


# Deep learning for automatic segmentation of the nuclear envelope in electron microscopy data, trained with volunteer segmentations

Helen Spiers<sup>1,2</sup>  | Harry Songhurst<sup>1,3</sup> | Luke Nightingale<sup>4</sup> | Joost de Folter<sup>4</sup> | The Zooniverse Volunteer Community | Roger Hutchings<sup>2</sup> | Christopher J. Peddie<sup>1</sup> | Anne Weston<sup>1</sup> | Amy Strange<sup>4</sup> | Steve Hindmarsh<sup>4</sup> | Chris Lintott<sup>1,2</sup> | Lucy M. Collinson<sup>1</sup> | Martin L. Jones<sup>1</sup>

<sup>1</sup>Electron Microscopy Science Technology Platform, The Francis Crick Institute, London, UK

<sup>2</sup>Department of Physics, University of Oxford, Oxford, UK

<sup>3</sup>Department of Computer Science, University of Manchester, Manchester, UK

<sup>4</sup>Scientific Computing Science Technology Platform, The Francis Crick Institute, London, UK

## Correspondence

Helen Spiers, Electron Microscopy Science Technology Platform, The Francis Crick Institute, London NW1 1AT, UK.  
Email: bowerssh@crick.ac.uk

## Funding information

Cancer Research UK, Grant/Award Number: FC001999; Medical Research Council: FC001999; Wellcome Trust: FC001999

## Abstract

Advancements in volume electron microscopy mean it is now possible to generate thousands of serial images at nanometre resolution overnight, yet the gold standard approach for data analysis remains manual segmentation by an expert microscopist, resulting in a critical research bottleneck. Although some machine learning approaches exist in this domain, we remain far from realizing the aspiration of a highly accurate, yet generic, automated analysis approach, with a major obstacle being lack of sufficient high-quality ground-truth data. To address this, we developed a novel citizen science project, Etch a Cell, to enable volunteers to manually segment the nuclear envelope (NE) of HeLa cells imaged with serial blockface scanning electron microscopy. We present our approach for aggregating multiple volunteer annotations to generate a high-quality consensus segmentation and demonstrate that data produced exclusively by volunteers can be used to train a highly accurate machine learning algorithm for automatic segmentation of the NE, which we share here, in addition to our archived benchmark data.

## KEYWORDS

cell biology, cellular imaging, citizen science, image processing, machine learning, segmentation, volume electron microscopy

## 1 | INTRODUCTION

Until recently, the study of cell morphology with electron microscopy (EM) was often restricted to qualitative illustration, as technological limitations prevented quantitative analysis of samples in three

dimensions. Development of novel volume EM methodologies, including serial blockface scanning electron microscopy (SBF SEM)<sup>1</sup> and focused ion beam SEM (FIB SEM),<sup>2</sup> has enabled automated acquisition of images through greater depths at high resolution,<sup>3</sup> with one microscope able to generate hundreds of gigabytes of aligned serial images per day.

However, our ability to analyse these data has not seen comparable advancement; segmentation of EM images remains a difficult and time-consuming manual process. Hence, to fully realize the analytical

The Zooniverse Volunteer Community: This publication has been made possible by the participation of volunteers in the Etch A Cell project. Their contributions are acknowledged at <https://www.zooniverse.org/projects/h-spiers/etch-a-cell/about/results>.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Traffic published by John Wiley & Sons Ltd.

potential of EM, there is a great need to develop fast, generalizable and accurate analysis solutions. Although some EM image analysis can be automated through application of methods such as machine learning,<sup>4-8</sup> these advances have mainly benefited specific domains such as connectomics,<sup>9,10</sup> where the segmentation problem is focused on tracing neurons and synapses in serial images from brain and nerves. This focus has generated a large amount of 'ground truth' data that have been successfully used in deep learning to generate algorithms to automate the task.

The same cannot be said of cell biology, where the segmentation challenge is more diverse, encompassing common organelles such as the nucleus, nuclear envelope (NE), mitochondria, endoplasmic reticulum and endosomes, as well as rare or transient organelles such as autophagosomes, secretory granules and phase-separated entities. As in connectomics, the production of ground truth segmentations has, to date, relied on the effort of the expert EM community. At present rates of data generation, this community alone is unable to generate sufficient ground truth segmentation data, representative of the appearance of the full range of organelles in different experimental conditions and biological model systems. To enable data analysis at a scale beyond the capacity of the research community, we engaged the help of a global community of willing volunteers through a novel online citizen science project, 'Etch a Cell' (<https://www.zooniverse.org/projects/h-spiers/etch-a-cell>), which asked members of the public to manually segment the NE, which was targeted for volunteer segmentation as it is the most easily identifiable subcellular structure for which reliable automatic segmentation was not widely available.

The NE is a double lipid bilayer found in most eukaryotic cells where it surrounds the nucleoplasm and encloses the genetic material of the cell. Alterations in the structure of the NE have been associated with disease<sup>11</sup> including cancer<sup>12,13</sup> and nuclear laminopathies.<sup>14</sup> However, despite the clear critical role of the NE in cell function, the nanoscale three-dimensional structure of this organelle has been poorly understood to date. In addition to its biological importance, segmentation of the NE is often a critical first step in the segmentation of a cell, as this structure provides important context to the three dimensional spatial distribution of other organelles.

Here, we present our method for establishing a high-quality consensus segmentation from multiple volunteer annotations on the same image. We demonstrate that exclusively volunteer produced data can be used to train a machine learning model for highly accurate automatic segmentation of the NE. Finally, we present a novel multi-axis modification of our machine learning algorithm that resulted in a marked improvement in model performance. We share all benchmark data and algorithms produced for the use of the wider research community.

## 2 | RESULTS

### 2.1 | Etch A Cell: An online citizen science project for NE segmentation

An online citizen science project, 'Etch A Cell' (EAC), was developed to enable large-scale segmentation of the NE in volume EM

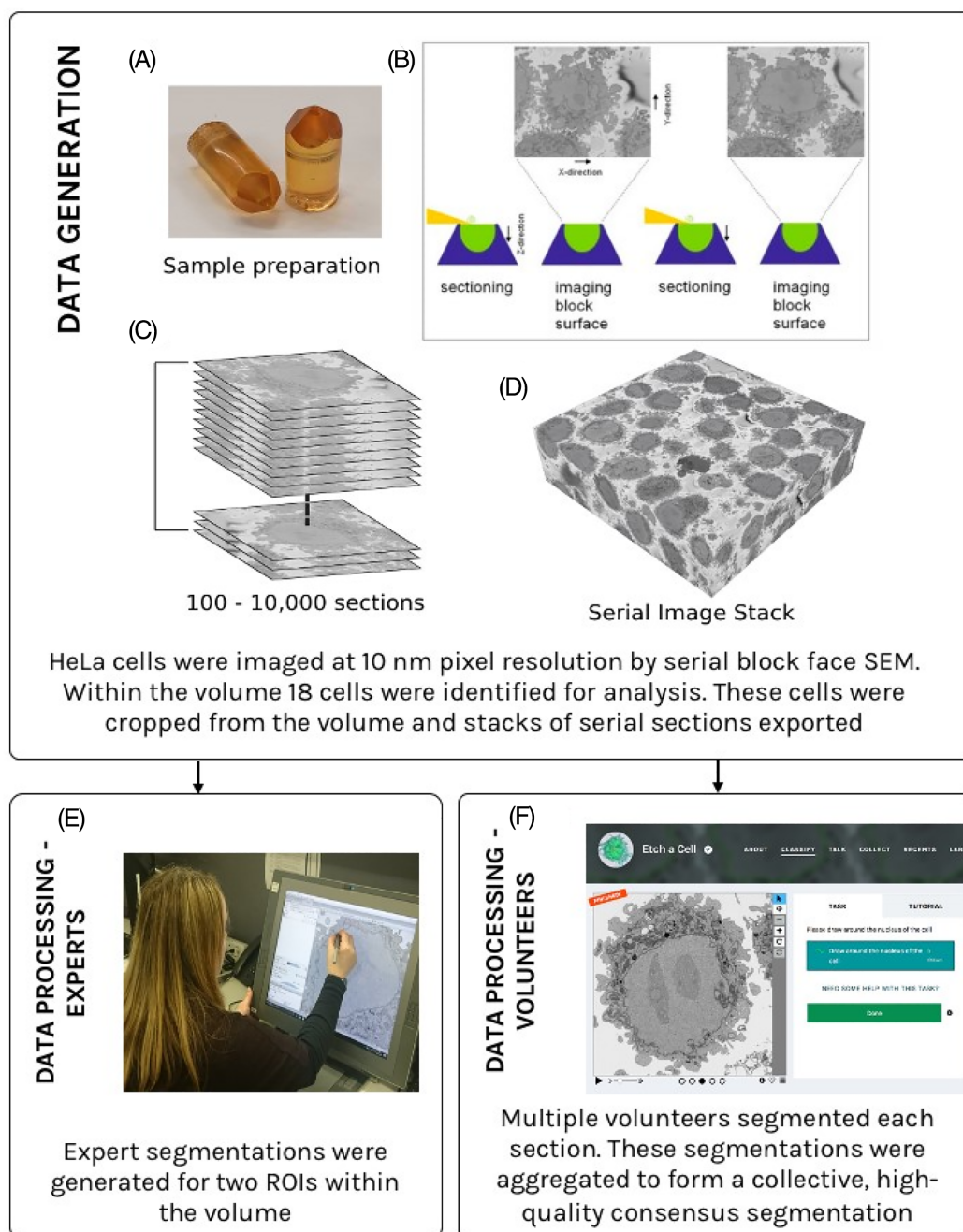
data through public engagement. Although online citizen science has been previously applied in similar contexts,<sup>15,16</sup> to our knowledge, this is the first application of non-expert, volunteer effort for the segmentation of organelles in EM data. To maximize the potential utility of the data produced for the research community, the commonly used HeLa cell line<sup>17</sup> was selected for analysis. A benchmark serial image data set was generated at 10-nm pixel resolution with SBF SEM (Figure 1 and Movie S1), and  $n = 18$  cells selected from this volume for volunteer segmentation (Table S1 details the unique cell ID assigned to each region of interest (ROI) and provides further descriptive information), resulting in a total of  $n = 4241$  slices for project inclusion after volume pre-processing (Figure S1AD and Section 4). Raw data have been made available via the EMPIAR repository (deposition ID: 137, accession code: EMPIAR10094, <https://www.ebi.ac.uk/pdbe/emdb/empiar/entry/10094/>).

### 2.2 | Development and deployment of EAC

Following an iterative design process lasting approximately 6 months (from October 2016 to the beginning of April 2017), EAC was launched on 6 April 2017 as a publicly available project on the Zooniverse online citizen science platform (<https://www.zooniverse.org/>). Volunteers could contribute segmentations through visiting an online classification interface (Figure S1E), where they were presented with a cell slice at random. A detailed tutorial provided task instructions (Figure S2), describing how to segment the NE with the novel 'Freehand Drawing Tool' and how to resolve segmentation ambiguities through viewing the two neighbouring slices either side of the target slice (Figure S1C-E and Section 4). Each slice was independently segmented by multiple volunteers with the number of classifications required set at  $n = 30$  (the 'retirement limit'). Each individual segmentation could consist of an arbitrary number of lines, which were recorded as an array of  $x,y$  pairs (Section 4, Figure S1F).

### 2.3 | An overview of volunteer interaction with EAC

In total,  $n = 104\,612$  classifications were submitted by volunteers before the EAC workflow was deactivated on 1 August 2019. As classifications could be made by unregistered volunteers, it was not possible to establish precisely how many individuals contributed, however, classifications submitted by logged in users were associated with  $n = 4749$  user IDs and  $n = 9444$  IP addresses, indicating between 5000 and 10 000 individuals contributed. As is often observed for online citizen science projects,<sup>18</sup> a large number of classifications were received shortly after project launch (Figure S3A), and the number of classifications submitted by each volunteer varied greatly (from 1 to 5451), indicating a broad range of engagement levels amongst the community contributing to this project (Figure S3B). Examining

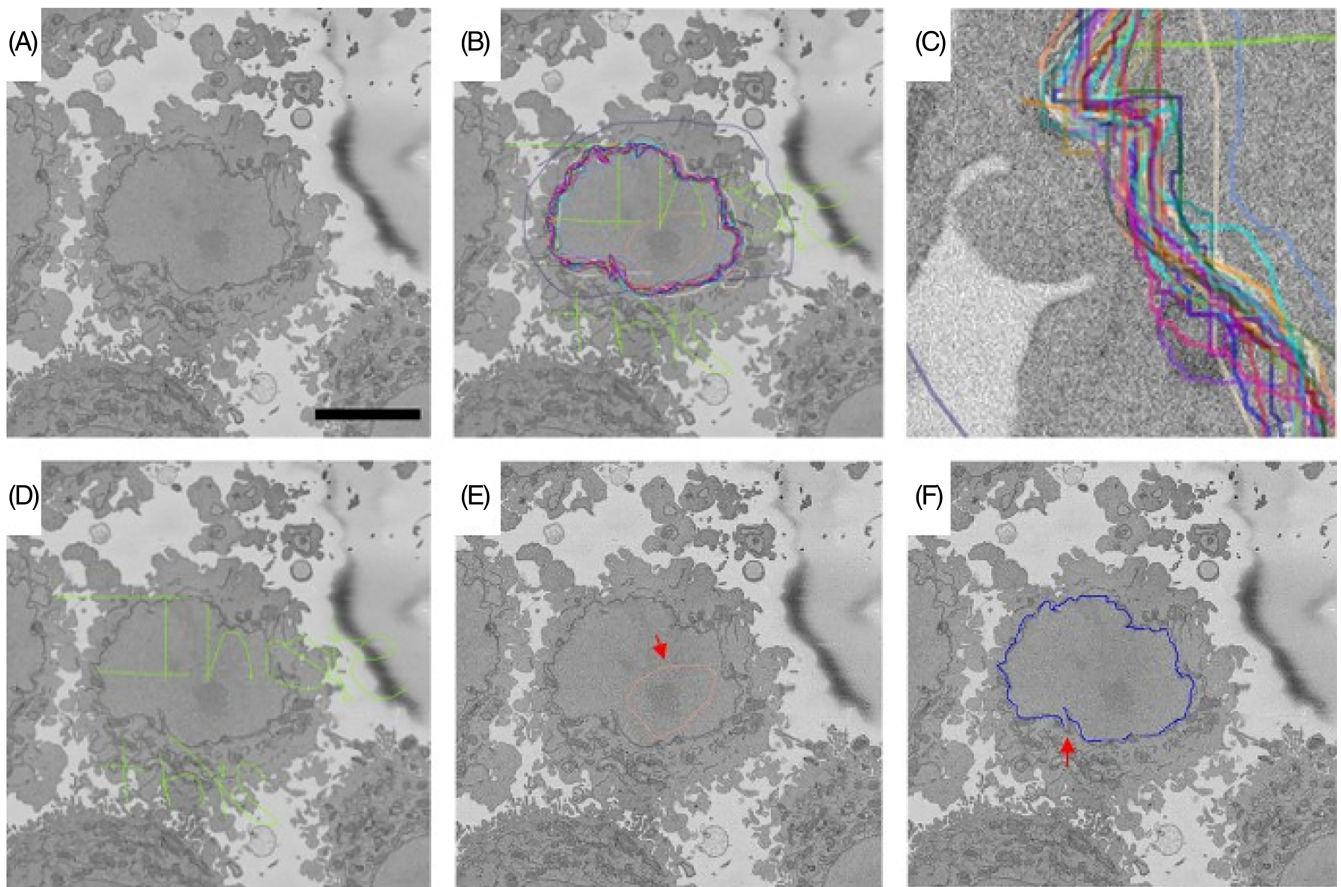


**FIGURE 1** Workflow for the acquisition and segmentation of serial EM images from benchmark samples. In this study we imaged resin-embedded HeLa cells at 10nm pixel resolution, A, using SBF SEM, B. This produced an image stack, C, of 518 sections (50 nm thickness,  $8192 \times 8192$  pixels, Movie S1) which were used to construct a 3D volume, D. ROIs from within this volume were segmented by both experts, E, and volunteers, F. Table S1 provides further information about the individual ROIs within the volume

the Lorenz curve for the distribution of volunteer classification contributions to the project (Figure S3C) and the corresponding Gini coefficient (0.83), indicated that a small number of volunteers contributed a large number of classifications, as is commonly observed in citizen science projects.<sup>18</sup> However, it should be noted that a significant proportion of the classifications submitted to the project were made by individuals only submitting a small number, reiterating the importance of all individual contributions.

## 2.4 | Forming consensus from multiple segmentations – aggregating volunteer annotations

To generate sufficiently high-quality data for downstream analyses, each individual slice within EAC was presented to multiple volunteers for segmentation. As expected, most volunteer segmentations were distributed on and around the NE (Figure 2A,C and Movie S2); however, distinct classes of segmentation error were observed, including



**FIGURE 2** Multiple segmentations were contributed by volunteers for each image in the EAC project. Each cell slice, A, was segmented by multiple volunteers [contributions from different volunteers shown in different colours, B, and zoomed in panel C]. The volunteer annotations vary in quality, with identifiable classes of error including; graffiti, D, ‘false-positive segmentation’ where non-NE is segmented, for example, a region within the NE, indicated with a red arrow, E, and ‘false-negative segmentation’, where NE pixels have been missed, indicated with a red arrow, F (Movie S2). For downstream analyses it is necessary to aggregate the multiple volunteer segmentations to establish a final ‘consensus’ NE segmentation for each slice. Panels are produced from slice number 70 from C001 (ROI 1656-6756-329) and 5  $\mu$ m scale bar is shown on panel A

‘graffiti’ (Figure 2D, possibly produced as an inadvertent consequence of well-intended classroom-based engagement using this project), ‘false-positive segmentation’ in which non-NE pixels are segmented (Figure 2E) and ‘false-negative segmentation’ where NE pixels are missed (Figure 2F). Of these error classes, the graffiti class was comparatively rare (Movie S2).

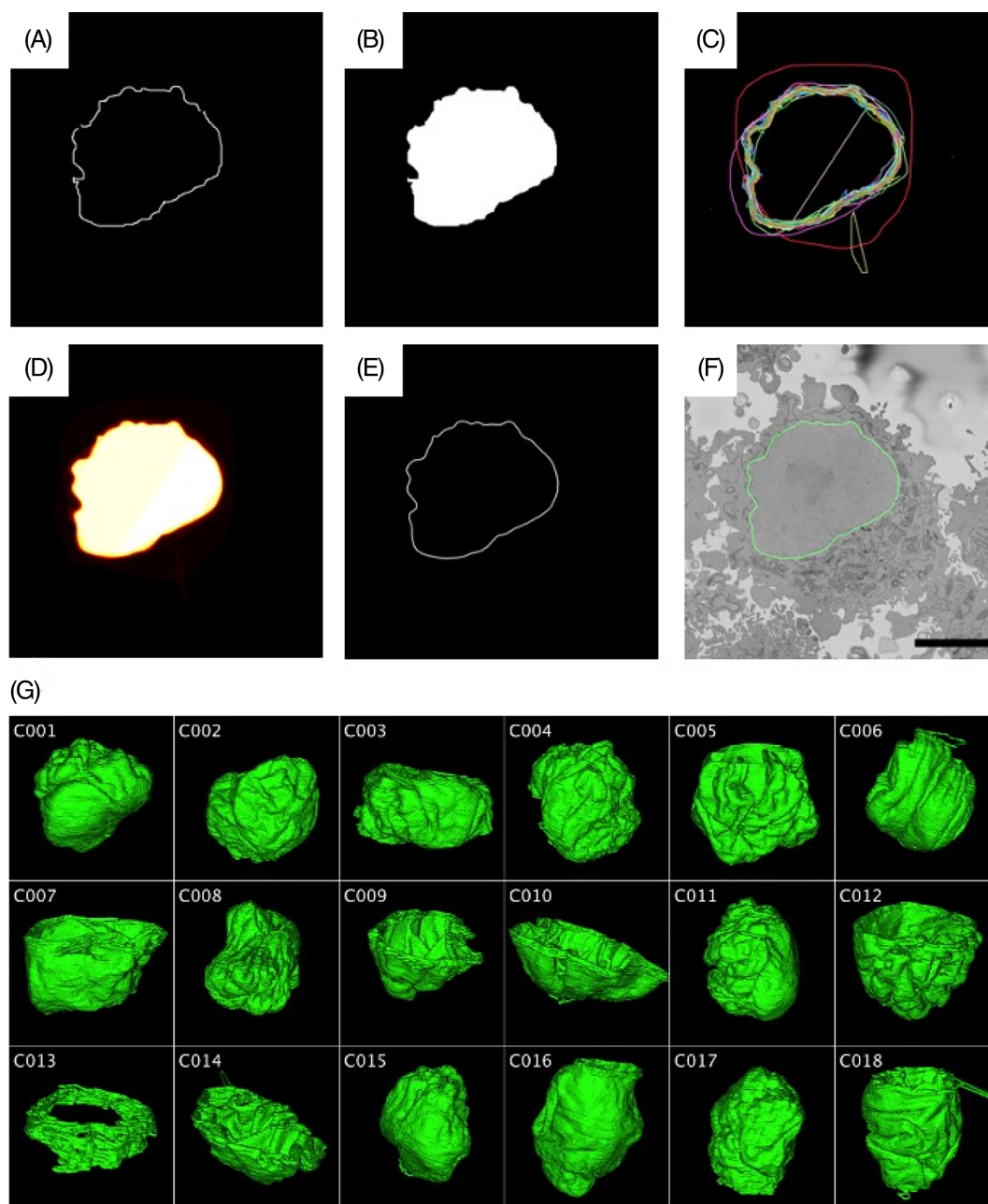
To remove outlying data and establish a ‘consensus’ segmentation for each slice, it was necessary to aggregate the multiple volunteer annotations. As the Freehand Drawing Tool was developed specifically for EAC, it was necessary to develop a novel aggregation approach. Because of the presence of noise, erroneous segmentations, and an unknown, variable number of line segments within the data, this was not trivial; hence, multiple novel aggregation approaches were developed and explored. Of these, the ‘CRIA’ algorithm was selected for our analytical pipeline, as it had a number of advantages compared with other approaches as shall be outlined.

Briefly, the CRIA algorithm procedure involved the following steps: first, closed loops were formed from each individual volunteer segmentation (Figure 3A–C), which could consist of

multiple separate lines (Figure S1F). The closed loops were produced through connecting separate lines after ordering them by minimizing distances. Next, interior areas were generated from the closed loops (Figure 3B). The interior areas were overlaid to generate a height map, with the ‘height’ reflecting the level of agreement between the separate volunteer segmentations for a single slice (Figure 3D). The consensus segmentation was determined through taking a mean ‘height’ level; hence the resulting, ‘final’, segmentation surrounds all the interior areas where half or more of the volunteer segmentations were in agreement (Figure 3E,F). This procedure was used to aggregate all volunteer segmentations (Figure 3G, Movie S3). The resulting aggregated data have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/Aggregations> and the CRIA code is available at <https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>.

In addition to producing high-quality consensus segmentations (Figure 3F,G) that showed high visual similarity to expert data (Figure 4A,B,D,E,G,H, Movies S4 and S5), this aggregation procedure

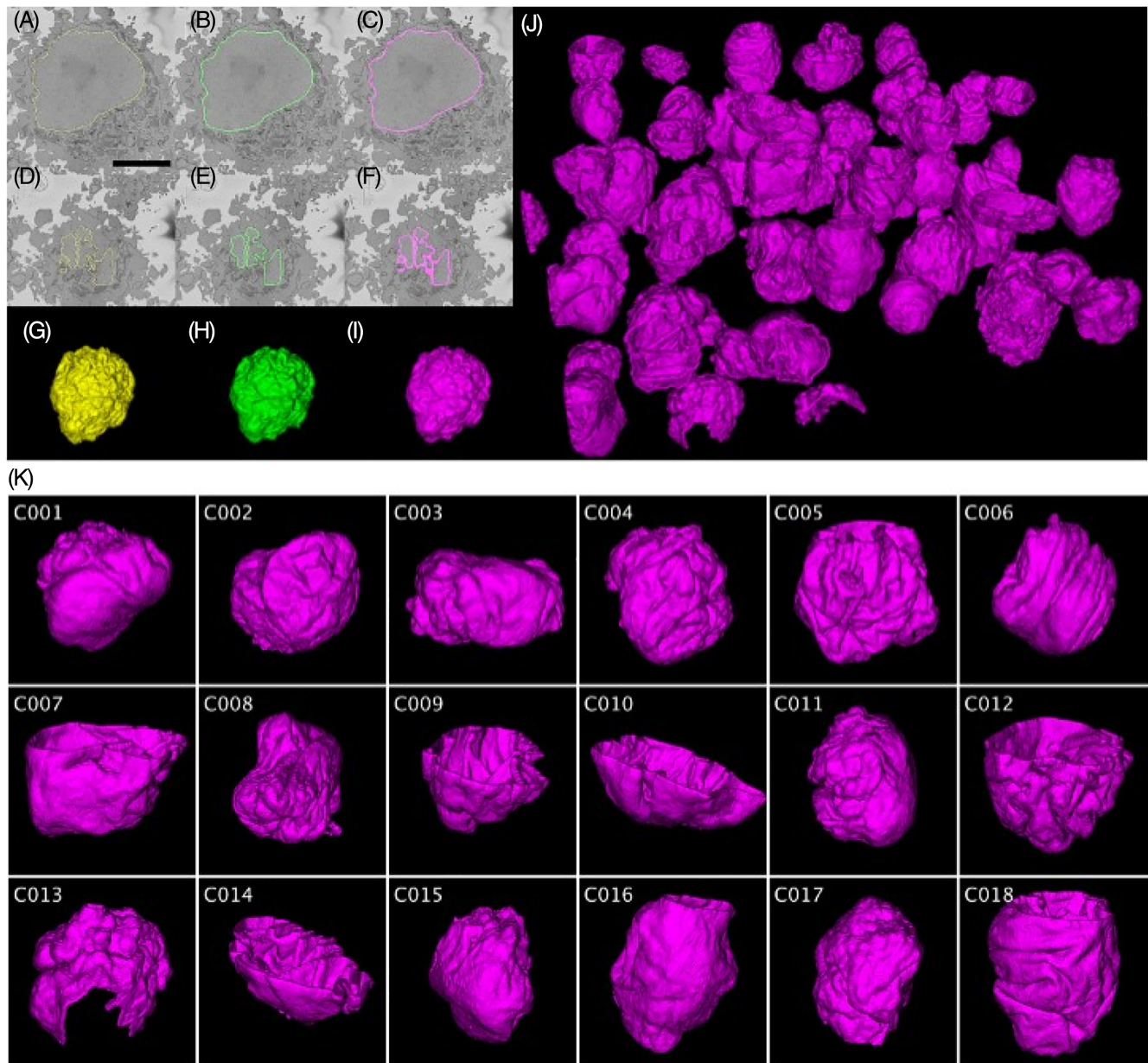




**FIGURE 3** The CRIA algorithm was developed for the aggregation of multiple volunteer segmentations. In this algorithm, each individual volunteer segmentation, A, was converted into a closed loop, B. This procedure was performed for all the segmentations associated with each slice of the ROI, as can be seen stacked in C. The closed loops were converted to interior areas and stacked, D. A final, consensus segmentation was determined as the outline of all interior areas where half or more of the volunteer segmentations were in agreement, E. This generated a high-quality, volunteer-produced segmentation (5µm scale bar), F. We show here the annotations and aggregation for slice number 150 from C001 (ROI 1656-6756-329). This process was applied to each slice of all  $n = 18$  volunteer-segmented ROIs, allowing generation of a 3D reconstruction of each ROI, G (Movie S3)

had a number of notable benefits. Importantly, the CRIA algorithm made use of all volunteer annotations in producing the final consensus segmentation, therefore no volunteer effort went unused. In contrast to other methods explored, the CRIA algorithm is fully automated and required no expert involvement, such as manual selection of high-quality segmentations to seed the algorithm. This is a critical advantage for a number of reasons; avoiding a requirement for manual

intervention minimizes the possibility of perturbing the final, consensus, segmentation through introducing subjectivity and bias associated with a single individual. Furthermore, eliminating the need for expert assessment removes a significant analytical bottleneck. Finally, and perhaps most notably, this approach has shown high-quality segmentations can be generated solely through collective non-expert effort.



**FIGURE 4** Consensus volunteer and machine-predicted NE segmentations are high quality. Visual inspection reveals a high similarity between expert, A, aggregated volunteer, B, and machine predicted, C segmentations [shown for slice number 150 from C001 (ROI 1656-6756-329)], and a high degree of overlap of these segmentations with the NE. Segmentations from slices found at the top and bottom of the volume (D, E, F) showed greater segmentation variability due to the presence of NE islands and membrane parallel to the cutting plane, which make these regions more challenging to segment [shown for slice number 40 from C001 (ROI 1656-6756-329)]; 5  $\mu$ m scale bar is shown on panel A. Despite this, 3D reconstruction of nuclei revealed a high similarity between expert, G, volunteer, H, and machine, I, segmented nuclei [shown for C001 (ROI 1656-6756-329)] (Movies S4 and S5). Automatic NE segmentation using our trained model applied to our full data volume captured many nuclei which had not previously been segmented through expert or volunteer effort, J (Movie S9), and the  $n = 18$  nuclei previously segmented by volunteers, K (Movie S10). Machine-predicted segmentations were produced with TAP

## 2.5 | Machine learning for NE segmentation

Aggregated volunteer NE segmentations were used to train a U-Net convolutional neural network (CNN) architecture<sup>19,20</sup> for automatic segmentation of the NE in SBF SEM data. Model performance was assessed through presenting the model with two previously unseen

ROIs, and comparing the resulting predicted NE segmentations with 'ground truth' (Table S1 and Section 4 provide further information about ROIs used for model training, validation and testing).

Two complementary forms of 'ground truth' data were available; expert generated segmentations (available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/Expert>) and aggregated volunteer

**TABLE 1** CNN performance metrics. Model performance was assessed by comparing the predicted NE segmentation to ground truth for two ROIs [ROI 1656-6756-329 (cell ID = C001) and ROI 3624-2712-201 (cell ID = C006)]. We report multiple metrics of model performance (Section 4) against two complementary modes of 'ground truth' data available (aggregated volunteer data and expert-produced segmentations)

Cell ID	ROI	F-measure	Recall	Precision	AHD (pixels)	AHD (nm)	F-measure area (single slice)
<i>Prediction compared to aggregated volunteer data</i>							
C001	ROI_1656-6756-329	0.700	0.792	0.628	1.638	16.377	0.995
C006	ROI_3624-2712-201	0.687	0.767	0.621	1.767	17.675	0.991
<i>Prediction compared to expert data</i>							
C001	ROI_1656-6756-329	0.340	0.697	0.225	3.129	31.287	0.993
C006	ROI_3624-2712-201	0.375	0.697	0.256	3.890	38.904	0.9924

segmentations (Section 4, Table S1 and Figure 4) providing a means to test two facets of model performance. In comparing the prediction with the aggregated volunteer data for each ROI, we were able to establish how well the model had learnt to perform the task of NE segmentation from the training data provided, which consisted of exclusively volunteer produced segmentations. Comparing model performance to expert data enabled assessment of how well the model (hence, indirectly, the volunteers) performed this task in comparison to experts.

The model performed well when compared to aggregated volunteer data. The average Hausdorff distance (AHD) between the predicted segmentation and the aggregated volunteer segmentation was 1.638 pixels (corresponding to a distance of 16.377 nm) for C001, and 1.767 pixels (17.675 nm) for C006. The F-measure, recall and precision of the model were 0.700, 0.792, 0.628, respectively, for C001 and 0.687, 0.767, 0.621 for C006 (Table 1). Although these metrics may initially seem poor in comparison to similar previous work,<sup>21,22</sup> it should be emphasized that we are examining the overlap between lines (the NE) rather than areas (the nucleus). Hence, for easier comparison of our model with previously reported metrics, we also provide the F-measure for the nucleus area for a single slice within each ROI. As anticipated, this metric shows a much higher model performance of 0.995 (C001) and 0.991 (C006).

Reflecting differences in volunteer and expert segmentation skill, it was expected that we would see reduced model performance (trained exclusively using volunteer produced data), when comparing against expert-produced 'ground truth' data. We observed an AHD of 3.129 pixels (31.287 nm) for C001 and 3.890 pixels (38.904 nm) for C006 between the prediction and expert data. The F-measure, recall and precision were 0.340, 0.697, 0.225, respectively, for C001, and 0.375, 0.697, 0.256 for C006 (Table 1). Although most of these metrics indicate good model performance, the F-measure and precision warrant further explanation. These measures are particularly poor in the case of comparing the model to the expert data due to an idiosyncrasy of the expert data. The width of the available expert data is narrower (30 nm) compared to both the aggregated and predicted width of the NE (70 nm), and because of this we see a degradation of the precision and F-measure metrics. This is because the model has assigned pixels as NE that do not correspond with pixels annotated by the expert, therefore, the false-positives rate is seemingly inflated. Unfortunately, it was not feasible to amend our expert ground truth

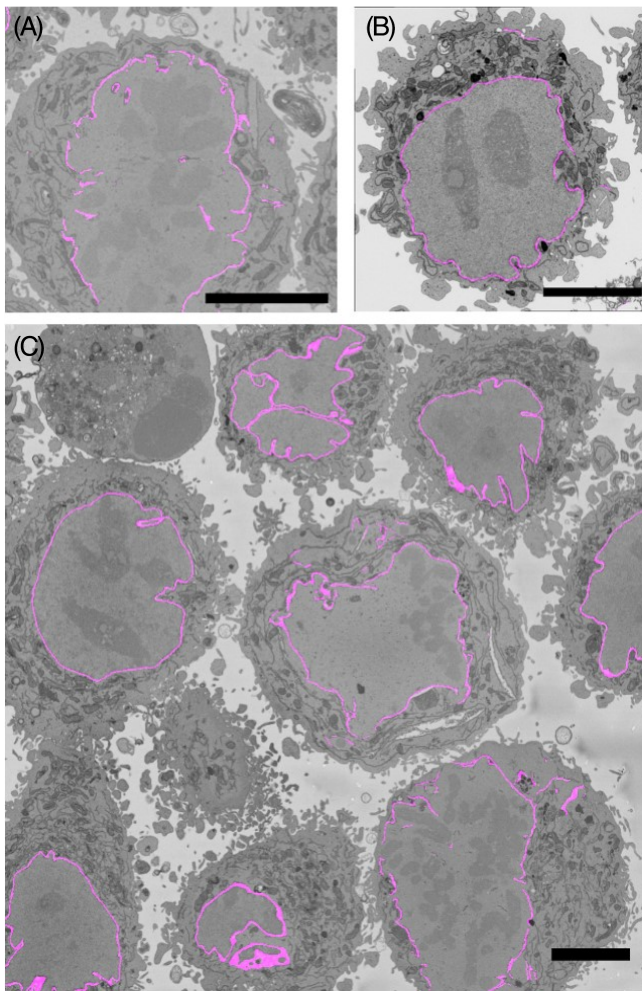
through either asking an expert to resegment the ROIs (this was not practical due to time constraints) nor was it recommendable to dilate the width of the expert segmentation (as this would introduce greater errors, e.g. incorrectly assigning cytoplasm pixels as NE). Despite this, upon visual inspection, it was found that the model performance was arguably superior to the expert segmentation as more relevant pixels appeared to be assigned to the NE by the model (Figure 4), which raises questions regarding the legitimacy of 'ground truth' data produced by a single expert, as shall be discussed later.

## 2.6 | Improved model performance with tri-axis prediction

Although the model performed well, expert visual inspection revealed some regions of under-segmentation (Figure S4A). These regions were not randomly distributed across the data, but were instead localized to sites at the top and bottom of the volume (the highest and lowest z slices, Figure S4B,C), presumably due to the higher degree of visual ambiguity in these regions caused by the presence of a greater number of NE islands and the membrane being oriented parallel to the SBF SEM imaging plane. To improve the automated segmentation, we sought to leverage additional information available in the volume. The data examined here were downsampled in the xy plane to 50 nm to be isotropic, therefore, it was possible to transpose the stack and run the model on each axis (Figure S4D-F). This resulted in three orthogonal NE predictions which were recombined to form a final segmentation, with pixels assigned to NE in all three predictions accepted (Figure S4G, Movies S6 and S7) and over-segmented pixels removed using a connected components analysis (Figure S4H). Visual inspection revealed a significantly improved segmentation (Figure S4I); however, it was not possible to quantify this improvement due to a lack of appropriate ground truth data.

The tri-axis prediction (TAP) approach was applied to the entire volume (Figure 4, Movies S1, S8 and S9), and took a total of 48 minutes to produce NE predictions for all nuclei within the volume (Figure 4J), including the  $n = 18$  nuclei already segmented with volunteer effort (Figure 4K and Movie S10). TAPs have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/Aggregations>. Serendipitously, a cell within the volume was undergoing mitosis,





**FIGURE 5** Model shows high NE segmentation performance in novel contexts. Applying TAP to the full data volume allowed us to observe that the model performed well when segmenting the partially broken down NE of a mitotic cell, A (Movie S11). TAP also showed good performance when applied to the same resin-embedded sample (which also contained both mitotic and interphase cells) imaged at higher resolution (5 nm) on the same microscope, B (Movie S12) and when applied to a HeLa cell from the same sample imaged by an alternative volume EM methodology, FIB SEM, C (Movie S13). All scale bars are 5  $\mu$ m

allowing us to observe that our model performed well in this challenging context in which the NE had partially broken down, despite not having been exposed to training data of this type (Figure 5A and Movie S11). This is in contrast to some other approaches for NE identification which rely on the presence of a clear boundary, such as flood or marker based watershed methods.<sup>23</sup> The ability of the algorithm to segment disassembled mitotic NE is particularly surprising given the NE effectively regresses to become ER during mammalian cell division. Further analysis of the features identified by the model may be useful in defining the transition of the NE to the ER and back during mitosis. TAP was also applied to an alternative region from the same resin-embedded sample imaged at higher resolution (5 nm) on the same microscope (which also contained both mitotic and

interphase cells, Figure 5B and Movie S12), and to a HeLa cell from the same sample imaged by an alternative volume EM methodology (FIB SEM) (Figure 5C and Movie S13). Visual inspection of these data sets showed good model performance indicating the model is generalizable to novel contexts; however, it should be acknowledged that some erroneous over-segmented pixels can be observed, particularly in the peripheral ER and edges of lipid droplets bordering the nuclear region in mitotic cells (Figure 5B and Movie S12), indicating there is scope for future improvement.

To provide a comparison of the approach presented here with an alternative, commonly used, image processing technique, we applied edge detection (Section 4) to a single ROI from the volume [cell ID = C001 (ROI 1656-6756-329)]. As expected, edge detection was found to be capable of detecting the NE reasonably well; however, it was unable to isolate this structure from the many other membranes present in the cell, such as the mitochondria and endoplasmic reticulum (Movie S14). The additional spatial context provided by the U-Net architecture allows the model to distinguish between similar edges based on the 3D structures of their neighbourhoods, and as outlined, the TAP modification enables segmentation of indistinct edges resulting from the membrane being aligned in the cutting plane.

Finally, considering the overarching objective of this research was to facilitate faster analysis of volume EM data, we calculated the time benefit of our predictive model. For an expert to manually segment a single ROI would have taken approximately 30 hours.<sup>24</sup> Prediction took approximately 1 minute per image stack for a single ROI, therefore, our model prediction represents 1800 times faster data analysis.

### 3 | DISCUSSION

We show here that volunteer effort through online citizen science can be effectively applied to the task of segmentation of organelles in electron micrographs, enabling data analysis at a scale not achievable by experts alone. We demonstrate the data produced is of sufficient quality for task automation through training a CNN capable of segmenting the NE at high accuracy. Although prior work has shown crowdsourced volunteer effort can be productively applied to comparable tasks, such as the marking of single particles from cryo-EM micrographs to generate 3D protein reconstructions,<sup>15</sup> and to the marking of whole cells,<sup>16</sup> to our knowledge this is the first study to demonstrate the ability of volunteers to effectively perform manual freehand segmentation of an organelle in volume EM data.

Such large-scale, systematic segmentation makes quantitative examination of organelle morphology feasible. This has the potential to drastically advance our understanding of NE morphology and function, in both normal and diseased states such as cancer<sup>12,13</sup> and nuclear laminopathies.<sup>14</sup> Yet, even with the collaboration of a community of citizen scientists it will not be possible to segment data at a scale proportional to current data production rates, and this challenge will become greater with further technological advancement. Hence, we sought to automate NE segmentation using volunteer produced segmentations as training data for a CNN,<sup>19</sup> resulting in a model able



to segment the NE to a high standard in a matter of minutes, rather than the hours, days or weeks required for manual segmentation. Critically, our model was trained exclusively with volunteer segmentations and required no expert microscopist input or intervention.

Although our model performed surprisingly well when applied to data produced under different conditions, there was scope for improved performance. Hence, we remain far from our aspiration of a highly-accurate, yet broadly applicable approach for the automated analysis of microscopy data for a single organelle, let alone each feature of interest in every volume acquired. It is anticipated that future work in this arena may be accelerated through application of approaches including transfer learning<sup>25</sup> and multiclass predictive models.<sup>26</sup> However, unless significant advances are seen in unsupervised machine learning approaches, it is expected that an adequately trained generic model will still require a large quantity of ground truth annotations, covering each organelle from a wide variety of imaging conditions and different cell types. Clearly, the lack of suitable training data remains a significant barrier to the development of a broad automated approach; however, as we have demonstrated here, large training data sets can be produced through volunteer engagement.

Many avenues exist for the extension of this research, including segmentation of other organelles, examination of data from other imaging modalities and the analysis of data from further cell lines or tissue types. We expect that such future projects would be significantly expedited by the workflows and analyses established in this first iteration of EAC. However, consideration will be required to effectively apply the methods and tools presented here to novel contexts. For example, it is anticipated that many of the organelles to be targeted in future citizen science projects, such as the mitochondria and endoplasmic reticulum, will likely display greater structural heterogeneity than the NE. We expect that this will pose two key challenges in the context of crowdsourcing segmentations through online citizen science. Firstly, this structural variability will impact the difficulty of manual segmentation of different organelles under different conditions, we will therefore need to refine and modify novel projects accordingly. For example, in designing our second EAC project for the segmentation of mitochondria ('EAC - Powerhouse Hunt' <https://www.zooniverse.org/projects/h-spiers/etch-a-cell-powerhouse-hunt>) it was necessary to adjust the field of view presented to the volunteer to ensure a reasonable number of mitochondria of an appropriate size were presented for segmentation. Other design modifications included providing a broad selection of mitochondrial image examples in project training materials. For organelles that can only be unequivocally identified by functional markers (e.g. autophagosomes labelled with green fluorescent protein) other study design adjustments will need to be explored such as presenting correlative light images in conjunction with EM images to guide segmentation. Beyond study design modifications, in such challenging instances it may also be possible to achieve higher data quality through engaging the help of a greater number of volunteers in providing segmentations for each image. Secondly, it is anticipated that the likely increased structural heterogeneity of future target organelles will require adjustment of our

aggregation pipeline, for example, to enable the aggregation of many independent structures within a single ROI. However, preliminary work to aggregate mitochondrial segmentations indicates that this, while being an important consideration, is not insurmountable: the analytical pipelines presented here may require modification; however, they do not require complete reinvention.

Although challenging from a study design perspective, the possibility of designing a portfolio of projects of varying difficulty provides a rich opportunity to engage volunteers through serving a greater variety of skills and interests. Reassuringly, citizen scientists have proven capable of performing a growing array of challenging tasks, from identifying supernovae<sup>27</sup> to visually assessing the quality of brain registration in functional magnetic resonance imaging studies.<sup>28</sup> We have been astonished that it is possible to train non-experts to recognize and segment complex organelles in minutes with just an online tutorial. We therefore remain confident in the abilities of our volunteer community to successfully perform novel segmentation tasks.

The challenge of motivating increased engagement and high-quality contributions will become increasingly important as our repertoire of citizen science projects expands and diversifies. Manual segmentation is a challenging task requiring a large time investment. We must therefore continue to develop novel modes of engaging our community and work to reduce the effort required to segment each slice. Mechanisms to achieve reduced volunteer effort per slice may include 'smart subject assignment'<sup>29,30</sup> – the intelligent passing of slices of appropriate difficulty to volunteers. Furthermore, it may be possible to actively retire slices from the project once an acceptable segmentation quality has been achieved – this would enable volunteer effort to be reduced for 'easier' images, and to instead be applied to images of greater difficulty.

Incorporating 'computers in the loop' may provide additional mechanisms for reducing segmentation effort. Future pipelines may include presenting volunteers with predicted segmentations for correction, rather than full segmentation. Feedback loops between computer-prediction and crowdcorrection could enable real-time model refinement, improve predictions and therefore progressively reduce need for volunteer correction, resulting in greater project efficiency. Predictive models need not be fully optimized to be useful; if a model is not yet able to accurately segment its target organelle, it may still provide valuable information that could be fruitfully leveraged, for example, the anticipated number of a particular organelle class and their approximate location. This would provide a mechanism for assessing volunteer ability, segmentation quality and subject difficulty.

We have demonstrated that experts can be removed from the task of manual segmentation, however, researcher time remains necessary to generate the infrastructure supporting this effort and for the continued refinement of multiple aspects of the analytical pipelines underlying these studies. More critically, researcher effort continues to be needed to interpret and assess the quality of volunteer or machine produced segmentations. This is a particularly challenging component of this work, as the required quality of a 'final segmentation' is often intimately linked to the research question being addressed. Related to the ubiquitous challenge of finalizing segmentations and

establishing 'truth' in this domain: it can be difficult to definitively assign pixels of noisy and nuanced micrographs to different regions, and much inter-expert (and intra-expert) variation can exist. The potential for pixel-assignment disagreement raises an interesting possibility regarding additional value of collectively producing segmentations; when multiple individuals annotate each slice, rather than a single expert, it is possible to produce a level of confidence that each pixel belongs to a certain region, rather than simply a binary designation. Such a segmentation-confidence map may be more reflective of the reality of cell morphology, where a subset of pixels may not definitely belong to a particular region. This may provide insight, with regions of variable confidence being of possible biological relevance, for example, we may expect nuclear pores in the NE to be less confidently designated as this organelle.

Future collaboration of the crowd and computing is poised to enable, for the first time, the large-scale, generalizable, yet accurate, quantification of multiple subcellular structures across many data modalities at the nanoscale.

## 4 | MATERIALS AND METHODS

### 4.1 | Cell model

HeLa cells were cultured in large Petri dishes in DMEM. Prior to fixation, the cells were dissociated from the dishes using trypsin, pelleted, and briefly washed in DMEM. Fixation was carried out by first resuspending the cells in an equal quantity of double-strength fixative [5% glutaraldehyde and 8% formaldehyde in 0.1 M phosphate buffer (pH 7.4), for a working concentration of 2.5% glutaraldehyde and 4% formaldehyde] for 15 minutes at room temperature, before pelleting and resuspending in single-strength fixative for 30 minutes at room temperature, and 30 minutes at 4°C. The cells were then embedded as a pellet in Durcupan resin following the method of the National Centre for Microscopy and Imaging Research.<sup>31</sup>

### 4.2 | Data acquisition

SBF SEM data were collected using a 3View2XP (Gatan, Pleasanton, California) attached to a Sigma VP SEM (Zeiss, Cambridge). Small portions of the cell pellets were mounted on pins using conductive epoxy resin (Circuitworks CW2400), trimmed to form an approximately 400 × 400 × 150 µm pillar, and coated with a 2 nm layer of platinum. Images were acquired at 8192 × 8192 pixels using a dwell time of 6 µs (10 nm reported pixel size, horizontal frame width of 81.99 µm) and 50 nm slice thickness. The SEM was operated at a chamber pressure of 10 pascals, with high current mode inactive. A 20-µm aperture was used with an accelerating voltage of 3 kV. Raw data consisted of a total of 518 images acquired sequentially, representing a depth of 25.9 µm and total volume of 17 4135 µm<sup>3</sup> (10 nm data set, Figure 1 and Movie S1). One image was excluded from analysis due to a technical fault resulting in loss of the cut material and the

production of a blank image. To enable further analyses and benchmarking, raw data have been made available via the EM-PIAR repository (deposition ID: 137, accession code: EMPIAR-10094, <https://www.ebi.ac.uk/pdbe/emdb/empiar/entry/10094/>). Digital micrograph (DM4) files were read into Fiji with the Bio-Formats library<sup>32</sup> and subsequently saved as TIFFs.

### 4.3 | Expert production of ground truth data

Ground truth segmentations for two ROIs were obtained by manual annotation in the Amira software package.<sup>33</sup> The two cells segmented by expert effort were cell ID = C001 (ROI 1656-6756-329) and cell ID = C006 (ROI 3624-2712-201) (Table S1). Expert segmentations have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BST448/Expert>.

### 4.4 | Volume pre-processing to produce images for online citizen science

Approximately, 40 cells were visualized within the full volume (Movie S1). Of these, cells with nuclei not intersecting the edge of the field of view in any slice were manually selected for study inclusion. This ensured each 2D slice presented to volunteers contained a complete NE, simplifying the labelling of this structure and reducing the likelihood of mis-identification. A total of  $n = 18$  appropriate cell volumes were selected using this criterion. Each selected cell was cropped from the full volume (Figure 1A-C) and exported as a sequence of TIFF images using the following procedure within Fiji software; a 3D ROI was created around each selected cell, with a size of 2000 × 2000 pixels in x,y. The number of slices (synonymous with 'subjects' in Zooniverse terminology) per cell ranged from 150 to 300 due to inherent differences in cell size and variability in cell completeness across the z-axis in the data volume (12 of the 18 cells were not complete in the z-axis, Table S1 and Movie S1). To ensure reasonable web browser download times, the raw images for each ROI were downsampled to 1000 × 1000 pixels in x,y, and converted from digital micrograph DM4 format to JPEG with quality of 90% to reach an image size of approximately 600 kb using the Fiji software package<sup>34</sup> with the Bio-Formats tool.<sup>32</sup>

### 4.5 | Development of EAC with the Zooniverse Project Builder

Cell images were presented to volunteers for NE segmentation via an online citizen science project, 'EAC'. This project was designed and deployed on the Zooniverse platform (<https://www.zooniverse.org/>) using the Project Builder (<https://www.zooniverse.org/lab>). The Zooniverse Project Builder is a free, web-browser-based toolkit that provides the core infrastructure necessary for designing, building and implementing all components of an online citizen science project,

including the project workflow and supporting materials such as the project tutorial (Figure S2). Project development took approximately 6 months, during this time the project workflow was designed and all supporting materials produced, including an in-depth tutorial to comprehensively explain the NE segmentation task (Figure S2). Prior to launch, the project was refined through a multi-step review process involving thorough assessment of the project by both Zooniverse volunteers and the Zooniverse research team. For key Zooniverse terms, refer to <https://help.zooniverse.org/getting-started/glossary/>.

#### 4.6 | EAC workflow design

The pre-processed slices were uploaded to EAC for volunteer segmentation using a python script (<https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>) that interfaced with the Zooniverse Panoptes API (<https://github.com/zooniverse/Panoptes>).

Slices were embedded within the project 'workflow'. The 'workflow' of a Zooniverse citizen science project refers to the series of tasks a volunteer is asked to complete when presented with data in the project's classification interface.

In the EAC project workflow, upon being presented with one of the uploaded cell slices at random, volunteers were asked to perform the task of segmenting the NE using a Freehand Drawing Tool applied directly to the image in a web browser (Figure S1E). Upon submission of the classification, the individual lines drawn by the volunteers were recorded as arrays of x,y pairs defining a line path (Figure S1F).

To support volunteers in the task of NE segmentation, a detailed project tutorial was provided on the classification interface (Figure S2). To enable more accurate annotation, pan and Zoom functionality was enabled. Furthermore, to provide the volunteer with a limited amount of three-dimensional context to help resolve segmentation ambiguities, the image to be segmented was presented as the central image within a 'flipbook' of five images, with the two neighbouring images on either side corresponding to the  $\pm 250$  and 500 nm planes in the z-dimension (Figure S1D,E).

To produce data of sufficient quality for downstream analyses, each individual image was presented to multiple volunteers to enable the generation of a 'consensus' from the aggregation of multiple annotations. The minimum number of required annotations is denoted the 'retirement limit' and was set at  $n = 30$  for this project. Therefore, each individual image received at least 30 volunteer segmentations. A small subset of images received more than this, as a small number of images continue to be presented to volunteers in the project classification interface after all available data have been segmented.

#### 4.7 | Citizen science data export

The EAC project workflow examined in this manuscript was deactivated on 1 August 2019 and the project data exported from the data exports page of the Zooniverse Project Builder as a comma separated value (CSV) file. A single classification could

consist of an arbitrary number of individual lines (a line being defined as a single continuous stroke annotation). Each segmentation is recorded in a JavaScript Object Notation formatted string within the annotation field of the CSV file. Each line within the annotation is represented as a series of (x,y) pairs defining a line path (Figure S1F). The unaggregated segmentation data have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/etch-a-cell-classifications.csv>.

#### 4.8 | Data aggregation with CRIA

Multiple volunteer segmentations were produced for each slice uploaded to the EAC project. It was therefore necessary to remove outlying data and establish a 'consensus' segmentation for each slice. The CRIA algorithm was developed to aggregate the volunteer segmentations. In this approach, first, each individual volunteer segmentation was converted into a closed loop. This procedure was performed for all segmentations associated with each slice of the ROI. Next, these closed loops were converted to interior areas and stacked. A consensus segmentation was determined by taking the outline of all interior areas where half or more of the volunteer segmentations were in agreement (Figure 3A,F). This procedure was repeated for every slice within each ROI. Aggregated volunteer data have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/Aggregations>. CRIA code is available at <https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>.

#### 4.9 | Model architecture

A U-Net CNN architecture<sup>19,20</sup> was trained with aggregated volunteer segmentations for the automatic segmentation of the NE. This architecture uses convolutional layers and an autoencoder-style compression path. Hyper-parameter optimization was performed through random search, resulting in selection of the following model parameters: patch size of (12, 256, 256), dropout rate of 0.3, 32 start filters, Adam optimizer with an initial learning rate of 0.0005, and batch normalization. Informed by the expected and visible width of the membrane in the data (60-80 nm), a NE width of 70 nm was selected.

Training data for the model consisted of the aggregated volunteer segmentations. A voxel resolution of 50 nm was selected. Table S1 details the ROIs used for model training, validation and testing. Two ROIs were selected for model validation: cell ID = C005 (ROI 2052-5784-112) and cell ID = C010 (ROI 3588-3972-1). Two ROIs were used for model testing: cell ID = C001 (ROI 1656-6756-329) and cell ID = C006 (ROI 3624-2712-201); these two ROIs had been manually segmented by both expert and volunteers. One ROI (cell ID = C013, ROI 1716-7800-517) was excluded from training due to having received insufficient data from the citizen science to perform data aggregation. Using a local high performance compute cluster (Section 4, Nvidia Tesla V100-SXM2-32GB) to train this model took

approximately 4 hours (1 hour for data pre-processing; 100 seconds per epoch = 3 hours in total).

The loss function used for the model was the smoothed dice coefficient (or F-measure), where

$$\text{Dice loss} = 1 - \text{smoothed dice coefficient}.$$

#### 4.10 | Model performance metrics

A commonly applied approach to assess the quality of a model prediction vs ground truth for image data is to directly map the pixels between the two images. We report the F-measure, which is similar to the Dice coefficient. This measures the coincidence of predicted cell membrane to ground truth membrane. The F-measure of the nucleus area (as opposed to the NE) is also reported to enable easier comparison with previous work. As this metric requires a closed area, this was performed on a qualified single slice near the centre of the cell. Finally, we also report the AHD, in both pixels and nm, between the predicted NE and the position of the ground truth NE. This metric takes an average of all minimal distances between pixels in the prediction ( $P$ ) and ground truth ( $G$ ):

$$\text{AHD}(G,P) = \frac{1}{\|G\|} \sum_{g \in G} \left\{ \min_{p \in P} \{d(g,p)\} \right\}$$

#### 4.11 | Tri-axis prediction

A multi-axis modification of our machine learning model was implemented to improve performance (Figure S4). Data were down-scaled in the xy plane to 50 nm to be isotropic, and the stack transposed to run the model over each axis (Movies S6 and S7). The resulting three orthogonal NE predictions were recombined to generate a final segmentation. All pixels assigned to NE in all three predictions were accepted, and connected components analysis was to remove over-segmented pixels (Section 4). This approach was named 'TAP'. TAP results for each ROI have been made available at <http://www.ebi.ac.uk/biostudies/files/S-BSST448/Predictions>. TAP code is available at <https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>.

#### 4.12 | Post-processing of 3D volumes

Connected components analysis<sup>35</sup> was implemented as a post-processing step at multiple points within the analyses presented. Within the single-axis implementation of the machine learning model, connected components analysis was used to remove objects below a threshold of 10 000 voxels. This threshold was selected as it resulted in the removal of small areas of erroneous over-segmented pixels,

while legitimate membrane was preserved due to its comparatively large size. Connected components analysis was also used to isolate the predicted NE segmentation for the target nuclei within each ROI, to discard any predicted NE associated with peripheral cells potentially present. In the TAP modification of the machine learning model, connected components analysis was similarly used to remove over-segmented pixels by removing objects below a threshold of 10 000 voxels and to identify and isolate the target nuclei within each ROI by selecting the largest connected component. The Python package, scikit-image<sup>36</sup> was used to automate these aspects of the data analysis pipeline (<https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>). Where NE segmentations produced by TAP are presented for the whole volume (e.g. Figure 4J), objects below a more stringent threshold of 100 000 voxels were removed using MorphoLibJ.<sup>23</sup> 3D renderings of segmentations were generated using Fiji's 3D Viewer plugin.<sup>37</sup>

#### 4.13 | Edge detection

Edge detection was performed with Fiji's "Find Edges" function (based on a 2D  $3 \times 3$  Sobel filter) after data rescaling and pre-processing with a median filter. Edge detection was applied to each slice of a single ROI (cell ID = C001 (ROI 1656-6756-329), Movie S14). Edge detection code is available at <https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>.

#### 4.14 | Computing resource

Data analysis was performed on available high performance computing. This included Google Colaboratory (<http://www.colab.research.google.com/>), Amazon Web Services (AWS) cloud computing service (<http://aws.amazon.com/>), and a local high performance compute cluster called 'CAMP' (Crick Data Analysis and Management Platform). For reproducibility and convenience, the final analytical pipeline was packaged and tested on AWS.

#### 4.15 | Code availability

All assets relating to the analysis and training have been made available on public repositories and a single automated pipeline for reproducing the work has been containerized using Docker to capture environment configurations. The agnosticity of the containerized pipeline has been tested by running on a public cloud instance (AWS). Further information regarding re-running the pipeline has been provided in the readme on GitHub. We provide both reproducibility instructions (using the original data) and instructions for applying the trained model to other data sets. Code is available at: <https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope>.



## ACKNOWLEDGMENTS

This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999). This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Helen Spiers, Lucy M. Collinson and Martin L. Jones conceived and directed the project. Helen Spiers wrote the manuscript, with input from all authors. Harry Songhurst, Luke Nightingale, Joost de Folter, Amy Strange, Helen Spiers and Martin L. Jones contributed data analysis. Roger Hutchings, Chris Lintott and Helen Spiers contributed to EAC project development. Christopher J. Peddie and Anne Weston produced and analysed volume EM data. The Zooniverse volunteers contributed many segmentations, without which this project would not have been possible.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/tra.12789>.

## ORCID

Helen Spiers  <https://orcid.org/0000-0001-7748-7935>

## REFERENCES

- Denk W, Horstmann H. Serial blockface scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* 2004;2:e329.
- Knott G, Marchman H, Wall D, Lich B. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J Neurosci.* 2008;28:2959–2964.
- Peddie CJ, Collinson LM. Exploring the third dimension: volume electron microscopy comes of age. *Micron.* 2014;61:9–19.
- Sommer C., Straehle C., Köthe U., Hamprecht F.A. Ilastik: Interactive learning and segmentation toolkit. Paper presented at: 2011 IEEE International Symposium on Biomedical Imaging: from nano to macro; 2011:230–233.
- Belevich I, Joensuu M, Kumar D, Vihinen H, Jokitalo E. Microscopy image browser: a platform for segmentation and analysis of multi-dimensional datasets. *PLoS Biol.* 2016;14:e1002340.
- Arganda-Carreras I, Kaynig V, Rueden C, et al. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics.* 2017;33:2424–2426.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*; 2012:25:1097–1105.
- Chamier L, Laine RF, Henriques R. Artificial intelligence for microscopy: what you should know. *Biochem Soc Trans.* 2019;47:1029–1040.
- Plaza SM, Funke J. Analyzing image segmentation for connectomics. *Front Neural Circ.* 2018;12:102.
- Januszewski M, Kornfeld J, Li PH, et al. High-precision automated reconstruction of neurons with flood-filling networks. *Nat Methods.* 2018;15:605–610.
- Burke B, Stewart CL. Life at the edge: the nuclear envelope and human disease. *Nat Rev Mol Cell Biol.* 2002;3:575.
- Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. *Nat Rev Cancer.* 2004;4:677.
- Chow K-H, Factor RE, Ullman KS. The nuclear envelope environment and its cancer connections. *Nat Rev Cancer.* 2012;12:196.
- Worman HJ. Nuclear lamins and laminopathies. *J Pathol.* 2012;226:316–325.
- Bruggemann J, Lander GC, Su AI. Exploring applications of crowdsourcing to cryo-EM. *J Struct Biol.* 2018;203:37–45.
- Kim JS, Greene MJ, Zlateski A, et al. Space-time wiring specificity supports direction selectivity in the retina. *Nature.* 2014;509:331–336.
- Landry JJ, Pyl PT, Rausch T, et al. The genomic and transcriptomic landscape of a HeLa cell line. *Genes Genomes Genetics.* 2013;3:1213–1224.
- Spiers H, Swanson A, Fortson L, et al. Everyone counts? Design considerations in online citizen science. *J Sci Commun.* 2019;18(1):1–32.
- Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015:234–241.
- He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770–778.
- Perez AJ, Seydohosseini M, Deerinc TJ, et al. A workflow for the automatic segmentation of organelles in electron microscopy image stacks. *Front Neuroanat.* 2014;8:126.
- Tek F. B., Kroeger T., Mikula S., Hamprecht F. A. Automated cell nucleus detection for large-volume electron microscopy of neural tissue. Paper presented at: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); IEEE; 2014:69–72.
- Legland D, Arganda-Carreras I, Andrey P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics.* 2016;32:3532–3534.
- Karabağ C, Jones ML, Peddie CJ, Weston AE, Collinson LM, Reyes Aldasoro CC. Segmentation and modelling of the nuclear envelope of HeLa cells imaged with serial block face scanning electron microscopy. *J Imag.* 2019;5:75.
- Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2009;22:1345–1359.
- Aly M. Survey on multiclass classification methods. *Neural Netw.* 2005;19:1–9.
- Wright DE, Lintott CJ, Smartt SJ, et al. A transient search using combined human and machine classifications. *Month Notice Roy Astron Soc.* 2017;472:1315–1323.
- Benhajali Y, Badhwar A, Spiers H, et al. A standardized protocol for efficient and reliable quality control of brain registration in functional MRI studies. *Front Neuroinform.* 2020;14:7.
- Fortson L., Wright D., Lintott C., Trouille L. Optimizing the human-machine partnership with zooniverse *arXiv preprint arXiv:1809.09738.* 2018.
- Trouille L, Lintott CJ, Fortson LF. Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human-machine systems. *Proc Nat Acad Sci.* 2019;116:1902–1909.
- Deerinc TJ, Bushong EA, Thor A, Ellisman MH. NCMIR methods for 3D EM: a new protocol for preparation of biological specimens for serial block face scanning electron microscopy. *Microscopy.* 2010;6–8.
- Linkert M, Rueden CT, Allan C, et al. Metadata matters: access to image data in the real world. *J Cell Biol.* 2010;189:777–782.

33. Stalling D, Westerhoff M, Hege H-C, et al. Amira: a highly interactive system for visual data analysis. *Visual Handbook*. 2005;38: 749-767.
34. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676.
35. Wu K, Otoo E, Shoshani A. Optimizing connected component labeling algorithms. *Medical Imaging 2005: Image Processing*. Vol 5747. International Society for Optics and Photonics; 2005:1965-1976..
36. Walt S, Schönberger JL, Nunez Iglesias J, et al. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453.
37. Schmid B, Schindelin J, Cardona A, Longair M, Heisenberg M. A high-level 3D visualization API for Java and ImageJ. *BMC Bioinformatics*. 2010;11:274.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Spiers H, Songhurst H, Nightingale L, et al. Deep learning for automatic segmentation of the nuclear envelope in electron microscopy data, trained with volunteer segmentations. *Traffic*. 2021;22(7):240–253. <https://doi.org/10.1111/tra.12789>