

RESEARCH

Open Access



# Segmentation and morphological analysis of amyloid fibrils from cryo-EM image data

Matthias Weber<sup>1\*</sup> , Matthias Neumann<sup>1</sup>, Matthias Schmidt<sup>2</sup>, Peter Benedikt Pfeiffer<sup>2</sup>, Akanksha Bansal<sup>2</sup>, Marcus Fändrich<sup>2</sup> and Volker Schmidt<sup>1</sup>

\*Correspondence:

[matthias.weber@uni-ulm.de](mailto:matthias.weber@uni-ulm.de)

<sup>1</sup>Institute of Stochastics, Ulm University, Ulm, Germany

Full list of author information is available at the end of the article

## Abstract

Fast assessment of the composition of amyloid fibril samples from cryo-EM data poses a serious challenge to existing image analysis tools. We develop a method for automated segmentation of single fibrils requiring only little user input during the training process. This is achieved by combining a binary segmentation based on a convolutional neural network with preprocessing steps to allow for easy manual generation of training data. Subsequent skeletonization turns the binary segmentation into a single-object segmentation. Then, we compute properties of shape and texture of each segmented fibril, including an estimation of the fibril width. We discuss the composition of the sample based on the distributions of these computed properties and outline how a classification of fibril morphologies might be performed using these properties.

**Keywords:** Cryo-EM image data; Amyloid fibril; Cross-over distance; Fibril width; Single-object segmentation; Convolutional neural network

## 1 Introduction

Endogenous proteins regularly form fibrillary structures which serve important cellular functions. However, formation of amyloid fibrils, i.e. abnormal protein filaments, gives raise to a group of diseases, so-called amyloid diseases [4]. Examples include Parkinson's or Alzheimer's disease and systemic amyloidosis. Studying the underlying mechanism includes the investigation of amyloid fibril samples by cryogenic electron microscopy (cryo-EM) [19, 23]. As amyloid fibrils are often helically twisted, the resulting 2D images show objects of periodically varying apparent width. The features regarding geometrical shape and texture of the fibrils observed in the 2D data contain plenty information about the underlying 3D structure of the fibrils. Software packages for EM-reconstruction like RELION [12], cryoSPARC [18] or EMAN2 [26] make use of this information to fully reconstruct the molecular 3D structure of amyloid fibrils. However, this process involves cumbersome hand-picking of fibrils and requires huge computational resources. Thus, for a fast characterization of a collection of fibrils, a different procedure is needed.

A viable approach to this is the statistical analysis of fibril shapes and textures, which are directly accessible from 2D image data. In a previous study, we proposed an automated method for the extraction of so-called cross-over points [29] from 2D data without

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

performing an actual segmentation of the entire fibrils. Based on a similar approach, a method for automated picking of fibril center lines was proposed [27]. While cross-over points and center lines are helpful tools for further analysis of fibrils, only a full segmentation allows for characterization by properties of shape and texture, e.g., the mean width or variance in intensity. To enable the detailed analysis of such properties, we will present an automated segmentation of fibrils from the cryo-EM images based on the application of a suitably chosen architecture of convolutional neural networks (CNNs). For high-quality micrographs where gray-scale thresholding is feasible, viable approaches for binary segmentation exist [25, 30]. When thresholding is infeasible, e.g., due to noise or the inner structure of fibrils, CNNs provide a promising approach. They have proven successful for segmentation of image data from various microscopic techniques [8–11]. However, the cryo-EM data investigated in the present paper poses challenges already for hand-labeling and further processing. Thus, we propose a preprocessing workflow alleviating many of these challenges and facilitating easy hand-labeling and processing. The employed kind of neural networks is typically used for phase-based segmentation, resulting in an image which distinguishes between fibrils and background, but not between individual fibrils. For their analysis, it is necessary to perform an additional segmentation step in order to extract single fibrils from cryo-EM data, where our method operates on the segmentation provided by the CNN using a skeletonization algorithm to detect and distinguish single fibrils.

Based on the provided segmentation, we show how to compute various properties of shape and texture of fibrils. These properties may usually be exploited for the classification of fibril morphologies.

The rest of this paper is organized as follows. In Sect. 2, materials and methods considered in the present paper are described. In particular, we describe the two-step segmentation procedure including the employed neural network and skeletonization methods as well as the properties of shape and texture by which we characterize each fibril. In Sect. 3, we describe, validate and discuss the results obtained by the segmentation procedure and provide a description of various properties computed from the segmented fibrils. Finally, Sect. 4 concludes.

## 2 Materials and methods

### 2.1 Sample description and cryo-EM

Previously described, de-novo designed heptapeptide sequence LHLHLRL, with terminal acetyl and amide groups, was used for fibrillation [21]. The peptide was dissolved in 10 mM HCl to get 0.6 mM stock solution. This was stored as 20  $\mu$ L aliquots at  $-80^{\circ}\text{C}$ . Fibrillation was set up using 91  $\mu$ L of buffer (25 mM Tris pH 8.0, 1 mM  $\text{ZnCl}_2$ ) and 9  $\mu$ L of peptide stock to get a final peptide concentration of 0.054 mM. This was incubated at room temperature for three days to get the final sample. Sample grids for cryo-EM data collection were prepared by applying a 3.5  $\mu$ L aliquot (0.054 mM) to glow-discharged holey carbon coated grids (400 mesh C-flat 1.2/1.3), blotted with filter paper and plunge-frozen in liquid ethane using a Vitrobot Mark 3 (Thermo Fisher Scientific).

The grid quality was checked during optimization with a JEM-2100 transmission electron microscope (JEOL) at 200 kV. The data set analyzed here was acquired using a K2-Summit detector (Gatan) in counting mode on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) at 300 kV. Data was collected using 1.04  $\text{\AA}$  pixel size,

40 frames and a defocus range around  $-0.8 \mu\text{m}$  to  $-2.0 \mu\text{m}$ . Movie frames were gain corrected using IMOD [14]. MOTIONCOR2 [33] was used for motion correction and dose weighing. Motion corrected images were further used for estimation of the contrast transfer function using gctf [31].

The cryo-EM data set used in the presented work consists of 1970 images. From 691 selected images, a total of 1000 fibrils were manually examined to determine the fibril properties. This was done on the basis of fibril width and cross-over distance which were measured using the program Fiji [22].

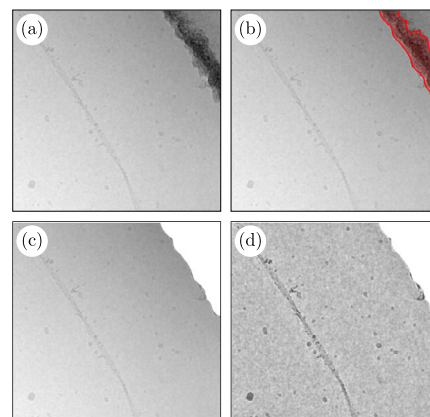
## 2.2 Binary segmentation

Most sample images contain more than one fibril, among various kinds of artifacts. The goal of this paper is the development of a procedure for the automated extraction of these fibrils from 2D cryo-EM images and for the statistical analysis of their shapes and textures. The first measure towards single-fibril segmentation consists of a binary segmentation which distinguishes a foreground phase, i.e. fibrils, from the background phase. For this, a neural network is used which is trained on hand-labeled data. To simplify hand-labeling and the application of the neural network, we perform a series of preprocessing steps, visualized in Fig. 1.

In a first step, we scale each image to 25% of its original size, as this drastically reduces computation time for subsequent processing while still preserving sufficient information for accurate segmentation. The resulting images have a pixel size of  $4.16 \text{ \AA}$ . While down-scaling reduces noise significantly, we apply a Gaussian smoothing with standard deviation of 2.5 px to further eliminate noise. Some images were taken close to the edge of the sample container. They show the container wall which we detect by applying a Sobel edge-detection [13]. After some morphological closing, the detected walls split the image into an inside and an outside part. Parts of the image outside the detected wall should not be considered for further analysis and, therefore, are removed from the image. In a final preprocessing step, we eliminate illumination gradients present in the images by subtracting a Gaussian filter with standard deviation of 50 px from the image. The Gaussian filter is computed only within the “inside part” of each image. Further measures are taken to avoid edge-effects from the Gaussian filter.

For binary segmentation of the resulting image data, various architectures of neural networks exist. A common network architecture is a sequence of convolutional and interme-

**Figure 1** Preprocessing steps applied to a sample image. The original image is downscaled and smoothed (a). Then, the container wall is detected (b) and removed (c). Finally, illumination gradients are removed (d)

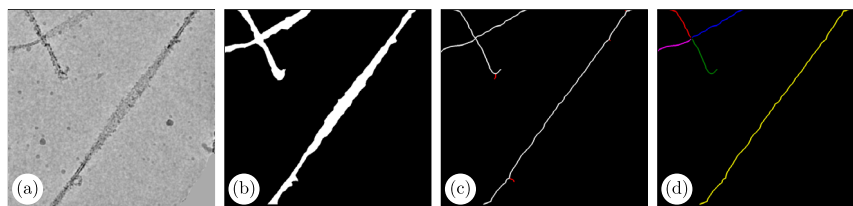


diary layers to project the input image into so-called latent space. This representation is supposed to contain the necessary information to perform the segmentation task for which the network is trained. To reconstruct the segmented image, transposed convolutions (or de-convolutions) are often used in combination with upsampling layers, resulting in a full-sized segmentation map. Examples of these networks include the popular U-net architecture [20] and Xception networks [6]. We use a U-net style adaption of the Xception architecture, i.e., a sequence of convolutional layers as described in [6] followed by upsampling and transposed convolution layers in a similar manner to achieve a full-sized segmentation map. The implementation was done in Keras/Tensorflow [1] and was based on [5]. It operates on image patches of size  $128 \times 128$  px. To generate training data, we perform hand-labeling of fibrils on the preprocessed images, where the preprocessing steps simplify this task considerably as contrast is improved and artifacts do not need to be considered manually. Hand-labeling is done by manually tracing the outline of each fibril and automatically filling in the rest of the fibril. Ambiguous cases where background is encompassed by a loop of overlapping fibrils are resolved by marking the background manually at any single point within it. The hand-labeled images are then cut into appropriately sized patches to train the neural network. In total, we used 58 hand-labeled images for training and 13 hand-labeled images for testing the neural network. Despite the low number of manually labeled images, a reliable segmentation is achieved.

Applying the trained CNN to the remaining preprocessed images yields grayscale images where the value of each pixel measures its likeliness to belong to a fibril. By setting a threshold of 0.5, we obtain a binary image where the true phase corresponds to fibrils as predicted by the CNN.

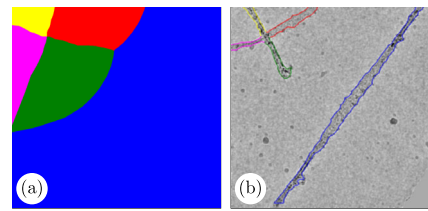
### 2.3 Single fibril segmentation

For statistical analysis, the foreground phase of the binarized images corresponding to the union set of all fibrils needs to be further segmented into single fibrils. Often, two or more fibrils cross each other at a single point, rendering analysis of their shape at this point infeasible. Thus, we resort to segmentation of the foreground phase into straight, non-overlapping segments of fibrils. To achieve this, we employ a skeletonization and subsequent processing of the skeleton to identify these segments as shown in Fig. 2. First, we apply a dilation with a disk of 10 px radius to the binarized images in order to eliminate noise-induced artifacts which otherwise would lead to errors in the skeleton. Then, we apply Zhang's algorithm [32] which calculates a skeleton of the foreground phase by successively removing boundary pixels without breaking the connectivity of the object. Similar



**Figure 2** Preprocessed image (a) and dilated binary segmentation map (b), obtained from application of the CNN, which is then skeletonized using Zhang's algorithm (c). Undesirable spurs (short branches) are shown in red. Pruning deletes these branches and, finally, the skeleton is split into non-branching parts (d). For visualization purposes all skeletons have been dilated by a few pixels

**Figure 3** By Euclidean distance transform, each pixel is labeled according to the nearest part of the skeleton (a). Using this information together with the binary segmentation map yields individual pieces of non-overlapping fibrils (b)



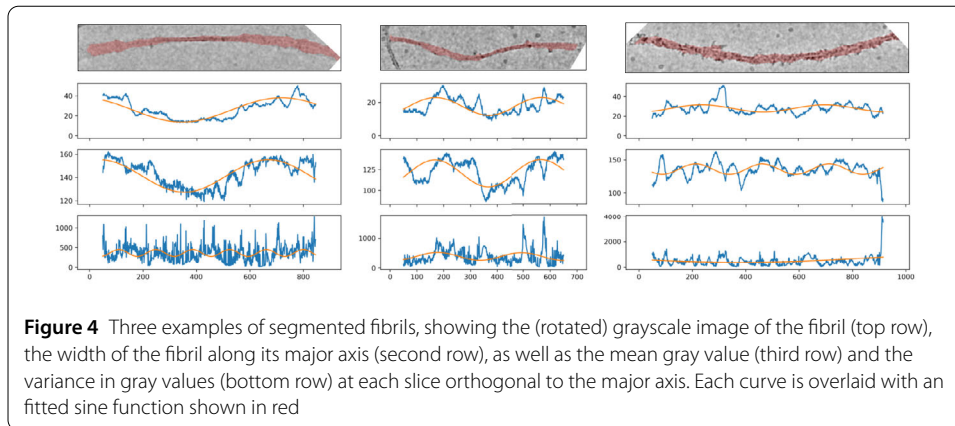
to other skeletonization algorithms like, e.g., Lee-thinning [15], the resulting skeleton suffers from undesirable branches which may be induced by surface roughness of the object. In our case, these spurs need to be removed, because they do not correspond to actual fibril segments.

To achieve a spur-free skeleton, so-called pruning algorithms are frequently applied, see e.g. [3]. However, standard pruning algorithms suffer from inconsistencies when two branches starting from the same point are shorter than a given threshold. These inconsistencies may result in both branches being removed, when one actually belongs to the (long) main branch, or both branches staying in place. We avoid this by successively removing the shortest branch/spur of the skeleton as long as it is shorter than a given threshold. This is achieved by determining branching points and end points of the skeleton using the hit-and-miss transformation [24] and computing the geodesic distance along each branch from its end point to the next branching point. Then, the shortest branch is removed from the skeleton if it is shorter than the given threshold and the process is repeated. Splitting the pruned skeleton at its branching points, the resulting parts correspond to the fibril segments of interest. These fibril segments are reconstructed by computing the Euclidean distance transform with respect to the skeleton segments. In doing so, we memorize the part of the skeleton to which the distance of a given point is smallest, see Fig. 3. The resulting map identifies the closest skeleton part for each pixel of the fibril. Finally, we rotate each fibril segment such that its major axis is parallel to the  $x$ -axis. As most fibril segments are fairly straight with no bends, this results in well-defined data for subsequent analysis.

## 2.4 Properties of shape and texture

For each fibril (i.e., fibril segment), we compute various properties of shape and texture, where we mask the original grayscale image with the segmented fibril. Then, we apply a one-dimensional Gaussian filter with standard deviation of 10 px along the major axis of the fibril to reduce noise and localized artifacts on the fibril. Furthermore, we apply a masked version of the Gaussian filter to eliminate edge effects. Many fibrils exhibit a grayscale gradient orthogonal to the major axis where the innermost part of the fibril is darkest and the outer parts are brighter. This is due to the molecular structure of the fibril consisting of twisted protofilaments [2]. Note that the one-dimensional Gaussian filter mostly conserves this gradient while still eliminating noise.

At each position along the major axis of a fibril, we compute the width of the fibril, the mean gray value and the variance in gray values on a slice orthogonal to the major axis. For each fibril, this yields three curves as shown in Fig. 4. It turns out that many fibrils exhibit a periodic pattern in all curves which is due to the helical structure of the fibrils also seen in the grayscale images. However, the sample seems to comprise different types of fibrils which also exhibit different patterns in these curves. Furthermore, an established



property of amyloid fibrils is the so-called cross-over distance [2, 16]. Formally, cross-overs are points on the fibril's skeleton where the fibril width (in its 2D projection) takes a local minimum. The distance between these points equals the pitch of the helical fibril structure. For measuring the cross-over distance, we fit sine functions to the width and gray values of each fibril. The wavelength of these functions roughly corresponds to the cross-over distance. By comparing the wavelengths and the fit quality, we can assess the reliability of the computed cross-over distance and identify fibrils which do not exhibit cross-overs.

Thus, for each fibril, we compute a number of scalar properties consisting of their width, the gray value at the cross-overs, the coefficient of variation of width along the major axis, and the cross-over distance.

### 3 Results

#### 3.1 Manual assessment of fibril morphologies

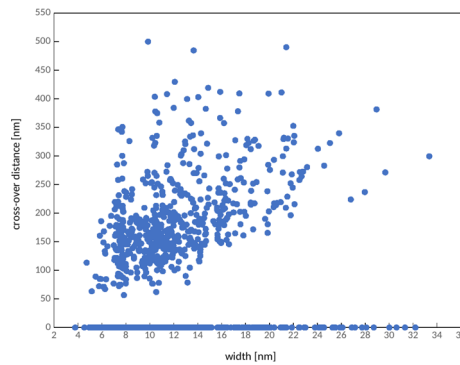
To investigate the possible presence of different fibril morphologies, 1000 fibrils were manually analyzed regarding width and cross-over distance. While the cross-over distance could not be measured for roughly 40% of the fibrils, the fibril width was obtained for all analyzed fibril structures. Based on a plot of the cross-over distance versus fibril width for  $\approx 60\%$  of the fibrils (Fig. 5) and the histogram of the fibril width for all analyzed fibrils (Fig. 6), it was not possible to readily subdivide the data set into well-resolved fibril morphologies. Nevertheless, visual analysis of the micrographs appeared to yield similar fibril morphologies at widths of approximately 5-6, 7-8, 9-10 and 11-12 nm (Fig. 7).

#### 3.2 Segmentation

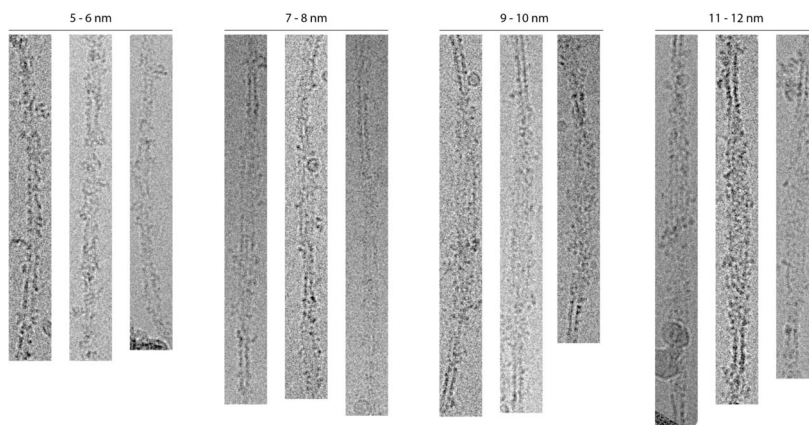
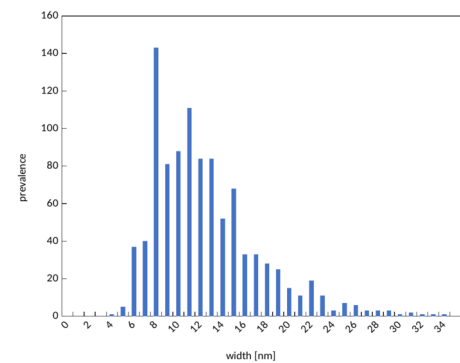
For the chosen approach to single-fibril segmentation, we note that segmentation quality can be assessed at various intermediate steps. The most crucial part of the segmentation procedure is the binary segmentation performed by the CNN, see Sect. 2.2. As we retained hand-labeled data which was not used for training, we can test the performance of the CNN segmentation using this data. A well-established measure for the quality of binary segmentation maps is the Sørensen-Dice coefficient  $\gamma \in [0, 1]$ , see [7], which is defined by

$$\gamma = \frac{2|T \cap S|}{|T| + |S|},$$

**Figure 5** Cryo-EM based measurement of fibril width and cross-over distance computed on 1000 fibrils. No cross-overs were found in the images for 413 fibrils, which are shown as dots on the horizontal axis



**Figure 6** Histogram of the distribution of fibril width present in the cryo-EM data set, computed by analyzing 1000 fibrils and using a binning of 1 nm

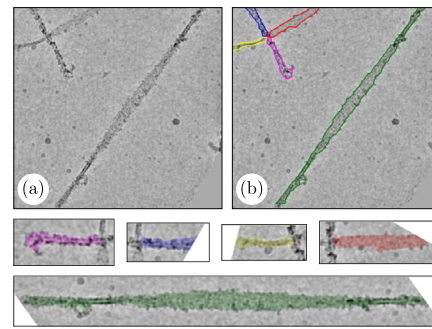


**Figure 7** Cryo-EM images of fibril morphologies present in the data set. Scalebar: 50 nm

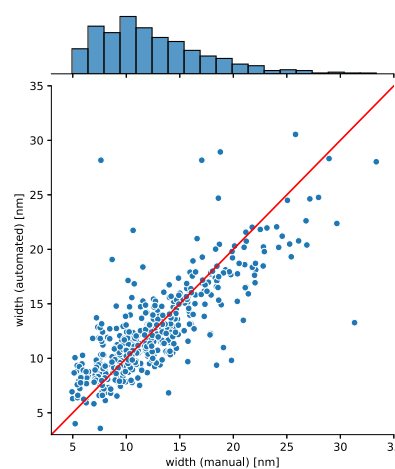
where  $|T \cap S|$  is the area of the intersection of ground truth  $T$  and segmentation map  $S$  and  $|T|$  and  $|S|$  are their respective areas. A value of  $\gamma = 1$  corresponds to  $T = S$  while a value of  $\gamma = 0$  indicates that no pixel is classified correctly in the segmentation map. Applying the trained neural network to the retained test data yields a mean Sørensen-Dice coefficient of  $\gamma = 0.83$ , indicating a reliable binary segmentation.

Assuming a sufficiently correct binary segmentation, the remaining steps can be assessed by visual inspection. However, still a possible pitfall is splitting the foreground phase into single fibril segments. Here, overlapping fibrils might not be split into sepa-

**Figure 8** Sample cutouts with fibril outlines overlaid on top, where each cutout is rotated such that the fibril's major axis matches the x-axis: Original (preprocessed) image (a) and single-fibril segmentation obtained by applying the CNN and the skeleton-based method (b)



**Figure 9** Scatterplot of manually measured widths (x-axis) and automatically determined widths (y-axis). The red line indicates the diagonal. Histograms of both variables are shown on the sides

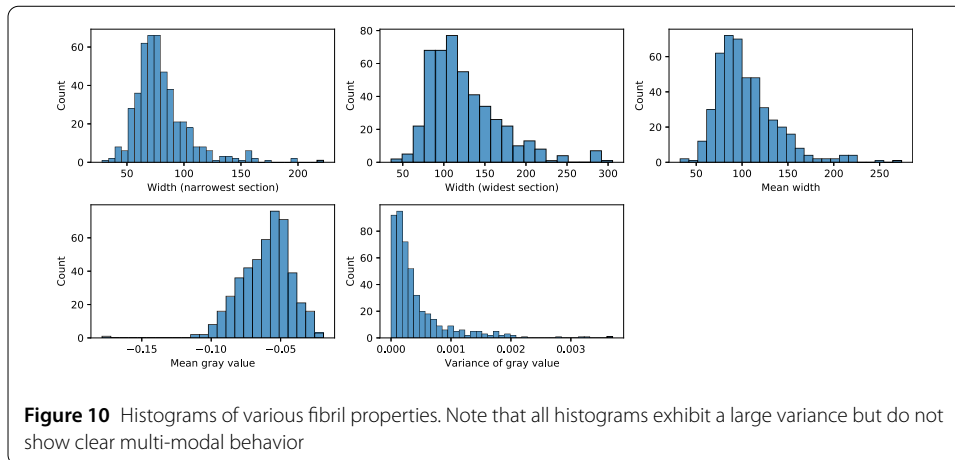


rate parts or straight fibrils without any overlap might be split into different segments. While the former would pose serious problems for further analysis and result in unreliable data, the latter would be a mere nuisance as we would only lose some information regarding connectivity. However, our method for single-fibril segmentation aims to avoid both errors and achieves good results, see Fig. 8. In total, we detected 2334 fibrils.

### 3.3 Validation

Based on the manually determined fibril widths, we performed further validation of the segmentation. For 500 of the manually measured fibrils, coordinates on the fibrils were stored and used to link the manually measured widths to the widths measured on the automated segmentation. As explained in Sect. 3.1, fibril widths may typically be used to classify fibril morphologies. Figure 9 shows the measured and automatically determined widths for all fibrils. The mean width obtained by manual measurement was  $124.3 \text{ \AA}$  while the mean width obtained by automated segmentation was  $124.9 \text{ \AA}$ . The median deviation of manually and automatically measured width was  $13.9 \text{ \AA}$ . Note that, as manual measurements were performed on images with a resolution reduced by a factor of 4, this corresponds to a difference of only 3.33 px. This result strengthens the proof of reliability established by computing the Sørensen-Dice coefficient, see Sect. 3.2.





### 3.4 Geometric properties of extracted fibrils

As described in Sect. 2.4, we computed several properties for each fibril. They are based on the width and mean gray values along each fibril, see Fig. 4. Histograms of various fibril properties are shown in Fig. 10. Note that the histograms of minimum, maximum and mean width of all fibrils, automatically extracted from image data, show a similar distribution, which is due to a strong positive correlation between these three properties. The correlation between minimum and maximum width is 0.76, between mean width and maximum width 0.97 and between mean width and minimum width 0.88. Properties based on the gray values which can be obtained from the single-fibril segmentation include the mean gray value and the variance of the gray value within each fibril. As the gray values observed in the cryo-EM data are linked to the type of protofilaments and their arrangement, they might be used as a proxy to determine fibril morphologies. When different morphologies are present in the sample, we would expect a broad variation in both properties and, more importantly, multi-modal distributions. While we observe a broad variation, the distributions do not appear to be multi-modal and thus, an appropriate classification of fibril morphologies based on these histograms is not feasible. This may be due to heterogeneity in the fibril morphologies. Similarly, a classification based on the properties derived from fibrils widths fails. In line with the attempt to manually classify different fibril morphologies (cf. Sect. 3.1, this suggests a high variability of the visual appearance of fibrils of the same morphology in cryo-EM data. As this heterogeneity is specific to the present data, a similar approach on other data may yield sufficient information for a fast classification of fibril morphologies using, e.g., a Gaussian mixture model [17] on any combination of the computed properties.

### 3.5 Discussion

The proposed method for automated segmentation of single fibrils leads to visually convincing results which is corroborated by the computed validation measures. Unlike many previously developed methods, see, e.g., [27, 28], the resulting segmentation does not only give the fibril center lines or rough outlines but also their precise shapes. While we mainly use this information to determine the fibril morphologies, further characteristics may be derived from the segmentation of single fibrils in order to characterize the obtained fibril morphologies. The methodology developed in the present paper does not use any preliminary classification of short fibril segments, as, e.g., in [18]. This pre-classification leads

to noise reduction before the final classification of morphologies is performed, which is beneficial compared to the methodology proposed in the present paper. However, an advantage of our approach is that it is independent of possible errors occurring at the stage of pre-classification, see the discussion in Sect. 4 of [18].

As described above, the single-fibril segmentation lends itself to the computation of various properties of fibrils. When classification of fibril morphologies based on fibril widths is feasible, the presented approach may be used to quickly and reliably determine the fibril widths and thus facilitating a fast classification. However, for the present data no such classification was possible due to heterogeneity in the fibril morphologies. As a further application, the single-fibril segmentation may be used as input for tools like, e.g., RELION [12] which require center lines or other segmentation-like input.

#### 4 Conclusion

In the present work, we proposed a method for mostly automated segmentation of amyloid fibrils from cryo-EM image data. While our method relied on binary segmentation by a convolutional neural network, preprocessing steps ensured that only little ground-truth data was needed. Furthermore, manual labeling of the preprocessed image data was fast and easy. The method yielded a reliable binary segmentation as shown by computing the Sørensen-Dice coefficient as well as by visual inspection. Further statistical analysis of the sample composition required a single-object segmentation achieved by employing a skeletonization scheme. This resulted in individually segmented fibrils for which we then computed various properties of shape and texture. In agreement with the attempt to manually classify different fibril morphologies, a high variability of the properties with no clear multi-modal distribution was observed, rendering an easy classification infeasible. However, insofar as manually measured values were available, the validation of the computed properties corroborated the reliability of the proposed method.

#### Funding

We thank the Deutsche Forschungsgemeinschaft for funding (FA 456/23-1 to MF and SCHM 997/30-1 to VS). All cryo-EM data were collected at the European Molecular Biology Laboratory, Heidelberg, funded by iNEXT (Horizon 2020, European Union). The work of MN was partially funded by the POLIS Cluster of Excellence (EXC 2154/1). Open Access funding enabled and organized by Projekt DEAL.

#### Availability of data and materials

The datasets used and analyzed during the current study are subject of ongoing research and are therefore not publicly available.

#### Declarations

##### Competing interests

The authors declare that they have no competing interests.

##### Author contribution

MW designed and implemented the segmentation algorithm, performed the analysis of the fibrils and wrote major parts of the manuscript. MN and MS contributed in performing the analysis and writing the manuscript. PF performed the manual assessment of fibrils and created the respective figures and text. AB performed the acquisition and initial analysis of the cryo-EM data. MF and VS supervised and aided in the conception and preparation of the manuscript. All authors read and approved the final manuscript.

##### Author details

<sup>1</sup>Institute of Stochastics, Ulm University, Ulm, Germany. <sup>2</sup>Institute of Protein Biochemistry, Ulm University, Ulm, Germany.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 June 2022 Accepted: 19 January 2023 Published online: 07 February 2023

## References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. <https://www.tensorflow.org/>. Software available from [tensorflow.org](https://www.tensorflow.org/).
2. Annamalai K, Gührs KH, Koehler R, Schmidt M, Michel H, Loos C, Gaffney PM, Sigurdson CJ, Hegenbart U, Schönland S, Fändrich M. Polymorphism of amyloid fibrils in vivo. *Angew Chem, Int Ed Engl*. 2016;55(15):4822–5.
3. Bai X, Latecki LJ, Liu WY. Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Trans Pattern Anal Mach Intell*. 2007;29(3):1–14.
4. Chiti F, Dobson CM. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu Rev Biochem*. 2017;86(1):27–68.
5. Chollet F. Image segmentation with a U-Net-like architecture. [https://keras.io/examples/vision/oxford\\_pets\\_image\\_segmentation/](https://keras.io/examples/vision/oxford_pets_image_segmentation/).
6. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p. 1251–8.
7. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302.
8. Evsevlev S, Paciornik S, Bruno G. Advanced deep learning-based 3D microstructural characterization of multiphase metal matrix composites. *Adv Eng Mater*. 2020;22(4):1901197.
9. Fend C, Moghiseh A, Redenbach C, Schladitz K. Reconstruction of highly porous structures from FIB-SEM using a deep neural network trained on synthetic images. *J Microsc*. 2021;281(1):16–27.
10. Furat O, Finegan DP, Diercks D, Usseglio-Viretta F, Smith K, Schmidt V. Mapping the architecture of single lithium ion electrode particles in 3D, using electron backscatter diffraction and machine learning segmentation. *J Power Sources*. 2021;483:229148.
11. Furat O, Wang M, Neumann M, Petrich L, Weber M, Krill CE, Schmidt V. Machine learning techniques for the segmentation of tomographic image data of functional materials. *Front Mater*. 2019;6:145.
12. He S, Scheres SHW. Helical reconstruction in RELION. *J Struct Biol*. 2017;198(3):163–76.
13. Kanopoulos N, Vasanthavada N, Baker RL. Design of an image edge detection filter using the Sobel operator. *IEEE J Solid-State Circuits*. 1988;23(2):358–67.
14. Kremer JR, Mastronarde DN, McIntosh R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol*. 1996;116(1):71–6.
15. Lee TC, Kashyap RL, Chu CN. Building skeleton models via 3-D medial surface/axis thinning algorithms. *CVGIP, Graph Models Image Process*. 1994;56:462–78.
16. Liberta F, Loerch S, Rennegarbe M, Schierhorn A, Westermark P, Westermark GT, Hazenberg BPC, Grigorieff N, Fändrich M, Schmidt M. Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nat Commun*. 2019;10(1):1104.
17. McLachlan GJ, Basford KE. *Mixture models: inference and applications to clustering*. New York: Dekker; 1988.
18. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. 2017;14(3):290.
19. Radamaker L, Lin YH, Annamalai K, Huhn S, Hegenbart U, Schönland SO, Fritz G, Schmidt M, Fändrich M. Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun*. 2019;10(1):1103.
20. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Berlin: Springer; 2015. p. 234–41.
21. Rufo CM, Moroz YS, Moroz OV, Stöhr J, Smith TA, Hu X, DeGrado WF, Korendovych IV. Short peptides self-assemble to produce catalytic amyloids. *Nat Chem*. 2014;6(4):303–9.
22. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–82.
23. Schmidt A, Annamalai K, Schmidt M, Grigorieff N, Fändrich M. Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril. *Proc Natl Acad Sci*. 2016;113(22):6200–5.
24. Soille P. *Morphological image analysis*. Berlin: Springer; 2004.
25. Sokolov PA, Belousov MV, Bondarev SA, Zhouavleva GA, Kasyanenko NA. FibrilJ: ImageJ plugin for fibrils' diameter and persistence length determination. *Comput Phys Commun*. 2017;214:199–206.
26. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*. 2007;157(1):38–46.
27. Thurber KR, Yin Y, Tycko R. Automated picking of amyloid fibrils from cryo-EM images for helical reconstruction with RELION. *J Struct Biol*. 2021;213(2):107736.
28. Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, Quentin D, Roderer D, Tacke S, Siebolds B, Schubert E, Shaikh T, Lill P, Gatsogiannis C, Raunser S. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol*. 2019;2(1):1–13.
29. Weber M, Bäuerle A, Schmidt M, Neumann M, Fändrich M, Ropinski T, Schmidt V. Automatic identification of crossovers in cryo-EM images of murine amyloid protein A fibrils with machine learning. *J Microsc*. 2020;277(1):12–22.
30. Yin Y, Prigent S, Torrent J, Rezaei H, Drasdo D, Doumic M. Automated quantification of amyloid fibrils morphological features by image processing techniques. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. 2019. p. 534–7.
31. Zhang K. Gctf: Real-time CTF determination and correction. *J Struct Biol*. 2016;193(1):1–12.
32. Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. *Commun ACM*. 1984;27:236–9.
33. Zheng S, Palovcak E, Armache JP, Verba K, Cheng Y, Agard D. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. 2017;14(4):331–2.