Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

Mask R-CNN provides efficient and accurate measurement of chondrocyte viability in the label-free assessment of articular cartilage

Hongming Fan [a], Pei Xu [b], Xun Chen [a], Yang Li [c], Zhao Zhang [a], Jennifer Hsu [a,b], Michael Le [a], Emily Ye [d], Bruce Gao [a], Harry Demos [e], Hai Yao [a,e,g], Tong Ye [a,f,*]

[a] Department of Bioengineering, Clemson University, SC, USA
[b] School of Computing, Clemson University, SC, USA
[c] School of Medicine, Yale University, New Haven, CT, USA
[d] College of Medicine, Medical University of South Carolina, Charleston, SC, USA
[e] Department of Orthopaedics & Physical Medicine, Medical University of South Carolina, Charleston, SC, USA
[f] Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC, USA
[g] Department of Oral Health Sciences, Medical University of South Carolina, Charleston, SC, USA

ARTICLE INFO

ABSTRACT

Objective: Chondrocyte viability (CV) can be measured with the label-free method using second harmonic generation (SHG) and two-photon excitation autofluorescence (TPAF) imaging. To automate the image processing for the label-free CV measurement, we previously demonstrated a two-step deep-learning method: Step 1 used a U-Net to segment the lacuna area on SHG images; Step 2 used dual CNN networks to count live cells and the total number of cells in extracted cell clusters from TPAF images. This study aims to develop one-step deep learning methods to improve the efficiency of CV measurement.

Method: TPAF/SHG images were acquired simultaneously on cartilage samples from rats and pigs using two-photon microscopes and were merged to form RGB color images with red, green, and blue channels assigned to emission bands of oxidized flavoproteins, reduced forms of nicotinamide adenine dinucleotide, and SHG signals, respectively. Based on the Mask R-CNN, we designed a deep learning network and its denoising version using Wiener deconvolution for CV measurement.

Results: Using training and test datasets from rat and porcine cartilage, we have demonstrated that Mask R-CNN-based networks can segment and classify individual cells with a single-step processing flow. The absolute error (difference between the measured and the ground-truth CV) of the CV measurement using the Mask R-CNN with or without Wiener deconvolution denoising reaches 0.01 or 0.08, respectively; the error of the previous CV networks is 0.18, significantly larger than that of the Mask R-CNN methods.

Conclusions: Mask R-CNN-based deep-learning networks improve efficiency and accuracy of the label-free CV measurement.

1. Introduction

Articular cartilage provides a smooth, lubricated surface to reduce shear and distribute loads during movement to protect underlying bone. This function is facilitated by a complex and well-organized extracellular matrix (ECM) maintained by chondrocytes [1,2]. Though they occupy a small percentage of cartilage, chondrocytes are critical in maintaining a balance between anabolism and catabolism of matrix constituents such as water, proteoglycans (PGs), and collagens. A disturbance to this balance can cause osteoarthritis (OA), the most common joint disease affecting an estimated 10% of men and 18% of women over 60 years of age worldwide [3]. Chondrocyte viability (CV), the fraction of viable chondrocytes in cartilage tissue, is an essential quantitative measure to evaluate cartilage injury [4,5], degeneration [6,7], and repair [8,9]. The CV of osteochondral allografts at the time of implantation is also an important factor that affects the long-term allograft survival rate [10–12]. Methods to measure CV primarily rely on individually labeling and counting cells through the introduction of dyes [13–15]. However, due to the potential cytotoxicity, dye-labeling does not allow measuring CV in *in vivo* applications, longitudinal studies, or allografts to be placed in patients.

Autofluorescence of intracellular fluorescent coenzymes, such as reduced forms of nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) and oxidized flavoproteins (FPs) [16], have been long used as a label-free means to study metabolic states of cells [17,18]. Previous studies found that intensities of NAD(P)H and FP emission underwent significant variation when cells were transitioning from viable to nonviable states [19–21]. Using two-photon excitation and freshly excised cartilage tissues from rat tibias, our studies [22] showed that live chondrocytes exhibited higher autofluorescence intensity from NAD(P)H than dead chondrocytes did. This observation suggested a label-free viability assay, in which live chondrocytes were identified with their bright and green color when images from NAD(P)H and FPs channels were merged and assigned with green and red pseudo colors, respectively. Both sensitivity and specificity of our label-free viability assay exceeded 90% in imaging cartilage on intact rat tibias [22].

Similar to the dye-labeling assay, our label-free viability assay needs to count live and dead cells individually to calculate CV once every cell is classified. The entire process of CV measurement requires human participation, and the analysis throughput is low as a result. Ideally, automated CV analysis needs to segment individual cells on TPAF images and classify each segmented cell as either live or dead based on appearance. However, individual cell segmentation is difficult because both cell and the ECM regions give out comparable levels of signals, and multiple chondrocytes are often nested in the lacuna, leaving a little gap between cells. As such, we previously proposed [23] (pCV-Nets) a strategy to avoid the requirement of individual cell segmentation using the following two-step cell cluster segmentation and classification: 1) utilizing SHG images to generate masks for lacunae, and 2) using the generated masks to extract cell clusters on TPAF images for classification. We used a U-Net model to segment chondrocyte clusters and two independent convolutional neural networks (CNNs) to identify the number of live or the total number of cells in each cluster. CV was determined by summing up the number of live and all cells while going over every cluster. Although a high accuracy was achieved in the viability analysis, the proposed method had an obvious disadvantage: three separate networks required three training data sets, causing high workload and computing costs. To improve the efficiency of CV measurement, this study aims to develop a single deep-learning network for individual chondrocyte segmentation and classification to improve efficiency and accuracy of the CV measurement.

Mask R-CNN [24] has been recently developed for instance segmentation that includes both segmentation and classification in a single architecture. The ROI Align technique and Fully Convolutional Network (FCN) [25] provide a pixel-level accuracy in segmentation and classification, suitable for microscopic image processing, where image contrast and the number of pixels covering a cell are both limited. In this study, we hypothesize that the Mask R-CNN architecture can provide a higher efficiency and maintains a high accuracy in the label-free CV measurement compared with the previous networks. For simplicity, we will use "previous CV networks" or pCV-Nets to indicate the previously developed multi-network deep learning method for CV analysis.

2. Materials and methods

2.1. Sample preparation

Rat and porcine cartilage samples were used in this study. Images of rat cartilage samples were acquired by a commercial two-photon microscope in the previous study [23]. Images of porcine cartilage were acquired specifically for this study on samples harvested from ten hind knee joints from adult Yorkshire pigs obtained from a local meat processing company. Muscles were stripped away from the joint, and the joint cavity was opened to reveal the articular cartilage surface. An average of 8 cartilage punches per plateau was harvested from tibia plateaus using 5 mm (ID) sample corers (18035-05, Fine Science Tools)

and were stored in Dulbecco's phosphate-buffered saline (DPBS, Corning) at 25 °C. To generate samples with a spectrum of CVs, 60 cartilage punches with no damage on surface (examined under our homebuilt two-photon microscope) were selected and randomly divided them into 4 groups with 15 samples/group: 1) the fresh sample group, which was imaged right after harvesting; 2) the 4 °C group, in which samples were kept in Dulbecco's phosphate-buffered saline (DPBS, Corning) at 4 °C in a refrigerator; 3) the cultured group, of which samples were cultured in mixed culture medium (DMEM w/Sodium Pyruvate, Penicillin-Streptomycin, Non-essential amino acids, and Fetal Bovine Serum) at 37 °C in a VWR CO2 incubator (10810-944, VWR Air Jacketed CO2 incubator) for 3 days; and 4) the frozen group, of which samples were frozen in Dulbecco's phosphate-buffered saline (DPBS, Corning) solution at −20 °C. All four sample groups were put in Petri dishes or 3D-printed sample holders and submerged in DPBS for imaging.

2.2. Label-free CV assay and TPAF/SHG imaging

As demonstrated previously [22], TPAF/SHG images were acquired and merged to form three-channel RGB pseudo-colored images by assigning red, green, and blue colors to FPs, NAD(P)H, and collagen channels, respectively. Bright, green chondrocytes were visually identified as live cells, while dim, red ones were dead cells. This assay is label-free and forms the basis for the automated CV measurement methods developed here.

In this study, rat samples were imaged by a commercial multiphoton laser scanning microscope (FV1200 inverted, Olympus Corporation, Tokyo, Japan) in the previous study [23] when our homebuilt microscope was under construction. The details of the microscope setting can be found elsewhere [22]. Briefly, a 30×, NA 1.05 silicone-oil immersion objective lens (UPLSAPO 30x, Olympus) was used to acquire images with a size of 1024 × 1024 pixels and a field of view (FOV) of 423 μm × 423 μm. A stack of 50 slices was acquired to cover a thickness of 50 μm of the cartilage tissue. TPAF and SHG imaging stacks were combined to form three-channel stacks using ImageJ (FIJI) [26] for viability analysis. An example of such a set of TPAF and SHG images is shown in Fig. 1(A)–(D).

The porcine samples were imaged by the home-built multi-channel two-photon microscope equipped with an ultrafast Ti:Sapphire laser (Chameleon Ultra II, Coherent Inc.), two GaAsP PMTs (PMT2101, Thorlabs), and one Multialkali (PMTSS, Thorlabs). The system was designed with a focus on label-free imaging and was described previously elsewhere [27]. This microscope allowed us to acquire images on all three channels (NAD(P)H, FPs, and SHG) simultaneously with a single excitation wavelength at 740 nm. Laser power was adjusted between 30–50 mW (measured at the exit pupil of the imaging objective) according to the image contrast. The transmission bands of the bandpass filters were 421–463 nm (NAD[P]H), 572–642 nm (FPs), and 352–388 nm (SHG). The imaging objective was a 16×, NA 0.8, long working distance water dipping lens (CFI75 LWD 16X W, Nikon). Typically, at each location, an image stack containing more than 30 slices with a step size of 2 μm was acquired from the surface to deeper layers of a sample. Each image has a size of 512 × 512 pixels or a FOV of 150 μm × 150 μm. TPAF and SHG imaging stacks were merged to form three-channel stacks using ImageJ (FIJI) [26] for further analysis. A set of TPAF and SHG example images is shown in Fig. 1(E)–(H).

2.3. Deep-learning CV analysis using previously developed multiple networks (pCV-Nets)

In our previously proposed CV measurement strategy (pCV-Nets), we used a U-Net plus Watershed method to find the masks for lacunae and two CNN networks to determine the total number of cells and the number of live cells in each cell cluster extracted from the TPAF image. Every CNN network classified each cell cluster into five categories according to the number of live cells or the total number of cells. The five categories
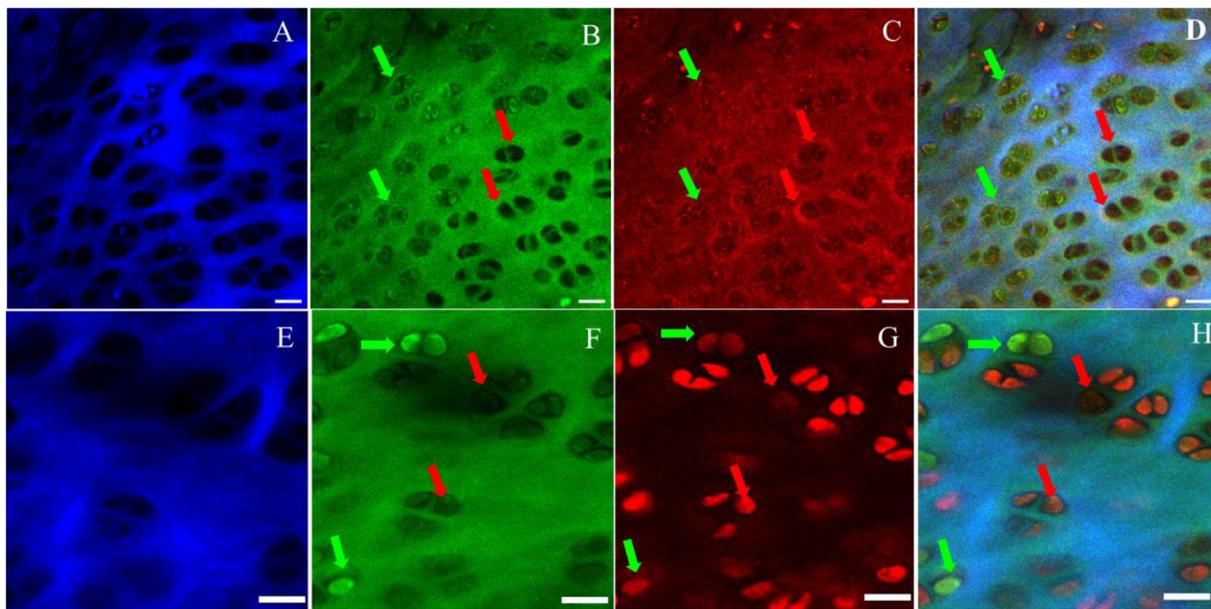
Fig. 1. Typical label-free images acquired by two-photon microscopes from rat and porcine cartilage tissues. The top (A–D) and bottom (E–H) rows are images acquired from rat and porcine cartilage, respectively. The blue, green, and red colors are assigned to SHG, NAD(P)H, and FPs channels. (D) and (H) are merged RGB images. Red arrows indicate dead cells, whereas green arrows indicate living cells. Scale bar: 20 μm.

were specified by 0–3, and 4 or more cells in a cluster, respectively. Three networks required three training data sets. The details of building all three networks were described elsewhere [23]. Once the total number of cells and the number of live cells in each cluster were determined, the CV was calculated by the following equation,

$$CV = \frac{\sum_{i=1}^{N} N_{Live}^{(i)}}{\sum_{i=1}^{N} N_{Total}^{(i)}} \tag{1}$$

where N_{Live} and N_{Total} refer to the number of live cells and the total number of cells in the i-th cell cluster; N is the total number of cell clusters segmented in a measured (imaged) area.

In this study, we implemented pCV-Nets only on rat images. The training dataset contained over 300 1024 × 1024 8-bit RGB images acquired previously from rat cartilage samples. Half of images in this training dataset was used to train U-Net (segmentation only) model in the previous study [23]; thus, we did not retrain the U-Net model for this study. The two CNN-based classification models were retrained based on the segmentation results. Additionally, a test dataset containing 120, 8-bit RGB 3-channel, 1024 × 1024 images acquired in the previous study was used to compare the performance between the pCV-Nets and Mask R-CNN based networks.

2.4. Deep-learning viability analysis using mask R-CNN architecture

Mask R-CNN [24,28] has been recently developed to significantly improve the accuracy of instance segmentation, which recognizes instances of objects and finds their boundaries at the pixel level in an image, and has quickly become an industry standard within several years. Adopting Mask R-CNN architecture, we developed a specific network called Mask R-CNN Chondrocyte Viability Network (MCV-Net) for measuring the CV via analyzing label-free images. MCV-Net outputs not only binary masks for each individual chondrocyte but also live or dead classification results for each cell. The output information is sufficient for calculating the CV of each imaged region.

Quantum yields of NAD(P)H and FPs are inherently low, resulting in low contrasts of TPAF images and potentially affecting the accuracy of

the viability analysis. To address this concern, we added the Wiener deconvolution [29], a widely used noise deduction method, in the pre-processing before MCV-Net and developed Wiener deconvolution MCV-Net (wMCV-Net). The Wiener deconvolution algorithm is described in detail in Supplementary Methods. Fig. 2 summarizes the major components of MCV-Net or wMCV-Net. (W)MCV-Net is implemented in two consecutive stages. In Stage 1, ResNet-50 and Region Proposal Network (RPN) equipped with deformable convolutions (DC) are implemented to efficiently create feature maps with a set of regions of interest (ROIs) from an input image. RPN-DC enhances the MCV-Net's ability to recognize chondrocytes in various sizes, shapes, and orientations. In Stage 2, the ROI Align layer first extracts fixed-size feature maps for all candidate chondrocytes proposed by Stage 1. Then, using the fixed-size feature maps produced by ROI Align, Fully Convolutional Network (FCN) [25] generates the bounding boxes encapsulating single cells and Fully Convolutional Layer (FCL) classifies each bounding box as live or dead cell as well as a mask that defines the cell region. Mask R-CNN replaces conventional ROI pooling with ROI Align, employing bilinear interpolation to protect the high-resolution features located at the cell border. Skip connections [30] are added between the convolutional layers in (w) MCV-Net to maintain the accuracy of the cell mask generation in processing images containing cells with a large variation in size, shape, or type. Various data augmentation, such as random cropping, rotation, flipping, and scaling, are applied to the training data to expand data volume and to diversify data distribution, which further enhances the adaptability of (w)MCV-Net.

The matplotlib 3.3, CUDA toolkit 10.2, Pytorch 1.8, and Torchvision 0.9 were used to implement the Mask-R-CNN model. On a specialized computer with an Intel i9-7920x CPU and an EVGA GeForce GTX 3080Ti GDDR6X graphics card, the training process took around 3 h. The CV was calculated simply by dividing the number of live cells by the total number of cells. The code will be available from the corresponding author upon request.

Models were trained and tested within species; for example, a model trained with rat images were not used for porcine images. For training rat or porcine models individually, training datasets were built by manually annotating chondrocytes in rat or porcine images as described in Supplementary Methods. The rat training dataset was the same as the one used for training the pCV-Nets. For the evaluation of the network
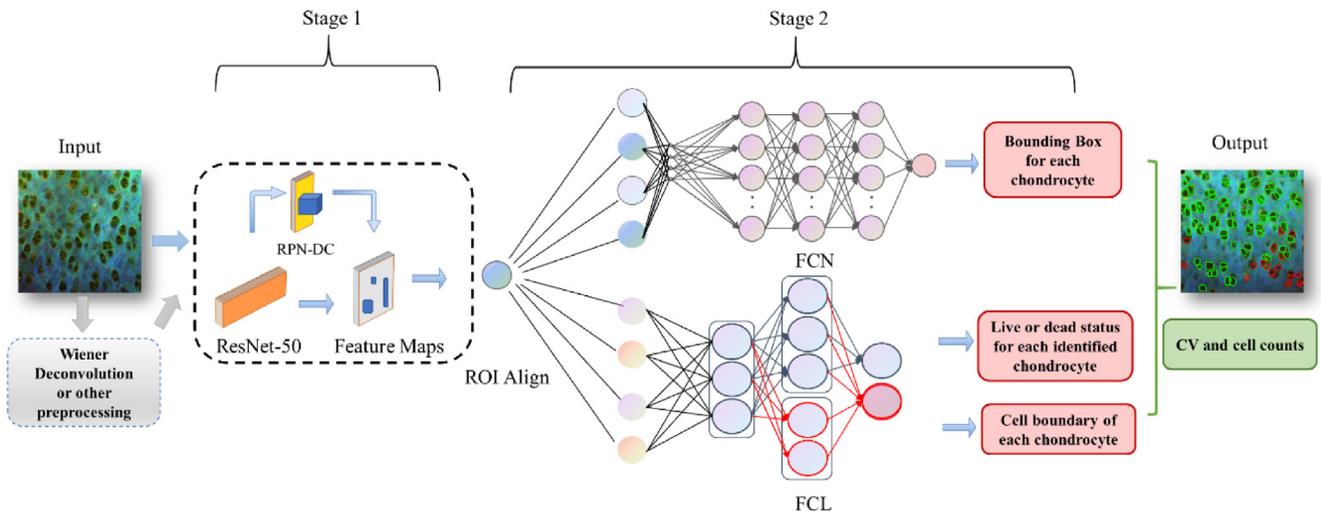
Fig. 2. The architectural diagram of MCV-Net or wMCV-Net for determining chondrocyte boundary and live or dead classes. Wiener deconvolution is an optional preprocessing for noise reduction before the network input. ResNet-50: residual network with a 50-layer convolutional neural network; RPN-DC: region proposal network with deformable convolution; FPN: feature pyramid network; FCL: fully connected layers; ROI: region of interest.

performance, we built a test dataset for samples from each species. The rat test dataset was the same as the one used for the pCV-Nets. The porcine test set consisted of 120, 8-bit RGB 3-channel, 512×512 images selected from 50 individual image stacks (typically 30–50 images/stack) acquired from different tibia cartilage samples. All images in the test dataset were annotated with LabelMe [31]. The live/dead cell status of each cell was validated by dye-labeling assay, as shown in Supplementary Methods.

2.5. Performance evaluation of the CV measurement methods and statistical analysis

The performance evaluation of the CV measurement methods included two aspects: the performance of networks and the error of the CV measurement. For evaluating the performance of networks, we adopted a standard set of evaluation metrics (precision, recall, and F1 score) commonly used in instance segmentation. Their definitions are as follows.

$$precision = \frac{True\ positive}{True\ positive + False\ positive} \qquad (2)$$

$$recall = \frac{True\ positive}{True\ positive + False\ negative} \qquad (3)$$

$$F1 = 2* \frac{precision*recall}{precision + recall} \qquad (4)$$

In the above equations, true-positive, false-positive, and false-negative indicate the number of masks falling in each category and were determined differently when we compared the performance between pCV-Nets and (w)MCV-Net. For pCV-Nets, the evaluation was done by going over every cell cluster segmented by the U-Net. True positive was only assigned to those clusters that correctly predicted the number of live and all cells according to the ground truth; otherwise, false positive was assigned. The false negative was assigned to cell clusters where the algorithm failed to detect existing cells. For (w)MCV-Net, the evaluation was done by studying the intersection over union (IoU) of every mask of each class (dead or live cells). IoU is a measure to quantify the intersection area between a predicted mask and a ground truth mask. The threshold of IoU was used to determine if a prediction was correct. A mask was defined as a true positive if its IoU is equal to or larger than 0.5 and otherwise as a false positive. A false negative was an

undetected mask. Precision reflects the fraction of correct predictions among all detected masks, while recall (also known as sensitivity) reflects the fraction of correct predictions among all masks in the ground truth. By calculating the harmonic mean of a classifier's precision and recall, the F1-score [32] is a single value for reflecting the overall performance.

Ultimately, we use the mean average precision (mAP) and mean average recall (mAR) [32,33] across classes to evaluate the performance of networks using the following equations:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} mAP_i \qquad (5)$$

$$mAR = \frac{1}{N} \sum_{i=1}^{N} mAR_i \qquad (6)$$

where mAP_i and mAR_j are the mean average precision and the mean average recall of the ith class, respectively, and N is the total number of classes. In pCV-Nets, only one class was needed to include the mask for cell clusters. In (w)MCV-Net, two classes (live and dead) of masks were detected. For each class, AP is defined as the area under the precision-recall curve (PR curve) at a threshold value of IoU. While the IoU threshold is adjusted from 0.5 to 0.95 with an increment of 0.05, 10 APs, known as AP@[0.5:0.05:0.95], can be found and their average value is defined as mAP of a class. The final mAP is calculated by averaging through all classes.

The error of the CV measurement was evaluated using the absolute error (AE) for each analyzed image and the mean absolute error (MAE) for an average over the number of test images. Additionally, the root mean square error (RMSE) was used to reflect the confidence of a measured CV. AE, MAE, and RMSE are defined as follows:

$$AE_i = |CV_G^i - CV_P^i| \qquad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} AE_i \qquad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (CV_G^i - CV_P^i)^2}{N}} \qquad (9)$$

where CV_G^i and CV_P^i are the ground truth and predicted CV for the i-th image, respectively; N is the number of images in a test data set.

We conducted a one-way analysis of variance (ANOVA) [34] using Kruskal–Wallis and Friedman tests as well as paired t tests. The processing group of the pig model (MCV-Net or wMCV-Net) was evaluated using the paired t tests [35]. The processing group (pCV-Nets, MCV-Net, and wMCV-Net) of the rat model served as the independent variables in one-way analyses of variance (ANOVA) with Kruskal–Wallis and Friedman tests. The dependent variables were accuracy, and Pearson's correlation coefficient and p-values of less than 0.05 were regarded as statistically significant in all situations. GraphPad 9 (GraphPad Software, Inc.) was used to perform statistical analysis.

3. Results

3.1. Performance of instance segmentation at the single-cell level

Representative segmentation results for rat and porcine images with their ground truth are shown in Fig. 3(A)–(F). The initial visual assessment shows that both MCV-Net and wMCV-Net can predict masks and cell classification; wMCV-Net seems to perform a little better with fewer missed or wrongly identified cells, and masks trace the cell edge more accurately. For quantitative assessment, we calculated mAP, mAR, and F1 score for each test image and performed statistical analysis of the outcome from 120 test images of each species. The F1 score was calculated using averaged mAP and mAR for each species. The statistical analysis is summarized in Fig. 4(A) and Table 1. The mAP, mAR, and F1 score of MCV-Net and wMCV-Net for either species exceed 0.9 except for mAP and F1 scores of the porcine test dataset using MCV-Net. This demonstrates Mask R-CNN based methods exhibit high performance in chondrocyte segmentation and classification. Further analysis shows that the noise reduction with Wiener deconvolution significantly improves the segmentation performance ($p < 0.001$). mAP- and mAR-IoU curves shown in Fig. 4(B) and (C) provide more evidence of the benefit of noise reduction. Both mAP- and mAR-IoU curves of wMCV-Net are broader and

taller than those of MCV-Net in either species. Table 1 summarizes the statistical analysis.

3.2. Performance comparison of the networks for CV measurement

The goal of this study is to test if Mask R-CNN based CV networks (MCV-Net and wMCV-Net) could provide at least the same performance as the previous CV networks (pCV-Nets) for CV measurement. For this purpose, we used the test datasets from rats to compare the accuracy of the CV measurement among the output of three above-mentioned networks. As shown in Fig. 5, there is a significant difference among the MAEs of pCV-Nets, MCV-Net, and wMCV-Net ($p \leq 0.001$). MCV-Net is significantly better than pCV-Nets. The results also demonstrate that noise reduction with Wiener deconvolution effectively improves the accuracy in CV measurement.

4. Discussion

In this report, we demonstrate a deep learning approach with improved efficiency for cell segmentation and classification using Mask R-CNN based networks (MCV-Net and wMCV-Net) for label-free CV measurement. (W)MCV-Net has two significant improvements over the previously proposed method [23] (pCV-Nets). Firstly, pCV-Nets use separate networks for segmentation and classification; each network needs its own training and test data sets, which increases the workload of both annotation and training. (W)MCV-Net is a single integrated architecture that can identify individual cells and, at the same time, classify them with live or dead status. This integrated architecture only needs a single annotated training and test dataset, making the training and viability analysis more efficient. Secondly, (w)MCV-Net performs instant pixel-level segmentation, which identifies each chondrocyte as a distinct object with the category that it belongs to and the boundary that separates it from the rest of the pixels in an image. In comparison, pCV-Nets
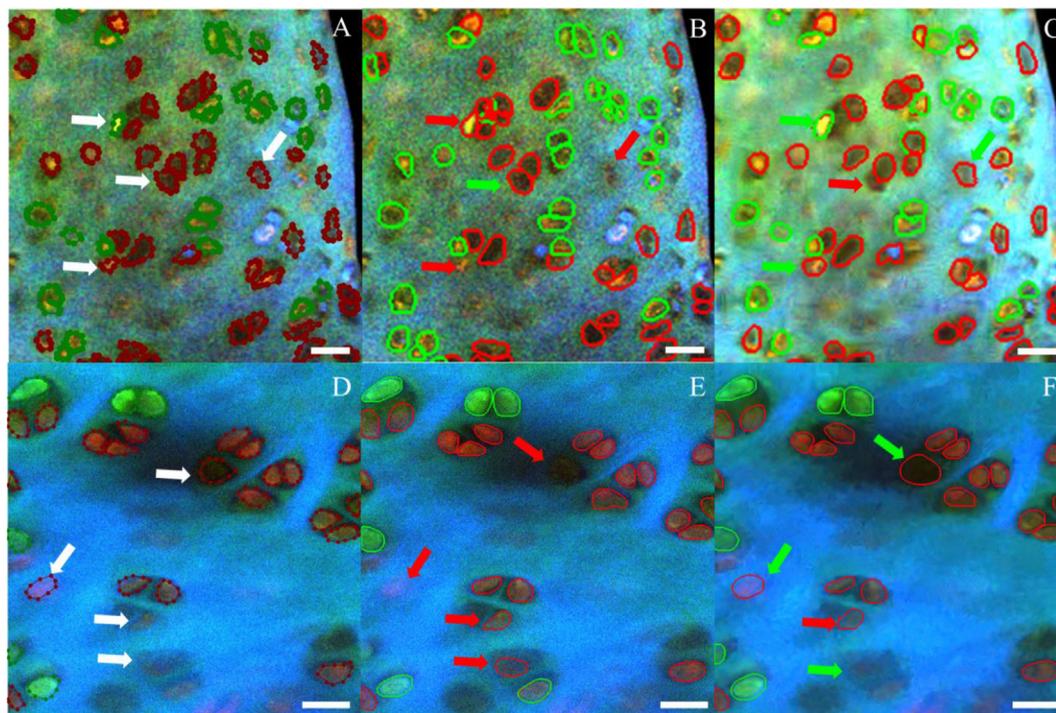

Fig. 3. Example RGB images used for annotation, training, and viability analysis for rat and porcine cartilage samples. (A) and (D) are the annotated ground truth of the rat and porcine cartilage images, respectively. Red and green dotted circles enclose the dead and live chondrocytes, respectively. (B) and (E) are the results of cell segmentation and classification using MCV-Net, while (C) and (F) are the results using wMCV-Net. Red and green solid circles are boundaries of dead and live chondrocytes, respectively. Red arrows indicate missed or incorrect cell classification, while green arrows represent the correct classification. For comparison, white arrows mark the ground truth for those cells pointed with red or green arrows.
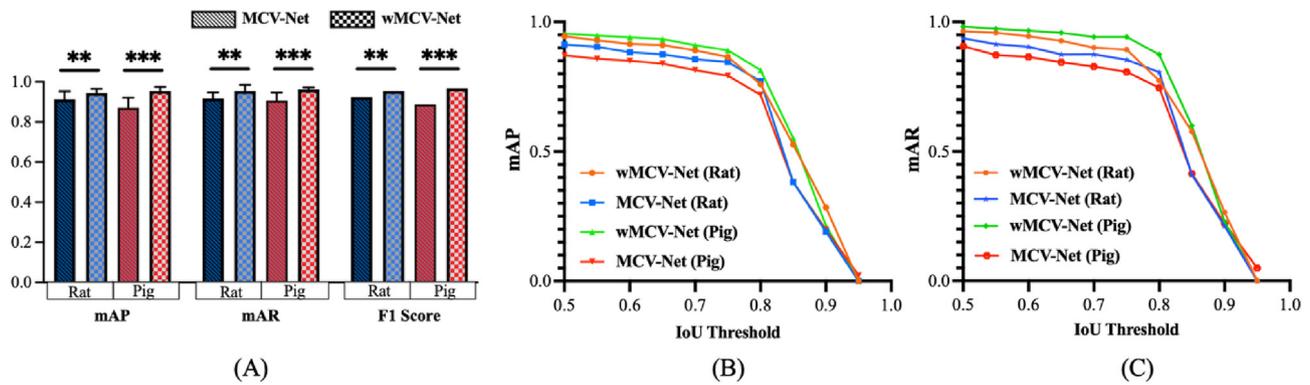
Fig. 4. Comparison of instance segmentation accuracies achieved by MCV-Net and wMCV-Net on rat and porcine cartilage samples. (A) mAP, mAR, and F1 scores with IoU threshold equal to 0.5. (B), (C) MAP-IoU and mAR-IoU curves of MCV-Net and wMCV-Net using specialized data (n = 240 test images, 120 for rat and porcine cartilage images individually). The IoU threshold measures the degree to which a predicted mask matches a ground-truth mask; a value of 1 denotes a pixel-perfect match, while a value of 0.5 denotes the proportion of successfully matched pixels to missed and false positive pixels. **: p-value < 0.01; ***: p-value < 0.001.

Table 1
Performance evaluation of different CV networks with images acquired from samples of different animals.

Networks (Species)	mAP	mAR	F1	MAE	RMSE
pCV-Net (Rat)	0.84 ± 0.07	0.85 ± 0.09	0.84	0.14 ± 0.06	0.34
MCV-Net (Rat)	0.91 ± 0.04	0.94 ± 0.03	0.92	0.08 ± 0.05	0.25
wMCV-Net (Rat)	0.94 ± 0.02	0.96 ± 0.03	0.95	0.01 ± 0.02	0.14
MCV-Net (Porcine)	0.87 ± 0.05	0.90 ± 0.04	0.88	0.06 ± 0.05	0.28
wMCV-Net (Porcine)	0.95 ± 0.02	0.96 ± 0.02	0.95	0.02 ± 0.02	0.10

available algorithms of instance segmentation did not perform well. The difficulties came from two aspects: the low image contrast associated with TPAF imaging and the limitation of algorithms before Mask R-CNN was published. As such, we developed pCV-Nets to get around the difficult task of instance segmentation. PCV-Nets utilize the relatively better contrast of SHG images to segment the region of lacunae and use the generated lacuna masks to extract cells on TPAF images for the CV analysis. Since each lacuna may contain multiple cells, the classification in pCV-Nets focuses on identifying the total or live cell numbers in each lacuna. Although this strategy is effective for CV analysis, further single-cell-based analysis is difficult.

(W)MCV-Net is a pixel-level instance segmentation method, which requires a pixel-level annotation for training and a relatively complicated architecture for finding the model. The improved efficiency and accuracy come with increased expense of annotation and training. For the CV measurements, (w)MCV-Net is perhaps an "overkill." However, it is worthwhile to have such a network that can provide pixel-level segmentation of chondrocytes for further cell-based analysis. Other than viability, classifications based on other features, for example, morphological features, are well suited for (w)MCV-Net. One advantage of the deep learning approach is its capability of automated feature extraction, effectively recognizing subtle morphological differences between categories. For example, morphological changes of chondrocytes may be used to associate with the stage of cartilage degeneration [36,37].

5. Conclusion

In this article, utilizing the Mask R-CNN architecture, we successfully demonstrate a new integrated network, (w)MCV-Net, which provides improved efficiency and accuracy of the label-free CV measurement than previously developed networks. The instance segmentation allows further quantitative cell-based analysis beyond the assessment of CV. An integrated imaging-CV measurement workflow is possible to be implemented with proper training, which may provide real-time, in situ CV measure of the cartilage. (W)MCV-Net is also possible to perform the cell viability measurement with other cell or tissue types though either re-training or transfer-learning is necessary.

Author contributions

HF built the desktop homebuilt two-photon microscope, acquired all porcine images, and analyzed data. HF and PX wrote codes for implementing Mask R-CNN based algorithms for CV analysis. XC and JH coded the U-Net based multiple networks for CV analysis. YL acquired all rat images. ML and ZZ helped with image annotation and data processing. EY, BG, HY, and HD provided insights of data interpretation and revision
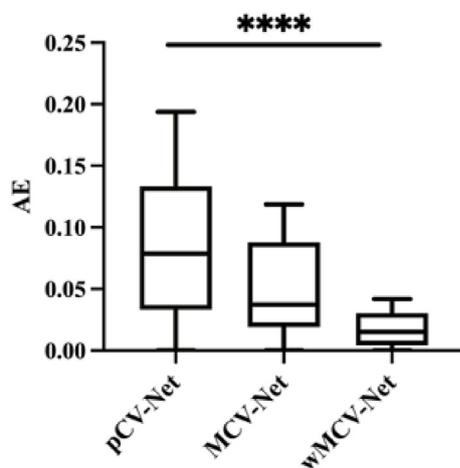

Fig. 5. Absolute errors of CV measurement using pCV-Net, MCV-Net, and wMCV-Net only with rat samples. ****: p-value < 0.0001.

perform semantic segmentation, which only outputs masks for cell clusters (multiple cells) but not individual cells. In addition, using the same testing data sets from rat samples, we also demonstrate that (w) MCV-Net provides significantly higher performance on recognizing individual cells and their categories than pCV-Nets. (W)MCV-Net shows higher mAP, mAR, and F1 scores, as seen in Table 1, leading to higher accuracy in CV measurement.

Cell counting, as it is in the CV measurement, generally requires instance segmentation though pixel-level segmentation is not necessarily required. When we started the deep learning CV measurement project,

of the draft. HF and TY conceived the original idea and drafted the manuscript. TY supervised the project.

Declaration of competing interest

The authors declare that there are no conflicts of interest related to this article.

Funding sources

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2023.100415.

References

[1] H. Muir, The chondrocyte, architect of cartilage. Biomechanics, structure, function and molecular biology of cartilage matrix macromolecules, BioEssays 17 (1995) 1039–1048, https://doi.org/10.1002/bies.950171208.

[2] W.M. Lai, J.S. Hou, V.C. Mow, A triphasic theory for the swelling and deformation behaviors of articular cartilage, J. Biomech. Eng. 113 (1991) 245–258, https://doi.org/10.1115/1.2894880.

[3] A.D. Woolf, B. Pfleger, Burden of major musculoskeletal conditions, Bull. World Health Organ. 81 (2003) 646.

[4] J.E. Jeffrey, D.W. Gregory, R.M. Aspden, Matrix damage and chondrocyte viability following a single impact load on articular-cartilage, Arch. Biochem. Biophys. 322 (1995) 87–96, https://doi.org/10.1006/abbi.1995.1439.

[5] K.D. Novakofski, R.M. Williams, L.A. Fortier, H.O. Mohammed, W.R. Zipfel, L.J. Bonassar, Identification of cartilage injury using quantitative multiphoton microscopy, Osteoarthr. Cartil. 22 (2014) 355–362, https://doi.org/10.1016/j.joca.2013.10.008.

[6] K. Kühn, D.D. D'Lima, S. Hashimoto, M. Lotz, Cell death in cartilage, Osteoarthr. Cartil. 12 (2004) 1–16, https://doi.org/10.1016/j.joca.2003.09.015.

[7] H.A. Kim, F.J. Blanco, Cell death and apoptosis in ostearthritic cartilage, Curr. Drug Targets 8 (2007) 333–345, https://doi.org/10.2174/138945007779940025.

[8] V. Rai, M.F. Dilisio, N.E. Dietz, D.K. Agrawal, Recent strategies in cartilage repair: a systemic review of the scaffold development and tissue engineering, J. Biomed. Mater. Res. A. 105 (2017) 2343–2354, https://doi.org/10.1002/jbm.a.36087.

[9] M. Brittberg, Cell carriers as the next generation of cell therapy for cartilage repair – a review of the matrix-induced autologous chondrocyte implantation procedure, Am. J. Sports Med. 38 (2009) 1259–1271, https://doi.org/10.1177/0363546509346395.

[10] R. Beaver, M. Mahomed, D. Backstein, A. Davis, D. Zukor, A. Gross, Fresh osteochondral allografts for post-traumatic defects in the knee. A survivorship analysis, J. Bone Jt. Surg. Br. 74-B (1992) 105–110, https://doi.org/10.1302/0301-620X.74B1.1732235.

[11] A. Bakay, L. Csönge, G. Papp, L. Fekete, Osteochondral resurfacing of the knee joint with allograft, Int. Orthop. 22 (1998) 277–281, https://doi.org/10.1007/s002640050260.

[12] R.F. LaPrade, J. Botker, M. Herzog, J. Agel, Refrigerated osteoarticular allografts to treat articular cartilage defects of the femoral condyles: a prospective outcomes study, J. Bone Jt. Surg. Am. 91 (2009) 805–811, https://doi.org/10.2106/JBJS.H.00703.

[13] A. Stoker, J. Stannard, J. Cook, Chondrocyte viability at time of transplantation for osteochondral allografts preserved by the Missouri osteochondral preservation system versus standard tissue bank protocol, J. Knee Surg. 31 (2018) 772–780, https://doi.org/10.1055/s-0037-1608947.

[14] B. Gantenbein-Ritter, E. Potier, S. Zeiter, M. van der Werf, C.M. Sprecher, K. Ito, Accuracy of three techniques to determine cell viability in 3D tissues or scaffolds, Tissue Eng. Part C Methods 14 (2008) 353–358, https://doi.org/10.1089/ten.tec.2008.0313.

[15] B. Rauch, R.B. Edwards, Y. Lu, Z. Hao, P. Muir, M.D. Markel, Comparison of techniques for determination of chondrocyte viability after thermal injury, Am. J. Vet. Res. 67 (2006) 1280–1285, https://doi.org/10.2460/ajvr.67.8.1280.

[16] S. Huang, A.A. Heikal, W.W. Webb, Two-photon fluorescence spectroscopy and microscopy of NAD(P)H and flavoprotein, Biophys. J. 82 (2002) 2811–2825, https://doi.org/10.1016/S0006-3495(02)75621-X.

[17] I. Georgakoudi, K.P. Quinn, Optical imaging using endogenous contrast to assess metabolic state, Annu. Rev. Biomed. Eng. 14 (2012) 351–367, https://doi.org/10.1146/annurev-bioeng-071811-150108.

[18] B. Chance, Optical method, Annu. Rev. Biophys. Biophys. Chem. 20 (1991) 1–30.

[19] R. Dittmar, E. Potier, M. van Zandvoort, K. Ito, Assessment of cell viability in three-dimensional scaffolds using cellular auto-fluorescence, Tissue Eng. Part C Methods 18 (2012) 198–204, https://doi.org/10.1089/ten.tec.2011.0334.

[20] L. Hennings, Y. Kaufmann, R. Griffin, E. Siegel, P. Novak, P. Corry, et al., Dead or alive? Autofluorescence distinguishes heat-fixed from viable cells, Int. J. Hyperth. 25 (2009) 355–363, https://doi.org/10.1080/02656730902964357.

[21] H.-W. Wang, Y.-H. Wei, H.-W. Guo, Reduced nicotinamide adenine dinucleotide (NADH) fluorescence for the detection of cell death, Anticancer Agents Med. Chem. 9 (2009) 1012–1017.

[22] Y. Li, X. Chen, B. Watkins, N. Saini, S. Gannon, E. Nadeau, et al., Nonlabeling and quantitative assessment of chondrocyte viability in articular cartilage with intrinsic nonlinear optical signatures, Exp. Biol. Med. 245 (2020) 348–359, https://doi.org/10.1177/1535370219896545.

[23] X. Chen, Y. Li, N. Wyman, Z. Zhang, H. Fan, M. Le, et al., Deep learning provides high accuracy in automated chondrocyte viability assessment in articular cartilage using nonlinear optical microscopy, Biomed. Opt. Express 12 (2021) 2759, https://doi.org/10.1364/BOE.417478.

[24] R. Girshick, Fast R-CNN, in: 2015 IEEE Int. Conf. Comput. Vis. ICCV, 2015, pp. 1440–1448, https://doi.org/10.1109/ICCV.2015.169.

[25] J. Long, E. Shelhamer, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, (n.d.) 10.

[26] J.L. West, I.D. Cameron, Using the Medical Image Processing Package, ImageJ, for Astronomy, (n.d.) 15.

[27] H. Fan, P. Xu, M. Le, J. Xsu, X. Chen, Y. Li, et al., Automated chondrocyte viability analysis of articular cartilage based on deep learning segmentation and classification of two-photon microscopic images, in: T.G. Brown, T. Wilson, L. Waller (Eds.), Three-Dimens. Multidimens. Microsc. Image Acquis. Process. XXIX, SPIE, San Francisco, United States, 2022, p. 22, https://doi.org/10.1117/12.2609880.

[28] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN. https://doi.org/10.48550/arXiv.1703.06870, 2018.

[29] A.C. Bovik, Handbook of Image and Video Processing, Academic Press, 2010.

[30] X. Mao, C. Shen, Y.-B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Adv. Neural Inf. Process. Syst., Curran Associates, Inc., 2016, in: https://proceedings.neurips.cc/paper_files/paper/2016/file/0ed9422357395a0d4879191c66f4faa2-Paper.pdf.

[31] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: a database and web-based tool for image annotation, Int. J. Comput. Vis. 77 (2008) 157–173, https://doi.org/10.1007/s11263-007-0090-8.

[32] R. Yacouby, D. Axman, Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models, in: Proc. First Workshop Eval. Comp. NLP Syst., Association for Computational Linguistics, Online, 2020, pp. 79–91, https://doi.org/10.18653/v1/2020.eval4nlp-1.9.

[33] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (2009) 427–437, https://doi.org/10.1016/j.ipm.2009.03.002.

[34] L. St»hle, S. Wold, Analysis of variance (ANOVA), Chemom. Intell. Lab. Syst. 6 (1989) 259–272, https://doi.org/10.1016/0169-7439(89)80095-4.

[35] T.K. Kim, T test as a parametric statistic, Korean J. Anesthesiol. 68 (2015) 540–546, https://doi.org/10.4097/kjae.2015.68.6.540.

[36] A. Karim, A.K. Amin, A.C. Hall, The clustering and morphology of chondrocytes in normal and mildly degenerate human femoral head cartilage studied by confocal laser scanning microscopy, J. Anat. 232 (2018) 686–698, https://doi.org/10.1111/joa.12768.

[37] Z. Abusara, I. Haider, E.K. Moo, S. Miller, S. Timmermann, W. Herzog, Chondrocyte morphology as an indicator of collagen network integrity, Connect. Tissue Res. 63 (2022) 319–328, https://doi.org/10.1080/03008207.2021.1922398.