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

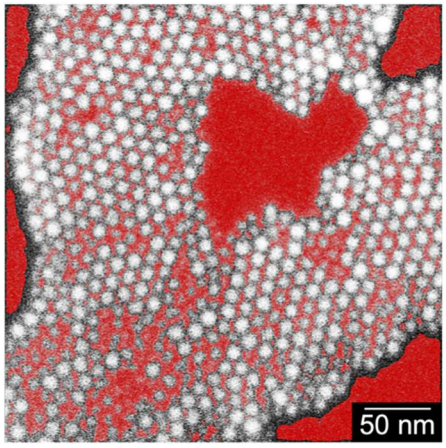

Figure S1. Level-set mask of PD1 using superlevel-set filtration on a preconditioned SEM image with contrast enhancement (histogram equalization) and noise removal (23×23 Gaussian blurring).

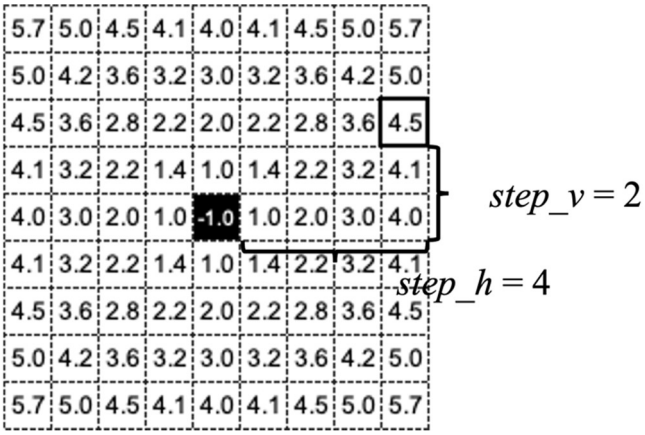

Figure S2. Illustration of the Euclidean distance transform applied to a binary image.

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 $-\sqrt{step\_h^2 + step\_v^2}$.

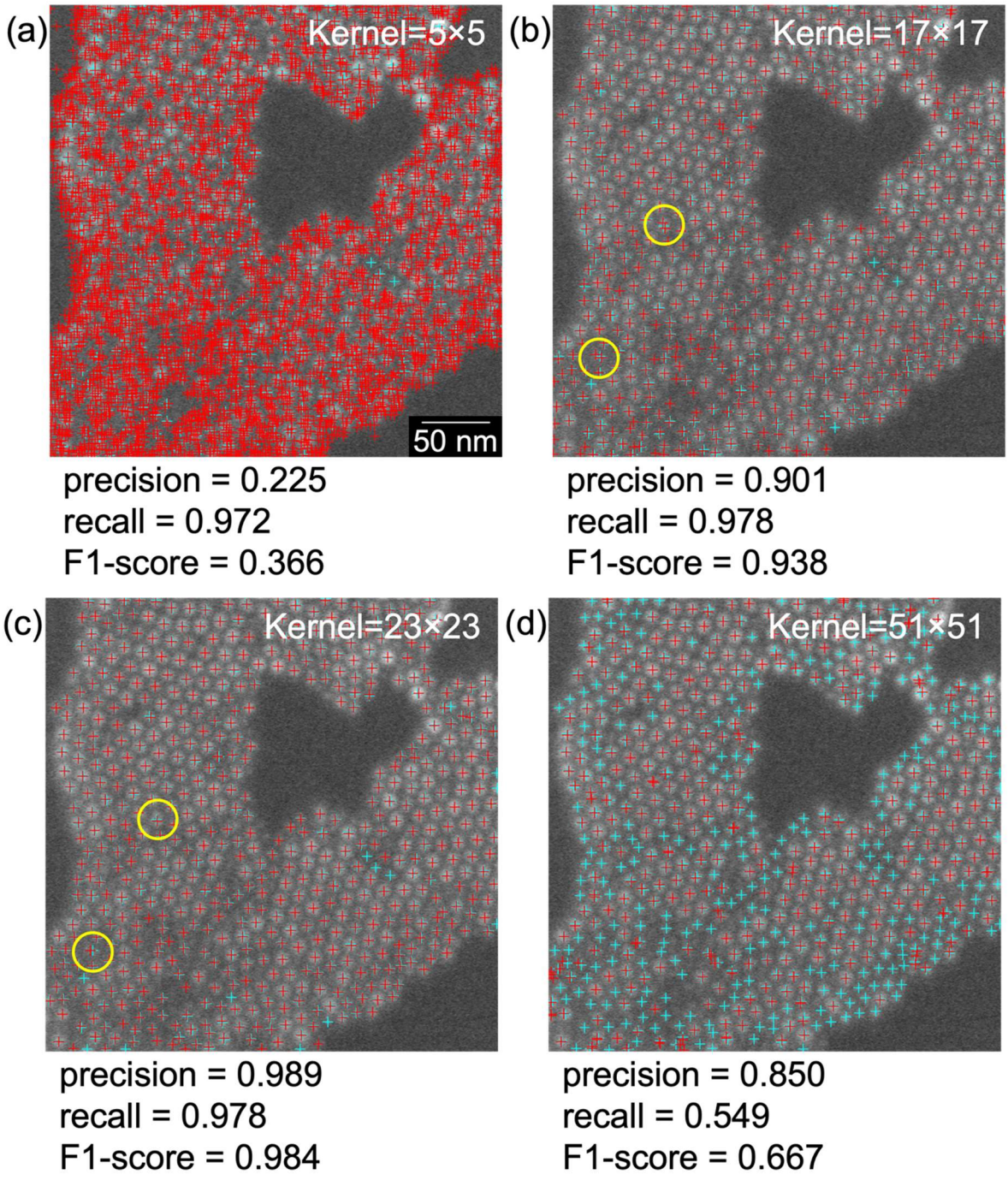

Figure S3. Detected particle positions in an SEM image with different Gaussian blur kernel sizes. Red crosses: detected positions, cyan crosses: manual annotations. (a) Results with a 5×5 kernel. (b) Results with a 17×17 kernel. (c) Results with a 23×23 kernel. (d) Results with a 51×51 kernel. Precision, 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.

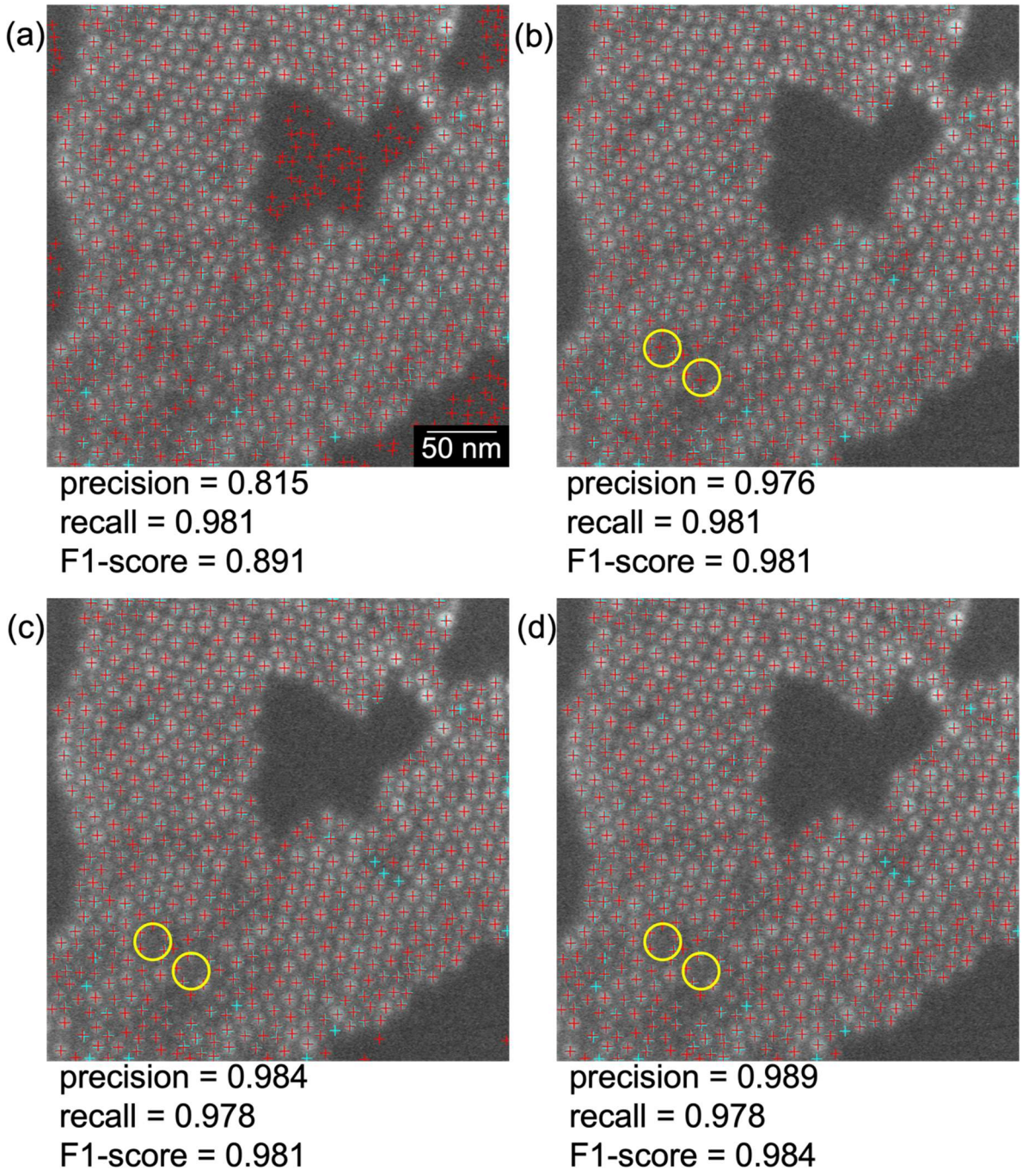

Figure S4. Detected positions in an SEM image under different refinement conditions. Red crosses: detected positions, cyan crosses: manual annotations. (a) No refinement. (b) Refinement using birth and death criteria. (c) Refinement using a level-set mask. (d) Refinement using a level-set mask as well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

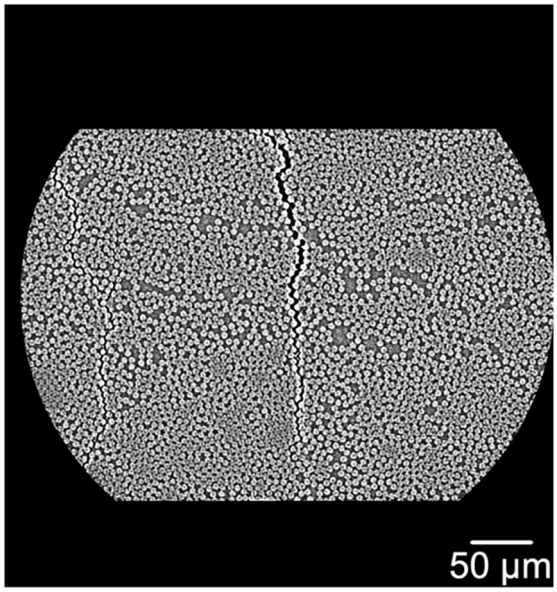

Figure S5. Example X-ray CT image showing the overall view.

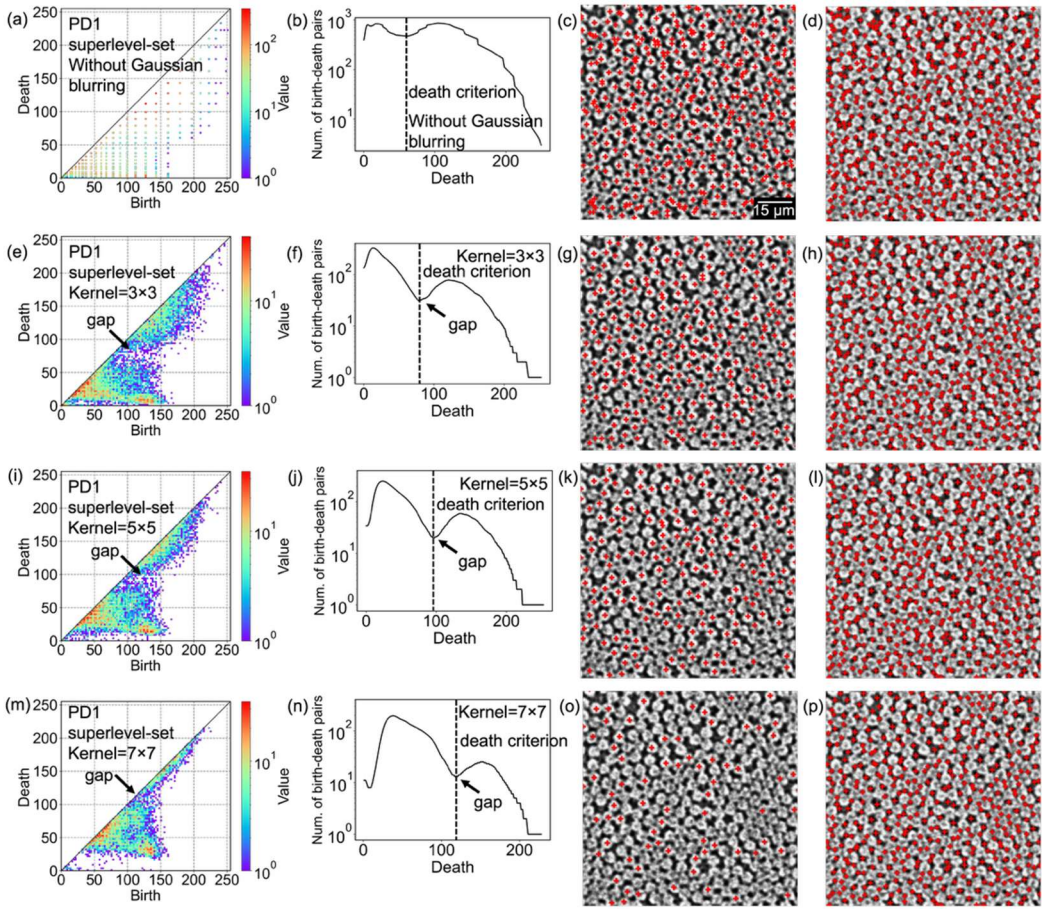

Figure S6. Void detection on a single X-ray CT image with different Gaussian blur kernel sizes. (a–d) Results without Gaussian blurring: PD1 from superlevel-set filtration, histogram of death values, and death points on fibers and resin. (e–p) Results for kernel sizes of 3×3, 5×5, and 7×7: PD1 from superlevel-set 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.

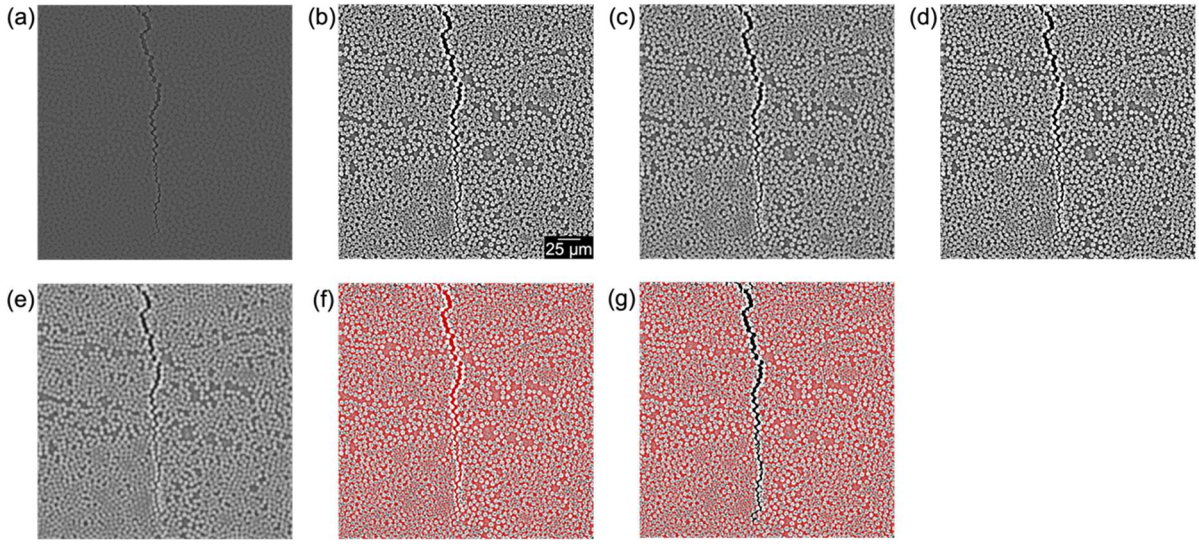

Figure S7. Identification and detection of a level-set mask. (a) Raw X-ray CT image. (b) Contrast enhancement via histogram equalization. (c) Noise removal using Gaussian blurring (kernel size: 5×5). (d) Voids filling based on superlevel-set filtration (death > 95). (e) Second-round noise removal using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

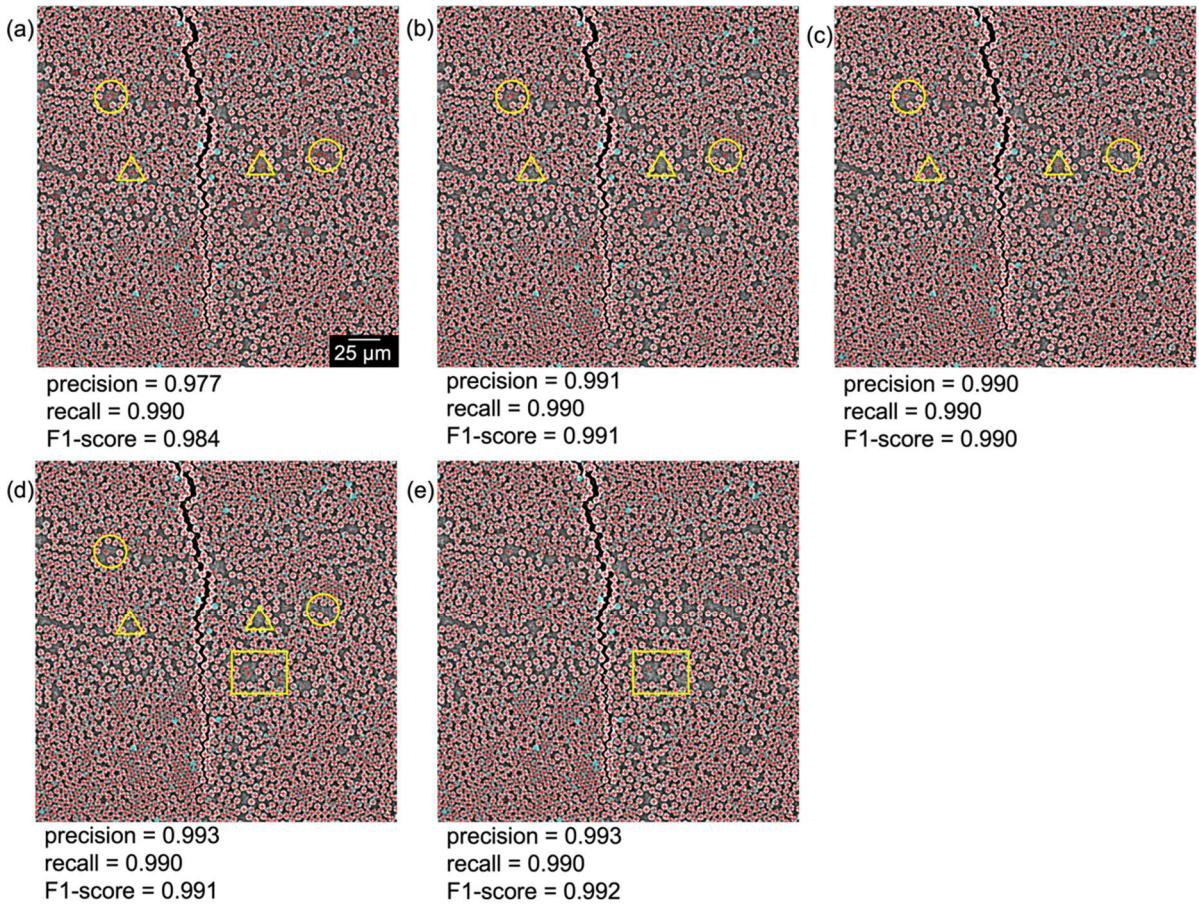

Figure S8. Detected positions in an X-ray CT image under different refinement conditions. Red crosses: detected positions, cyan crosses: manual annotations. (a) No refinement. (b)

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.

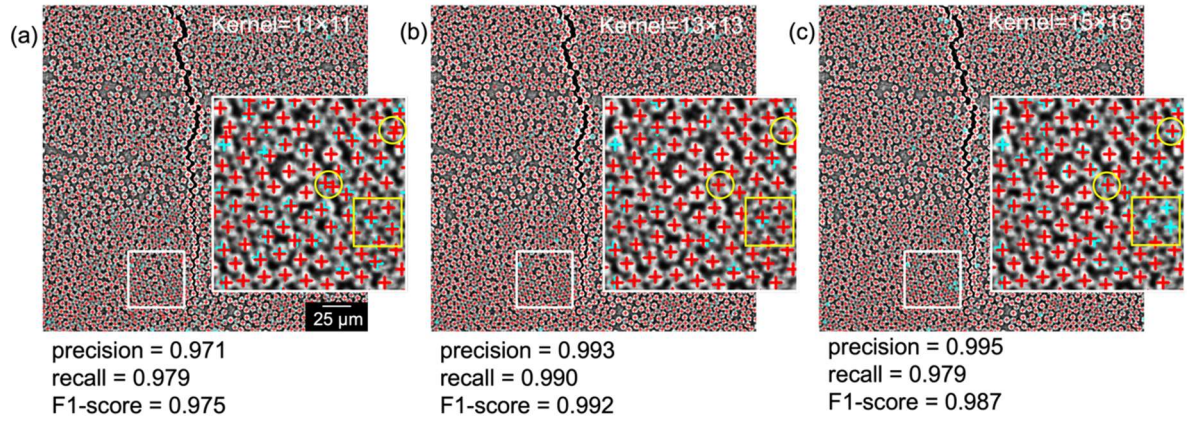

Figure S9. Total detected positions in X-ray CT images with different Gaussian blur kernel sizes. Red crosses: detected positions, cyan crosses: manual annotations. (a) Results with an 11×11 kernel. (b) Results with a 13×13 kernel. (c) Results with a 15×15 kernel. Precision, recall, and F1-score are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

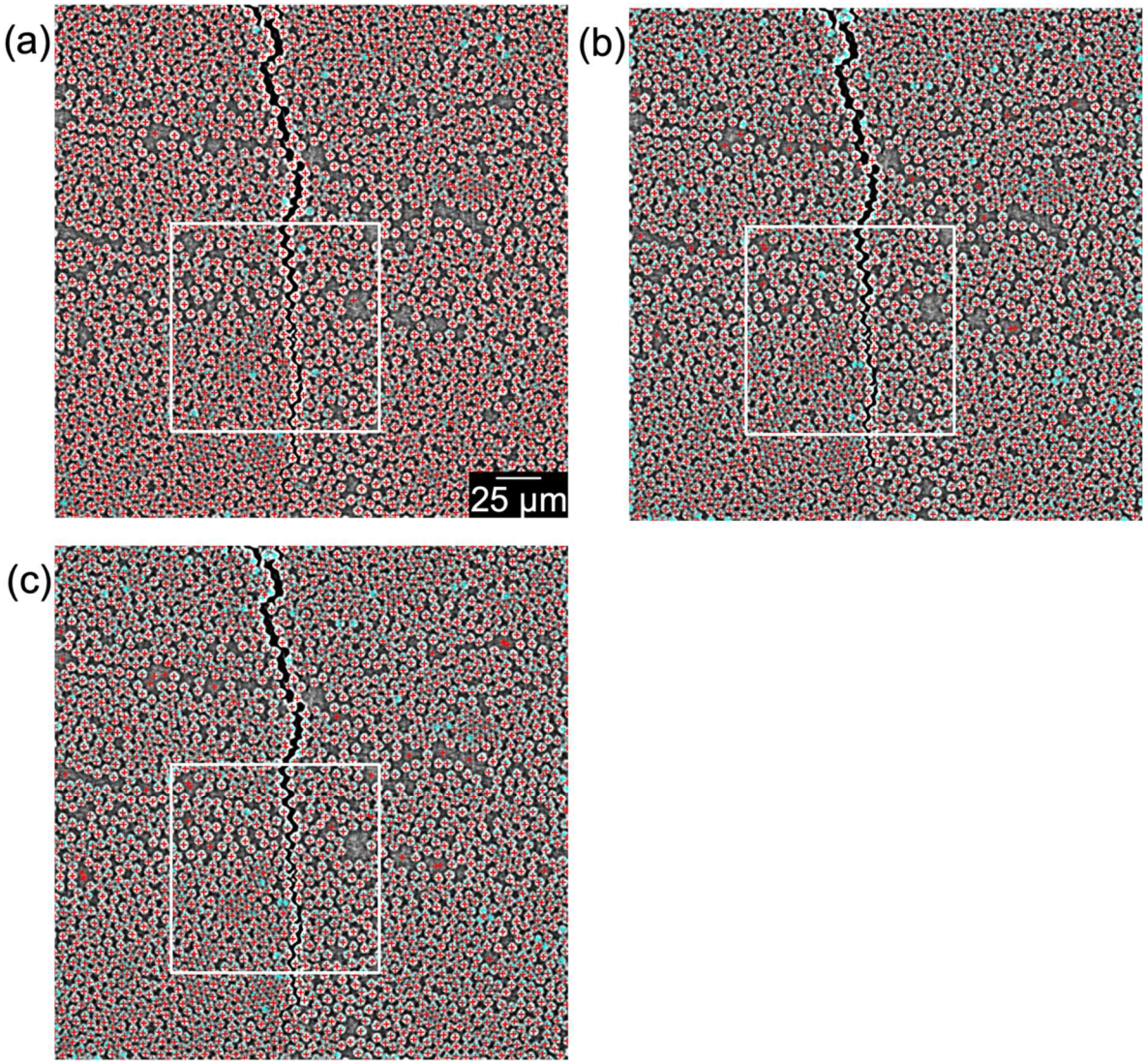

Figure S10. Full-image fiber detection results obtained using (a) our method, (b) the watershed algorithm, and (c) the local thickness algorithm. Red crosses: detected positions, cyan crosses: manual annotations. White rectangles indicate the cropped regions shown in Figure 10.

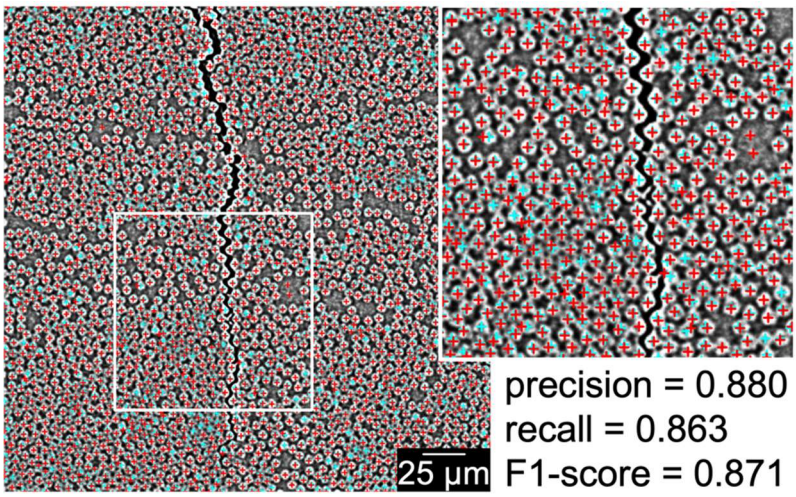

Figure S11. Fiber detection using the watershed algorithm with grayscale-based distance transforms. Red crosses: detected positions, cyan crosses: manual annotations. Precision, recall, and F1-score are calculated using the manual annotations as the reference.

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.

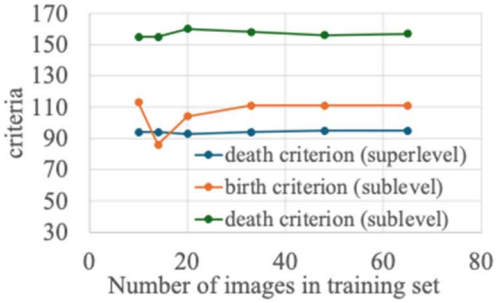

Figure S12. Convergence analysis of parameters with varying the number of sample images.

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).

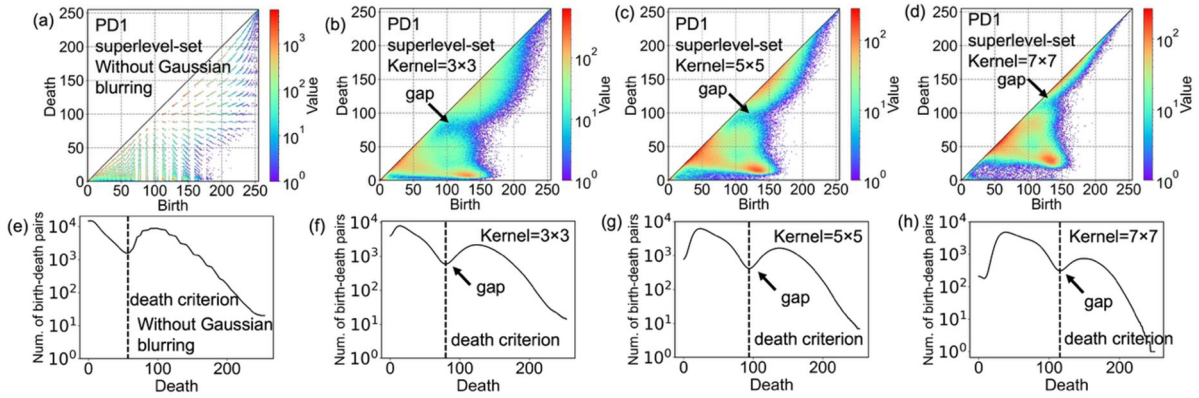

Figure S13. Death criteria calculated from 49 X-ray CT images under different Gaussian blur kernel sizes. (a) PD1 using superlevel-set filtration on images without noise removal. (b–d) PD1 using superlevel-set filtration on preconditioned images with 3×3, 5×5 and 7×7 Gaussian blur kernel sizes. (e–f) 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.

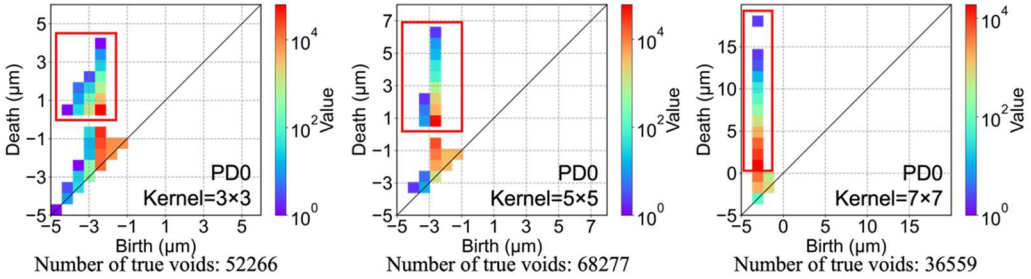

Figure S14. PD0 computed from the total filling voids across 49 X-ray CT images. The PD0 color scale shows the density of birth–death pairs in a 20×20 birth–death grid. Birth–death pairs with negative death values correspond to excessive void detections, whereas those with positive death values represent 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.

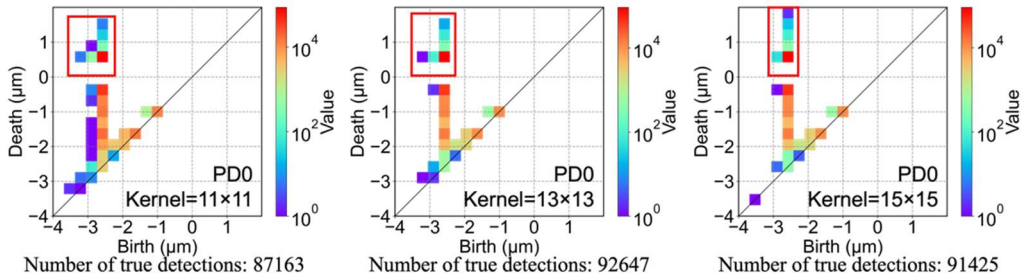

Figure S15. PD0 computed from the total detected fiber positions across 49 X-ray CT images. The PD0 color scale shows the density of birth–death pairs in a 20×20 birth–death grid. Birth–death pairs with negative death values correspond to excessive fiber detections, whereas those with positive death 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.

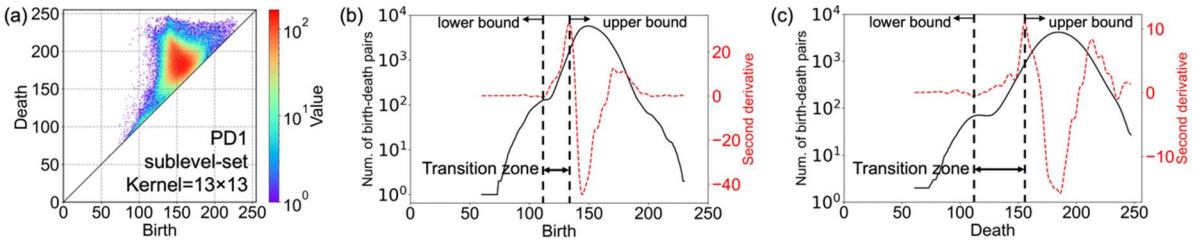

Figure S16. Determination of the birth and death criteria to eliminate spurious positions within resin regions of X-ray CT images. (a) PD1 using sublevel-set filtration on 49 preconditioned images (13×13 Gaussian blurring). The PD1 color scale shows the density of birth–death pairs in a 255×255 birth–death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

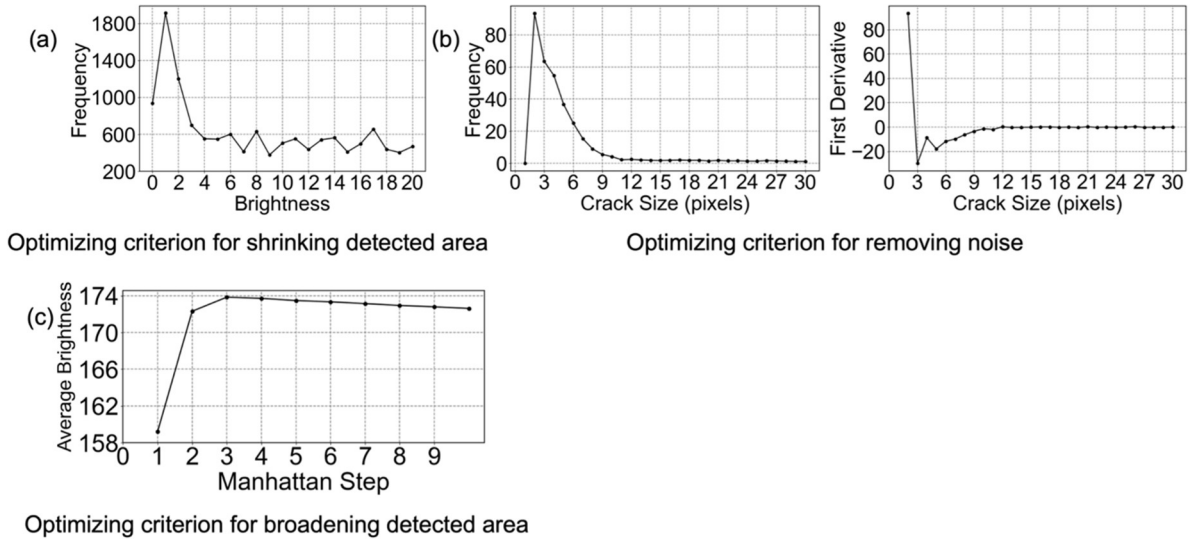

Figure S17. Optimization of criteria for crack detection. (a) Determination of the grayscale criteria (≤ 5) for shrinking the detected region (shown in Figure 12c) based on the brightness histogram. (b,c) Optimization of the noise removal criteria (area ≥ 11 pixels) based on the crack-size histogram 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).