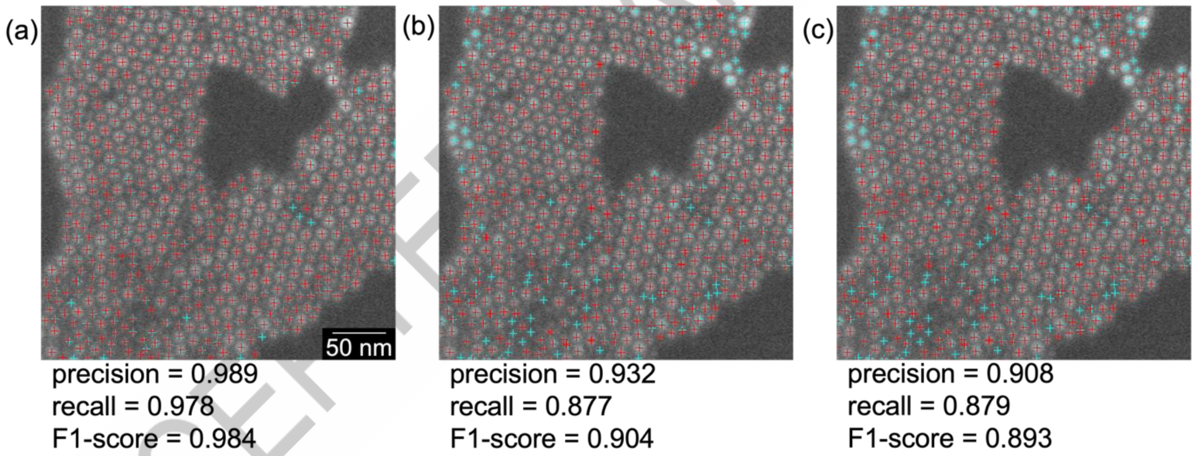


Figure 4. Particle positions identified in an SEM image of nanoparticles. Red crosses: detected positions, cyan crosses: manual annotations. Positions are detected by (a) our method, (b) the watershed algorithm, and (c) the local thickness algorithm. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

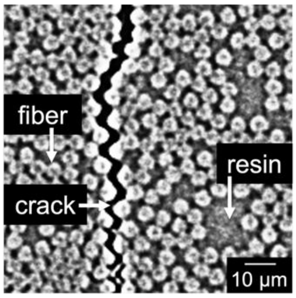


Figure 5. Illustration of an X-ray CT image of CFRP with a resolution of 0.7 μm, where 1 pixel corresponds to 0.5 μm.

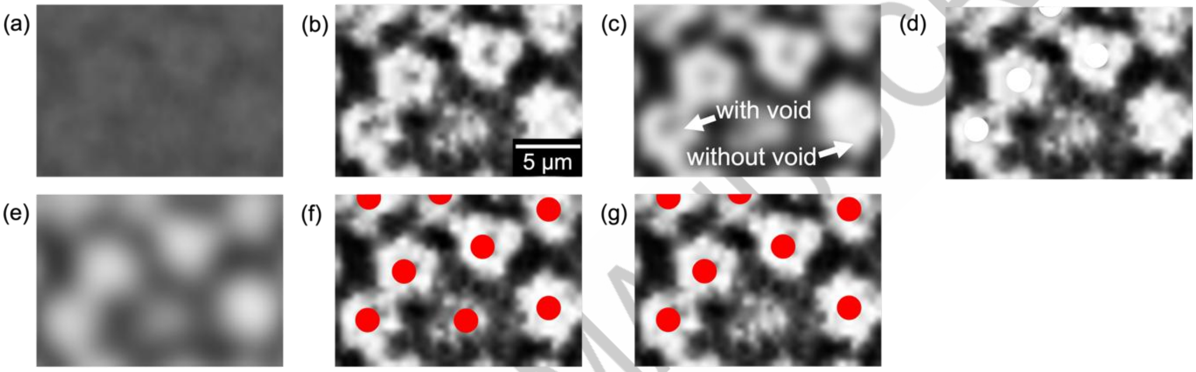


Figure 6. Detection process of fiber positions. (a) Raw data obtained by X-ray CT with a 0.7 μm resolution (1 pixel = 0.5 μm). (b) Contrast enhancement via histogram equalization. (c) Gaussian blurring (kernel size: 5×5). (d) Voids filling based on superlevel-set filtration (death > 95). (e) Second-round Gaussian blurring (kernel size: 13×13). (f) Fiber position detection based on sublevel-set filtration. (g) Removal of spurious detections within resin based on criteria: (death ≤ 156 and birth ≤ 111) or (death ≤ 156 and coordination number < 5).

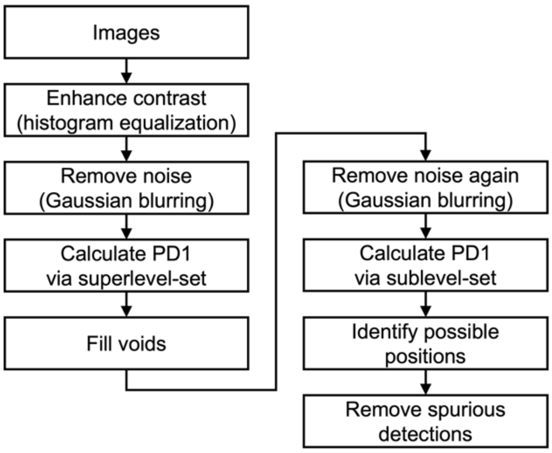


Figure 7. Workflow for detecting fiber positions in an X-ray CT image of CFRP.

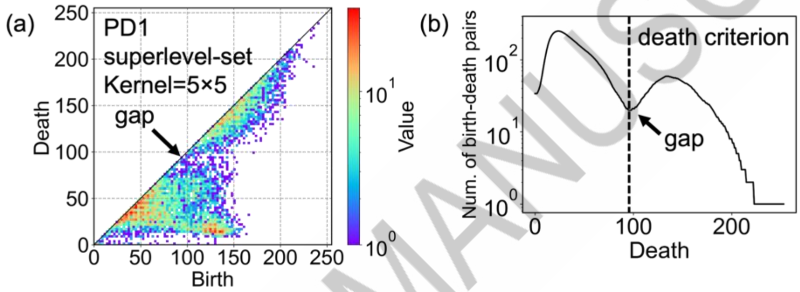


Figure 8. Determination of the death criterion to identify voids within fibers from an X-ray CT image. (a) PD1 using superlevel-set filtration on a single preconditioned image (5×5 Gaussian blurring). The color scale shows the density of birth–death pairs in a 100×100 birth–death grid. (b) Histogram of death values with the corresponding death criterion.

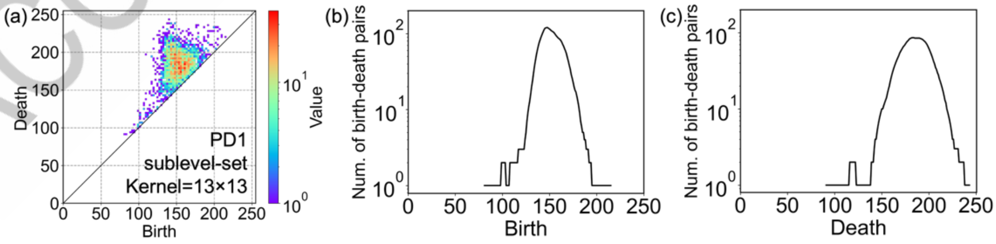


Figure 9. Determination of the birth and death criteria to eliminate spurious positions within resin regions of an X-ray CT image. (a) PD1 using sublevel-set filtration on a single preconditioned image (13×13 Gaussian blurring). The color scale shows the density of birth–

death pairs in a 100×100 birth–death grid. (b) Histogram of birth values, with the birth criterion determined as the midpoint between two peaks. (c) Histogram of death values, with the death criterion determined as the midpoint between two peaks.

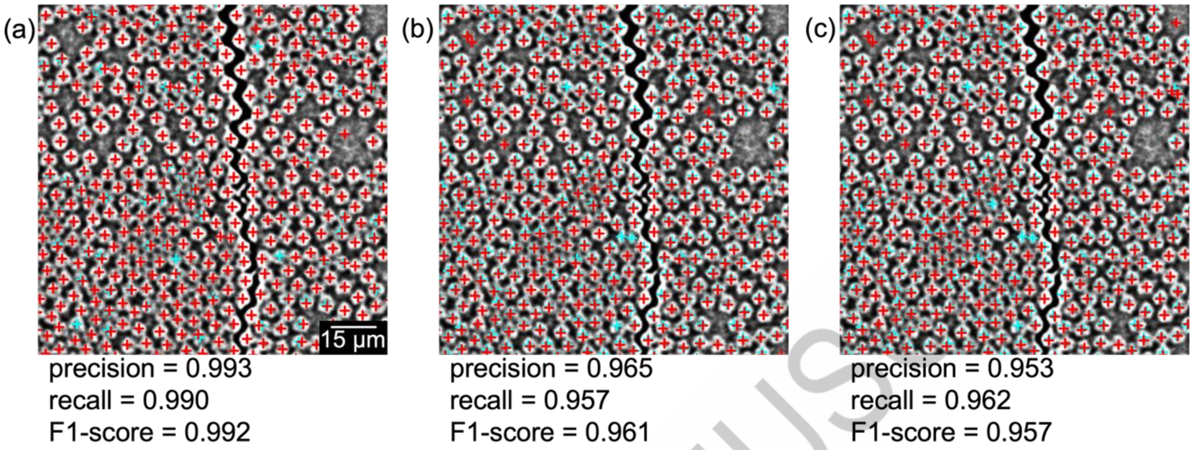


Figure 10. Fiber positions identified in an X-ray CT image of CFRP. Red crosses: detected positions, cyan crosses: manual annotations. Positions are detected by (a) our method, (b) the watershed algorithm, and (c) the local thickness algorithm. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

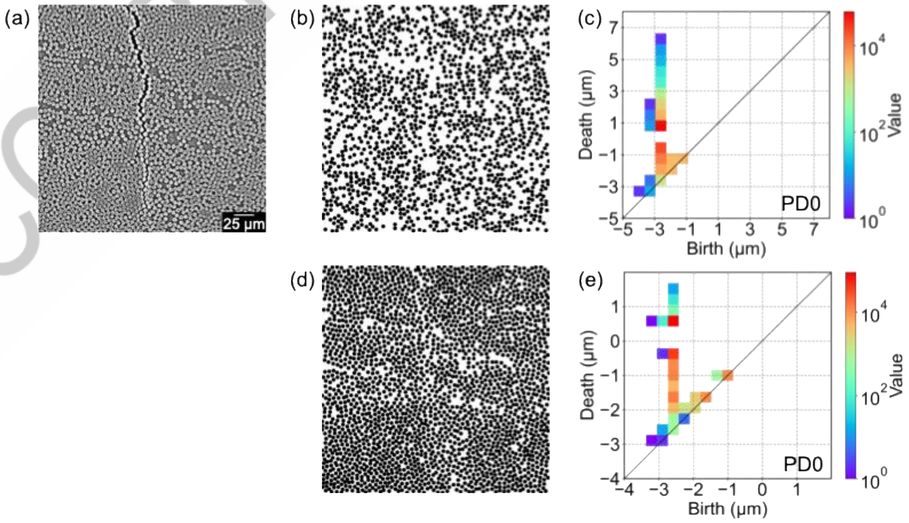


Figure 11. Procedure for evaluating true void and fiber detections. (a) X-ray CT image. (b) Void positions detected by PD1 using superlevel-set filtration on preconditioned images, with

each position shown as a black disk of radius 2.5 μm. (c) Corresponding PD0 from the detected void positions across 49 images. (d) Fiber positions detected by PD1 using sublevel-set filtration on preconditioned images, with each position shown as a black disk of radius 2.5 μm. (e) Corresponding PD0 from the detected fiber positions across 49 images. The PD0 color scale shows the density of birth–death pairs in a 20×20 birth–death grid. Negative death values denote excessive detections, while positive death values denote true detections.

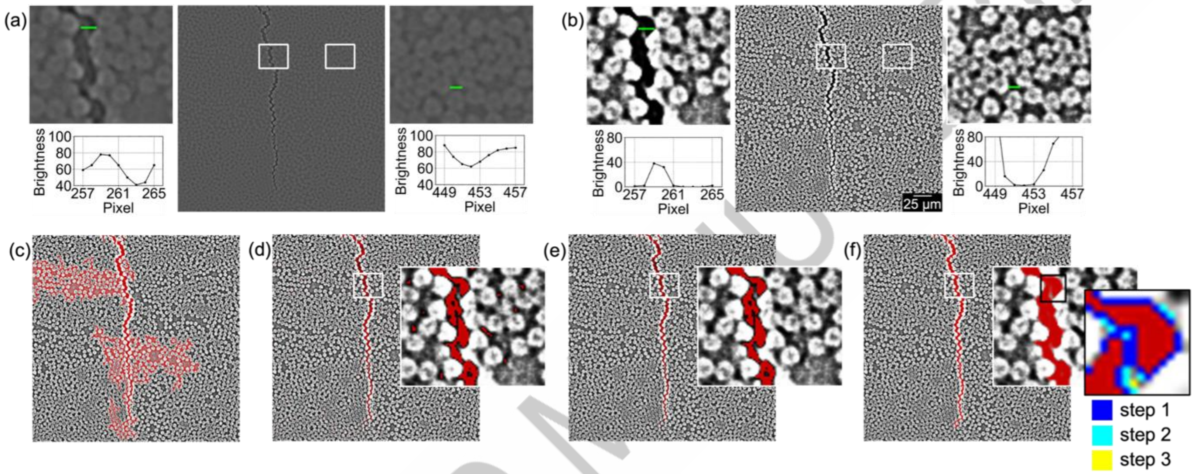
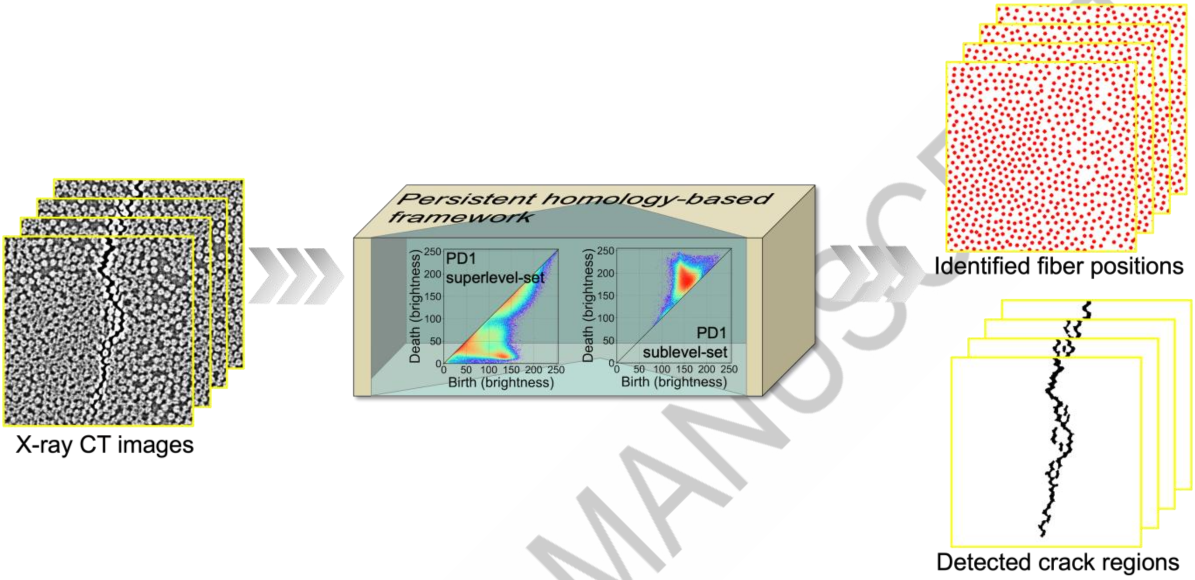


Figure 12. Detection process of crack regions. (a) Raw X-ray CT image with visible cracks. (b) Contrast enhancement via histogram equalization. (c) The level-set mask of PD1 from superlevel-set filtration (extracted at death = 0). (d, e) Refinement by applying criteria of grayscale (threshold ≤ 5) and area (≥ 11 pixels). (f) Final crack expansion using three Manhattan steps, restricted within the level-set mask.

IMPACT STATEMENT

To enable quantitative quality assessment of composite materials such as carbon fiber-reinforced plastics (CFRPs), we developed an automated and innovative framework that utilizes persistent homology to accurately and robustly extract fiber positions and cracks from X-ray CT images of CFRPs.



graphical abstract

Supplemental Material for: "Identification of microscopic structures in CFRP from X-ray CT based on topological data analysis"

Xichan Gao[a]*, Kazuto Akagi[a]*, Daiki Kido[b,c], Mayrene Uy[b], Masao Kimura[b,c]

[a]The Advanced Institute for Materials Research (AIMR), Tohoku University, Sendai, Japan; [b]Institute of Materials Structure Science (IMSS), High Energy Accelerator Research Organization (KEK), Tsukuba, Japan; [c]School of High Energy Accelerator Science, Dep. Materials Structure Science, The Graduate University for Advanced Studies (SOKENDAI), Tsukuba, Japan

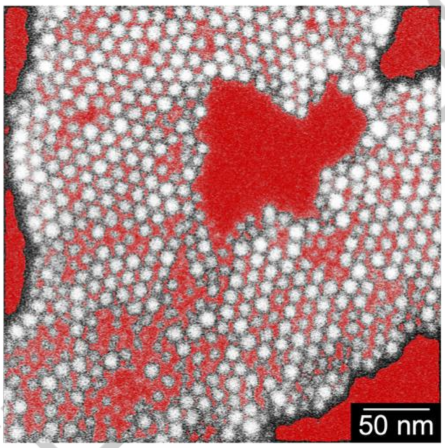
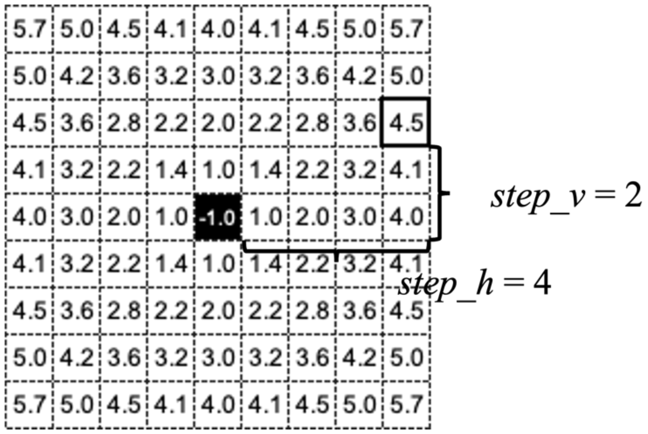


Figure S1. Level-set mask of PD1 using superlevel-set filtration on a preconditioned SEM image with contrast enhancement (histogram equalization) and noise removal (23×23 Gaussian blurring).



5.7	5.0	4.5	4.1	4.0	4.1	4.5	5.0	5.7
5.0	4.2	3.6	3.2	3.0	3.2	3.6	4.2	5.0
4.5	3.6	2.8	2.2	2.0	2.2	2.8	3.6	4.5
4.1	3.2	2.2	1.4	1.0	1.4	2.2	3.2	4.1
4.0	3.0	2.0	1.0	-1.0	1.0	2.0	3.0	4.0
4.1	3.2	2.2	1.4	1.0	1.4	2.2	3.2	4.1
4.5	3.6	2.8	2.2	2.0	2.2	2.8	3.6	4.5
5.0	4.2	3.6	3.2	3.0	3.2	3.6	4.2	5.0
5.7	5.0	4.5	4.1	4.0	4.1	4.5	5.0	5.7

Figure S2. Illustration of the Euclidean distance transform applied to a binary image.

The rule of distance transform is summarized as follows:

1. All black pixels adjacent to white pixels are marked as -1.
2. The distance from a white pixel to the nearest black pixel marked as -1 is computed as $\sqrt{step\_h^2 + step\_v^2}$.
3. The distance from a black pixel to the nearest white pixel marked as 1 is computed as $-\sqrt{step\_h^2 + step\_v^2}$.

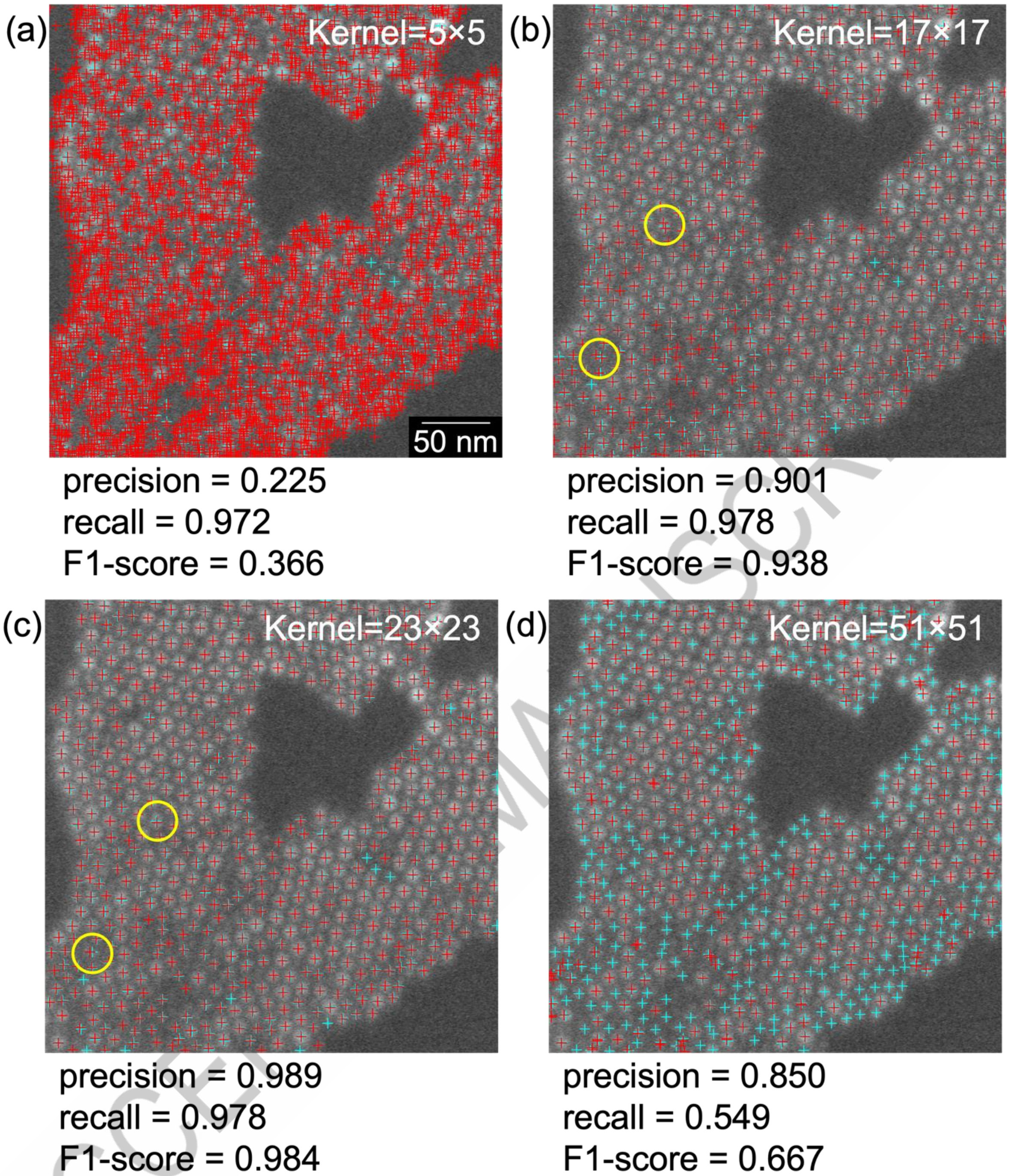


Figure S3. Detected particle positions in an SEM image with different Gaussian blur kernel sizes. Red crosses: detected positions, cyan crosses: manual annotations. (a) Results with a 5×5 kernel. (b) Results with a 17×17 kernel. (c) Results with a 23×23 kernel. (d) Results with a 51×51 kernel. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

The kernel size of Gaussian blur influences particle detection. As shown by the detected particle positions—particularly those highlighted by yellow circles—an undersized kernel size results in over-detection, where multiple regions within a single particle are spuriously identified as separate particles (i.e., excessive detection). In contrast, an oversized kernel size causes many particles to be missed.

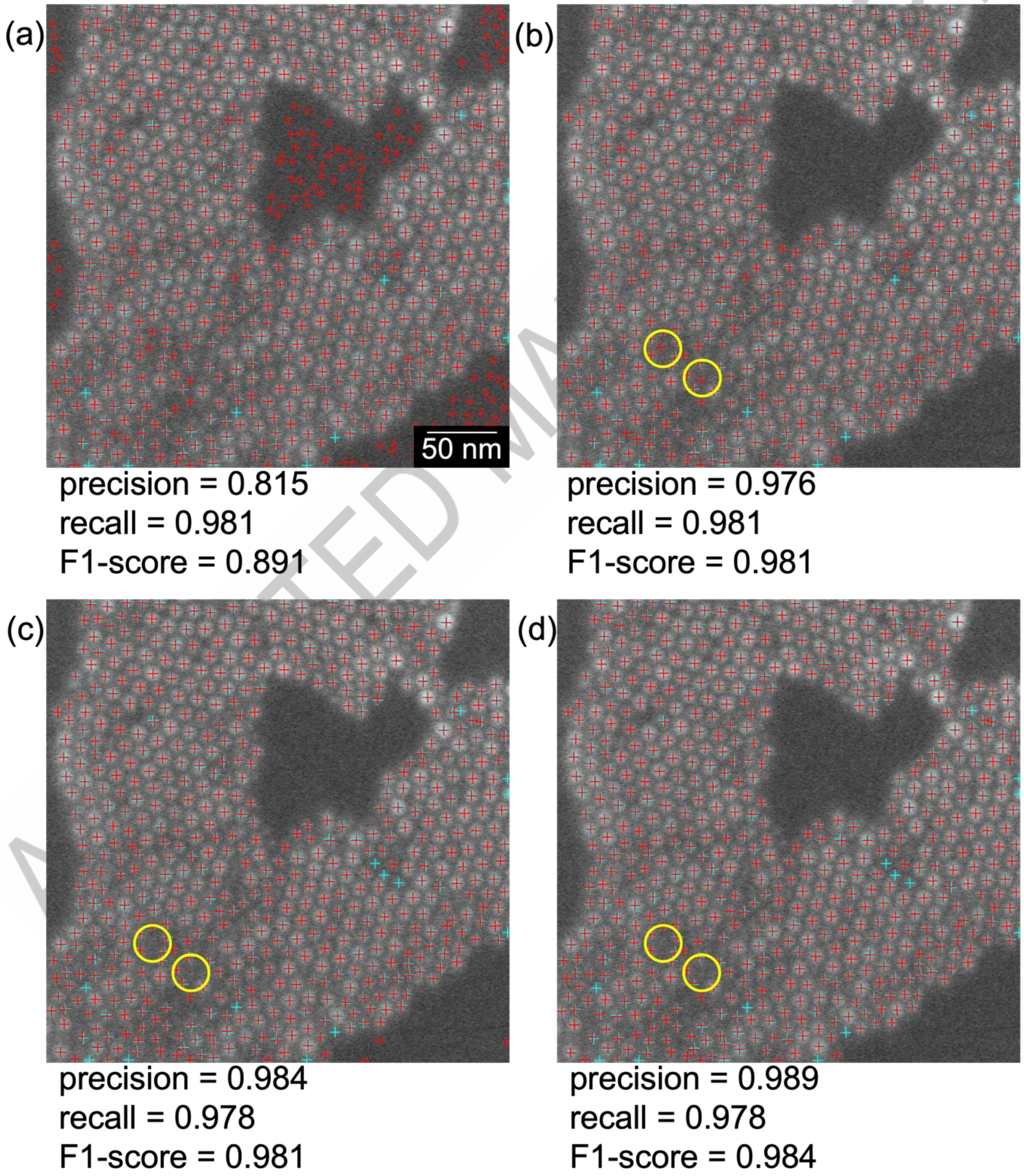


Figure S4. Detected positions in an SEM image under different refinement conditions. Red crosses: detected positions, cyan crosses: manual annotations. (a) No refinement. (b) Refinement using birth and death criteria. (c) Refinement using a level-set mask. (d) Refinement using a level-set mask as well as birth and death criteria.

As shown by the detected positions, applying the birth and death criteria removes most spurious points. In addition, as highlighted by the yellow circles, applying a level-set mask further eliminates many spurious detections.

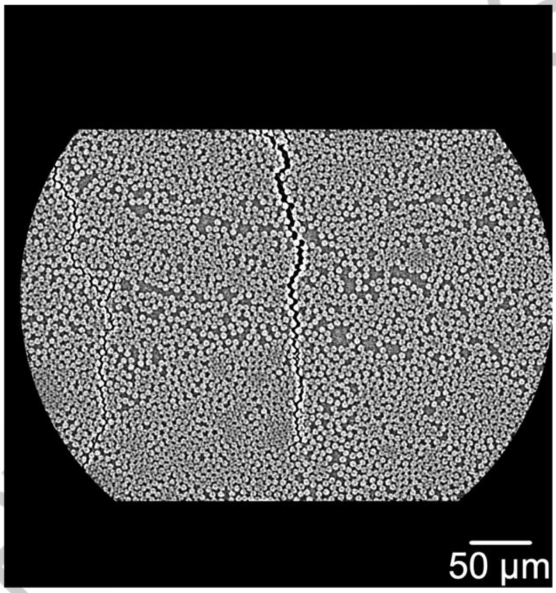


Figure S5. Example X-ray CT image showing the overall view.

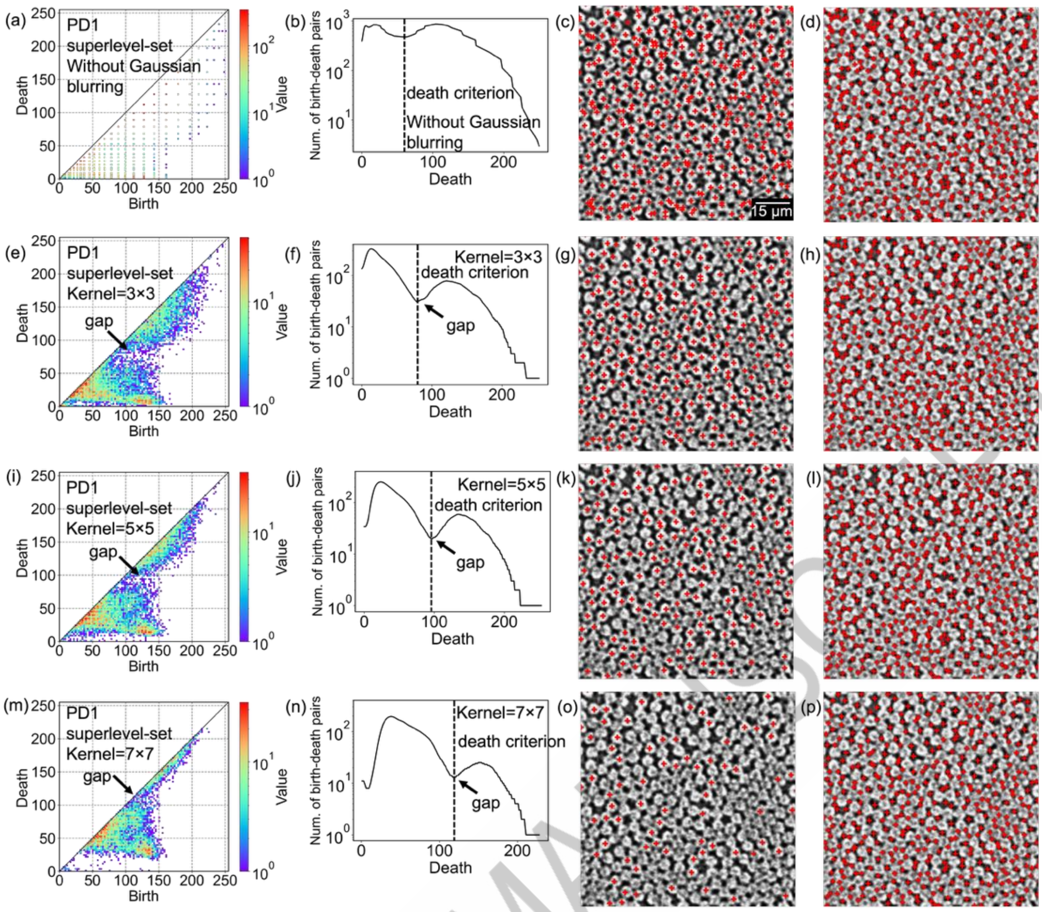


Figure S6. Void detection on a single X-ray CT image with different Gaussian blur kernel sizes. (a–d) Results without Gaussian blurring: PD1 from superlevel-set filtration, histogram of death values, and death points on fibers and resin. (e–p) Results for kernel sizes of 3×3, 5×5, and 7×7: PD1 from superlevel-set filtration (e, i, m), corresponding histograms (f, j, n), death points on fibers (g, k, o), and death points in resin (h, l, p). The PD1 color scale shows the density of birth–death pairs in a 100×100 birth–death grid. The unit of birth and death values is the intensity (brightness) in the 8-bit grayscale image.

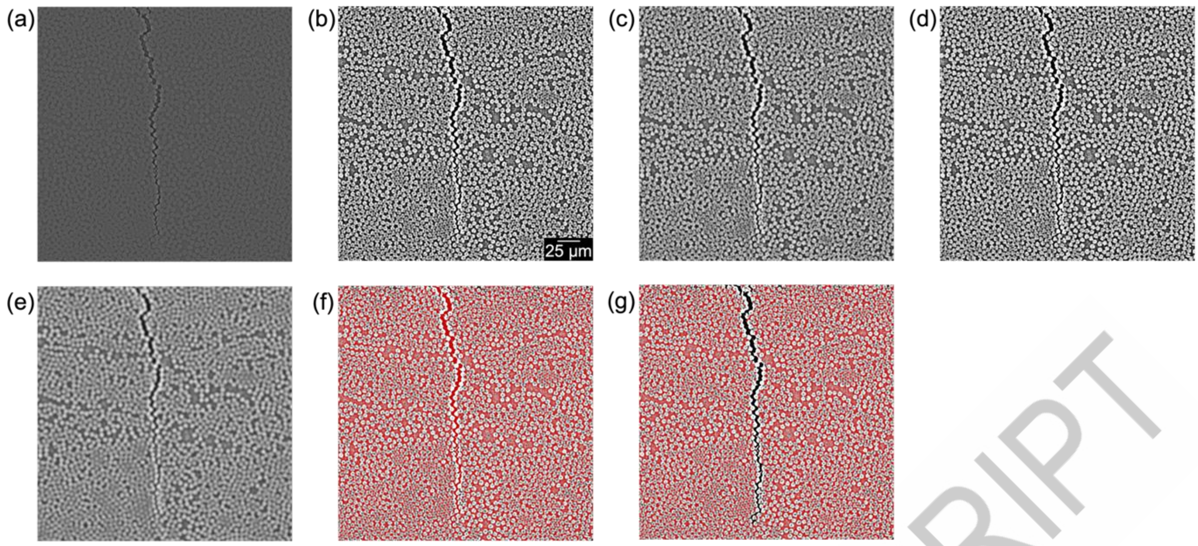


Figure S7. Identification and detection of a level-set mask. (a) Raw X-ray CT image. (b) Contrast enhancement via histogram equalization. (c) Noise removal using Gaussian blurring (kernel size: 5×5). (d) Voids filling based on superlevel-set filtration (death > 95). (e) Second-round noise removal using Gaussian blurring (kernel size: 13×13). (f) Level-set mask of PD1 based on superlevel-set filtration. (g) Final level-set mask (resin regions) after excluding identified crack regions.

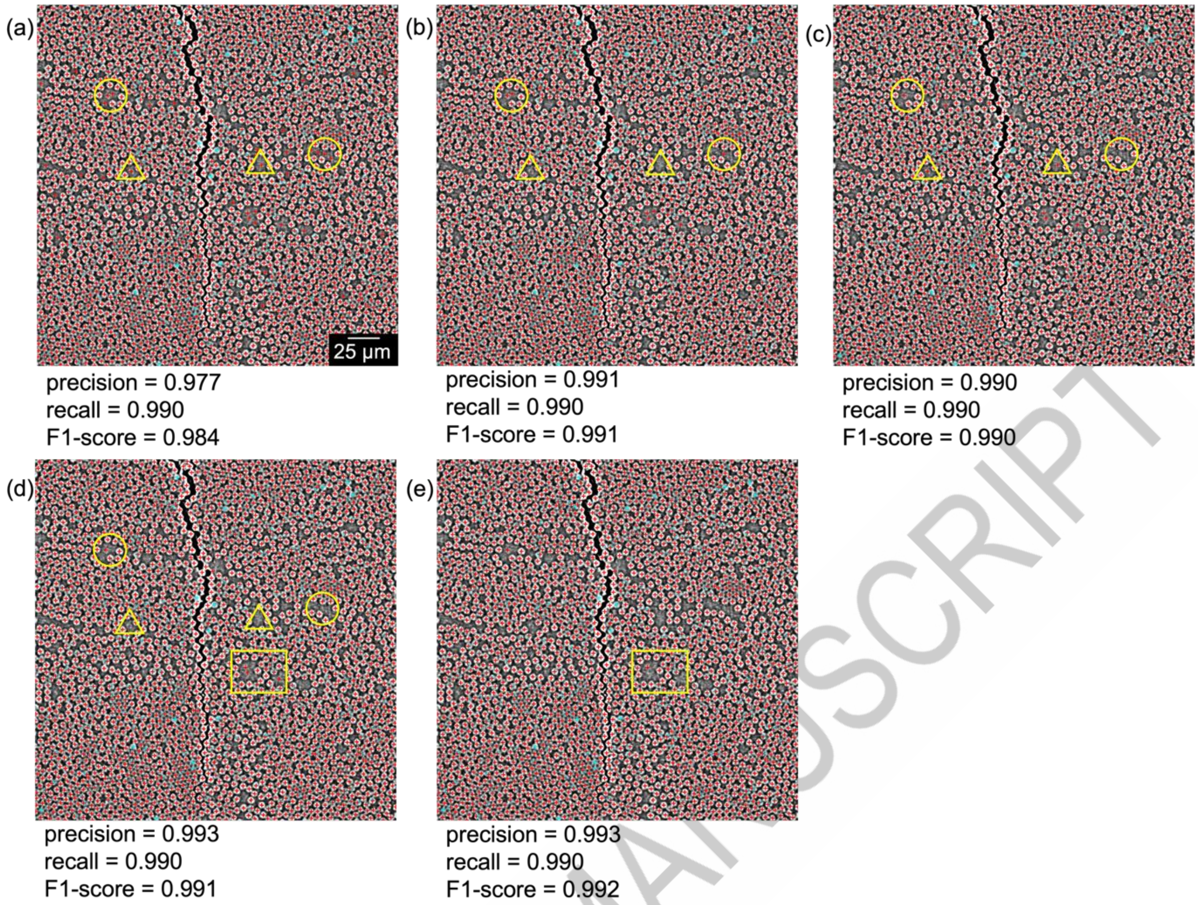


Figure S8. Detected positions in an X-ray CT image under different refinement conditions. Red crosses: detected positions, cyan crosses: manual annotations. (a) No refinement. (b) Refinement using birth and death criteria. (c) Refinement using a level-set mask. (d) Refinement using a level-set mask as well as birth and death criteria. (e) Refinement using a level-set mask, birth and death criteria, and coordination number (CN). Precision, recall, and F1-score are calculated using the manual annotations as the reference.

As shown by the detected positions, applying either a level-set mask or birth and death criteria removes most spurious points. Comparison of Figure S8a–d shows that the removals from the level-set mask (yellow circles) and from the birth and death criteria (yellow triangles) differ. Applying the CN constraint eliminates some remaining spurious detections (yellow rectangle), however, comparison of Figure S8d and Figure S8e indicates that the results are already highly accurate without CN refinement, while incorporating the CN option yields a slight improvement in accuracy.

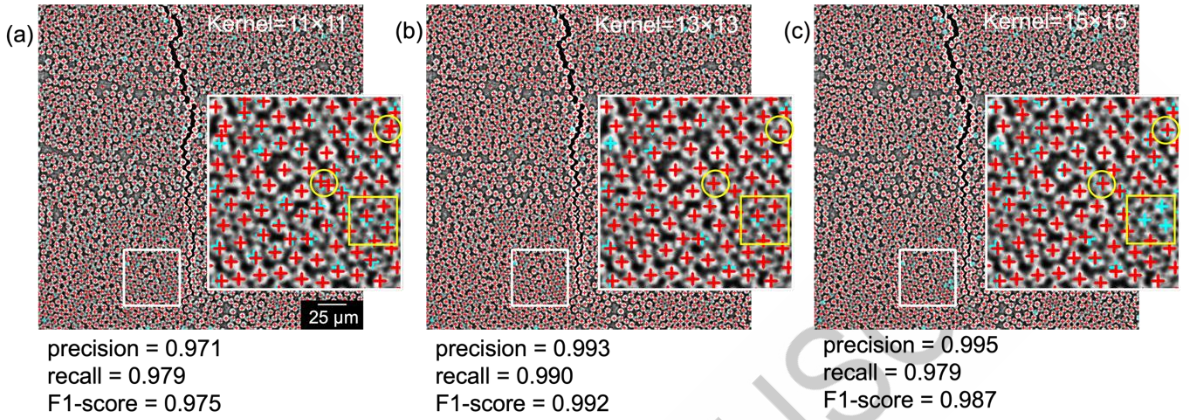


Figure S9. Total detected positions in X-ray CT images with different Gaussian blur kernel sizes. Red crosses: detected positions, cyan crosses: manual annotations. (a) Results with an 11×11 kernel. (b) Results with a 13×13 kernel. (c) Results with a 15×15 kernel. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

The kernel size of Gaussian blur influences fiber detection. As observed from the detected fiber positions, a smaller kernel size (11×11) leads to over-detection, where multiple regions within the same fiber are spuriously identified as separate fibers (i.e., excessive detection), particularly those highlighted within yellow circles. In contrast, a larger kernel size (15×15) results in missed detections, with some fibers not being identified, as indicated by the yellow rectangles.

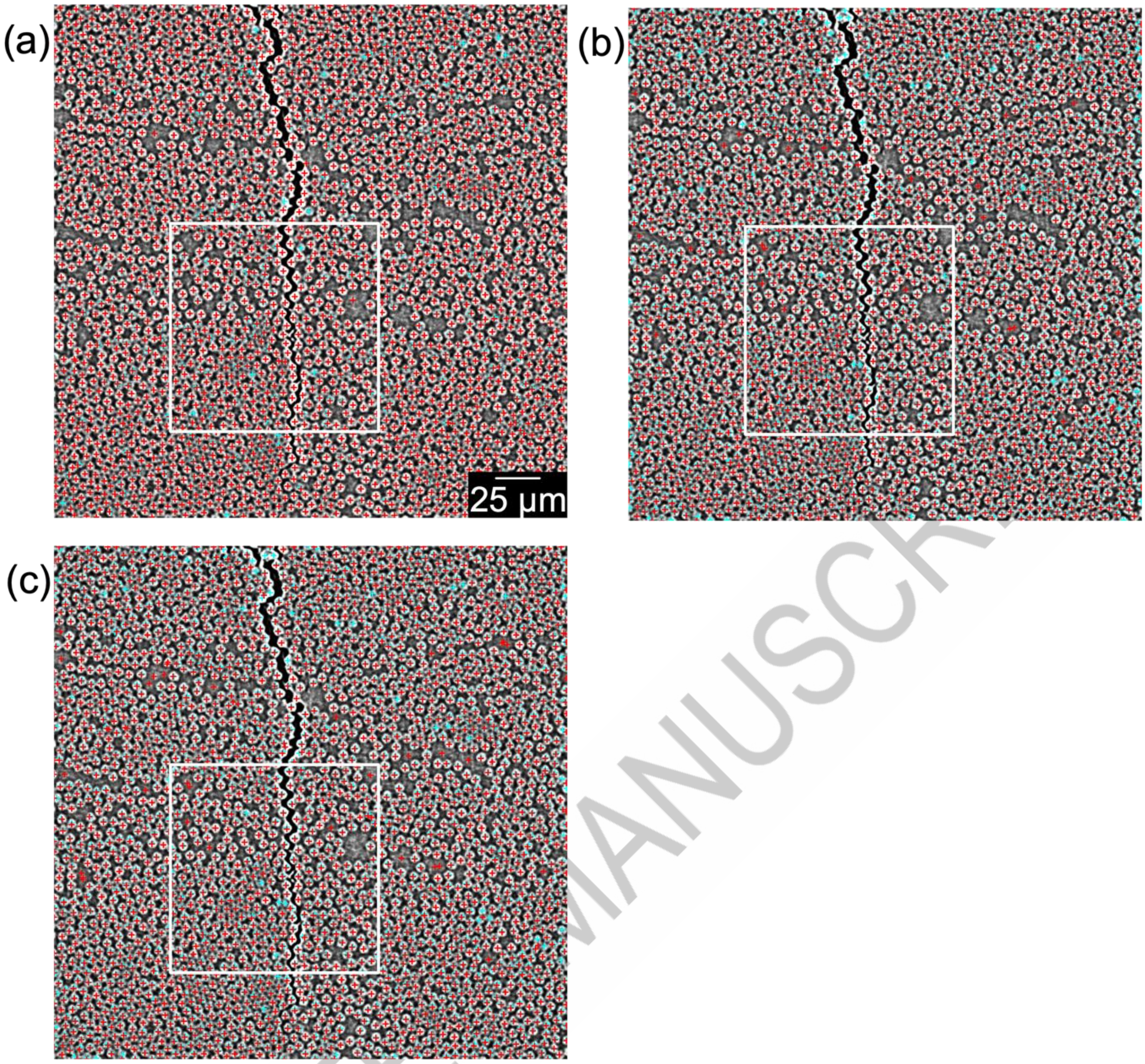


Figure S10. Full-image fiber detection results obtained using (a) our method, (b) the watershed algorithm, and (c) the local thickness algorithm. Red crosses: detected positions, cyan crosses: manual annotations. White rectangles indicate the cropped regions shown in Figure 10.

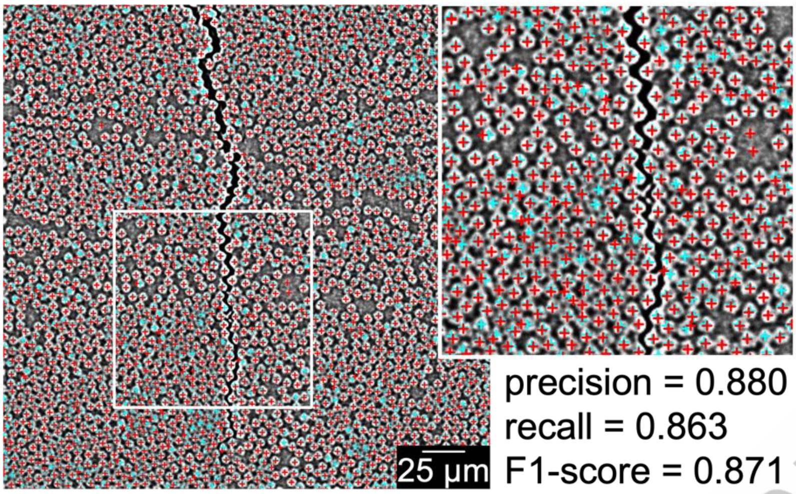


Figure S11. Fiber detection using the watershed algorithm with grayscale-based distance transforms. Red crosses: detected positions, cyan crosses: manual annotations. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

For the watershed algorithm with grayscale-based distance transforms, fiber detection starts from the raw X-ray CT image, followed by Gaussian blurring (kernel size: 13×13) to remove noise. The Euclidean distance transform is then computed on the grayscale foreground, defined by intensities above the Otsu threshold, and local maxima are identified as seeds. To reduce spurious minima and prevent over-segmentation, h-minima suppression (0.001) is applied. Seeds are expanded by watershed flooding, and region filtering (lower limit: 30 pixels, upper limit: 215 pixels) to remove unrealistically small or large objects. Final particle positions are defined as the maximum of the distance transform within each segmented region. The parameters are optimized through a grid search.

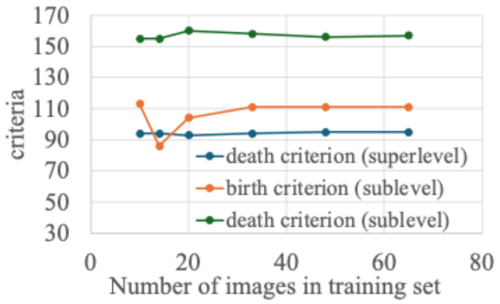


Figure S12. Convergence analysis of parameters with varying the number of sample images.

The control parameters in our method include:

- Kernel size for Gaussian blur.
- Death criterion from superlevel-set-based PD1 for separating voids in fibers and resin regions.
- Kernel size for the second-round Gaussian blur.
- Birth and death criteria from sublevel-set-based PD1 for removing spurious detections.

Kernel sizes for first- and second-round noise removal remain unchanged across different sample sizes, whereas the birth and death criteria stabilize at a sample size of 49 images (20-frame interval).

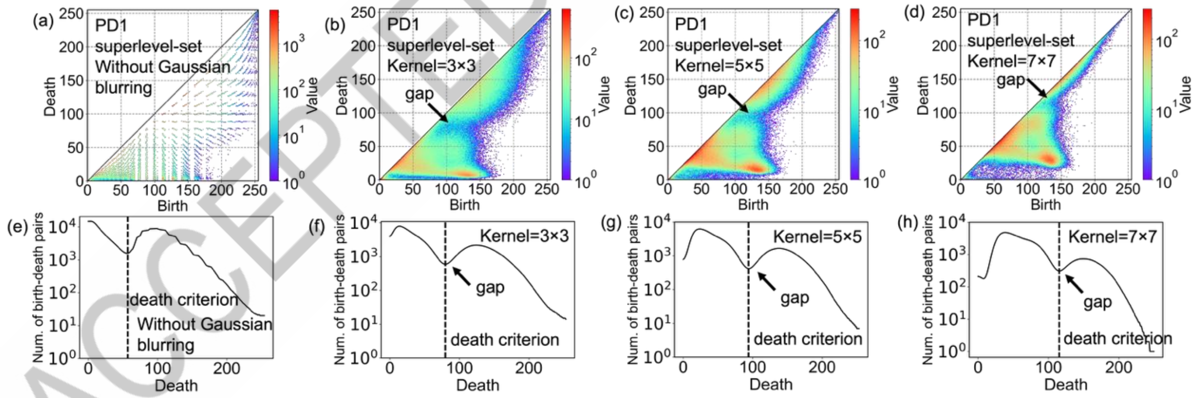
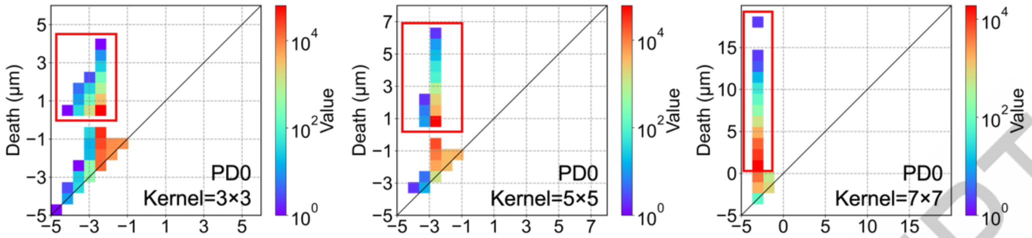
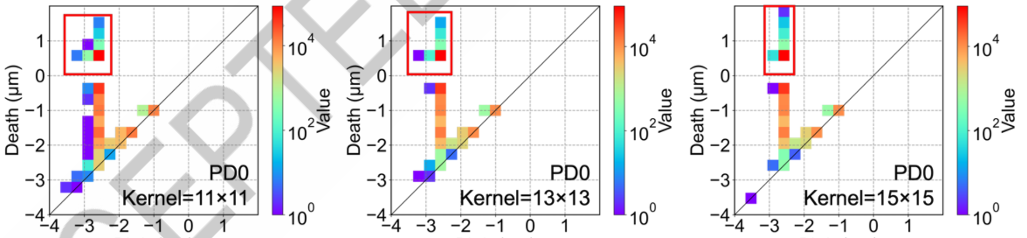


Figure S13. Death criteria calculated from 49 X-ray CT images under different Gaussian blur kernel sizes. (a) PD1 using superlevel-set filtration on images without noise removal. (b–d) PD1 using superlevel-set filtration on preconditioned images with 3×3, 5×5 and 7×7 Gaussian blur kernel sizes. (e–f) Histograms of death values with the corresponding death criteria from (a–d). The PD1 color scale shows the density of birth–death pairs in a 255×255 birth–death grid. The unit of birth and death values is the intensity (brightness) in the 8-bit grayscale image.



Number of true voids: 52266 Number of true voids: 68277 Number of true voids: 36559

Figure S14. PD0 computed from the total filling voids across 49 X-ray CT images. The PD0 color scale shows the density of birth–death pairs in a 20×20 birth–death grid. Birth–death pairs with negative death values correspond to excessive void detections, whereas those with positive death values represent true voids. The numbers of true voids are listed (highlighted within red rectangles). A kernel size of 5×5 is identified as optimal, as it yields the largest number of true voids.



Number of true detections: 87163 Number of true detections: 92647 Number of true detections: 91425

Figure S15. PD0 computed from the total detected fiber positions across 49 X-ray CT images. The PD0 color scale shows the density of birth–death pairs in a 20×20 birth–death grid. Birth–death pairs with negative death values correspond to excessive fiber detections, whereas those with positive death values represent true detections. The numbers of true detections are listed (highlighted within red rectangles). A kernel size of 13×13 is identified as optimal, as it yields the largest number of true detections.

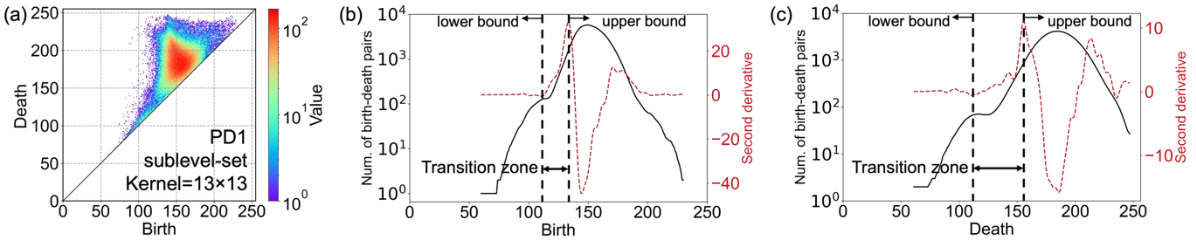
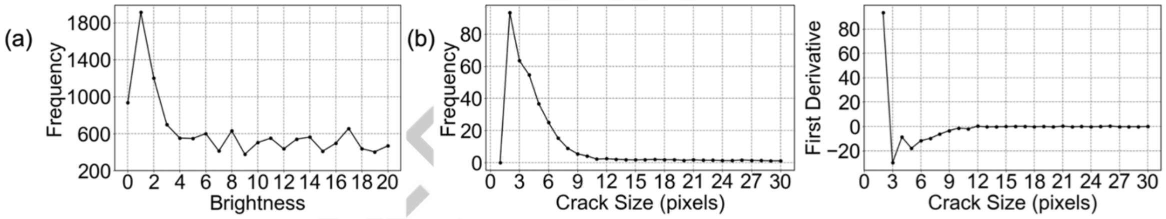
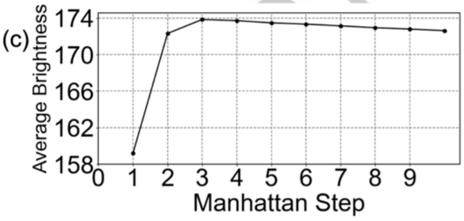


Figure S16. Determination of the birth and death criteria to eliminate spurious positions within resin regions of X-ray CT images. (a) PD1 using sublevel-set filtration on 49 preconditioned images (13×13 Gaussian blurring). The PD1 color scale shows the density of birth–death pairs in a 255×255 birth–death grid. (b) Histogram of birth values and the corresponding second derivative curve. (c) Histogram of death values and the corresponding second derivative curve.



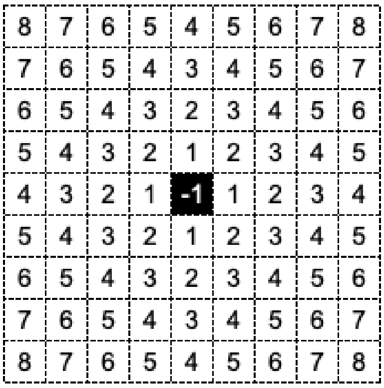
Optimizing criterion for shrinking detected area

Optimizing criterion for removing noise



Optimizing criterion for broadening detected area

Figure S17. Optimization of criteria for crack detection. (a) Determination of the grayscale criteria (≤ 5) for shrinking the detected region (shown in Figure 12c) based on the brightness histogram. (b,c) Optimization of the noise removal criteria (area ≥ 11 pixels) based on the crack-size histogram and its first derivative. (d) Optimization of the area expansion criterion based on the relationship between average brightness and Manhattan step count.



8	7	6	5	4	5	6	7	8
7	6	5	4	3	4	5	6	7
6	5	4	3	2	3	4	5	6
5	4	3	2	1	2	3	4	5
4	3	2	1	-1	1	2	3	4
5	4	3	2	1	2	3	4	5
6	5	4	3	2	3	4	5	6
7	6	5	4	3	4	5	6	7
8	7	6	5	4	5	6	7	8

Figure S18. Rules of the Manhattan distance transform for a binary image.

Outline of scripts using the Python API for obtaining a level-set mask:

```
import cv2

import numpy as np

import homcloud.interface as hc


### Pre-conditioning

image = cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)

image_histogram_eq = cv2.equalizeHist(image)

image_blur = cv2.GaussianBlur(image_histogram_eq, (xx, xx), 0) # kernel size = xx


### Detect of voids in fibers

# Calculate persistent trees with superlevel-set filtration

hc.BitmapPHTrees.for_bitmap_levelset(image_blur,          mode='superlevel',

save_to="grayscale-tree.pdgm")


# Get PD1

phtrees = hc.PDList("grayscale-tree.pdgm").bitmap_phtrees(1)


# Prepare a white canvas

h, w = image.shape[:2]

canvas = np.zeros((h, w), 255, dtype=np.uint8)


# Perform inverse-analysis and fill the inside of the rings with black

for node in phtrees.nodes:

    for pixel in node.volume():

        canvas[pixel[1], pixel[0]] = 0


# Save as an image

cv2.imwrite("output.png",                                              canvas)
```

Outline of scripts using the Python API for detecting fiber and particle positions:

```
import cv2
import numpy as np
import homcloud.interface as hc

### Pre-conditioning
image = cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)
image_histogram_eq = cv2.equalizeHist(image)
image_blur = cv2.GaussianBlur(image_histogram_eq, (xx, xx), 0) # kernel size = xx

### Detect of voids in fibers
# Calculate persistent homology with superlevel-set filtration
hc.PDList.from_bitmap_levelset(image_blur, mode='superlevel',
save_to="superlevel.pdgm")

# Get PD1
pd = hc.PDList("superlevel.pdgm").dth_diagram(1)

# Vizualize PD1
pd.histogram(x_bins=255).plot(colorbar={"type": "log"})

# Set threshold to remove spurious points in resin matrix
death_threshold = xx
pairs = [pair for pair in pd.pairs() if pair.deathtime() > xx]

# Draw white disks at the voids
for pair in pairs:
    death_position = pair.death_position
    x = death_position[1]
    y = death_position[0]
    cv2.circle(image_histogram_eq, (x, y), xx, 255, thickness=-1) # radius of
disk = xx

### Pre-conditioning
image_blur2 = cv2.GaussianBlur(image_histogram_eq, (xx, xx), 0) # kernel size = xx

### Detect of fibers positions
# Calculate persistent homology with sublevel-set filtration
hc.PDList.from_bitmap_levelset(image_blur2, mode='sublevel',
save_to="sublevel.pdgm")

# Get PD1
pd = hc.PDList("sublevel.pdgm").dth_diagram(1)

# Visualize PD1
pd.histogram(x_bins=255).plot(colorbar={"type": "log"})

# Set threshold to remove spurious points in resin matrix
birth_threshold = xx
death_threshold = yy
pairs = [pair for pair in pd.pairs() if not (pair.birthtime() <= xx and
pair.deathtime() <= yy)]

# level-set mask is also applicable here, if necessary

# Get the fiber positions as pixel coordinates
for pair in pairs:
    death_position = pair.death_position
    print(death_position)
```

Table S1. Optimized parameters for the watershed and local thickness algorithms.

	Watershed algorithm				Local thickness algorithm			
	Kernel size	hole size (pixels)	h-minima	area thresholds (pixels)	Kernel size	hole size (pixels)	h-maxima	diameter thresholds (pixels)
particles	5×5	5	1.1	(50, 1225)	5×5	3	1.1	(5, 33)
CFRP	3×3	2	0.7	(55, 205)	3×3	2	0.6	(7, 14)

Kernel size: for Gaussian blurring; hole size: for area-based hole filling (holes smaller than the threshold are filled); h-minima: suppression of spurious minima; area thresholds: regions retained only within the area range (parentheses indicate lower and upper bounds); h-maxima: removal of shallow peaks; diameter thresholds: regions retained within the diameter range (parentheses indicate lower and upper bounds).