Development of Convolutional Neural Network Based Instance Segmentation Algorithm to Acquire Quantitative Criteria of Mouse Development

Yuta Tokuoka[1], Takahiro G Yamada[1], Noriko F Hiroi[2], Tetsuya J Kobayashi[3], Kazuo Yamagata[4] & Akira Funahashi[1]

[1]*Department of Biosciences and Informatics, Keio University, Kanagawa, Japan.*

[2]*Faculty of Pharmacy, Sanyo-onoda City University, Yamaguchi, Japan.*

[3]*Institute of Industrial Science, The University of Tokyo, Tokyo, Japan.*

[4]*Faculty of Biology-Oriented Science and Technology, KINDAI University, Wakayama, Japan.*

Multicellular organisms develop from the fertilized eggs through numbers of cell divisions. While the pattern of cell division has specific rules for healthy development, some fluctuations for this pattern are allowable. In order to uncover this robust mechanism of development, the position of cells in each embryo must be analyzed quantitatively. In the embryonic developmental biology, various studies are attempted to acquire the quantitative criteria from time-series three-dimensional microscopic images by image analysis such as segmentation, to understand the mechanism of development. Unfortunately, due to the fact of inaccuracy of nuclei detection and segmentation, it is a hard task to evaluate an embryo quantitatively from bioimages automatically. Based on these demands of quantitative analysis, we developed QCA Net, which accurately performs nuclear segmentation of three-dimensional fluorescence microscopic images for early-stage mouse embryos. QCA Net is based on Convolu-

tional Neural Network. We trained QCA Net using a part of one early-stage mouse embryo. As a test, QCA Net performed segmentation of different 11 mouse embryos images. We succeeded in accurately acquiring the shape of the nucleus without fusion of nuclear regions. Besides, we achieved accurate extraction of the time-series data of nuclear number, volume, surface area, and center of gravity coordinates as the quantitative criteria of mouse development, from the segmentation images acquired by QCA Net. To our surprise, these results suggested that QCA Net recognized and distinguished the nucleus and a polar body formed in meiosis process. We consider that QCA Net can drastically contribute to performing segmentation of various bioimages in the embryonic developmental biology. The various quantitative criteria obtained from segmented images can uncover various unknown mechanisms of embryonic development, such as the robust mechanism of development.

The recent improvement of the microscopy and bioimaging technologies established technologies for live cell imaging[1–7]. This technology helped to get sufficient amount and quality of time-series three-dimensional (3D) fluorescence microscopic images of the development. In the embryonic developmental biology, various studies are attempted to acquire the quantitative criteria such as abnormal chromosome numbers, the synchrony of cell division, and the rate of development[8–10]. These studies analyzed time-series 3D microscopic images of the developmental embryo, whose cell nucleus is labeled fluorescently, utilizing image analysis such as segmentation. The previous segmentation methods in bioimage processing consisted of many types of image processing such as filtering, thresholding, morphological operation, watershed, and mask processing[4,10–13]. These methods required parameter value to execute image processing. Even though the optimal value of parameters depends on the features of each image and the imaging condition of the microscopy system, the value is arbitrarily set by the analyst, and optimization of this value tends to be neglected.

Keller et al. implemented DSLM-SI (digital scanned laser light sheet fluorescence microscopy with incoherent structured illumination microscopy) to perform time-lapse observation of *Drosophila* early-stage embryos[4]. They performed nuclear segmentation of time-series images acquired by DSLM-SI. Their algorithm was mainly based on image processing. The images obtained by DSLM-SI have a high signal-to-noise ratio (SNR). Besides, *Drosophila* embryos are transparent compared to that of the model organisms such as a mouse. However, the segmentation accuracy was drastically decreasing as the development of embryo was proceeding (95% (24.5 hours post fertilization (h.p.f.)), 73% (4.57 h.p.f.), and 54% (711.5 h.p.f.)). These results suggested that image analysis of developed embryos is difficult. The reason for making image analysis difficult is that time-series 3D fluorescence microscopic images have complicated three image features. Firstly, the fluorescence intensity decreases along the z-axis because the inside of the embryo is not just transparent. Secondly, the fluorescence intensity decreases along the time axis due to the fading of the fluorescent substance. Thirdly, the high spatial resolution cannot be achieved because of keeping a balance with cytotoxicity and photographing speed. These complicate image features are difficult for an analyst to grasp correctly. Therefore, the selection of appropriate image processing is the hard task. In other words, the current low accuracy is attributed to that the change in the spatiotemporal features of time-series 3D fluorescence microscopic image is not correctly grasped.

In recent years, the analysis methods for various bioimages using Convolutional Neural Network (CNN) which is one of deep learning algorithms in machine learning methodology have been proposed[14–20]. CNN showed better performance than the previous methods in the field of image analysis[21]. One of the critical advantages of CNN is to extract the image features useful for analysis automatically. CNN has also been applied to the segmentation methods of bioimage, and its performance was superior to the previous methods[14–18]. Çiçek et al. implemented 3D U-Net based on CNN and performed segmentation of Xenopus kidney tissue[15]. They created training data by annotating each image with ”kidney tubule”, ”inside kidney tubule”, and ”background” in voxel-wise manually. As a result of learning this training data, IoU (Intersection over Union) which is the evaluation metric of segmentation achieved 0.723, and it was shown that 3D U-Net could perform segmentation with high accuracy. Ho et al. performed segmentation of 3D fluorescent microscopy images which included nuclei of rat kidney labeled by the developed 3D CNN algorithm[18]. As a result of learning, since the voxel accuracy achieved 0.922, it was shown that highly accurate segmentation could be achieved. However, there remains a problem that a part of segmented nuclei regions is fused with the other regions. These object fusing is disturbing to acquire the quantitative criteria from bioimage.

These segmentation algorithms are based on Fully Convolutional Networks (FCN)[22], which consist of only convolution layers in CNN, and segmentation methodology by FCN is called semantic segmentation. Since semantic segmentation gives the same label to objects of the same class (Fig. 1 Semantic Segmentation), there arises the problem that regions are fused when neighboring/overlapping objects are segmented[23]. For objects with organization level, semantic segmentation is appropriate, but it is not suitable for objects with a cell or intracellular organelle level. On the other hand, the segmentation methodology called for instance segmentation adds a different label on objects of the same class (Fig. 1 Instance Segmentation). Therefore, instance segmentation is suitable for segmentation of cells and nuclei. However, because it is necessary to recognize individual objects, instance segmentation is a difficult task compared with semantic segmentation in general. Also, many instance segmentation methods using CNN proposed in the field of general image recognition are complicated and difficult to apply to other fields[24–26]. Moreover, almost all of these methods are targeting 2D images. For these reasons, there was no practical algorithm to execute instance segmentation for 3D bioimages until now.

Therefore, we developed QCA Net as a new CNN based segmentation algorithm, which does not cause object fusions. Since QCA Net is a simple structure combining the conventional semantic segmentation algorithms, it is an algorithm that can be easily applied to the analysis of bioimages. We trained QCA Net by a part of one early-stage mouse embryo. As a test, QCA Net performed instance segmentation of different 11 mouse embryos images. We succeeded in an accurate acquisition of the shape of the nucleus without fusion of nuclear regions. Besides, we succeeded in an accurate extraction of the quantitative criteria of mouse development, from the instance segmentation images acquired by QCA Net. The extracted quantitative criteria are time-series data of nuclear number, volume, surface area, and center of gravity coordinates. Surprisingly, QCA Net did not only perform nuclear segmentation accurately but also performs segmentation of nuclei alone, excluding polar bodies that are difficult to distinguish in image processing.

Results

Quantitative Criterion Acquisition Network (QCA Net) Our implemented algorithm was defined as Quantitative Criterion Acquisition Network (QCA Net) that performs instance segmentation of 3D fluorescence microscopic images (Fig. 2). QCA Net consists of two subnetworks, which are Nuclear Segmentation Network (NSN) and Nuclear Detection Network (NDN). NSN learned the task of nuclear segmentation, and NDN learned the task of nuclear identification. The inputs of QCA Net are time-series 3D fluorescence microscopic images. QCA Net executes instance segmentation for the images one by one. When time-series 3D fluorescence microscopic images are input, instance segmentation is performed at each time. The lower part of Fig. 2 showed an example of instance segmentation for the 3D fluorescence microscopic image at four-cell stage.

The segmentation task at each time consists of the following procedure. Initially, the input of 3D fluorescence microscopic image is preprocessed. Next, NSN performs semantic segmentation of nuclear region, and NDN performs semantic segmentation of nuclear center region in parallel. Finally, the estimated nuclear region by NSN is segmented by marker-based watershed from the identified nuclear center region by NDN. Executed by these processes, instance segmentation at a one-time point is achieved.

As shown in the upper part of Fig.2, inputting time-series images to QCA Net sequentially, it outputs instance segmentation images at each time. QCA Net can extract the quantitative criteria for mouse development from time-series instance segmentation images. These criteria are the time-series data of the nuclear number, volume, surface area, and center of gravity coordinates. Details of hyperparameters and learning methods of QCA Net were described in Methods.

Qualitative evaluation of QCA Net To evaluate the performance of QCA Net, we compared the segmentation accuracy between QCA Net and 3D U-Net[15], which is semantic segmentation algorithm of 3D bioimage. QCA Net and 3D U-Net learned the same dataset which was sampled from one early-stage mouse embryo. The creation of datasets was described in Methods.

We qualitatively compared segmentation results obtained by QCA Net and 3D U-Net (Fig. 3). QCA Net and 3D U-Net performed segmentation of mouse embryos used for learning. The 3D fluorescence microscopic images of the segmentation target are at 16 cell stage and 50 cell stage in mouse embryo whose cell nuclei fluorescently labeled (Fig. 3a, d). These mouse embryos were different from the embryos used by learning. QCA Net performed instance segmentation of these embryos (Fig. 3b, e). In qualitative point of view, the nuclear region was divided correctly, and QCA Net accurately performed instance segmentation. Similarly, 3D U-Net performed semantic segmentation of these embryos (Fig. 2c, f). We evaluated that the segmented nuclear region was fused, and 3D U-Net did not perform segmentation accurately.

The polar body has a nucleus, so it is fluorescently labeled. Therefore, it should be difficult to exclude it from the segmentation result by general image processing. To achieve the purpose of acquiring the quantitative criteria of mouse development, we did not add the polar bodies to training data as the ground truth. We confirmed that the polar body was excluded in the segmentation result (Fig. 3b). It suggested that QCA Net can distinguish and recognize the nucleus and the polar body. On the other hand, although 3D U-Net learned the same training data as QCA Net, the polar body was not excluded in the segmentation result (Fig. 3c). This point is described in Recognition of polar bodies section.

QCA Net performed segmentation of time-series images of the mouse embryo used for learning (Supplementary Video 1). Early-stage mouse embryos form blastocysts during the developmental process. The cells are very close together during the blastocysts stage, so image analysis is complicated and challenging. The segmentation result showed that QCA Net performed segmentation accurately although nuclei closely gathered during the blastocysts stage.

QCA Net and 3D U-Net performed segmentation of 10 different mouse embryos which were not included in the learning datasets (Supplementary Fig. 1). The 3D fluorescence microscopic images of the segmentation target were at 8, 16, and 32 cell stage in mouse embryo whose cell nuclei fluorescently labeled (Supplementary Fig. 1a,d,g,j). Despite of the different mouse embryos from the learned embryo, QCA Net accurately performed segmentation of the nuclear region (Supplementary Fig. 1b,e,h,k). Also, QCA Net excluded the polar bodies indicated by the blue arrowheads in Supplementary Fig. 1a,d,g, and precisely performed segmentation of the nuclei region. On the other hand, there was much fusion of nuclear region in the segmentation result by 3D U-Net (Supplementary Fig. 1c,f,i,l). This result also showed that the polar body was segmented like the nucleus (Supplementary Fig. 1c,f,i blue arrowheads). The segmented polar body was one of the causes for the false positive error. Based on these results, it was qualitatively shown that QCA Net could perform instance segmentation with high accuracy.

QCA Net performed segmentation of time-series images of 10 different mouse embryos which were not included in the learning datasets (Supplementary Video 2). The development of mouse embryos had complicate characteristics such as the rate of development, nucleus arrangement, shape, and fluorescence intensity. However, the segmentation result showed that QCA Net could perform instance segmentation against these complicate characteristics robustly.

Quantitative evaluation of QCA Net We quantitatively compared segmentation accuracy of learned QCA Net and 3D U-Net. In a correct answer of the segmentation, a voxel of a correct object region was classified as an object region (True Positive, TP), or a voxel of a right background region was classified as a background region (True Negative, TN). On the other hand, in an error answer of the segmentation, a voxel of a right background region was classified as an object region (False Positive, FP), or a voxel of a correct object region was classified as a background region (False Negative, FN). According to these classifications, IoU which is the evaluation metrics of segmentation is defined by $\mathrm{IoU} = \frac{TP}{TP+FP+FN}$, where TP, FP, FN denote the number of true-positive voxels, false-positive voxels, and false-negative voxels respectively. IoU was conventionally used in the segmentation task because IoU was a metrics for comprehensively measuring the false-positive rate and the false-negative rate. However, since the IoU was calculated for each image, it could not evaluate whether or not accurate segmentation was performed without fused nuclei. So the IoU was not suitable metrics for evaluating instance segmentation. Therefore, we used a metrics called MUCov[24] based on the IoU for the evaluation of instance segmentation. The MUCov is a metric calculated by dividing the total sum of the IoU for each segmented nucleus by the nuclear number. It is defined by $\mathrm{MUCov} = \sum_i \frac{1}{N} \max_j \mathrm{IoU}(y_i, y_j^*)$, where N, y, y^*, i, j denote the segmented nuclear number, the segmented nuclear region, the ground truth of nuclear region, the label attached to the segmented nuclear ($i = 1, \dots, N$), and the label attached to the ground truth of nuclear region respectively.

We describe the result of evaluating the segmentation accuracy by QCA Net as instance segmentation, and QCA Net without NDN and 3D U-Net as semantic segmentation. To evaluate both semantic and instance segmentation, we used the metrics of IoU and MUCov. We evaluated the segmentation accuracy of QCA Net, QCA Net without NDN, and 3D U-Net using images of 18 time points mouse embryos used for learning (Table. 1). The value of each metrics was calculated for each time point, and its average and standard deviation were shown in Table. 1.

The IoU of QCA Net exceeded the IoU of 3D U-Net. 3D U-Net was reported that this could accurately perform semantic segmentation. This result showed that QCA Net was superior to the current existed algorithm of segmentation. The main reason could be considered that QCA Net performed parameter tuning using the Bayesian optimization[27] and hyperparameters of QCA Net used for learning (e.g., the optimization method, regularization) was different from 3D U-Net. Detail of tuning and learning methods was described in Methods. Also, the IoU of QCA Net slightly exceeded the IoU of QCA Net without NDN because QCA Net succeeded in exclusion of false-positive nuclei using identification results by NDN.

The MUCov of QCA Net exceeded the MUCov of both QCA Net without NDN and 3D U-Net. We considered that the MUCov of QCA Net without NDN and 3D U-Net were low caused by the fusion of nuclear regions in the segmentation (Fig. 3 and Supplementary Fig. 1). NDN performed nuclear identification with high accuracy, so QCA Net might precisely divide the fused nuclear region based on the result of nuclear identification by NDN. The accuracy of instance segmentation by QCA Net was affected by the accuracy of nuclear identification by NDN. We evaluated the nuclear identification accuracy of NDN using F-measure (Supplementary Note 1 and Supplementary Table 1). F-measure is the metric that comprehensively measures the false-positive rate and false-negative rate to evaluate nuclear identification. The average F-measure of 11 different mouse embryos was 0.915 in NDN, while the average F-measure was 0.856 in the previous study[10]. Although the nuclear identification algorithm of the previous study performed nuclear identification in same mouse embryos with high accuracy, the accuracy of NDN exceeded its accuracy. So this result showed that the nuclear identification of NDN was superior to the current existed algorithm regarding the accurate segmentation.

Based on these results, it was quantitatively shown that QCA Net performed instance segmentation with high accuracy. Details of the evaluation at each time point were shown in Supplementary Table 2 and 3.

Acquiring quantitative criteria of mouse development by QCA Net To analyze the mechanism of development, we need to acquire the location of the cells in individual embryos accurately. Based on the location of the cells, various vital criteria for development can be extracted. In this section, we describe the result of extracting the quantitative criteria of mouse development from the time-series segmentation images acquired by QCA Net. As the quantitative criteria of the mouse development, we extracted the time-series data of the nuclear number, volume, surface area, and center of gravity coordinates because these are the critical quantitative criteria used in standard embryonic development analysis (Fig. 4).

Initially, we evaluated the time-series data of the nuclear number (Fig.4a). Each series represented the result by QCA Net and the ground truth. The F-measure of nuclear identification accuracy indicated by this result was 0.932, and it was higher than the 0.898 which was the F-measure of the previous study[10]. From this result, it was shown that QCA Net extracted the time-series data of nuclear number with high accuracy.

Secondly, we evaluated the time-series data of the nuclear volume (Fig. 4b). Each series represented the mean and the standard deviation of the nuclear volume. This result showed the periodical tendency that the restoration of nuclear volume existed after its sharp decrease (Fig. 4b green arrowheads). The same tendency was also shown in the nuclear volume of *Zebrafish* embryo in the previous study[28]. The previous study discussed that this tendency was caused by nucleus aggregation and nuclear membrane collapse in anaphase in the physiological point of view. Based on our result and previous study, QCA Net could correctly catch specific characteristic of anaphase in development. Also, the nuclear volume at the two-cell stage from the pronuclear stage was about 7,000 μm^3 (Fig. 4b from 0 days up to about 1.3 days). The previous study reports that the volume of the mouse embryo at the two-cell stage was approximately 56,000 μm^3 [29]. Therefore, the scale of the nucleus volume was within the scale of the mouse embryo, and the nuclear volume of this result was a reasonable value. These results showed that the time-series data of the nuclear volume was accurately extracted.

Thirdly, we evaluated the time-series data of the nuclear surface area (Fig. 4c). Each series represents the mean and the standard deviation. As with the time-series data of the nuclear volume, the time-series data of the nuclear surface area showed the tendency mentioned above. Since the nucleus is spherical, we considered that the result showed the same tendency similar with the nuclear volume. Therefore, QCA Net accurately extract the time-series data of the nuclear surface area.

Finally, we evaluated the time-series data of the nuclear center of gravity coordinates (Fig. 4d). The color shifts from cold color to warm color as the development of embryo was proceeding in Fig. 4d. As the developmental process, the internal space expands and changes from morula to blastocyst. Also, the exterior of the cell forming the spherical outer wall is called trophectoderm[30]. The trophectoderm becomes the source of the extraembryonic tissue. During the developmental process, it was observed that the internal space was expanded and the blastocyst was formed (Fig. 4d). The time-series data of the nuclear center of gravity coordinates extracted by QCA Net accurately showed this phenomenon.

Besides, we evaluated the quantitative criterion of 10 different mouse embryos not involved in learning (Supplementary Fig. 2, 3, 4, and 5). These results showed the similar results to the result of evaluation for the learned images. From these results, QCA Net could also acquire the quantitative criteria from mouse embryos not involved in learning.

Recognition of polar bodies In early-stage embryos, polar bodies are formed in the process of meiosis in oocytes. The polar body has the nucleus but hardly has cytoplasm. Also, the polar body slowly degenerates at the developmental process and disappears naturally. Therefore, since polar bodies may not be concerned with normal developmental process, we considered that they should be excluded from segmentation targets. However, polar bodies tend to be extracted by image processing because fluorescent proteins by microinjection of mRNA encoding target fluorescent protein are introduced to polar bodies. Based on this fact, it is challenging to perform segmentation only the nucleus excluding the labeled polar bodies in standard image processing.

QCA Net succeeded in performing segmentation of nuclei without polar bodies (Fig. 5). We speculated the reason that QCA Net recognized the polar body which is difficult to be distinguished from the nucleus. It is meaningful both biologically and image engineering to understand how to recognize the polar body and exclude it from the segmentation result.

The two cases that QCA Net recognizes polar bodies and excludes them from segmentation results could be considered. The first case was that NSN and NDN excluded polar bodies (Fig. 5a,b,c). In both NSN and NDN, only nuclear regions were correctly identified, and polar bodies were excluded. The second case was that NSN identified polar bodies, but NDN excluded polar bodies (Fig. 5d,e,f). Even if NSN performed segmentation of the polar bodies as a false-positive error when the NDN did not identify the polar bodies, the segmentation region of the polar bodied was excluded in post-processing.

We considered the high accuracy of segmentation in QCA Net was achieved by the second case. QCA Net independently identified nuclei by NSN and NDN. The watershed process using the result of NDN was performed to divide the nuclear region segmented by NSN. Therefore, when NDN did not identify the false-positive error of NSN, it was excluded in post-processing. Even if NSN performed segmentation of the polar bodies, NDN did not identify them and excluded them from the segmentation result. In this way, QCA Net could distinguish between nuclear and polar bodies that are difficult to be distinguished in image processing. These results showed that QCA Net could perform higher quality of analysis for various bioimages.

Discussion

Segmentation is an important and challenging task of bio-image analysis to uncover the biological phenomenon. Although the segmentation accuracy was improved by the method based on the deep learning algorithm[14–18], these methods were based on semantic segmentation, and many problems, such as fusion of objects, remained by the semantic segmentation methodology. In this study, we focused on this point and developed QCA Net which was the instance segmentation algorithm for high-density cells or intracellular organelle in 3D images. QCA Net was a simple structure so that the algorithm could be readily applied to various targets such as a microscopic image of a bright field, phase contrast, and differential interference contrast. Particularly in nuclear detection by NDN, we considered that NDN could estimate the central nuclear region because cells and intracellular organelle have a patterned and straightforward shape. Since the method of NDN was simple and powerful, in the case of bioimage analysis, there is no need to use a complicated object detection algorithm as used in the field of general image recognition[31,32]. This nuclear detection algorithm was inspired by the idea based on the standard segmentation methodology[11,13].

In qualitatively and quantitatively, the methodology of instance segmentation by QCA Net showed the effectiveness especially for high-density objects as compared with the previous methodology of semantic segmentation. The segmentation accuracy of QCA Net showed that IoU was 0.817 and MUCov was 0.801, which exceeded the segmentation accuracy of 3D U-Net[15] (IoU was 0.665, MUCov was 0.334) which was the semantic segmentation algorithm of the representative bioimage. The nuclear identification accuracy of NDN showed that F-measure was 0.915, which exceeded the nuclear identification accuracy of the previous study[10] (F-measure was 0.856) which was the algorithm to identify nuclei in mouse embryos. Also, the quantitative criteria of mouse development extracted by QCA Net were consistent with the previously biological knowledge in many respects. By QCA Net, it is expected that the quality and throughput of analysis in the embryo development field will be significantly improved.

Finally, we describe the recognition of polar bodies in QCA Net. Relationships of polar bodies in the development have been discussed for a long time[33,34]. However, the precise explanation why polar bodies exist in embryo was hardly reported. We showed that QCA Net recognized polar bodies, so we considered that it was possible to trace only polar bodies in the developmental process by applying QCA Net. Therefore QCA Net will be a powerful infrastructure tool in the field of developmental biology in the future. On the other hand, the way to recognize the polar bodies was not evident in this study. The main reason being that the regression of deep learning was so complicated that the reason how this regression could correctly cope with the specific task was veiled. Therefore, some researches tried to analyze learned features[35,36]. Besides, the previous study reported that each layer in the neural network had the role of specific image processing[37]. Based on these researches, the regression by deep learning could be replaced as the combination of specific image processing. If this combination is revealed and the layer which has the role to distinguish nuclear and polar body is determined, the way of recognition for polar bodies will be clear. It is more important than nothing to leave the black box in deep learning. In a study using deep learning, the attitude to deepen understanding in this way is more important than anything else.

Methods

Fluorescence image. The images for learning and evaluation were the time-series images of 11 mouse early embryos obtained by a confocal microscope. They were 5,522 images observed by 3D fluorescence microscope in mouse embryos developing from the pronuclear stage to the maximum of 53 cell stages. Supplementary Table. 4 showed the condition of image acquisition. Each mouse embryo showed different developmental rates and developmental stages (Supplementary Fig. 6).

Ground Truth creation. We manually created the ground truth from fluorescence microscopic images at different 18-time points in 1 embryo using Fiji[38]. NSN and NDN learned the task of nuclear segmentation and identification from the created ground truth. This embryo developed from the pronuclear stage to 50 cell stage. Since polar bodies are not concerned with normal developmental process, we excluded polar bodies from the ground truth. The ground truth to learn the task of nuclear identification was a sphere region with a diameter of 5 voxels. The sphere region was created based on the nuclear center of gravity coordinates. Supplementary Table. 5 and Supplementary Fig. 7 showed the sampled 18-time point and these cell stage. The ground truth, which compared the time-series data of the nuclear number, was created at each time point of 11 mouse embryos.

QCA Net overview. QCA Net consists of Nuclear Segmentation Network (NSN) that learned nuclear segmentation task and Nuclear Detection Network (NDN) that learned nuclear identification task (Fig. 2). QCA Net performs instance segmentation of the time-series 3D fluorescence microscopic images at each time point, and the quantitative criteria for mouse development are extracted from the acquired time-series segmentation image.

We implemented QCA Net on Python 2.7 and used Chainer[39] which is an open source framework. We used NVIDIA Tesla K40 (745 MHz, 4.29 TFLOPS) and NVIDIA Tesla P100 (1189 MHz, 9.3 TFLOPS) for calculation of learning and segmentation. P100 is on Reedbush-H which is a calculation server of the University of Tokyo Information Infrastructure Center. The performance of each GPU in parenthesis represents the operating frequency and single precision floating point performance, respectively.

Pre-processing in QCA Net. We performed four pre-processing, (a) Normalization, (b) Mirror padding, (c) Bicubic interpolation, (d) Data augmentation, to the input 3D fluorescence microscopic images.

We performed (a) Normalization with the objective of preventing divergence of values and gradient disappearance in learning. The value of each voxel to be normalized (I') is defined by

$$I' = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{1}$$

Where I defines a value of each voxel to be normalized, I_{max} defines maximum voxel value in the image, and I_{min} defines minimum voxel value in the image. The value of I' is obtained for all the voxels in the image, and the value range of the voxel value is $[0, 1]$.

(b) Mirror padding is one of the methods padding around the image. Mirror padding is performed by acquiring voxel values inside from m pixels from the edge of the image and extrapolating this mirror image to the outer edge. We performed (b) Mirror padding to fit the patch area within the image even if the voxel of interest is out of the image. The patch size of QCA Net was 128 voxels, so the size of mirror padding was 64 voxels of half of the patch size.

Since the x, y, and z-axis resolution in the microscopic image to be analyzed was $0.8 : 0.8 : 1.75 \mu m$ (Supplementary Table. 4), it was necessary to align it to the actual scale ratio of $1 : 1 : 1$.

We interpolated 2.1875-times in the z-axis direction as (c) Bicubic interpolation. Although the learning and the evaluation target are acquired under the same condition, interpolation to the actual scale ratio is essential when applying to the image acquired under different conditions.

We performed (d) Data augmentation, and expanded the number of data four times for each training image by flip on the x-axis, the y-axis, and both the x and y-axes. In general, when the number of samples of training data is small, a method of processing data and expanding the number of samples to prevent over learning is used. In this study, the number of samples of the ground truth was small, so we expanded the data. Since the bias of the luminance in the z-axis direction, which is the feature of the time-series 3D fluorescence microscopic image, is always constant, we did not expand data to the z-axis direction.

Nuclear Segmentation Network. NSN is a network that performs semantic segmentation of nuclear regions from 3D fluorescence microscopic images. We used SGD as an optimization method for learning. The structure of the network is based on 3D U-Net[15], and parameter tuning suitable for the dataset was performed by the Bayesian optimization in SigOpt. The tuning methods were detailed in Tuning hyperparameter of NSN & NDN Section. NSN had 16,850 parameters and was a structure smaller than 3D U-Net (Supplementary Table. 6).

Nuclear Detection Network. NDN is a network that performs semantic segmentation of nuclear center region from 3D fluorescence microscope image. We used Adam[40] as an optimization method for learning. The structure of the network was based on 3D U-Net[15], and parameter tuning suitable for the data set was performed by the Bayesian optimization in SigOpt. NDN had 235,934 parameters and was a structure larger than 3D U-Net (Supplementary Table. 7).

Post-processing in QCA Net. We performed two post-processing, (a) Re-interpolation and (b) Marker-based watershed, on the semantic segmentation image output from NSN and NDN. (a) Re-interpolation restores the resolution of the image interpolated for segmentation and identification. (b) Marker-based watershed divides the semantic segmentation region by watershed with the center region of the identified nuclei as a marker. This process enabled QCA Net to execute instance segmentation.

Tuning hyperparameter of NSN & NDN. Hyperparameters in NSN and NDN were optimized by the Bayesian optimization. To perform the optimization, we used SigOpt which is the optimization platform. The total number of hyperparameters targeted for optimization is 10 (Supplementary Table. 8).

We describe the hyperparameters for optimization. Up & downsampling denotes the number of times to perform upsampling and downsampling. Initial channels denote the number of channels in the first convolution layer. The number of channels in the subsequent convolutional layer is based on the structure of 3D U-Net which is doubled in the upsampling section and is halved in the downsampling section. Kernel size denotes the common kernel size in all convolution layers. Weight decay denotes the parameter of the L2 norm.

In this experiment, we used SGD and Adam as the optimizer. The initial learning rate, which is a parameter of SGD, denotes the initial value of the learning coefficient, and the decay learning rate denotes a parameter for multiplying the initial learning rate when the accuracy of the reference evaluation metrics is lowered. The reference evaluation metrics of NSN was IoU and NDN was F-measure. Adam's parameters $\alpha, \beta_1, \beta_2, \epsilon$ are parameters incorporated into the expression that determines learning rate of Adam. Epoch denotes the number of learning iteration, and an attempt to input all the training data and make it learn is defined as one epoch. Since Epoch is the value which we empirically understand learning converges, we did not add to this optimization target, and it was 50 in both models. Since Epoch is a parameter that directly affects the amount of computation, we consider that if computation resources are abundant, it should be added to the optimization target.

Model architecture & learning condition of NSN. We describe hyperparameters of NSN determined by the Bayesian optimization. Supplementary Table 9, 10 shows the hyperparameter of the determined NSN and the model architecture of NSN was decided based on this result (Supplementary Table. 6). The output function of NSN called as softmax is defined by

$$y_k = \frac{\exp(x_k)}{\Sigma_{j=1}^{K} \exp(x_j)} \tag{2}$$

where K denotes the number of class, which is nucleus or background region, x denote each input from the final layer, and y denotes output value. Also, the objective function of NSN called as softmax cross entropy is defined by

$$E = -\sum_{n=1}^{N} \sum_{k=1}^{K} d_{nk} \log y_k \tag{3}$$

where d denotes the ground truth and N denotes the number of learning data. We evaluated learning the model determined based on each optimizer (Supplementary Table. 11 and Supplementary Fig. 8). As a result, NSN trained using SGD was able to perform nuclear segmentation with high accuracy, so we adopted NSN trained using SGD for QCA Net.

Model architecture & learning condition of NDN. We describe the hyperparameter of NDN determined by the Bayesian optimization. Supplementary Table 12, 13 shows the hyperparameter of the determined NDN and the model architecture of NDN was decided based on this result (Supplementary Table. 7). Also, softmax function and softmax cross entropy similar to NSN was used for output function and objective function. We evaluated by learning the model determined based on each optimizer (Supplementary Table. 14 and Supplementary Fig. 9). As a result, NDN trained by Adam was shown to be able to perform nuclear identification with high accuracy, so we adopted Adam trained NDN for QCA Net.

Extraction of quantitative criteria from segmentation image. We describe the method of extracting the nuclear number, volume, surface area, center of gravity coordinates from the segmentation images. The nuclear number was extracted by counting the number of labels of the segmentation images. The nuclear volume was extracted by converting the voxel number of the segmented nuclear region for each label into the actual scale. The nuclear surface area was extracted by converting the voxel number of the nuclear region that is in contact with the background region to the actual scale. The nuclear center of gravity coordinates was calculated by calculating the center of gravity of the segmented nuclear region for each label.

1. Yamagata, K., Suetsugu, R. & Wakayama, T. Long-term, six-dimensional live-cell imaging for the mouse preimplantation embryo that does not affect full-term development. *Journal of Reproduction and Development* **55**, 343–350 (2009).

2. Yamagata, K. *et al.* Fluorescence cell imaging and manipulation using conventional halogen lamp microscopy. *PloS one* **7**, e31638 (2012).

3. Keller, P. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* **322**, 1065–1069 (2008).

4. Keller, P. J. *et al.* Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nature methods* **7**, 637–642 (2010).

5. Fercher, A., O'Riordan, T. C., Zhdanov, A. V., Dmitriev, R. I. & Papkovsky, D. B. Imaging of cellular oxygen and analysis of metabolic responses of mammalian cells. In *Live Cell Imaging*, 257–273 (Springer, 2010).

6. Tomer, R., Khairy, K., Amat, F. & Keller, P. J. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nature methods* **9**, 755–763 (2012).

7. Abe, T., Aizawa, S. & Fujimori, T. Live imaging of early mouse embryos using fluorescently labeled transgenic mice. In *Imaging and Tracking Stem Cells*, 101–108 (Springer, 2013).

8. Bao, Z. *et al.* Automated cell lineage tracing in caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2707–2712 (2006).

9. Mizutani, E. *et al.* Abnormal chromosome segregation at early cleavage is a major cause of the full-term developmental failure of mouse clones. *Developmental biology* **364**, 56–65 (2012).

10. Bashar, M. K., Komatsu, K., Fujimori, T. & Kobayashi, T. J. Automatic extraction of nuclei centroids of mouse embryonic cells from fluorescence microscopy images. *PLoS One* **7**, e35550 (2012).

11. Chinta, R. & Wasser, M. Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live drosophila embryos. *Cytometry Part A* **81**, 52–64 (2012).

12. Bashar, M. K., Yamagata, K. & Kobayashi, T. J. Improved and robust detection of cell nuclei from four dimensional fluorescence images. *PloS one* **9**, e101891 (2014).

13. Rajasekaran, B., Uriu, K., Valentin, G., Tinevez, J.-Y. & Oates, A. C. Object segmentation and ground truth in 3d embryonic imaging. *PloS one* **11**, e0150853 (2016).

14. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (Springer, 2015).

15. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 424–432 (Springer, 2016).

16. Xing, F., Xie, Y. & Yang, L. An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging* **35**, 550–566 (2016).

17. Van Valen, D. A. *et al.* Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology* **12**, e1005177 (2016).

18. Ho, D. J., Fu, C., Salama, P., Dunn, K. W. & Delp, E. J. Nuclei segmentation of fluorescence microscopy images using three dimensional convolutional neural networks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, 834–842 (IEEE, 2017).

19. Buggenthin, F. *et al.* Prospective identification of hematopoietic lineage choice by deep learning. *Nature Methods* **14**, 403–406 (2017).

20. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. Label-free prediction of three-dimensional fluorescence images from transmitted light microscopy. *bioRxiv* 289504 (2018).

21. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105 (2012).

22. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440 (2015).

23. Noh, H., Hong, S. & Han, B. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, 1520–1528 (2015).

24. Silberman, N., Sontag, D. & Fergus, R. Instance segmentation of indoor scenes using a coverage loss. In *European Conference on Computer Vision*, 616–631 (Springer, 2014).

25. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2980–2988 (IEEE, 2017).

26. Bai, M. & Urtasun, R. Deep watershed transform for instance segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2858–2866 (IEEE, 2017).

27. Brochu, E., Cora, V. M. & De Freitas, N. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599* (2010).

28. Puah, W. C., Chinta, R. & Wasser, M. Quantitative microscopy uncovers ploidy changes during mitosis in live drosophila embryos and their effect on nuclear size. *Biology open* **6**, 390–401 (2017).

29. Pogorelova, M., Yashin, V., Pogorelov, A. & Golichenkov, V. Quantitative tomography of mouse early embryo. In *Doklady Biological Sciences*, vol. 418, 61–63 (Springer, 2008).

30. Fleming, T. P. A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Developmental biology* **119**, 520–531 (1987).

31. Liu, W. *et al.* Ssd: Single shot multibox detector. In *European conference on computer vision*, 21–37 (Springer, 2016).

32. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788 (2016).

33. Verlinsky, Y. *et al.* Analysis of the first polar body: preconception genetic diagnosis. *Human reproduction* **5**, 826–829 (1990).

34. Xia, P. Intracytoplasmic sperm injection: correlation of oocyte grade based on polar body, perivitelline space and cytoplasmic inclusions with fertilization rate and embryo quality. *Human Reproduction (Oxford, England)* **12**, 1750–1755 (1997).

35. Zeiler, M. D., Taylor, G. W. & Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2018–2025 (IEEE, 2011).

36. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. *See https://arxiv. org/abs/1610.02391 v3* **7** (2016).

37. Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, 609–616 (ACM, 2009).

38. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676–682 (2012).

39. Tokui, S., Oono, K., Hido, S. & Clayton, J. Chainer: a next-generation open source framework for deep learning. In *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS)*, vol. 5 (2015).

40. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Acknowledgements The research was funded by JSPS KAKENHI Grant Numbers 16H04731. Computations were primarily performed using the computer facilities at The University of Tokyo (Reedbush). The Bayesian optimization was performed using SigOpt (https://sigopt.com).

Author contriButions Y.T. and A.F. designed the study and the conceptual idea. Y.T. implemented the algorithm of QCA Net. T.J.K. provided the Ground Truth of nuclear identification. K.Y. provided the dataset of mouse embryo. Y.T., T.Y., N.F.H., and A.F. wrote the manuscript with suggestions from the other authors.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to A.F. (email: funa@bio.keio.ac.jp).

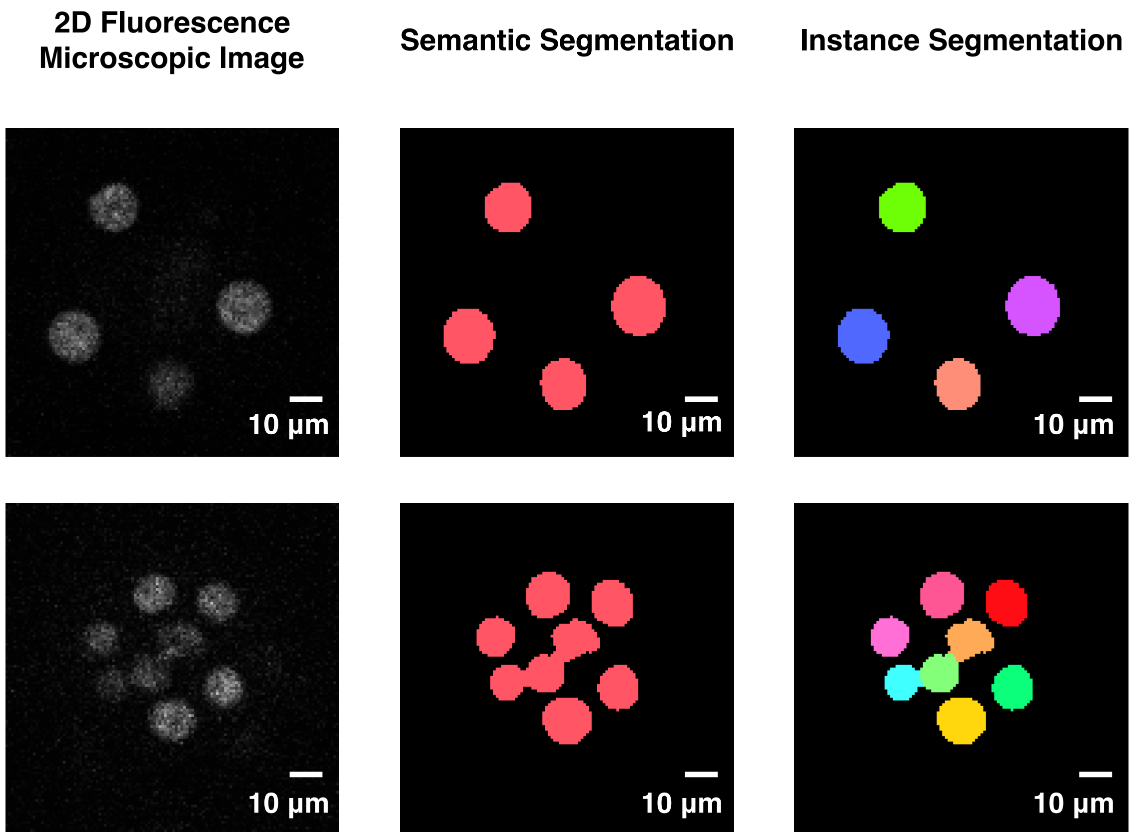

Figure 1: **Conceptual diagram of the different segmentation method**
The conceptual diagram in the case where the semantic segmentation and the instance segmentation are performed on 2D fluorescence microscopic image. In the 2D fluorescence microscopic image, all objects subject to segmentation are the same class ”nucleus”. The semantic segmentation gives the same labeling on objects of the same class, while the instance segmentation performs different labeling on objects of the same class. When sufficient spatial distance exist between objects as shown in the upper part of the figure, segmentation is accurately performed in the semantic segmentation and the instance segmentation. When objects are adjacent/overlapping as shown in the lower part of the figure, the semantic segmentation fuses the object region because segmentation is performed with the same label. On the other hand, the instance segmentation does not fuse the object region because segmentation is performed with the different labels.

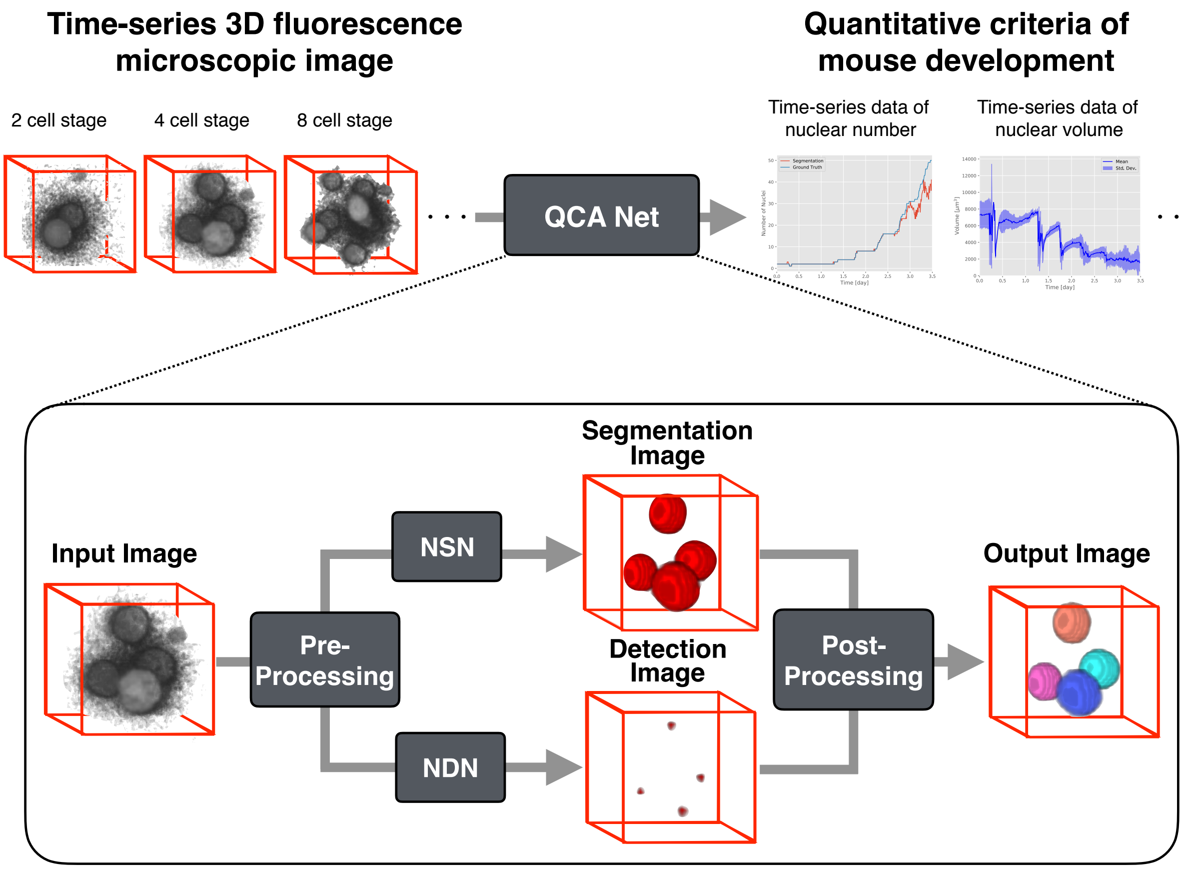

Figure 2: **Flow Diagram of our implementation algorithm QCA Net**
The conceptual diagram of QCA Net that extracts quantitative criteria for mouse development from time-series 3D fluorescence microscopic images of early-stage mouse embryos as input. QCA Net performs instance segmentation at each time of time-series images. In the process of QCA Net, the pre-processing is firstly performed to the input of the 3D fluorescence microscopic image at each time. The image after pre-processing is parallelly processed in Nuclear Segmentation Network (NSN) performing the nuclear segmentation and Nuclear Detection Network (NDN) performing the nuclear identification. The segmented nuclear region by NSN is divided by marker-based watershed in post-processing, using the identified nuclear center region by NDN. Through these process, QCA Net achieves the instance segmentation at a one-time point of time-series images. Performing these process at each time, QCA Net can acquire the time-series instance segmentation image. Finally, QCA Net extracts the quantitative criteria of mouse development such as the time-series data of the nuclear number, volume, and so on, from the time-series instance segmentation images.

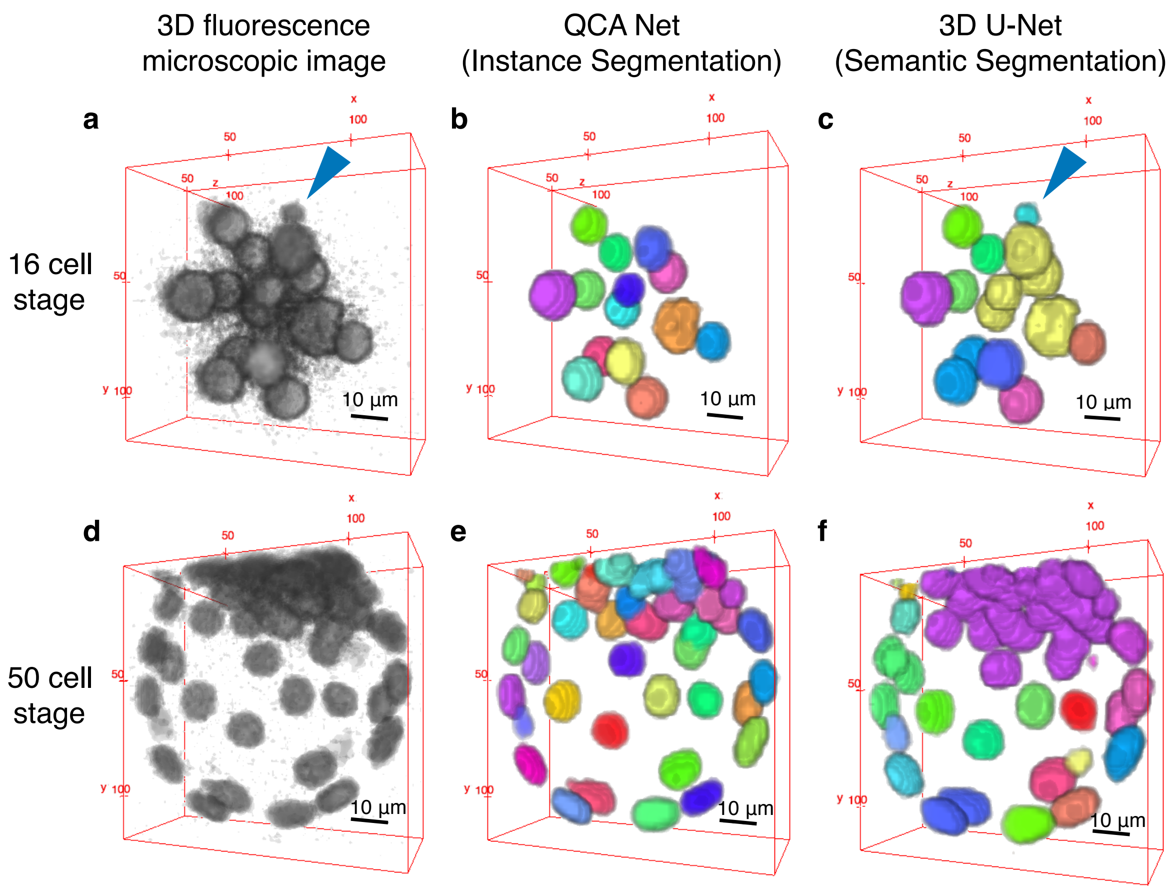

Figure 3: **Comparison of segmentation results of QCA Net and 3D U-Net**
The result of the segmentation by QCA Net and 3D U-Net to the 3D fluorescence microscopic images of one mouse embryo at each cell stage. Each color represents the individual segmented nuclear region. (a, d) The 3D fluorescence microscopic images of one mouse embryo about 2.4 days (16 cell stage) and about 3.5 days (50 cell stage) from the pronuclear stage. (b, e) The result of the instance segmentation by QCA Net to (a, d). (c, f) The result of the semantic segmentation by 3D U-Net to (a, d). It qualitatively shows that QCA Net performs accurate segmentation of nuclear region without the fusion of nuclear region. The blue arrowhead shows the polar body, and QCA Net success in not performing the segmentation of the polar body. On the other hand, the fusion of nuclear regions and the segmentation of the polar body occurs in 3D U-Net.

Table 1: **Quantitative evaluation of segmentation**
The result that the segmentation accuracy of QCA Net, QCA Net without NDN, and 3D U-Net are compared by the IoU and the MUCov. The IoU was used as a metrics of semantic segmentation, and the MUCov was used as a metrics of instance segmentation. The evaluated objects are the 18-time point embryos which were also used learning as the ground truth, and the values of each metrics represent the mean and the standard deviation.

Model	IoU	MUCov
QCA Net (instance segmentation)	0.817 (0.121)	0.801 (0.117)
QCA Net w/o NDN (semantic segmentation)	0.813 (0.121)	0.647 (0.130)
3D U-Net (semantic segmentation)	0.665 (0.123)	0.334 (0.120)

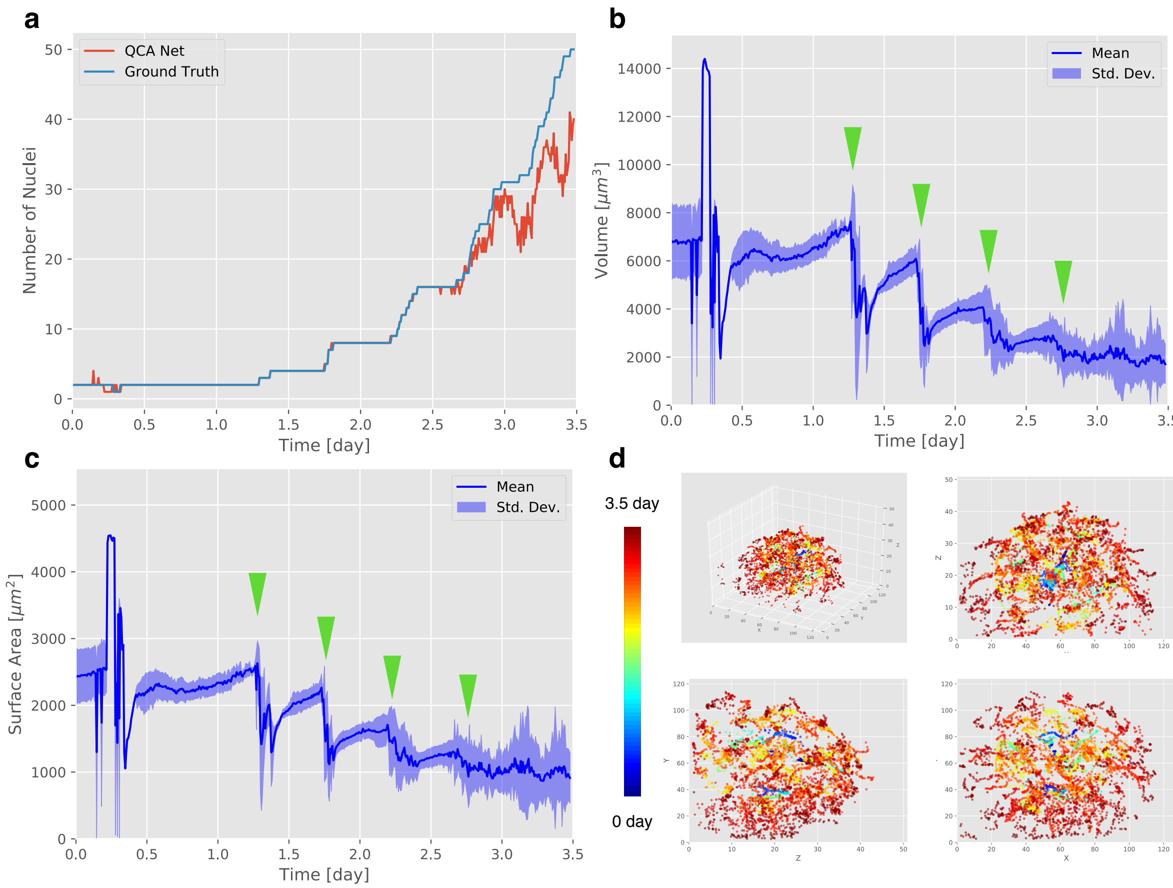

Figure 4: **Quantitative criteria of mouse development extracted by QCA Net**
(a) The time-series data of the nuclear number. Each series represents the result of QCA Net (red) and Ground Truth (blue). (b) The time-series data of the nuclear volume. The tendency that the nuclear volume rapidly decreases and then returns is periodically observed (green arrowheads). This tendency may indicate anaphase. (c) The time-series data of the nuclear surface area. Similar to the time-series data of the nuclear volume, this result also captures the features of anaphase (green arrowheads). (d) The time-series data of the nuclear center of gravity coordinates. The color shifts from cold color to warm color as the developmental process. The upper left figure shows the results displayed in 3D, and the upper right, the lower left, and the lower right chart show the cross sections of the XZ, YZ, and XY axes, respectively. During the developmental process, the internal clearance is widening, which shows that it forms the blastocyst.

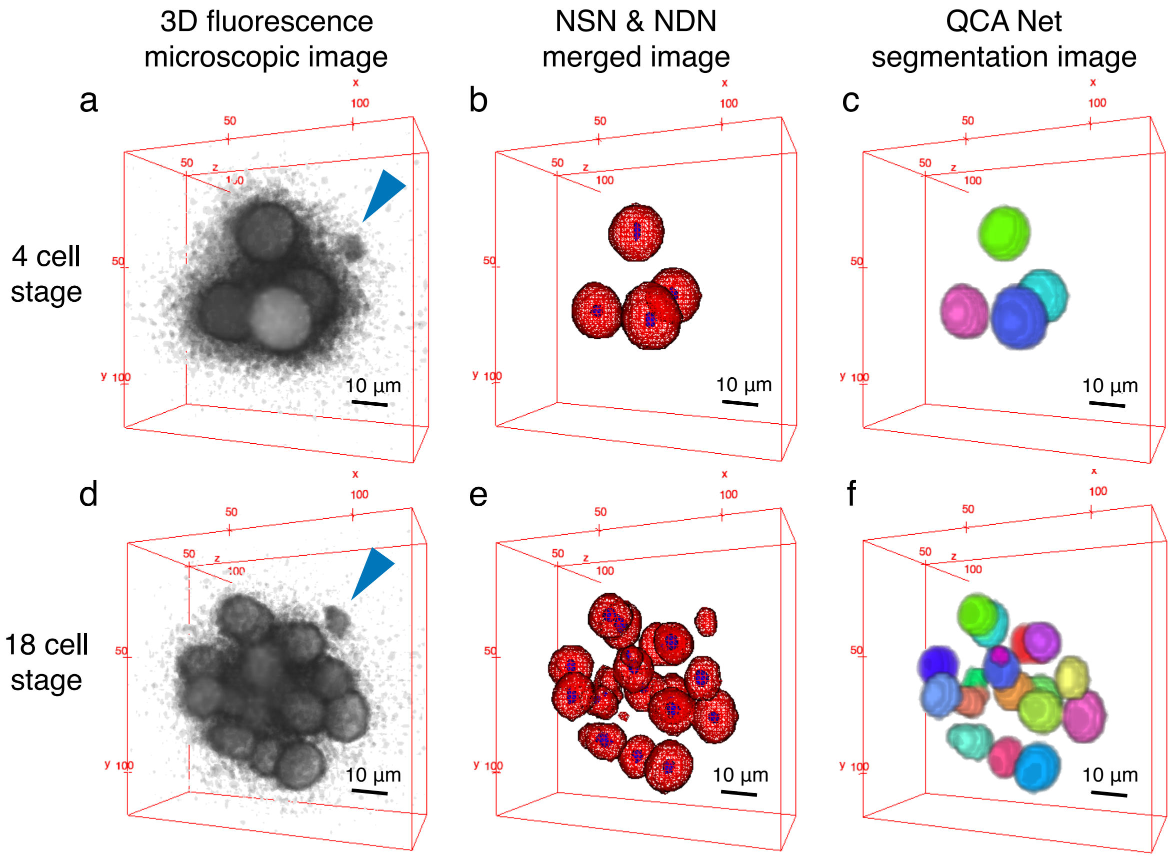

Figure 5: **Case where polar bodies are recognized and excluded from segmentation targets** Two cases in which segmentation is performed excluding the polar bodies. (a) The 3D fluorescence microscopic image of the mouse embryo at four-cell stage. The blue arrowhead represents the polar body. (b) The nuclear segmentation result of (a) by NSN (red mesh) and the nuclear identification result of (a) by NDN (blue mesh). (c) The instance segmentation result of (a) by QCA Net. These results show that NSN and NDN exclude the polar body. (d) The 3D fluorescence microscopic image of the mouse embryo at 18 cell stage. The blue arrowhead represents the polar body. (e) The nuclear segmentation result of (d) by NSN (red mesh) and the nuclear identification result of (d) by NDN (blue mesh). (f) The instance segmentation result of (d) by QCA Net. NSN performs segmentation of the polar body, but NDN excludes identification of the polar body. As a result, the polar body is excluded from the result of instance segmentation.